

**TESIS**

**ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP  
PEMBELAJARAN ONLINE DI MASA PANDEMI COVID-19  
MENGUNAKAN METODE NAIVE BAYES, K - NEAREST NEIGHBOR  
DAN PEMBOBOTAN TF-IDF**



Disusun oleh:

**Nama : Ahmad Hallmi**  
**NIM : 20.77.1259**  
**Konsentrasi : Informatics Technopreneurship**

**PROGRAM STUDI S2 TEKNIK INFORMATIKA  
PROGRAM PASCASARJANA UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA**

**2021**

**TESIS**

**ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP  
PEMBELAJARAN ONLINE DI MASA PANDEMI COVID-19  
MENGUNAKAN METODE NAIVE BAYES, K - NEAREST NEIGHBOR  
DAN PEMBOBOTAN TF-IDF**

**ANALYSIS OF TWITTER USER SENTIMENT ON ONLINE LEARNING  
DURING COVID-19 PANDEMIC USING NAIVE BAYES METHODS, K -  
NEAREST NEIGHBOR AND TF-IDF WEIGHTING**

Diajukan untuk memenuhi salah satu syarat memperoleh derajat Magister



Disusun oleh:

**Nama : Ahmad Halimi**  
**NIM : 20.77.1259**  
**Konsentrasi : Informatics Technopreneurship**

**PROGRAM STUDI S2 TEKNIK INFORMATIKA  
PROGRAM PASCASARJANA UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA**

**2021**

**HALAMAN PENGESAHAN**

**ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP PEMBELAJARAN  
ONLINE DI MASA PANDEMI COVID-19 MENGGUNAKAN METODE NAIVE  
BAYES, K - NEAREST NEIGHBOR DAN PEMBOBOTAN TF-IDF**

**ANALYSIS OF TWITTER USER SENTIMENT ON ONLINE LEARNING DURING  
COVID-19 PANDEMIC USING NAIVE BAYES METHODS, K - NEAREST  
NEIGHBOR AND TF-IDF WEIGHTING**

Diperiapkan dan Disusun oleh

**Ahmad Halimi**

**20.77.1259**

Telah Dibacakan dan Dipertahankan dalam Sidang Ujian Tesis  
Program Studi S2 Teknik Informatika  
Program Pascasarjana Universitas AMIKOM Yogyakarta  
pada hari NamaHari, tanggal ujian tesis

Tesis ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Magister Komputer

Yogyakarta, 1 November 2021

**Rektor**

**Prof. Dr. M. Suyanto, M.M.**  
**NIK. 190302001**

## HALAMAN PERSETUJUAN

**ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP PEMBELAJARAN  
ONLINE DI MASA PANDEMI COVID-19 MENGGUNAKAN METODE NAIVE  
BAYES, K - NEAREST NEIGHBOR DAN PEMBOBOTAN TF-IDF**

**ANALYSIS OF TWITTER USER SENTIMENT ON ONLINE LEARNING DURING  
COVID-19 PANDEMIC USING NAIVE BAYES METHODS, K - NEAREST  
NEIGHBOR AND TF-IDF WEIGHTING**

Dipersiapkan dan Disusun oleh

**Ahmad Halimi**

**20.77.1259**

Telah Diujikan dan Dipertahankan dalam Sidang Ujian Tesis  
Program Studi S2 Teknik Informatika  
Program Pascasarjana Universitas AMIKOM Yogyakarta  
pada hari NamaHari, tanggal ujian tesis

**Pembimbing Utama**

**Prof. Dr. Kusriani, M.Kom**  
**NIK. 190302106**

**Anggota Tim Penguji**

**Dr. Andi Sunyoto, M.Kom**  
**NIK. 190302052**

**Pembimbing Pendamping**

**M. Rudyanto Arief, M.T**  
**NIK. 190302098**

**Alva Hendi Muhammad, S.T.,**  
**M.Eng., Ph.D**  
**NIK. 190302493**

**Prof. Dr. Kusriani, M. Kom**  
**NIK. 190302106**

Tesis ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Magister Komputer

Yogyakarta, 1 November 2021  
**Direktur Program Pascasarjana**

**Prof. Dr. Kusriani, M. Kom.**  
**NIK. 190302106**

## HALAMAN PERNYATAAN KEASLIAN TESIS

Yang bertandatangan di bawah ini,

**Nama mahasiswa : Ahmad Halimi**  
**NIM : 20.77.1259**  
**Konsentrasi : Informatics Technopreneurship**

Menyatakan bahwa Tesis dengan judul berikut:  
**Analisis Sentimen Pengguna Twitter Terhadap Pembelajaran Online Di Masa Pandemi Covid-19 Menggunakan Metode Naive Bayes, K - Nearest Neighbor Dan Pembobotan TF-IDF**

Dosen Pembimbing Utama : Prof. Dr. Kusriani, M. Kom  
Dosen Pembimbing Pendamping : M. Rudyanto Arief, M.T

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Tim Dosen Pembimbing
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi

Yogyakarta, 1 November 2021  
Yang Menyatakan,



Ahmad Halimi

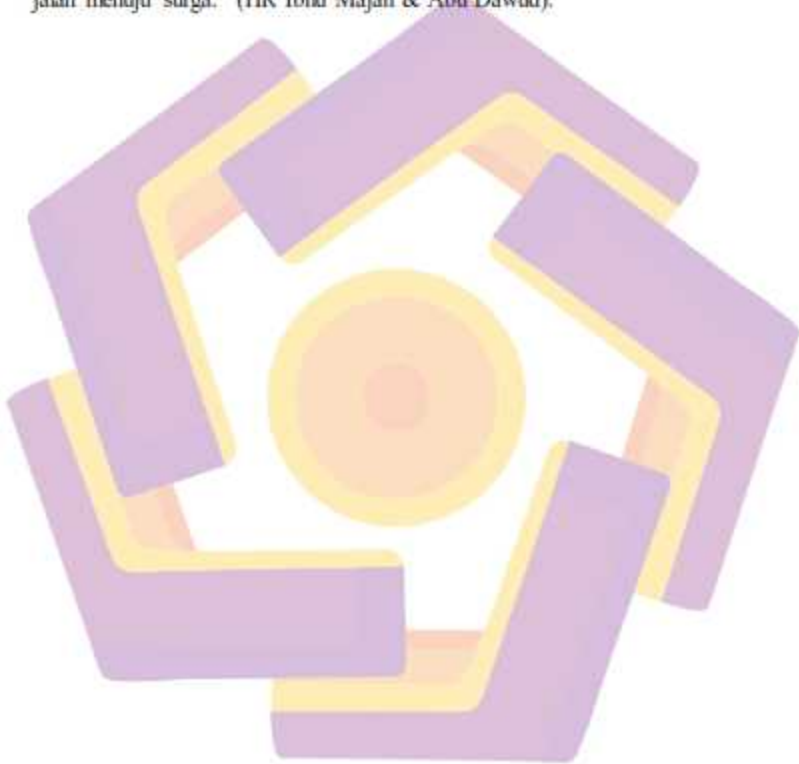
## HALAMAN PERSEMBAHAN

Puji Syukur penulis panjatkan kepada Allah SWT, yang telah memberikan kesehatan, rahmat dan hidayah, sehingga penulis masih diberikan kesempatan untuk menyelesaikan tesis ini. Semoga dapat diterima sebagai suatu amal kebaikan penelitian ini.

Penelitian ini saya persembahkan kepada kedua Orang Tua tersayang Ibu dan Bapak untuk segala bentuk dukungannya, doa dan kebaikan yang dilakukan sehingga bisa menyelesaikan studi Magister Teknik Informatika di Universitas Amikom Yogyakarta dan juga para pembaca semoga semua yang terdapat dalam naskah laporan penelitian tesis ini dapat memberikan wawasan tambahan dan kontribusi keilmuan yang baik dan bermanfaat.

## HALAMAN MOTTO

Nabi Muhammad SAW juga pernah bersabda, "Barang siapa yang menapaki suatu jalan dalam rangka menuntut ilmu, maka Allah akan memudahkan baginya jalan menuju surga." (HR Ibnu Majah & Abu Dawud).



## KATA PENGANTAR

Puji Syukur penulis panjatkan kepada Allah SWT, yang telah memberikan kesehatan, rahmat dan hidayah, sehingga penulis masih diberikan kesempatan untuk menyelesaikan tesis yang berjudul **“Analisis Sentimen Pengguna Twitter Terhadap Pembelajaran Online Di Masa Pandemi Covid-19 Menggunakan Metode Naive Bayes, K - Nearest Neighbor Dan Pembobotan TF-IDF”** yang digunakan untuk memenuhi salah satu persyaratan untuk memperoleh gelar Magister Teknik Informatika di Universitas Amikom Yogyakarta.

Dengan segala hormat, Penulis menyadari bahwa tesis dapat diselesaikan berkat dukungan dan bantuan dari berbagai pihak, oleh karena itu penulis berterima kasih kepada semua pihak yang secara langsung maupun tidak langsung memberikan kontribusi dalam menyelesaikan Tesis ini. Pada kesempatan ini penulis ingin menyampaikan ucapan terima kasih:

1. Kepada Orang Tua Saya, Bapak Muhamamad Hasan dan Ibu Yuyun Rahmawati selaku memberi doa dan dukungannya yang tulus.
2. Kepada Bapak Prof. Dr. M. Suyanto, M.M selaku Rektor Universitas Amikom Yogyakarta.
3. Kepada Ibu Prof. Dr. Kusriani, M. Kom, selaku Direktur Pascasarjana Universitas Amikom Yogyakarta dan juga selaku dosen pembimbing utama yang telah banyak memberi wawasan baru dalam menulis serta memberikan masukan yang membangun untuk penelitian ini.

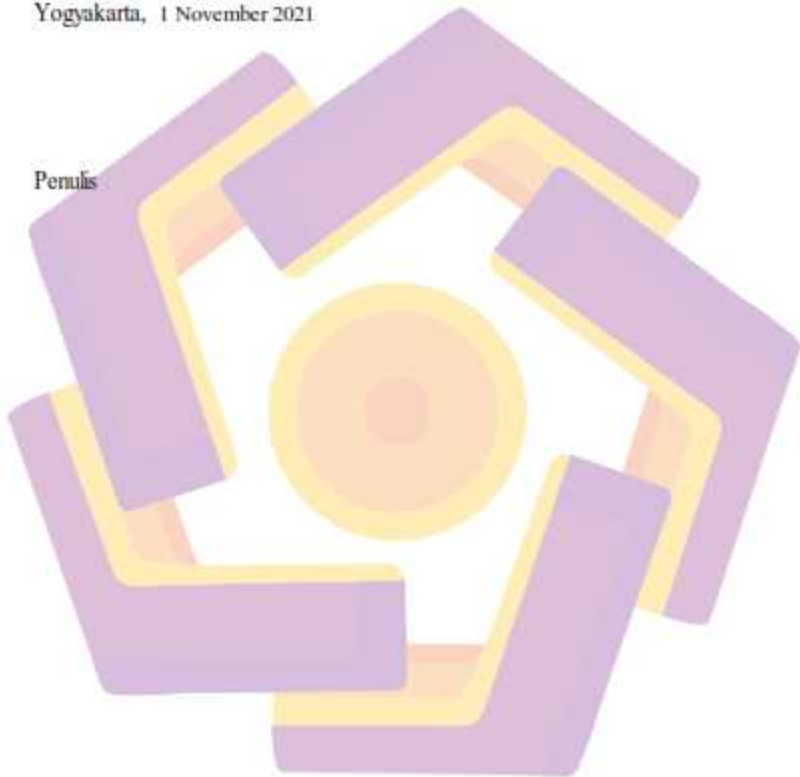


4. Kepada Ibu Prof. Dr. Ema Utami, S.Si., M.Kom selaku Wakil Direktur Pascasarjana Universitas Amikom Yogyakarta.
5. Kepada Bapak M. Rudyanto Arief, M.T selaku dosen pembimbing pendamping yang telah banyak mengarahkan secara teknis dan memberi saran dan wawasan dalam membangun penelitian ini.
6. Kepada Bapak Dr. Andi Sunyoto, M.Kom selaku dosen penguji satu yang telah memberi wawasan dan pengarahan terhadap penelitian yang saya.
7. Kepada Bapak Alva Hendi Muhammad, S.T., M.Eng., Ph.D selaku dosen penguji kedua yang telah memberi wawasan dan pengarahan terhadap penelitian yang saya.
8. Kepada Segenap Dosen dan Staff Magister Teknik Informatika Universitas Amikom Yogyakarta yang telah memberikan ilmu, wawasan, bantuan dan pengalaman baru selama perkuliahan.
9. Kepada Rekan seperjuangan MTI PJJ Angkatan 2020 yang selalu memberikan pengalaman baru dan waktu untuk bertukar pikiran dalam perkuliahan.
10. Kepada Istri Nuri Firdaus Amalyah, S.Ak selalu mendukung dalam proses penelitian ini sampai selesai.
11. Kepada bapak Kamil Malik, M.Kom, Gulpi Qorik Oktagalu Pratamasunu, S.Pd., M.Kom, Fathorozi Nur Fajri, M.Kom, Fathur Rizal, M.Kom, dan segenap dosen Fakultas Teknik Universitas Nurul Jadid yang selalu bisa memberi motivasi dan selalu bisa berbagi pengalaman dalam penelitian.

Penulisan menyadari bahwa dalam penyusunan tesis ini ada kekurangan. Oleh karena itu penulis dengan senang hati menerima kritik dan saran yang membangun dari pembaca. Akhir kata, penulis berharap semoga tesis ini dapat memberi ke manfaat bagi yang membacanya.

Yogyakarta, 1 November 2021

Penulis



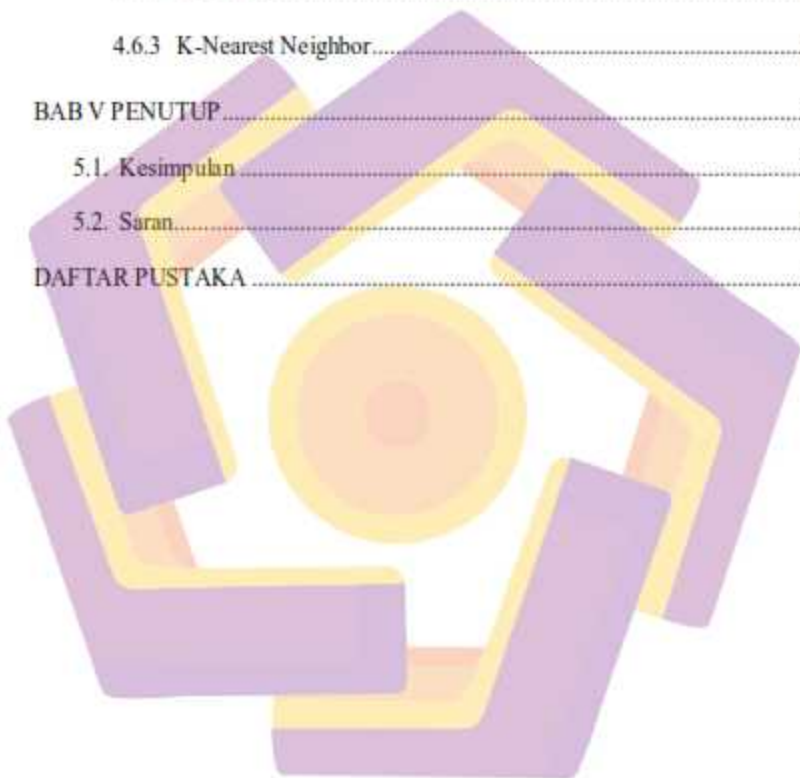
## DAFTAR ISI

HALAMAN PENGESAHAN.....	iii
HALAMAN PERSETUJUAN.....	iv
HALAMAN PERNYATAAN KEASLIAN TESIS.....	v
HALAMAN PERSEMBAHAN.....	vi
HALAMAN MOTTO.....	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	xi
DAFTAR TABEL.....	xv
DAFTAR GAMBAR.....	xvii
INTISARI.....	xix
<i>ABSTRACT</i> .....	xx
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang Masalah.....	1
1.2. Rumusan Masalah.....	10
1.3. Batasan Masalah.....	10
1.4. Tujuan Penelitian.....	11
1.5. Manfaat Penelitian.....	11
BAB II TINJAUAN PUSTAKA.....	12
2.1. Tinjauan Pustaka.....	12
2.2. Keaslian Penelitian.....	18
2.3. Landasan Teori.....	22

2.3.1 Analisis Sentimen.....	22
2.3.2 Emosi.....	23
2.3.3 <i>Information retrieval system</i> .....	24
2.3.4 Twitter Dan Twitter Api.....	28
2.3.5 Klasifikasi Teks.....	31
2.3.6 Pembobotan Kata (TF-IDF).....	32
2.3.7 <i>K-nearest neighbor</i> .....	35
2.3.8 Naïve Bayes.....	37
<b>BAB III METODE PENELITIAN.....</b>	<b>42</b>
3.1. Jenis, Sifat, Dan Pendekatan Penelitian.....	42
3.1.1 Jenis Penelitian.....	42
3.1.2 Sifat Penelitian.....	42
3.1.3 Pendekatan Penelitian.....	42
3.2. Metode Pengumpulan Data.....	43
3.2.1 Data Primer.....	43
3.2.2 Data Sekunder.....	44
3.3. Metode Analisis Data.....	44
3.4. Alur penelitian.....	46
3.4.1 Identifikasi Masalah.....	47
3.4.2 Pengumpulan Data.....	47
3.4.3 Preprocessing.....	48

3.4.4 Lexicon.....	51
3.4.5 Pembobotan TF-IDF.....	52
3.4.6 Klasifikasi Sentimen dan Hasil Pengujian.....	52
<b>BAB IV HASIL PENELITIAN DAN PEMBAHASAN.....</b>	<b>55</b>
4.1. Pengumpulan Data.....	55
4.1.1 Crawling Data Tweet.....	55
4.1.2 Membersihkan Data.....	56
4.1.3 Klasifikasi Sentimen Responden.....	57
4.2. Preprocessing.....	59
4.2.1 Case Folding.....	60
4.2.2 Tokenization.....	61
4.2.3 Spell Checker.....	63
4.2.4 Stopword.....	64
4.2.5 Normalization.....	65
4.2.6 Stemming.....	66
4.3. Lexicon.....	69
4.3.1 Klasifikasi Sentimen Lexicon.....	69
4.3.2 Klasifikasi Opini.....	72
4.4. Pembobotan TF-IDF.....	73
4.5. Klasifikasi Sentimen.....	74
4.5.1 Naïve Bayes.....	75

4.5.2 K-Nearest Neighbor.....	78
4.6. Hasil Pengujian.....	81
4.6.1 Data Training dan Data Testing.....	83
4.6.2 Naïve Bayes.....	84
4.6.3 K-Nearest Neighbor.....	88
BAB V PENUTUP.....	93
5.1. Kesimpulan.....	93
5.2. Saran.....	94
DAFTAR PUSTAKA.....	95



## DAFTAR TABEL

Tabel 2. 1. Matriks Literatur Review Dan Posisi Penelitian Analisis Sentimen Pengguna Twitter Terhadap Pembelajaran <i>Online</i> Di Masa Pandemi Covid-19 Menggunakan Metode Naive Bayes, K - Nearest Neighbor Dan Pembobotan TF-IDF .....	18
Tabel 2. 5. Hasil perhitungan TF .....	33
Tabel 2. 6. Hasil perhitungan DF .....	34
Tabel 2. 7. Hasil perhitungan IDF .....	34
Tabel 2. 8. Hasil perhitungan TF-IDF .....	35
Tabel 2. 9. Dataset sentimen k-nearest neighbor .....	36
Tabel 2. 10. Pehitungan jarak k pada k-nearest neighbor .....	37
Tabel 2. 11. Hasil label tetangga terdekat pada k-nearest neighbor .....	37
Tabel 2. 12. Dataset sentimen naive bayes .....	39
Tabel 2. 13. Perhitungan probability .....	40
Tabel 2. 14. Perhitungan probability sentimen .....	40
Tabel 2. 15. Hasil perhitungan probability sentimen .....	41
Tabel 2. 16. Hasil label sentimen .....	41
Tabel 3. 1. Perhitungan Klasifikasi Opini .....	52
Tabel 4. 1. Crawling Data Tweet .....	55
Tabel 4. 2. Pembersihan Data .....	57
Tabel 4. 3. Klasifikasi Sentimen Responden .....	58
Tabel 4. 4. Case Folding .....	60

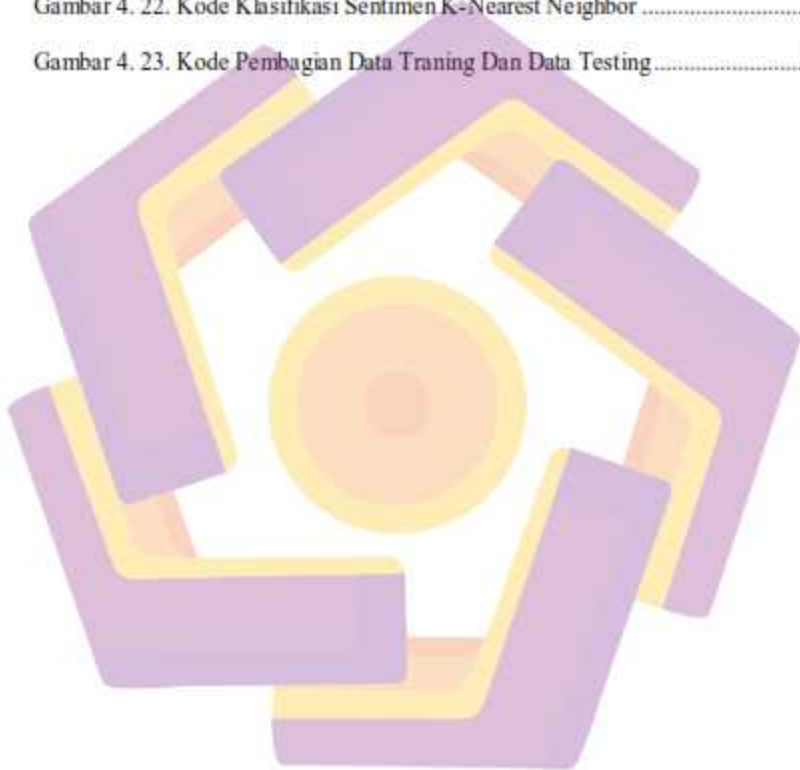
Tabel 4. 5. Tokenization.....	61
Tabel 4. 6. Spell Checker.....	63
Tabel 4. 7. Stopword.....	64
Tabel 4. 8. Normalization.....	66
Tabel 4. 9. Stemming.....	67
Tabel 4. 10. Klasifikasi Lexicon.....	69
Tabel 4. 11. Pembobotan Kata TF-IDF.....	73
Tabel 4. 12. Data Sentimen Responden dan Lexicon.....	75
Tabel 4. 13. Hasil Klasifikasi Naive Bayes.....	75
Tabel 4. 14. Total Perbandingan Lexicon dan Responden di Naive Bayes.....	76
Tabel 4. 15. Hasil Klasifikasi K-Nearest Neighbor.....	79
Tabel 4. 16. Total Perbandingan Lexicon dan Responden di K-Nearest Neighbor.....	80
Tabel 4. 17. Confusion Matrix.....	82
Tabel 4. 18. Pembagian Data Training dan Data Testing.....	83
Tabel 4. 19. Pengujian Naive Bayes.....	84
Tabel 4. 20. Naive Bayes Complement Responden.....	85
Tabel 4. 21. Naive Bayes Complement Lexicon.....	86
Tabel 4. 22. Pengujian K pada K-Nearest Neighbor.....	88
Tabel 4. 23. K-Nearest Neighbor Nilai K-3 Responden.....	89
Tabel 4. 24. K-Nearest Neighbor Nilai K-3 Lexicon.....	90



## DAFTAR GAMBAR

Gambar 2. 1. Sistem pemodelan pengolahan data .....	25
Gambar 2. 2. Implementasi klasifikasi teks .....	31
Gambar 3. 1. Alur Penelitian.....	46
Gambar 3. 2. Lexicon.....	51
Gambar 4. 1. Kode Crawling Data Tweet.....	56
Gambar 4. 2. Kode Menggabungkan data dan Membersih Data Kembar .....	57
Gambar 4. 3. Klasifikasi sentimen Manual.....	58
Gambar 4. 4. Kode Tampil Data Mentah.....	59
Gambar 4. 5. Kode Case Folding.....	60
Gambar 4. 6. Kode Bersihkan url, simbol dan lain-lainnya.....	62
Gambar 4. 7. Kode Tokenization.....	62
Gambar 4. 8. Kode Spell Checker.....	64
Gambar 4. 9. Kode Stopword.....	65
Gambar 4. 10. Kode Normalization.....	66
Gambar 4. 11. Kode Stemming.....	67
Gambar 4. 12. Wordcloud Tweet.....	68
Gambar 4. 13. Kode Wordcloud .....	68
Gambar 4. 14. Klasifikasi Responden dan Klasifikasi Lexicon.....	70
Gambar 4. 15. Kode Klasifikasi Lexicon.....	71
Gambar 4. 16. Klasifikasi Opini Lexicon .....	72
Gambar 4. 17. Perbandingan Klasifikasi Sentimen Naive Bayes .....	76

Gambar 4. 18. Hasil K lasifikasi Sentimen Naive Bayes.....	77
Gambar 4. 19. Kode Klasifikasi Sentimen Naive Bayes.....	78
Gambar 4. 20. Perbandingan K lasifikasi Sentimen K-Nearest Neighbor.....	79
Gambar 4. 21. Hasil K lasifikasi K-Neareset Neighbor .....	80
Gambar 4. 22. Kode Klasifikasi Sentimen K-Nearest Neighbor .....	81
Gambar 4. 23. Kode Pembagian Data Traning Dan Data Testing.....	84



## INTISARI

Semejak kebijakan pembelajaran online berlangsung, membuat masyarakat Indonesia meluapkan emosinya ke media sosial twitter dengan berbagai opini. Tujuan penelitian ini untuk analisis sentimen pengguna twitter terhadap pembelajaran online di masa pandemi covid-19 dengan metode *naïve bayes*, *k-nearest neighbor* dengan menggunakan pembobotan kata TF-IDF. Data twitter yang akan di *crawling* dari tanggal 16 maret 2020 sampai 10 april 2021 dengan menghasilkan data *tweet* sebanyak 17.373 dan bersihnya sebanyak 15.759. Di penelitian akan melakukan klasifikasi sentimen dan akurasi dengan *naïve bayes* dan *k-nearest neighbor* menggunakan pembobotan TF-IDF dari data *tweet* telah *crawling*. Data *tweet* yang telah *crawling*, dilakukan klasifikasi sentimen dengan responden sentimen sebanyak negatif 8148, positif 5556, netral 2055 dan *lexicon* sebanyak negatif 7671, positif 5698, netral 2390. Proses pengujian dilakukan pembersihan dan perbaikan data *tweet* dengan *preprocessing*, kemudian dilakukan klasifikasi sentimen *naïve bayes* dan *k-nearest neighbor* dengan pembobotan TF-IDF. Hasil dari klasifikasi sentimen *naïve bayes* pada data responden dan *lexicon* memiliki tingkat presentase kemiripan hasil sentimen sebanyak 84 % dan 16% tidak sama serta *k-nearest neighbor* pada data *lexicon* memiliki tingkat presentase kemiripan 72% sama dan 28% tidak sama dan responden juga miliki tingkat presentase kemiripan 73% sama dan 27 % tidak sama. Hasil terbaik akurasi di peroleh dari metode *naïve bayes* dan *k-nearest neighbor* dengan membagi data training dan data testing sebagai pengujiannya. Nb\_respoden *naïve bayes complement* dengan akurasi 74,08%, presisi 69,72% dan recall 64,51% serta di pengujian *k-neareast neighbor* nilai k-3 pada knn\_respoden dengan akurasi 78,46%, presisi 76,47% dan recall 76,64%.

Kata kunci: daring, analisis sentimen, twitter, naïve bayes, k-nearest neighbor, covid-19

## ABSTRACT

Since the online learning policy took place, Indonesian people vented their emotions on Twitter with various opinions. The purpose of this study is to analyze the sentiments of twitter users towards online learning during the covid-19 pandemic using the naive Bayes, k-nearest neighbor method using the TF-IDF word weighting. Twitter data that will be crawled from March 16, 2020 to April 10, 2021, generates 17,373 tweets and a net 15,759. In this study, sentiment classification and accuracy will be carried out with naive Bayes and k-nearest neighbor using TF-IDF weighting from tweet data that has been crawled. Tweet data that has been crawled, sentiment classification is carried out with sentiment responden as negative 8148, positive 5556, neutral 2055 and lexicon as negative 7671, positive 5698, neutral 2390. The testing process is carried out cleaning and repairing tweet data by preprocessing, then naive Bayes sentiment classification is carried out and k-nearest neighbor with TF-IDF weighting. The results of the naive Bayes sentiment classification on the respondent's data and lexicon have a similarity percentage level of sentiment results as much as 84% and 16% are not the same and k-nearest neighbor on lexicon data have a similarity percentage level of 72% the same and 28% are not the same and respondents also have a similarity level. the percentage of similarity is 73% the same and 27% is not the same. The best accuracy results are obtained from the naive Bayes method and k-nearest neighbor by dividing the training data and testing data. Nb\_responden naive bayes complement with 74.08% accuracy, 69.72% precision and 64.51% recall and in the k-nearest neighbor test the value of k-3 on knn\_responden with 78.46% accuracy, 76.47% precision and 76 recall, 64%.

**Keyword:** daring, sentiment analysis, twitter, naive bayes, k-nearest neighbor, covid-19

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang Masalah

Semenjak tahun 2020 virus covid-19 masuk ke Indonesia, membuat masyarakat takut terhadap penyebaran virus covid-19. Berdasarkan informasi yang diperoleh dari satgas covid-19 pada tanggal 02 Desember 2020, masyarakat yang terpapar virus covid-19 sudah mencapai 549.508 orang serta yang kasus aktif 73.429 orang (Kompas, 2020). Dampak penyebaran virus covid-19 sangat cepat yang disebabkan dari kontak fisik antar manusia, sehingga pemerintah Indonesia membuat suatu kebijakan untuk mencegah penyebaran virus covid-19 tidak terlalu meluas di kalangan masyarakat dengan cara sosial berskala besar (PSBB) dan memberlakukan pembelajaran secara daring atau *online* (Kemdikbud, 2020b). Dengan kebijakan PSBB telah berlangsung membuat masyarakat tidak bisa melakukan aktivitas secara normal seperti pergi bekerja dan sekolah.

Peningkatan angka terjangkit virus covid-19 di Indonesia tinggi membuat Kementerian Pendidikan dan Kebudayaan melakukan strategi untuk mencegah penyebaran di kalangan pelajar dan tenaga pengajar dengan menerbitkan surat keputusan nomor 36962/MPK.A/HK/2020 tentang pencegahan wabah virus corona pada satuan pendidikan tentang pembelajaran secara *online* atau daring kepada Kepala Dinas Pendidikan seluruh Indonesia dan Kepala Sekolah seluruh Indonesia untuk mengikuti aturan yang ditetapkan untuk menghindari dari penyakit virus covid-19 yang berlangsung (Kemdikbud, 2020a). Dikarenakan kurangnya

sosialisasi mengenai pembelajaran secara daring dan kesiapan siswa atau mahasiswa terkait sarana pembelajaran seperti laptop, *handphone* dan internet, membuat masyarakat berargumen terkait pembelajaran daring tersebut yang dinilai kurang efektif.

Pembelajaran secara *online* atau daring juga ramai diperbincangkan oleh pengguna media sosial lainnya seperti twitter dengan mengutarakan opini-opini mereka ke dalam tulisan pada *tweet* untuk memberitahu kondisi yang dialaminya saat ini. Banyak opini-opini pengguna twitter terhadap pembelajaran daring yang sempat menjadi trending topik di Indonesia dengan isi *tweet* beragam (Natasuwarna, 2020). Maka diperlukan kesimpulan terhadap pembelajaran secara *online* di masa pandemi dengan analisis sentimen pengguna twitter untuk memberikan informasi terkait pembelajaran *online* yang saat ini ramai diperbincangkan.

Analisis sentimen merupakan proses pengolahan data teks yang memiliki makna yang belum bisa diartikan kedalam sebuah informasi. Secara umum, sentimen dapat diartikan ke dalam kondisi positif, negatif, dan netral dari suatu data teks yang telah dikelola keasliannya (Fitriyyah, Saffiadi, & Pratama, 2019). Maka untuk mengetahui sentimen pengguna twitter terhadap pembelajaran secara *online* atau daring, diperlukan langkah-langkah dari pengumpulan data *tweet* pengguna dari kyanan *platform* twitter, kemudian *preprocessing* data supaya bersih dari kata-kata yang tidak dibutuhkan, dan tahapan akhir melakukan klasifikasi dengan metode yang digunakan seperti *naïve bayes*, *k-nearest neighbor* dan lain-lainnya (Suryono & Taufiq Luthfi, 2021).

Penelitian sebelumnya yang dilakukan oleh Samsir, Ambiyar, Unung Verawardina, Firman Edi dan Ronal Watrianthos yang berjudul "Analisis Sentimen Pembelajaran dari Masa Pandemi Covid-19". Tujuan penelitian tersebut adalah untuk menganalisis opini rakyat Indonesia terhadap pembelajaran secara daring di masa pandemi covid-19 yang datanya diperoleh dari twitter dan dianalisis menggunakan metode *naïve bayes*. Studi kasus penarikan data *tweet* dengan kata kunci "#belajardarirumah", "pembelajaran daring", "belajar", "online", "kuliah", "daring" dan *tweet* tersebut berbahasa Indonesia, kemudian dilakukan *preprocessing* untuk seleksi data serta membersihkan data dari *noise* karena tidak memiliki struktur beraturan, setelah tahap *preprocessing* sudah selesai tahap selanjutnya ekstra fitur dengan melakukan pembobotan kata TF-IDF untuk mempermudah klasifikasi pada *naïve bayes*. Hasil yang diperoleh dari klasifikasi yaitu 30% sentimen positif, 69% sentimen negatif, dan 1% netral dari periode november 2020. Dengan tingginya sentimen negatif yang dihasilkan, memberikan kesimpulan bahwa opini masyarakat Indonesia tidak senang terhadap pembelajaran *online*. Penelitian ini hanya berfokus persentase hasil klasifikasi sentimen saja dan tidak ada penjelasan akurasi yang diperoleh dari metode *naïve bayes*, maka diperlukan hasil akurasi untuk mengetahui tingkat keberhasilan data yang di klasifikasi (Samsir, Ambiyar, Verawardina, Edi, & Watrianthos, 2021).

Penelitian sebelumnya yang dilakukan oleh Andri Oktavianto dan Satria Fadil Persada yang berjudul "Persepsi Publik tentang Pembelajaran Daring di Indonesia Menggunakan *Elk Stack* dan *Python* untuk Analisis Sentimen di Twitter". Tujuan penelitian ini untuk mengetahui pembelajaran daring telah diterima baik

oleh masyarakat khususnya mayoritas pelajar di Indonesia, dengan pengambilan data twitter dari bulan Maret – Juni 2019 sebanyak 133.477 *tweet* yang telah dikumpulkan menggunakan *elk stack*. Selanjutnya dilakukan *preprocessing* untuk membersihkan data supaya proses klasifikasi sentimen lebih akurat dan proses klasifikasi menggunakan *naïve bayes*, maka hasil yang diperoleh yaitu 98,3% positif dan 1,8% negatif untuk sentimennya, serta akurasi yang diperoleh dari *naïve bayes* 73%. Kurangnya dari penelitian ini yaitu, di *preprocessing data* tidak ada proses *case folding* untuk mengubah kata *capital* ke *lowercase*, *stopword* di *preprocessing data* membuang kata yang tidak bermakna, *stemming* di *preprocessing data* memperbaiki kata imbuhan yang muncul, dan normalisasi untuk mengubah kata yang tidak baku menjadi baku, dengan kemungkinan tersebut dapat meningkat akurasi yang diperoleh (Oktavianto & Persada, 2021).

Penelitian sebelumnya yang dilakukan oleh Amar P. Natusuwarna tentang analisis sentimen keberlanjutan pembelajaran daring dengan menggunakan *support vector machine* pada media sosial twitter. Tujuan penelitian ini adalah untuk analisis sentimen dengan menerapkan metode *support vector machine* serta membandingkan dua metode yaitu TF dan TF-IDF sehingga memperoleh hasil nilai evaluasi yang tinggi dari *k-fold* pada *k-fold cross validation*. Pengambilan data dalam penelitian ini yaitu diambil dari media sosial terkait komentar masyarakat terhadap mendikbud di saat pembelajaran secara daring keberlanjutan di rapat kerja Bersama Komisi X DPR pada tanggal 02 Juni 2020. Dataset yang telah dikumpulkan sebanyak 100 *tweet* positif dan 100 *tweet* negatif, serta dalam pengolahan data tersebut menggunakan *software rapidminer* untuk proses *support*



*vector machine* dan seleksi dengan TF dan TF-IDF. Hasil pengujian data dalam penelitian ini dibagi menjadi data training dan data testing yang menghasilkan nilai evaluasi dari *8-fold cross validation* dengan akurasi sebesar 86,00%, presisi sebesar 87,38%, dan recall sebesar 85,02%. Penelitian ini hanya berfokus pada perbandingan pada TF dan TF-IDF pada metode *support vector machine* dengan *k-fold cross validation* dan dibutuhkan perbandingan dengan metode lainnya sebagai perbedaan pada penelitian, sekian itu dataset yang digunakan hanya sebanyak 200 data sehingga kurang tepat jika menggunakan metode *cross validation* sebagai pengujian kualitas metode *support vector machine* di penelitian ini (Natasuwarna, 2020).

Penelitian sebelumnya yang dilakukan oleh Kezia Sekarayu Setyawati, Andreas Handoyo, Henry Noviamus Palit tentang analisis sentimen pembelajaran jarak jauh di Universitas Kristen Petra dengan metode *naïve bayes classifier*. Tujuan penelitian ini adalah untuk mengetahui tanggapan mahasiswa universitas Kristen Petra terhadap pembelajaran jarak jauh dengan cara pengumpulan data kuesioner yang dibuat pada setiap semester, serta pertanyaan yang diajukan kemudian diberi label topik dan label sentimen sehingga jawaban mahasiswa dikategorikan berdasarkan dosen, materi, fasilitas pendukung, dan media pembelajaran. Lalu dilakukan *preprocessing* data dengan membuang kalimat yang kembar, memperbaiki kata yang singkat, membuang kata tidak bermakna serta pengambilan kata dasar. Ketika proses selesai maka dilanjutkan dengan pemodelan klasifikasi *naïve bayes classifier* yang menghasilkan analisis sentimen di kenakan semester dari tahun 2019/2020 ke 2020/2021 yang bernilai positif dari segi materi

dan media pembelajaran, serta akurasi yang diperoleh yaitu 89% untuk klasifikasi sentimen dan 80% untuk klasifikasi topik. Kekurangan dari penelitian ini adalah tidak bisa melakukan proses *preprocessing* kata negasi seperti “tidak baik” diganti menjadi “buruk” serta tidak ada penjelasan besarnya tingkat persentase sentimen positif, negatif dan netral dari data kuesioner yang dikelola (Setyawati, Handojo, & Palit, 2021).

Penelitian sebelumnya yang dilakukan oleh Desi Musfiroh, Ulfa Khaira, Pradita Eko Prasetyo Utomo, Tri Suratno tentang analisis sentimen terhadap perkuliahan daring di Indonesia dari dataset twitter menggunakan *lexicon*. Tujuan penelitian ini untuk mengetahui opini masyarakat terhadap perkuliahan daring yang menjadi topik hangat di media sosial twitter yang bisa menjadi sumber informasi atau sebagai acuan evaluasi pelaksanaan perkuliahan daring di Indonesia. Penarikan data dalam penelitian ini dilakukan dengan menggunakan kata kunci “kuliah daring” melalui api twitter yang dibuat, sehingga data yang diperoleh kurang lebih sebanyak 6000 *tweet*, kemudian dilakukan *preprocessing* data supaya *tweet* tersebut bersih dari tanda baca, selain kata huruf, data kembar, dan memperbaiki kata ejaan, serta di tahap akhir dilakukan perhitungan polaritas dan klasifikasi sentimen dengan *inset lexicon* di setiap kata pada *tweet* satu per satu. *Inset lexicon* memiliki makna pada sejumlah kata bahasa Indonesia yang bersifat positif atau negatif yang disertai pembobotan kata. Dalam pengujian klasifikasi akurasi dengan teknik *cross-validation* yang menjadi data training dan data testing, maka diperoleh hasil dari 3 label yaitu 63.4% *tweet* negatif, 27.6% *tweet* positif, dan 8.9% *tweet* netral dengan perbandingan data training 80% dan data testing 20% yang menghasilkan akurasi

79.2%, presisi sebesar 72.9%, recall sebesar 62.8%, dan *f-measure* sebesar 67.4%. Kekurangan pada penelitian ini adalah minimnya kamus kata, menangani kata yang singkatan dari bahasa, serta membuang *emoticon* atau memanfaatkan *emoticon* sebagai keberhasilan pada sentimen dan membuat kalimat opini yang mengandung iklan atau berita (Musfiroh, Khaira, Eko, Utomo, & Suratno, 2021).

Penelitian sebelumnya yang dilakukan oleh Angelina Puput Giovani, Ardiansyah, Tuti Haryanti, Laela Kurniawati, Windu Gata tentang analisis sentimen aplikasi ruang guru dengan metode klasifikasi menggunakan twitter. Tujuan penelitian ini dilakukan adalah untuk membandingkan beberapa metode klasifikasi pada studi kasus di penelitian yang berkaitan pada komentar pengguna media sosial twitter terhadap aplikasi ruang guru. Ruang guru merupakan salah satu aplikasi *e-learning* berbasis pembelajaran digital yang saat ini digunakan sebagai media pembelajaran di masa pandemi covid-19. Untuk mengetahui sentimen pengguna twitter terhadap ruang guru dilakukan dengan beberapa langkah. Langkah pertama yaitu pengambilan data sebanyak 513 dari twitter, langkah kedua proses *preprocessing* data supaya *tweet* tersebut bersih dari tanda baca, selain kata huruf, data kembar, dan memperbaiki kata ejaan maka sentimen yang diperoleh yaitu sentimen positif sebanyak 338 *tweet* dan negatif sebanyak 175 *tweet*, langkah ketiga yaitu dengan penerapan metode klasifikasi *naïve bayes*, *support vector machine*, *k-nearest neighbor* dan *feature selection* dengan metode *particle swarm optimization* pada data yang telah dibersihkan serta menghasilkan akurasi *support vector machine* 76,93%, *naïve bayes* 65,40%, *k-nearest neighbor* 70,42% tanpa menggunakan *particle swarm optimization* dan jika menggunakan *particle swarm*

*optimization support vector machine* (ps) 78,55%, *naïve bayes* (ps) 67,32%, *k-nearest neighbor* (ps) 77,21%, dengan bantuan *particle swarm optimization* tingkat akurasi dari metode ini naik lebih baik. Penelitian ini memiliki kelemahan dalam proses *preprocessing* data yaitu tidak ada perbaikan kata yang tidak baku menjadi baku (Giovani, Ardiansyah, Haryanti, Kurniawati, & Gata, 2020).

*Naïve bayes* merupakan klasifikasi yang efektif untuk mendapatkan hasil klasifikasi yang tepat dan efisien dalam proses daya pikir untuk memanfaatkan input secara relatif cepat, dan *naïve bayes* menggunakan teori probabilitas untuk mencari kemungkinan peluang terbesar dari klasifikasi dapat dilihat dari frekuensi data training. Di proses klasifikasi yang dilakukan dengan cara memberi nilai ke setiap dokumen berdasarkan term kemunculan kata (Hemanto, Mustopa, & Kuntoro, 2020). Kekurangan dari *naïve bayes* jika probabilitas kondisinya nilai nol, maka diasumsikan bahwa setiap variabel independen membuat akurasi berkurang karena ada korelasi pada variabel dengan variabel yang lainnya. Sehingga dibutuhkan menggunakan satu probabilitas pada klasifikasi. Kekurangan dari *naïve bayes* yaitu jika nilai probabilitas itu nol maka hasil klasifikasi tersebut juga nol, keakuratannya klasifikasi tidak diukur dengan satu probabilitas serta dibutuhkan bukti-bukti lain untuk menyakinkan sebagai pengambilan keputusan (Andika, Azizah, & Respatiwan, 2019).

*K-nearest neighbor* adalah metode klasifikasi termasuk *supervised learning* dengan mempelajari data yang dikelompokan berdasarkan kedekatan jarak objek dengan objek yang lainnya. Untuk hitung jarak antara titik dengan menggunakan *euclydean distance* sebagai pengukuran jarak matriks dengan memberi titik ke

variabel  $K$  adalah jarak, nilai  $X$  adalah data training serta nilai  $Y$  adalah data testing. Kelemahan *k-nearest neighbor* yaitu kompleksitas kesamaan data yang memiliki besar apabila dimensinya tinggi serta data yang noise mudah dipengaruhi. Salah satu kelemahan dari *k-nearest neighbor* konvensional yang memiliki nilai  $K$  terbaik. Nilai  $K$  merupakan nilai yang digunakan untuk menentukan jumlah tetangga yang terdeteksi, maka penentuan nilai  $K$  sangat krusial sehingga begitu pengaruh dengan penentuan hasil akhir dari klasifikasi. Solusinya dengan menghitung validitas suatu data training untuk mendapatkan hasil optimal dari nilai  $K$ , selanjutnya diberikan nilai bobot pada suatu tetangga supaya mampu melakukan pembobotan lebih tinggi untuk mempengaruhi polaritas dari data testing dengan lebih kuat (Sari, 2020).

TF-IDF (*term frequency – inverse document frequency*) merupakan metode yang digunakan untuk perhitungan pembobotan pada setiap kata yang paling umum digunakan sebagai *information retrieval*. TF-IDF ini sangat efisien dan mudah untuk mendapatkan hasil akurasi. TF (*term frequency*) diasumsikan mempunyai rasio sesuai dengan kemunculan kata pada dokumen, serta dapat memperbaiki nilai recall, informasi retrieval dan tidak untuk nilai presisi. IDF (*inverse document frequency*) berfokus untuk melihat kemunculan *term* dari keseluruhan dokumen, maka *term frequency* yang jarang muncul pada seluruh dokumen dianggap sangat berharga dan diasumsikan banding terbalik dengan *term frequency* yang banyak kemunculan memiliki nilai berharga (Septian, Fahrudin, & Nugroho, 2019).

Berdasarkan penelitian sebelumnya, pada penelitian yang akan dilakukan untuk analisis sentimen pengguna twitter terhadap pembelajaran *online* di masa

pandemi covid-19 dengan metode *naïve bayes*, *k-nearest neighbor* dengan menggunakan pembobotan kata TF-IDF. Perbedaan dari penelitian untuk mengetahui perbedaan sentimen dari proses klasifikasi *naïve bayes* dan *k-nearest neighbor* serta pengujian pada pengolahan *preprocessing* untuk mengetahui hasil perbedaan yang peroleh dengan mengelompokan sentimen positif, negatif dan netral serta menghasilkan perbandingan nilai akurasi, presisi dan recall dari data tersebut.

### 1.2. Rumusan Masalah

Berdasarkan penjelasan di latar belakang masalah, maka rumusan masalah pada penelitian ini adalah sebagai berikut:

1. Bagaimana hasil sentimen terhadap *tweet* pengguna twitter menggunakan metode *naïve bayes*, *k-nearest neighbor* dengan menggunakan pembobotan TF-IDF pada pembelajaran masa pandemi covid-19?
2. Berapa tingkat akurasi yang dihasilkan menggunakan metode *naïve bayes*, *k-nearest neighbor* dengan menggunakan pembobotan TF-IDF pada pembelajaran masa pandemi covid-19?

### 1.3. Batasan Masalah

Batasan masalah dalam penelitian ini adalah sebagai berikut:

1. *Tweet* yang di analisis sentimen hanya kalimat *tweet* berbahasa Indonesia.
2. Penarikan data dari tanggal 16 Maret 2020 sampai tanggal 10 April 2021.
3. Keyword digunakan hanya berkaitan dengan pembelajaran *online* di masa pandemi covid-19.
4. Jumlah data di peroleh sekitar 4000 *tweet* yang belum di proses.

5. Metode yang digunakan hanya *naïve bayes*, *k-nearest neighbor* dengan menggunakan pembobotan TF-IDF.
6. Data akan dikelola menggunakan bahasa pemrograman *python* pada *anaconda jupyter notebook*.
7. *Tools* yang digunakan *library pandas*, *sntwitter*, *numpy*, *sastrawi* dan *preprocessor*
8. Pengelompokan emosional pengguna positif, negatif dan netral terhadap Pembelajaran masa pandemi covid-19.
9. Data *tweet* tidak mengandung multimedia hanya berupa teks.

#### **1.4. Tujuan Penelitian**

Tujuan dari penelitian ini antara lain:

1. Untuk mengetahui hasil sentimen pengguna media sosial twitter terhadap pembelajaran *online* di masa pandemi covid-19 dengan metode *naïve bayes*, *k-nearest neighbor* dengan menggunakan pembobotan TF-IDF.
2. Untuk mengetahui perbedaan tingkat akurasi dari metode *naïve bayes*, *k-nearest neighbor* terhadap pembelajaran *online* di masa pandemi covid-19

#### **1.5. Manfaat Penelitian**

Manfaat dari penelitian yang dilakukan adalah sebagai berikut.

1. Mampu mengetahui emosional pengguna twitter terhadap kebijakan pemerintah tentang pembelajaran *online* di masa pandemi covid-19.
2. Dapat menjadi sebuah keputusan strategi pembelajaran *online* di masa depan.
3. Menjadi referensi terhadap penelitian lain di masa depan yang berkaitan dengan analisis sentimen pembelajaran *online* masa pandemi covid-19.

## BAB II

### TINJAUAN PUSTAKA

#### 2.1. Tinjauan Pustaka

Pada penelitian yang dilakukan oleh Danielle Berkovic (Berkovic, Ackerman, Briggs, & Ayton, 2020) analisis sentimen *tweet* oleh penderita *arthritis* selama pandemi covid-19. Bukti yang muncul menunjukkan bahwa orang dengan penyakit radang sendi (*arthritis*) melaporkan peningkatan rasa sakit fisik dan tekanan psikologi selama pandemi covid-19. Pengguna twitter sehari melonjak sebesar 23% selama masa pandemi, dikarekakan tidak bisa melakukan aktivitas dengan normal hanya bisa berdiam diri di rumah saja. Pada penelitian ini adalah untuk mengidentifikasi sentimen emosional yang positif atau negatif dari pengguna twitter di masa pandemi covid-19. Dalam pengumpulan data sentimen menjadi 2 topik yaitu covid-19 dan radang sendi (*arthritis*) dengan memiliki hashtag yang berbeda-beda. Hashtag yang digunakan “#coronavirus”, “#covid-19”, “#isolation”, dan “#sosiakdistancing” serta *arthritis* hashtag “#arthritis”, “#spoonie”, “#rheumatologist”, dan “#rheumatology”. Hasil total 149 *tweet* telah dianalisis mayoritas *tweet* adalah perempuan dari Amerika Serikat. Secara detailnya jenis kelamin laki-laki 21%, perempuan 70% serta tidak diketahui 9% kemudian negara Amerika Serikat 45,5% Inggris raya dan Irlandia Utara 26%, Kanada 12%, tidak diketahui 13%. Kesimpulan *tweet* oleh pengguna dengan penyakit radang sendi lebih banyak khawatir selama masa pandemi covid-19.



Penelitian yang akan lakukan adalah mengumpulkan data yang akan sebagai proses klasifikasi sentimen pada pembelajaran *online* di masa pandemi covid-19 dengan hashtag pada *tweet* yang pengguna lakukan serta menerapkan metode *naïve bayes* dan *k-nearest neighbor* sebagai hasil sentimen positif, negatif maupun netral.

Pada penelitian selanjutnya yang dilakukan oleh Manguri, Ramadhan, Mohammed Amin.(H. Manguri, N. Ramadhan, & R. Mohammed Amin, 2020). Analisis sentimen tentang wabah covid-19 di seluruh dunia. Menurut WHO pada penelitian yang telah dibuat ini, lebih dari 20 juta orang telah tertular penyakit covid-19 hingga 157 ribu orang telah meninggal karena wabah ini. Banyaknya berita mengeluhkan penyakit covid-19 membuat masyarakat resah dan ketakutan hingga tidak bisa melakukan aktivitas normalnya. Banyak *tweet* opini masyarakat tentang wabah ini lalu dilakukan klasifikasi data *tweet* telah kumpulan dari tanggal 9 april sampai 15 april 2020 klasifikasi dengan hashtag "covid-19" dan "coronavirus". Pada proses pengolahan klasifikasi menggunakan metode *naïve bayes classifier* untuk mengetahui sentimen masyarakat yang *objective*, *subjective* dan *neutral*. Hasil pengujian data yang terkumpul 530232 *tweet* menghasilkan akurasi sentimen yang *objective* 60%, *subjective* 26% serta *neutral* 14%. Kesimpulan reaksi masyarakat berbeda-beda dari hari ke hari dari posting perasaan mereka di media sosial khususnya twitter.

Penelitian yang akan lakukan adalah untuk mengetahui hasil akurasi dari metode *naïve bayes* dan *k-nearest neighbor*, apakah hasil sentimen peroleh baik atau tidak dalam melakukan klasifikasi, seperti *review* yang telah dijelaskan dengan kumpulan data *tweet* yang begitu banyak.

Penelitian selanjutnya yang dilakukan oleh Jimmy Samuel (Samuel, Ali, Rahman, Esawi, & Samuel, 2020) opini publik covid-19 dari klasifikasi *tweet* dengan *machine learning*. Pada krisis pandemi covid-19 menimbulkan ketakutan massal dan fenomena panik, yang dipicu oleh informasi yang tidak akurat tentang covid-19. Sehingga pada penelitian ini mengidentifikasi sentimen publik yang terkait dengan pandemi khususnya dari *tweet* pengguna twitter dengan ciri khas hashtag yang bermacam-macam seperti “#coronavirus”, “#realdonaldtrump”, dan sebagainya untuk melakukan klasifikasi sentimen perlunya metode *naïve bayes*. Hasil uji klasifikasi ke akurasi menggunakan metode *naïve bayes* 74% dan klasifikasi sentimen yang positif 36 *tweet* serta negatif 68 *tweet*. Kesimpulan penyebaran virus covid-19 akan selalu membuat warga Amerika Serikat cemas dan panik sampai masa pandemi ini selesai.

Penelitian yang akan lakukan adalah untuk mengetahui beberapa *tweet* yang mengandung sentimen positif, negatif ataupun netral pada suatu kondisi publik dengan metode *naïve bayes* dan *k-nearest neighbor*.

Penelitian selanjutnya oleh Muhammad Dwison Alizah (Alizah, Nugroho, Radiyah, & Gata, 2020) sentimen analisis terkait *lockdown* pandemi covid-19. Covid-19 telah ditetapkan sebagai pandemi oleh WHO yang berdampak sangat besar. Salah satu upaya untuk pencegahan penyebaran virus covid-19 agar tidak meluas. Dibutuhkan kebijakan dari pemerintah daerah untuk melakukan *lockdown* supaya penyebaran virus corona ini tidak meluas. Akan tapi pada kenyataannya ini bukan solusi terbaik untuk semua pihak. Maka pada penelitian ini analisis prediksi sentiment terhadap kebijakan tentang *lockdown* dari *tweet* netizen akan menjadi

data klasifikasi kemudian dikelola dengan menggunakan *naïve bayes* dan *support vector machine* sebagai metodenya. Untuk melakukan proses mengklasifikasikan dibutuhkan tahapan *preprocessing* dari *tokenization*, *stopword*, normalisasi dan *stemming* dari data *tweet* sebanyak 15.494 serta dilakukan pembagian data latih sebanyak 10.845 dan 4649 sebagai data uji. Hasil pengujian klasifikasi sentimen menggunakan metode *naïve bayes* yaitu bernilai positif 718 *tweet*, negatif 1397 serta netral 1655 dan metode *support vector machine* positif 886, negatif 1357, serta netral 1835. *Naïve bayes* mampu mencapai 81% akurasi, sedangkan *support vector machine* 87%.

Penelitian yang akan peneliti lakukan adalah tahapan *preprocessing* data *tweet* yang di perlu oleh dari *crawling tweet* dan dilakukan *preprocessing* yaitu *case folding*, *tokenization*, *stopword*, *stemming*, *lexicon*, dan *normalization*. Ketika *preprocessing* data selesai, maka selanjutnya melakukan metode sentimen dengan *naïve bayes*, *k-nearest neighbor* serta pembobotan TF-IDF untuk mendapatkan hasil sentimen dan akurasi yang benar.

Pada penelitian selanjutnya oleh Muhammad Syarifuddin (Syarifuddin, 2020) analisis sentimen opini publik mengenai covid-19 pada twitter pada aspirasi, opini, dan kritik masyarakat sangat sulit disampaikan secara langsung karena suatu hal. Tapi media sosial tidak memungkinkan untuk sebagai tempat ungkapan opini terhadap wabah ini. Maka dari penelitian akan analisis sentimen opini masyarakat menyelaraskan dan memberi pandangan baru mengenai isu tentang covid-19. Untuk mengetahui kecenderungan opini bersifat positif atau negatif, dapat dilihat dari banyak opini netizen. Untuk melakukan data klasifikasi *tweet* dari opini netizen

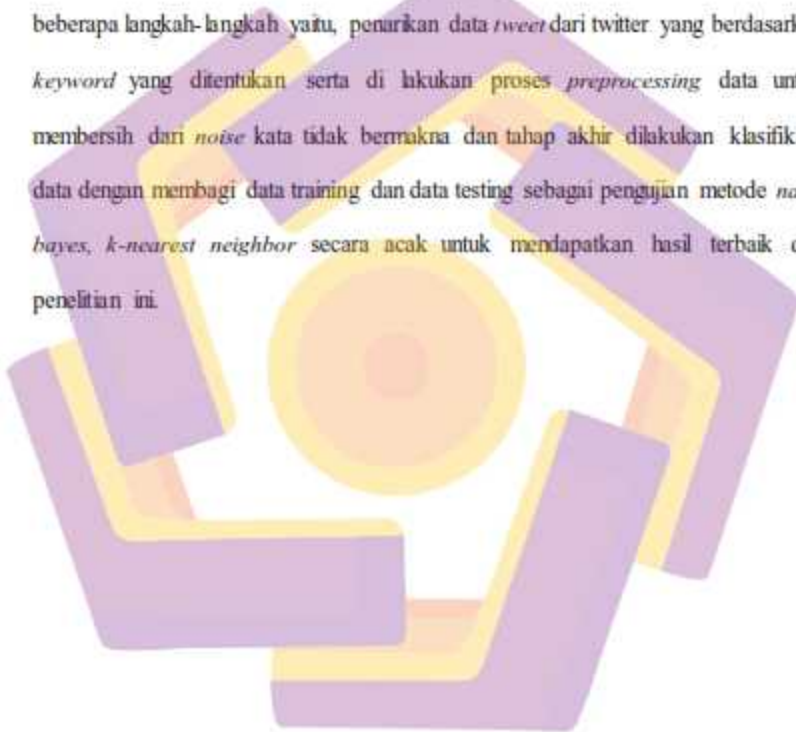
dibutuhkan *preprocessing* dan metode *naïve bayes* dan *k-nearest neighbor*. Dari hasil pengujian metode *naïve bayes* sentimen yaitu bernilai positif 66.40% dan negatif 58.94% sedangkan metode *k-nearest neighbor* sentiment yang positif 61.83% dan negatif 53.02%. Kesimpulan dari kemunculan wabah covid-19 penyakit membuat resah dan takut terhadap masyarakat dari *tweet* pengguna media sosial twitter.

Penelitian yang akan lakukan adalah untuk mengetahui apa yang dirasakan oleh masyarakat khusus pengguna twitter terhadap pembelajaran *online* di masa pandemi covid-19. Dari hasil analisis sentimen lakukan oleh peneliti dapat dijadikan sebuah evaluasi untuk Menteri Pendidikan Dan Kebudayaan (MENDIKBUD) sebagai tindakan perbaikan kebijakan pada masyarakat terutama para siswa dan mahasiswa yang mengalaminya.

Pada penelitian selanjutnya oleh Lazuardi (Lazuardi, Munandar, Harsiti, Mutaqin, & Hays, 2020) analisis sentimen opini publik tentang kesejahteraan pendidik honorer. Perkembangan internet di Indonesia sangat pesat dari segi informasi dari pengguna media sosial. Salah satu media sosial yang paling populer adalah twitter. Bahkan pemerintah saat ini tidak luput dari komentar netizen terhadap kinerjanya. Pada penelitian ini opini netizen tentang seorang guru honorer di Sekolah Dasar Kabupaten Pandeglang yang tinggal sebelah toilet sekolahnya. Pada kasus ini dapat perhatian oleh masyarakat di media sosial, terutama *tweet* tentang sentimen terhadap pelayanan publik khususnya dalam meningkatkan kesejahteraan pendidik. Untuk melakukan klasifikasi sentimen opini masyarakat menggunakan metode *naïve bayes* dan dikelompokkan menjadi dua yaitu data uji

10% dan data latih 90%. Hasil dari pengujian menunjukkan akurasi analisis sentimen terhadap guru honorer tinggal sebelah toilet sebanyak 88,24% dari data *tweet* yang positif 19 postingan, negatif 16 postingan serta netral 16.

Penelitian yang akan dilakukan adalah untuk mengelompokkan sentimen masyarakat pengguna twitter terhadap pembelajaran secara *online* yang dibagi beberapa langkah-langkah yaitu, penarikan data *tweet* dari twitter yang berdasarkan *keyword* yang ditentukan serta dilakukan proses *preprocessing* data untuk membersihkan dari *noise* kata tidak bermakna dan tahap akhir dilakukan klasifikasi data dengan membagi data training dan data testing sebagai pengujian metode *naïve bayes*, *k-nearest neighbor* secara acak untuk mendapatkan hasil terbaik dari penelitian ini.



## 2.2. Keaslian Penelitian

Tabel 2. 1. Matrks Literatur Review Dan Posisi Penelitian Analisis Sentimen Pengguna Twitter Terhadap Pembelajaran *Online* Di Masa Pandemi Covid-19 Menggunakan Metode Naive Bayes, K - Nearest Neighbor Dan Pembobotan TF-IDF

No	Judul	Peneliti, media publikasi, dan tahun	Tujuan penelitian	Kesimpulan	Saran atau kelemahan	Perbandingan
1	SENTIMEN ANALISIS PUBLIK TERHADAP JOKO WIDODO TERHADAP WABAH COVID-19 MENGGUNAKAN METODE MACHINE LEARNING	Sisferi Hikmawan, Amsal Pardamean, Siti Nuri Khasannah, Jurnal Kajian Ilmiah, 2020(Hikmawan, Pardamean, & Khasannah, 2020)	Melakukan klasifikasi terhadap sentimen pada kata kunci "covid" dan "jokowi", apakah ini positif, negatif ataupun netral dari data twitter yang menggunakan k-nn, nb, svm	Dari hasil pengujian menunjukkan, akurasi dari <i>naive bayes</i> sebesar 84,58%, <i>support vector machine</i> sebesar 92,93%, dan <i>k-nn</i> sebesar 83,70%	Ke masa depannya diperlukan dataset lebih banyak untuk memaksimalkan hasil <i>processing</i> pada bahasa Indonesia tidak baku	Penelitian ini mengklasifikasikan sentimen twitter terhadap wabah covid-19 dan jokowi. Penelitian yang diajukan yaitu analisis sentimen pembelajaran <i>online</i> masa pandemi covid-19 dengan metode nb dan k-nn
2	TWITTER SENTIMENT ANALYSIS ON WORLDWIDE COVID-19 OUTBREAKS	Kamran H. Manguri, Rebaz N. Rumadhan, Pshko R. Mohammed Amin. Kurdistan Journal Of Applied Research, 2020(H. Manguri et al., 2020)	Mengetahui bagaimana individu di negara tersebut bereaksi terhadap wabah penyakit covid-19 dan informasi yang relevan yaitu data <i>tweet</i>	Kata kunci corona dan covid-19 untuk polaritas secara signifikan lebih tinggi dari 50 persen dan objektif 64 persen	Menemukan sumber informasi tidak hanya di twitter saja. Tetapi bisa juga di media sosial seperti <i>facebook</i> , <i>instagram</i> , dan dapat diadopsi sebagai model sentiment emosi kasus serupa	Penelitian ini mengidentifikasi emosional terhadap wabah covid 19. Penelitian yang diajukan yaitu analisis terhadap emosional dalam pembelajaran <i>online</i> masa pandemi covid-19

Tabel 2. 1. Matriks Literatur Review Dan Posisi Penelitian (lanjutan)

No	Judul	Peneliti, media publikasi, dan tahun	Tujuan penelitian	Kesimpulan	Saran atau kelemahan	Perbandingan
3	"CORONAVIRUS COULD KILL ME #STAYHOMESAVELIVES": A CONTENT AND SENTIMENT ANALYSIS OF TWEETS BY PEOPLE WITH ARTHRITIS DURING COVID-19.	Danielle Berkovic, <i>Journal Of Medical Internet Research</i> , 2020(Berkovic et al., 2020)	Mengidentifikasi topik yang penting bagi individu dengan radang sendi selama pandemi covid-19, dan untuk mengeksplorasi konteks emosional dari <i>tweet</i> oleh penderita radang sendi selama fase awal pandemi	<i>Tweet</i> oleh pengguna menyoroti mengkhawatirkan selama masa pandemi covid-19 mencakup gejala meningkat dalam konteks informasi	Hasil penelitian ini analisis <i>tweet</i> mayoritas perempuan dan berasal dari Amerika Serikat menunjukkan kecemasan seputar obat dan peningkatan gejala fisik	Penelitian ini analisis konten dan sentimen <i>tweet</i> hashtag #stayhomesavelives. Penelitian yang diajukan mencari kata hashtag pembelajaran daring / pembelajaran jarak jauh
4	SENTIMENT ANALYSIS OF PUBLIC OPINIONS ON THE WELFARE OF HONORARY EDUCATORS USING NAIVE BAYES	D R Lazuardi, T A Munandar, H Harsiti, Z Mutaqin, R N Hays, <i>Top Conference Series: Materials Science And Engineering</i> , 2020(Lazuardi et al., 2020)	Mengetahui sentimen pengguna terhadap pemerintah kabupaten pandeglang berdasarkan kasus pemberitaan guru honorer di Kabupaten Pandeglang	Analisis sentimen menunjukkan terdapat 16 posting twitter yang menunjukkan netral, 19 posting menunjukkan positif dan 16 posting sisanya menunjukkan kelas negatif	Tingkat akurasi yang dihasilkan dari model analisis sentimen yang terbentuk adalah 88,24%	Penelitian ini dalam pengumpulan data hanya 51 <i>tweet</i> dari netizen. Penelitian yang diajukan akan menggunakan data <i>tweet</i> lebih banyak

Tabel 2. 1. Matriks Literatur Review Dan Posisi Penelitian (lanjutan)

No	Judul	Peneliti, media publikasi, dan tahun	Tujuan penelitian	Kesimpulan	Saran atau kelemahan	Perbandingan
5	ANALISIS SENTIMEN APLIKASI RUANG GURU DI TWITTER MENGGUNAKAN ALGORITMA KLASIFIKASI	Angelma Puput Giovani, Ardiansyah, Tuti Haryanti, Laela Kurniawati, Windugata, Jurnal Teknoinfo.2020 (Giovani et al., 2020)	Membahas langkah-langkah melakukan processing analisis sentimen pada aplikasi ruang guru apakah positif atau negatif	Hasil dan pengujian maka hasil svm(pso) mempunyai akurasi dan performa tinggi dari pada nb, nb (pso), svm, sym (pso), knn dan k-nn (pso).	Dalam penelitian ini sumber data yang digunakan hanya dari satu sumber media sosial, yaitu twitter	Penelitian ini membandingkan metode nb, svm-k-nn. Penelitian diajukan lebih fokus perbandingan nb dan k-nn dengan topik dan dataset yang berbeda
6	ANALISIS SENTIMEN UJARAN KEBENCIAN PEMILIHAN PRESIDEN 2019 MENGGUNAKAN ALGORITMA NAIVE BAYES	Mufia Chalida, M. Didik R. Wahyudi, Jnanaloka,2020(Chalida, 2019)	Dalam pengumpulan data menggunakan metode multistage random dari analisis kecenderungan opini pengguna dengan <i>tweet</i> di twitter mengenai ujaran kebencian terhadap pemilihan presiden 2019. Pengambilan data dari kota-kota besar di Indonesia	Dari data telah didapatkan sekitaran 5055 <i>tweet</i> pengguna dengan hashtag #pilpres 2019. Kemudian uji analisis sentimen ujaran kebenciannya, menghasilkan klasifikasi sentimen tidak relevan sebanyak 11,3%, negatif sebanyak 35,4%, netral sebanyak 26,7%, sentimen positif sebanyak 26,6%	Perlu dilakukan pengamatan lebih dalam apakah postingan-postingan tersebut dilakukan oleh akun-akun tertentu untuk tujuan kampanye hitam menjatuhkan masing-masing kandidat	Penelitian ini terhadap proses perolehan data sampai 5055 data dengan hashtag #pilpres 2019. Pada penelitian yang diajukan akan menggunakan banyak hashtag untuk peroleh dataset lebih banyak



Tabel 2. 1. Matriks Literatur Review Dan Posisi Penelitian (lanjutan)

No	Judul	Peneliti, media publikasi, dan tahun	Tujuan penelitian	Kesimpulan	Saran atau kelemahan	Perbandingan
7	PENGARUH METODE PENYEIMBANGAN KELAS TERHADAP TINGKAT AKURASI ANALISIS SENTIMEN PADA TWEETS BERBAHASA INDONESIA.	Ivan Nathaniel Husada, Hapnes Toba, Jurnal Teknik Informatika Dan Sistem Informasi, 2020 (Husada & Toba, 2020)	Analisis sentimen berbahasa Indonesia dengan kelas yang tidak seimbang pada dataset yang diambil dari media sosial twitter.	Hasil yang didapat adalah hampir semua model analisis sentimen dengan arsitektur neural network memiliki skor yang lebih baik dibandingkan model dengan metode klasifikasi <i>naive bayes</i> .	Penelitian yang lebih lanjut akan dapat menggunakan dataset berbahasa Indonesia dengan data yang lebih banyak dengan label sentimen yang lebih sesuai dengan agar dapat meminimalisir dari manual labeling.	Penelitian ini perbandingan metode <i>naive bayes</i> dan nn. Pada penelitian diajukan lebih fokus metode <i>naive bayes</i> , dan k-nn untuk dapatkan akurasi yang lebih baik
8	COVID-19 PUBLIC SENTIMENT INSIGHTS AND MACHINE LEARNING FOR TWEETS CLASSIFICATION	Jim Samuel, G. G. Md. Nawaz Ali, Md. Mokhlesur Rahman, Ek Esawi, Yana Samuel, Information, 2020 (Samuel et al., 2020)	Mengidentifikasi sentimen publik yang terkait dengan pandemi menggunakan <i>tweet</i> khusus coronavirus di Amerika Serikat metode klasifikasi, dalam konteks analitik tekstual	Dengan metode <i>naive bayes</i> . Akurasi klasifikasi yang kuat sebesar 91% untuk <i>tweet</i> pendek, dengan metode <i>naive bayes</i> . Serta juga mengamati bahwa metode klasifikasi logistic regression memberikan akurasi sebesar 74%	Dengan kedua metode tersebut menunjukkan kinerja lebih lemah untuk <i>tweet</i> yang lebih panjang	Penelitian ini mengidentifikasi sentimen <i>tweet</i> khusus coronavirus. Penelitian diajukan lebih memperbanyak data <i>tweet</i>

### 2.3. Landasan Teori

#### 2.3.1 Analisis Sentimen

Analisis sentimen merupakan ungkapan opini kejadian yang terjadi saat ini. Dalam bidang penelitian untuk melakukan sentimen kepada penilaian, perbaikan dan sikap emosional seseorang terhadap kondisi seperti pelayanan produk, organisasi, permasalahan, kejadian fenomenal (Pawar, Jawale, & Kyatanavar, 2016). Analisis sentimen dicoba buat memandang komentar terhadap suatu permasalahan yang ada atau bisa juga digunakan buat identifikasi kecenderungan perihal yang lagi jadi topik pembicaraan. Analisis sentimen dalam riset ini merupakan proses pengelompokkan *tweet* pengguna di twitter tentang pendidikan Dalam Jaring (DARING), Pendidikan Jarak Jauh (PJJ), serta hashtag tertentu mengetahui emosi seseorang yang memiliki dampak positif, negatif maupun netral.

Pengaruh serta manfaat dari analisis sentimen menimbulkan riset penelitian analisis sentimen sangat tumbuh. Di Amerika setidaknya terdapat 40- 60 industri besar semacam *microsoft, google, hewlett-packard, sap*, serta *sas* memfokuskan pada pelayanan analisis sentimen (Olson & Delen, 2008) manfaat analisis sentimen dalam produk bisa mengenali komentar dari netizen terhadap suatu produk dengan bagaikan perlengkapan bantu untuk mengenali reaksi terhadap suatu produk. Kemudian bisa melaksanakan keputusan dalam melakukan strategis berikutnya.

Analisis sentimen adalah klasifikasi realitas tidak gampang dalam proses klasifikasi biasa, sebab penggunaan bahasa terdapat kata yang ambigu dalam kata itu sendiri. Pada analisis sentimen memiliki tingkat analisis berbeda yaitu:

1. Tingkat kalimat

Tingkat kalimat menganalisis suatu kalimat serta memastikan masing-masing kalimat bernilai sentimen positif, negatif serta netral. Umumnya netral tidak terdapat opini (Wenando, Hayami, & Anggrawan, 2020).

2. Tingkat dokumen

Tingkat dokumen akan melakukan tahapan mengklasifikasikan tersebut, apakah memiliki sentimen positif maupun negatif. Dalam menganalisis dokumen hanya berisi opini tentang satu entitas dan tidak bisa diterapkan ke dokumen yang memiliki lebih dari satu entitas (Wenando et al., 2020).

3. Tingkat entitas serta aspek

Tingkat aspek memiliki dua tahapan yang belum memastikan apa sebenarnya orang diminati dan tidak diminati, melainkan langsung pada gagasan opini komentar dari sentimen positif serta negatif (Wenando et al., 2020).

### 2.3.2 Emosi

Emosi adalah reaksi suatu kejadian mampu pada seseorang. Makna ucapan tersebut menunjukkan terhadap kecenderungan perihai yang bersifat mutlak dalam diri makhluk hidup. (Goleman, 2015) berkata kalau emosi menunjukkan pada suatu kondisi serta bentuk yang unik pada suatu kondisi psikologis, biologis serta rasa kecenderungan untuk berperan. Emosi adalah respon terhadap suatu gairah dari

kondisi luar dan dalam diri orang, seperti emosi senang, cinta, bahagia mendesak pergantian kondisi perasaan seseorang dari segi fisiologi kelihatan tertawa. Berbanding terbalik pada kondisi seseorang yang sedih mendesak berperilaku marah, kesal, menangis.

Twitter kerap kali digunakan buat mengatakan emosi menimpa suatu perihal yang positif maupun negatif. Pengenalan emosi pada *tweet* bisa dilakukan menggunakan analisis sentimen bisa yang dimanfaatkan buat menggali opini pengguna tentang sesuatu topik.

Emosi yang dimiliki manusia dikategorikan jadi 5 emosi dasar cinta, bahagia, marah, khawatir, serta pilu. Emosi cinta bahagia merupakan emosi positif. Emosi kesal, khawatir, serta pilu merupakan emosi negatif terhadap penelitian pada kosakata emosi di Indonesia. Menghasilkan 2 kategori yaitu emosi positif dan negatif. Pengelompokan terhadap kosakata emosi nampak pada list berikut:

1. Positif

Bahagia, suka, terpesona, tertarik, cinta, peduli, semangat, tegas dan lain - lainnya.

2. Negatif

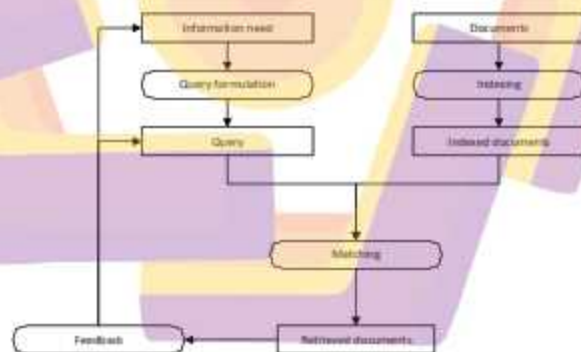
Cemburu, curiga, histeris, iri, dengki, kecewa, tersinggung, muak, benci, emosi, kesal, dendam, dan lain - lainnya.

### 2.3.3 *Information retrieval system*

*Information retrieval system* ataupun sistem pemerolehan data merupakan suatu program perangkat lunak yang menaruh serta mengendalikan data pada dokumen sering kali berupa bacaan serta multimedia. Pada sistem ini tidak ini akan

secara akurat mengembalikan data ataupun jawaban mengenai pertanyaan akan tetapi memberitahukan posisi dari dokumen, yang kemungkinan mempunyai data yang diharapkan sebagai dokumen yang relevan. Suatu sistem pemerolehan data yang baik akan dianggap dokumen relevan. Tetapi sistem tidak akan melakukan pencarian yang relevansi terkait subjektif opini pengguna. Dalam pelaksanaan pengguna dapat mempunyai suatu *query* persis dengan sistem serta memperhitungkan dokumen relevansi yang didapatkan secara berbeda seperti menyukainya serta tidak menyukainya (Hiemstra, 2009).

Terdapat 3 proses yang wajib didukung oleh sistem pemerolehan data, awal menggambarkan isi dari dokumen, menggambarkan data inginkan oleh pengguna, serta menggambarkan perbandingan dari kedua tersebut.



Gambar 2. 1. Sistem pemodekn pengolahan data

Dalam gambar 2.1 menggambarkan alur dokumen disebut dengan proses pendeskripsian yang hasilkan dalam bentuk dokumen. Sistem pemerolehan data memakai penuh suatu metode yang mengidentifikasi per kata dalam bacaan kapital

atau lainnya dan mengubah dalam bentuk bacaan huruf kecil. Dalam proses pendeskripsian dapat menyamakan simpanan dokumen ke dalam sistemnya. Namun kadang kala di dokumen tersebut cuma disimpan sebagian contoh judul dan abstrak saja, dan ditambahkan ke data posisi dokumen sesungguhnya.

Pengguna bukan hanya mencari sesuatu hiburan semata tetapi pengguna memiliki kebutuhan terhadap informasi data. Proses penggambaran akan kebutuhan informasi data kerap sebagai *formulation process query* yang menghasilkan suatu yang lebih luas untuk mendapatkan tampilan diskusi dari sistem ke pengguna lebih mengerti data yang dibutuhkan, yang bisa dikata umpam balik informasi.

Dari proses dari *query* serta penggambaran dokumen dijelaskan bagian *process matching*. *Matching process* ataupun proses penyesuaian umumnya menciptakan catatan dokumen yang sistematis. Akuisisi yang sistematis diinginkan dapat menaruh relevansi pada dokumen catatan paling atas untuk menyingkat waktu baca pengguna.

Berikut proses-proses information retrieval (Worth, 2010):

#### 1. Tokenization

Suatu dokumen unit mempunyai lapisan perkata khusus dari tugas tokenisasi untuk memecah lapisan kata jadi terpisah yang disampaikan suatu token. Berikut contohnya: Input: magister, amikom, yogyakarta, angkatan tahun 2020; Output: |magister| |amikom| |yogyakarta| |angkatan| |tahun| |2020| Setelah melakukan tokenisasi tanda baca titik(.), koma (,), seru(!), pagar atau hashtag (#), dan add (@) dan symbol lainnya. Terdapat case -

folder, 48 yaitu mengecilkan wujud huruf dari yang kapital seperti kata "AMIKOM" jadi "amikom"

## 2. Stopword

*Stopword* merupakan *preprocessing text mining* pada saat melakukan eliminasi suatu kata. Kata-kata telah di eliminasi merupakan kata yang sangat kerap timbul dalam sesuatu dokumen yang mempunyai penilaian sedikit serta membantu memilah dokumen diperlukan tahapan dalam menghilangkan kata sambungan maupun perkataan yang sering muncul tapi tidak bermakna. Yaitu ini, itu, yaitu, maka, walaupun, dan lain-lainnya.

## 3. Normalization

Merupakan suatu proses persamaan dalam na kata eja yang mempunyai arti yang sama, contoh kata yang dicari 'india' ketika mengharapkan akan memperoleh 'I N. D. I. A' (Worth, 2010). Dalam bahasa Indonesia banyak ditemukan kata tidak jelas seperti 'slama' dalam artian 'selama'. Sehingga dalam menggunakan kata ejaan yang benar harus pusing pada Kamus Besar Bahasa Indonesia (KBBI).

## 4. Stemming

Tujuan dari *stemming* merupakan buat kurangi infleksi (pembuatan kata baru) akibat akumulasi sesuatu imbuhan akhiran serta sisipan sehingga merendahkan sesuatu kata ke bentuk kata dasarnya (Worth, 2010). Stemming merupakan langkah dari *preprocessing* bacaan yang digunakan sebagai menciptakan kata dasar tanpa ada imbuhan kata apapun. Seperti kata

49 “mbingungkan” menjadi “bingung” dan “membuat” menjadi “buat” dan lain sebagainya.

#### 2.3.4 Twitter Dan Twitter Apl

##### 1. Twitter

*Twitter* dibangun oleh Jack Dorsey pada tahun 2006 di San Francisco dan pertama kali diluncurkan dunia maya pada bulan Juli 2006 dengan situs bernama [twitter.com](https://twitter.com) sampai saat ini masih dipakai oleh pengguna media sosial. *Twitter* merupakan *mikroblog* dengan setiap pengguna mampu mengirim serta membaca pesan *tweet* yang berupa teks, untuk pada teks bisa kurang dari 200 karakter yang muncul pada beranda pengguna *twitter*.

Menurut penelitian (Fammy, Bambang, & Hapsari, 2020) *twitter* salah satu menjadikan media sosial populer saat ini tidak kalah dengan penggunaanya dari media sosial seperti *instagram*, *facebook* dan lain-lainnya. Media sosial *twitter* juga tengah digandrungi kembali setelah sebelumnya sempat populer di tahun 2011. Kali ini media sosial *twitter* kembali digunakan oleh pengguna khususnya di kalangan remaja. Dengan adanya *twitter* sebuah kejadian dapat menjadi fenomena yang viral apabila dapat banyak perhatian dari pengguna media sosial *twitter*. *Twitter* sering kali digunakan sebagai ungkapan emosi pribadi mengenai suatu hal yang positif maupun negatif. Emosi seseorang memiliki peran dalam dirinya yaitu senang, cinta, duka, sedih, marah (Saputri, Mahendra, & Adriani, 2019).

Pada *twitter* untuk membalas *tweet* pengguna lainnya harus menggunakan tanda add (@) sebagai memanggil nama pengguna lainnya atau membalas dari



*tweet* sebelumnya. Tak hanya tanda add (@) sebagai fitur balasan *tweet* seseorang di twitter masih banyak fitur-fitur yang harus dikenalkan yaitu:

a. Beranda

Beranda biasa digunakan untuk melihat *tweet* seseorang yang kirim oleh following atau yang telah diikuti.

b. Profil

Profil ini yang akan dilihat oleh seluruh pengguna untuk mengetahui biodata pribadi dan *tweet* yang telah di buat.

c. Pengikut

Pengikut atau *followers* adalah pengguna lainnya yang telah membuat kita menjadi kawan di twiternya. Menjadi pengikut tersebut bisa membaca *tweet*, membaca profil, dan juga komentar pada *tweet* yang telah dibuat.

d. mengikuti

Mengikuti atau *following* merupakan seseorang yang telah mengikuti pada pengguna twitter lainnya agar *tweet* telah dikirim oleh pengguna lainnya yang ikuti masuk dalam beranda.

e. Mention

Mention merupakan perbincangan antar pengguna yang di balas oleh pengguna lainnya untuk mengajak berbicara.

f. Favorit

*Tweet* dari pengguna lainnya yang bagus atau keren akan ditandai supaya ketika kembali ke halaman sebelumnya tidak hilang.

g. Pesan langsung

Pada pesan langsung digunakan sebagai pengiriman pesan seperti *chat*, sms antara pengguna lainnya.

h. *Hashtag*

*Hashtag* atau simbol pagar (#) yang ditulis topik perbincangan agar pengguna twitter lainnya bisa melakukan pencarian sejenis topik yang diinginkan.

i. List

Pada list ini dapat dikelompokkan dari pengguna twitter yang telah dijadikan pengikut dalam satu *group*. Sehingga pengguna telah mereka ikuti bisa melihat secara keseluruhan.

j. Terkini

Topik terkini ialah sebuah kejadian dalam waktu bersamaan yang sedang banyak dibicarakan oleh pengguna twitter seperti pandemi covid dengan hashtag #dirumahsaja.

## 2. Twitter API

Twitter mempunyai *Application Programming Interface* (API) dapat digunakan developer untuk mengembangkan aplikasi sesuai dengan kebutuhan masing-masing. Sejenis *Application Programming Interface* (API) adalah salah cara komunikasi komputer dengan pengguna lainnya. Lalu mempermudah untuk mendapatkan sebuah data informasi pengguna harus memberi izin pada perangkat aplikasi untuk memberi akses informasi data yang detail, seperti email, nomor *handphone*, dan lain-lainnya. Untuk dapatkan akses twitter api

maka pengguna harus terlebih dahulu melakukan pendaftaran di situs [developer.twitter.com](https://developer.twitter.com). Setelah selesai daftar sebagai pengembang akan mendapatkan kunci pengguna, akses pengguna, akses token serta akses token rahasia yang dipakai sebagai syarat untuk autentikasi dari pelayanan yang akan dikembangkan. Tujuan dari autentikasi yaitu untuk memberi akses kepada pengembangan dalam melakukan pencarian mengenai suatu contoh objek yang trend pada twitter.

### 2.3.5 Klasifikasi Teks

Klasifikasi teks adalah sesuatu pekerjaan digunakan sebagai penetapan sebuah teks telah ditetapkan ke bebas teksnya. Dalam melakukan klasifikasi teks bisa pakai pengaturan menyusun teks, mengatur teks dan mengkategorikan kata. Seperti tulisan dapat diatur dari tema, lalu keluhan tiket dapat diatur bersumber yang relevansi, dan percakapan dapat diatur dari suatu bahasa. Peralatan melakukan klasifikasi pengambilan data dibagikan input, menganalisis data, dan habis itu secara otomatis menetapkan peranda tag yang relevan, semacam proses gambar 2.2 dibawah ini (Monkeylearn, 2019).



Gambar 2. 2. Implementasi klasifikasi teks

Klasifikasi teks *machine learning* belajar membuat klasifikasi yang bersumber pada pengamatan sebelumnya. Dengan menggunakan contoh yang telah diberi

label bagaikan informasi kaitan, metode *machine learning* bisa menekuni asosiasi yang berbeda antara bagian teks, beserta keluaran istimewa adalah tags diharapkan buat inputan ialah teks, suatu tags yakni klasifikasi maupun jenis yang sudah di tentukan tadinya teks tertentu bisa masuk ke dalamnya. Misalnya, apabila kita sudah mendefinisikan kamus kita buat mempunyai kata- kata berikut "ini, adalah, itu, tidak, luar biasa, kurang baik, bola basket", serta kita mau membuat vektor bacaan" ini luar biasa," kita hendak mempunyai representasi vektor berikut dari bacaan itu: (1, 1, 0, 0, 1, 0, 0) (Monkeylearn, 2019).

### 2.3.6 Pembobotan Kata (TF-IDF)

*Term Frequency – Inverse Document Frequency* singkatan dari TF-IDF adalah sebagai metode menentukan hal penting dari sebuah kata terhadap data yang telah di ekstraksi. Metode akan berjalan dengan menghitung jumlah kata yang unik dari kata serta dibandingkan total kata dengan metode yang diterapkan (Haranto, Pinandito, & Kharisma, 2018). Pada proses TF-IDF dijadikan dua kelompok TF (*Term Frequency*) dan proses IDF (*Inverse Document Frequency*)

#### 1. TF (*Term Frequency*)

Memastikan faktor pembobotan kata pada suatu dokumen yang didasarkan jumlah kemunculan pada dokumen tersebut. Jumlah menghasilkan sesuatu kata yang diperhitungkan dalam beri pembobotan pada kata. Terus menjadi besar kumpulan dokumen, akan semakin besar juga bobotnya dalam pemberian nilai pada suatu dokumen.

## 2. IDF (*Inverse Document Frequency*)

Perihal dibutuhkan sebab *term* yang banyak timbul sebagai dokumen, mampu dianggap sebagai *term universal* hingga tidak berarti nilainya. Meskipun faktor kejarangan membawa kata *term scarcity* wajib dicermati dalam pemberian bobot pada koleksi dokumen. Tahapan untuk pembobotan TF-IDF adalah:

$$X(y, u) = TF(d, p) * idf$$

Pada tahapan ini rumus menjelaskan:

$X_{yu}$  : pembobotan dokumen ke -d pada kata ke - p.

$TF_{dp}$  : pencarian sebuah kata pada dokumen.

$idf_p$  : inversed document frequency ( $\log v/df$ ).

$V$  : total dokumen.

$DF$  : banyaknya dokumen yang muncul berdasarkan kata kunci.

Langkah-langkah proses pembobotan kata sebagai berikut:

- a. Menghitung jumlah TF (*Term Frequency*) setiap kata pada dokumen pada tabel 2.5.

Tabel 2. 2. Hasil perhitungan TF

Kata	TF (term frequency)					
	D6	D1	D2	D3	D4	D5
Belajar		1	1	1	1	
Rumah		1		1	1	
Saja		1				
Tetap	1	1		1		
Semangat	1	1		1		

b. Menghitung DF (*Document Frequency*) pada tabel 2.6.

Tabel 2. 3. Hasil perhitungan DF

Kata	TF (TERM FREQUENCY)						DF
	D6	D1	D2	D3	D4	D5	
Belajar		1	1	1	1		4
Rumah		1		1	1		3
Saja		1					1
Tetap	1	1		1			3
Semangat	1	1		1			3

c. Menghitung IDF (*Inverse Document Frequency*) pada tabel 2.7.

Tabel 2. 4. Hasil perhitungan IDF

Kata	IDF
	Log (v/df)
Belajar	$\log (6/4)^{10} = 0,176$
Rumah	$\log (6/3)^{10} = 0,301$
Saja	$\log (6/1)^{10} = 0,778$
Tetap	$\log (6/3)^{10} = 0,301$
Semangat	$\log (6/3)^{10} = 0,301$

d. Menghitung TF-IDF (*Term Frequency - Inverse Document Frequency*) pada tabel 2.8.

Tabel 2. 5. Hasil perhitungan TF-IDF

TF (TERM FREQUENCY)					
D6	D1	D2	D3	D4	D5
0	0.176	0.176	0.176	0.176	0
0	0.301	0	0.301	0.301	0
0	0.778	0	0	0	0
0.301	0.301	0	0.301	0	0
0.301	0.301	0	0.301	0	0

### 2.3.7 *K-nearest neighbor*

*K-nearest neighbor* adalah metode yang sering dipakai untuk melakukan klasifikasi data dan teks. Pada metode *k-nearest neighbor* ini melakukan klasifikasi pada objek yang berdasar dari data yang pendekatannya jarak ke *object* atau sering dikatakan dengan *neighborhood*. *Supervised learning* adalah suatu pembelajaran yang outputnya diharapkan sudah diketahui sebelumnya. Data yang sudah ada akan menjadi kriteria data yang dimasukkan.

*K-nearest neighbor* akan digunakan sebagai menghitung jarak terdekat dari analisis sentimen ke sampel data uji untuk menentukan klasifikasi dengan adanya data sampel tersebut. Sehingga dalam memastikan hasil klasifikasi *k-nearest neighbor* dapat dilihat dari jarak dekat object dengan setiap kelompok. Untuk menjalankan proses *k-nearest neighbor* seperti ini:

1. Memastikan jumlah *object* k.
2. Menghitungkan jarak *object* dengan setiap kelompok data. Dan diperhitungkan

dengan rumus *euclidean distance*  $P(o, i) = \sqrt{\sum_{k=1}^n (O_k - I_k)^2}$   $P$  merupakan jarak dari  $o$  dan  $i$  untuk pengujian serta testing data. Baru mendapatkan hasil pengklasifikasian.

Kalau sudah selesai semua dari tahapan tersebut. Maka pengklasifikasian menggunakan *k-nearest neighbor* selanjutnya di validasi untuk cocokkan dengan data pakar dan nilai akurasi.

Langkah – langkah perhitungan *k-nearest neighbor* terdiri dari 2 label dengan skala kuantitatif  $x_1$  dan  $x_2$  serta dengan label  $y$  yang berisi sentimen positif dan negatif pada tabel 2.9.

Tabel 2. 6. Datset sentimen k-nearest neighbor

X1	X2	Y
7	7	Negatif
7	4	Negatif
3	4	Positif
1	4	Positif

1. Tentukan variabel  $k$  = untuk jumlah tetangga terdekat. Misalkan dengan memberi nilai  $k = 3$ .
2. Pada data yang baru di hitung jarak antara semua data training.
3. Memberi urutan jarak yang ditentukan tetangga terdekat dengan nilai minimum  $k$  berdasarkan jaraknya pada tabel 2.10.



Tabel 2. 7. Pehitungan jarak k pada k-nearest neighbor

X1	X2	Kuadrat jarak dengan data baru (3,7)
7	7	$(7-3)^2 + (7-7)^2 = 16$
7	4	$(7-3)^2 + (4-7)^2 = 25$
3	4	$(3-3)^2 + (4-7)^2 = 9$
1	4	$(7-3)^2 + (7-7)^2 = 13$

4. Periksa label dari tetangga terdekat pada tabel 2.11.

Tabel 2. 8. Hasil label tetangga terdekat pada k-nearest neighbor

X1	X2	Kuadrat jarak dengan data baru (3,7)	Peringkat jarak minimum	Hasil tetangga terdekat	Label tetangga terdekat (y)
7	7	$(7-3)^2 + (7-7)^2 = 16$	3	True	Buruk
7	4	$(7-3)^2 + (4-7)^2 = 25$	4	False	-
3	4	$(3-3)^2 + (4-7)^2 = 9$	1	True	Baik
1	4	$(7-3)^2 + (7-7)^2 = 13$	2	True	Baik

### 2.3.8 Naïve Bayes

*Naïve bayes* memiliki metode berakar pada teorema *naïve bayes* yang mana dilakukan sebagai klasifikasi. Tahapan klasifikasi dapat menggunakan metode statistic dan probabilitas yang ditemukan oleh ilmuwan Inggris, Thomas Bayes. Bisa memprediksi pandangan ke masa depan berdasarkan kejadian dari masa lalu, maka disebut dengan metode teorema *naïve bayes*. *Naïve bayes* memiliki ciri khas yaitu berasumsi begitu kuat kepada independensi suatu kondisi atau kejadian masing-masing (Aprilah, Kurniawan, Baydhowi, & Haryati, 2021).

Berdasarkan David L Olson dan Dursun Delen (Olson & Delen, 2008) untuk *naïve bayes* sendiri memiliki masing-masing kondisi yang menghitung probabilitas

dan keputusan dengan ketentuan yang dianggap benar di faktor informasi objek. Pada metode ini memiliki atribut object yang diasumsikan sebagai independent. Dalam melakukan perhitungan probabilitas yang meliputi jumlah akhir dari frekuensi pada tabel master keputusan.

Pada pengujian *naïve bayes* hendak menggunakan *library* alami *language toolkit (nltk)* tujuan untuk miliki otomatis klasifikasi *tweet* sentimen yang positif maupun negatif (Ratnawati, 2018). Dalam proses melaksanakan klasifikasi berlandaskan informasi sebelumnya. *Naïve bayes classifier* menggunakan probabilitas sebelumnya, yakni nilai yang dipercayai baik saat sebelum jalani uji coba kepada tiap label pada waktu training serta lebih efektif dalam pendidikan *supervised* ataupun terawasi. Pada *naïve bayes* hendak menggunakan tata cara dalam sebagian wujud aplikasi, parameter ialah *maximum likelihood* ataupun kemiripan kata paling tinggi (Ratnawati, 2018) buat mengklarifikasi informasi yang hendak dihitung adalah  $P(X|Y)$ , hipotesa yang benar buat informasi sampel Y yang dilihat bisa diterapkan pada persamaan.

$$P(M|K) = \frac{p(K|M) \cdot P(M)}{p(K)}$$

Penjelasan:

- M : sampel data kelas tidak ketahu.
- K : kelas khusus untuk data hipotesis M.
- $P(M|K)$  : probabilitas hipotesis k pada berlandaskan kondisi M.
- $P(M)$  : probabilitas K.
- $P(K|M)$  : probabilitas hipotesis m berlandaskan kondisi K.
- $P(K)$  : probabilitas M.

Langkah-langkah untuk perhitungan peluang sebuah kata persamaan dengan *conditional probability* pada studi kasus dalam analisis sentimen menggunakan metode *naïve bayes* untuk mengetahui isi kata yang bersifat positif atau negatif yang berdasarkan dari *tweet* pengguna twitter. Maka untuk menghindari nilai 0 pada *conditional probability* digunakan teknik *laplace smoothing* untuk menghasilkan nilai:

$$P(w_i + c_j) = \frac{\text{Jumlah}(w_i, c_j) + 1}{(\sum_{w \in v} \text{jumlah}(w_i, c_j)) + |v|}$$

Keterangan:

1. Jumlah  $(w_i, c_j) + 1$  merupakan jumlah kemunculan kata pada setiap sentimen serta diberi nilai 1 untuk menghindari nilai 0.
2.  $(\sum_{w \in v} \text{jumlah}(w_i, c_j))$  merupakan total jumlah dari kemunculan kata pada seluruh masing-masing sentimen.
3.  $|v|$  merupakan jumlah kata unik yang muncul didalam seluruh sentimen.

Tabel 2. 9. Dataset sentimen naïve bayes

Label	Kalimat	Sentimen
1	Belajar <i>online</i> bosan sekali	Negatif
2	Semangat belajar <i>online</i> teman	Positif
3	Rindu belajar sekolah senang	Positif
4	Tidak enak, tidak ada uang jajan	Negatif

Dari dataset pada tabel 2.12 yang telah dikumpulkan sebanyak 22 *tweet* dan di pecah menjadi sentimen positif 12 dan negatif 10 dari data *tweet*. Selanjutnya dilakukan mencari kata unik dari keseluruhan dataset dan mencari *conditional probability* masing-masing kata pada setiap sentimen:

## 4. Positif

$$\frac{0 + 1}{12 + 10} + \frac{1}{32} = 0,03125$$

## 5. Negatif

$$\frac{0 + 1}{10 + 10} + \frac{1}{20} = 0,05$$

Maka dari nilai dataset yang di cari *conditional probability* adalah sebagai berikut:

Tabel 2. 10. Perhitungan probability

Sentimen	P(M K)			
	Semangat	Belajar	Online	Termin
Positif	0,03125	0,03125	0,03125	0,03125
Negatif	0,05	0,05	0,05	0,05

Tabel 2. 11. Perhitungan probability sentimen

Label	Kalimat	Sentimen
1	Tidak semangat belajar rumah saja	?

Pada tabel 2.13 dan tabel 2.14 di setiap kata tersebut akan dicari kecocokan kata dengan *conditional probability* dari dataset, jika tidak ada kesamaan kata maka akan digunakan teknik *lapcase smoothing* untuk ditambah nilai 1 sehingga nilai probabilitas tidak akan jadi 0 waktu pencocokan kata seperti di tabel 2.15:

Tabel 2. 12. Hasil perhitungan probability sentimen

Sentimen	Tidak	Semangat	Belajar	Rumah	Saja
Positif	1	0,03125	0,03125	1	1
Negatif	1	0,05	0,05	1	1

Setelah itu proses perhitungannya akan seperti di tabel 2.16:

Tabel 2. 13. Hasil label sentimen

Label		
	Positif	Negatif
1	0,0625	0,1

Lalu dicari nilai paling tinggi diantara sentimen porabilitas yang positif dan negatif pada kata “tidak semangat belajar dirumah saja” di label 1 yang bersifat negat

## BAB III

### METODE PENELITIAN

#### 3.1. Jenis, Sifat, Dan Pendekatan Penelitian

##### 3.1.1 Jenis Penelitian

Pada jenis penelitian yang dilakukan menggunakan penelitian kualitatif, yang memerlukan pemahaman terhadap suatu kondisi permasalahan yang alami dengan memperoleh data-data kemudian dilakukan analisis dan mendapatkan kesimpulan. Penelitian kualitatif secara umum digunakan untuk meneliti kehidupan masyarakat yang berkaitan analisis sentimen pembelajaran secara *online* di masa pandemi covid-19.

##### 3.1.2 Sifat Penelitian

Sifat penelitian yang dilakukan yaitu deskriptif. Deskriptif merupakan metode analisis yang telah dilakukan dengan mengumpulkan data-data yang bersangkutan dengan penelitian dilakukan dengan memberi parameter dan jumlah dataset yang telah ditentukan dan dilakukan proses metode *naïve bayes* dan *k-nearest neighbor* untuk mendapatkan hasil akurasi dan sentimen dari data *tweet* tersebut.

##### 3.1.3 Pendekatan Penelitian

Pada pendekatan penelitian yang dilakukan menggunakan pendekatan kuantitatif sesuai alur penelitian telah dibuat untuk mendapatkan hasil dari penjelasan di batasan masalah yaitu sentimen positif, negatif, dan netral serta akurasi yang diperoleh.

### 3.2. Metode Pengumpulan Data

Sumber data yang dibutuhkan pada penelitian ini terdiri dari dua pengumpulan data primer dan sekunder:

#### 3.2.1 Data Primer

Data primer merupakan proses pengumpulan data yang akan dilakukan dengan cara wawancara dan observasi pada studi kasus yang penelitian lakukan:

##### 1. Observasi

Observasi sebagai proses pengumpulan data dengan cara mengamati fenomena terjadi di media sosial twitter yang berkaitan dengan *tweet* pembelajaran secara *online* di masa pandemi. Dari tahapan ini proses pengumpulan data dari media sosial twitter pada tanggal 16 Maret 2020 sampai tanggal 10 April 2021 yang berkaitan kata-kata ataupun *hashtag* yang sering digunakan pada *tweet* pengguna secara dan proses pengumpulan menggunakan sistematis kisaran waktu 8 jam serta memperoleh data *tweet* sebanyak 17.373.

##### 2. Wawancara

Wawancara sebagai proses pengumpulan data dengan tanya jawab dengan berapa-berapa narasumber secara acak. Tujuan wawancara untuk mengetahui sentimen pada data *tweet* yang telah dikumpulkan dari langkah-langkah observasi serta proses wawancara ini butuh waktu selama 2 minggu dari untuk memperoleh hasil sentimen dari 4 responden.

### 3.2.2 Data Sekunder

Data Sekunder merupakan pengumpulan data yang di peroleh dari membaca buku, jurnal, dan penelitian yang berkaitan analisis sentimen dari tahapan proses pengumpulan data cara pengujian data sampai mendapatkan hasil yang benar.

### 3.3. Metode Analisis Data

Penelitian ini dilakukan proses analisis data pengguna twitter terhadap pembelajaran *online* di masa pandemi covid-19 yang diperoleh dari tahapan pengumpulan data. Maka tahapan analisis data yang dilakukan penelitian tersebut sebagai berikut:

1. Untuk mengetahui hasil sentimen yang didapatkan dari proses metode *naïve bayes* dan *k-nearest neighbor* dengan menggunakan pembobotan TF-IDF:
  - a. Mengumpulkan data yang telah didapatkan dari media sosial twitter dengan memberi sentimen secara manual dari masing-masing *tweet* yang telah dilakukan wawancara dari tahapan pengumpulan data.
  - b. Melakukan proses preprocessing data dari *case folding, cleansing data, tokenization, spell checker, stemming*, dan *normalization* agar hasil datanya baik dan dapat dilanjut ketahap klasifikasi.
  - c. Melakukan proses klasifikasi sentimen menggunakan sistematis dari *tweet* yang telah di preprocessing hasil ke dalam proses lexicon.
  - d. Proses pembobotan kata menggunakan TF-IDF dan dilanjutkan ke dalam proses klasifikasi dari dua metode *naïve bayes* dan *k-nearest*



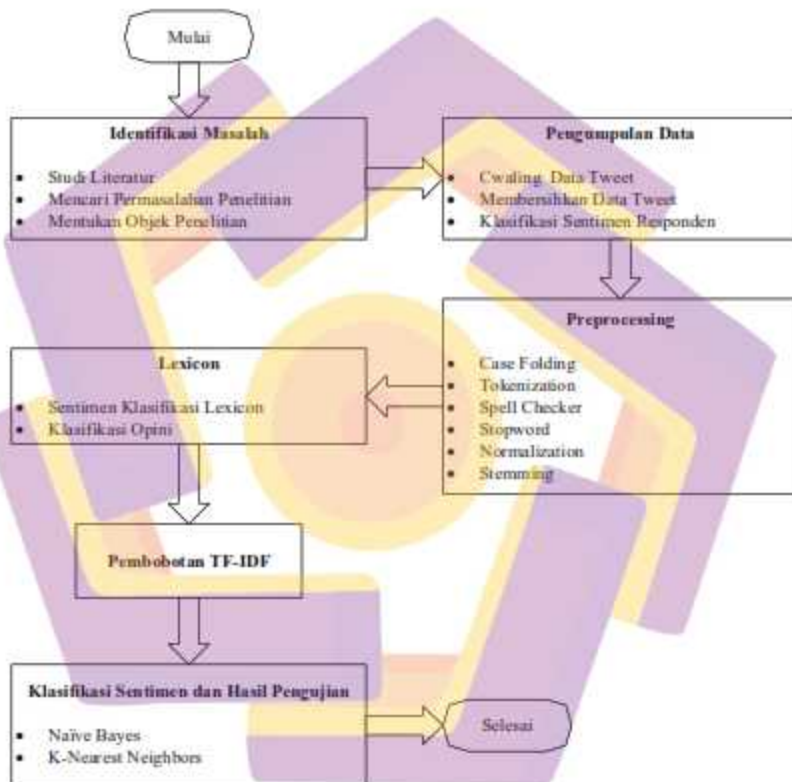
*neighbor* untuk mengetahui hasil prediksi dari sentimen yang manual dari 4 responden dan 1 sistematis dari lexicon.

2. Proses untuk mendapatkan tingkat akurasi pada analisis sentimen dengan menghitung nilai akurasi, presisi dan recall dengan masing-masing metode *naïve bayes* dan *k-nearest neighbor*. Untuk peroleh tingkat akurasi perlunya membandingkan jumlah data yang dimiliki, jika nilai akurasi tinggi maka hasilnya bagus, akan tetapi jika hanya melihat nilai akurasi saja. Perlunya menghitung presisi dan recall dengan data relevan.



### 3.4. Alur penelitian

Pada penelitian ini dibutuhkan langkah-langkah yang benar supaya penelitian ini dapat berjalan dengan efektif. Berikut alur penelitian ini dapat dilihat pada gambar bawah ini.



Gambar 3. 1. Alur Penelitian

### 3.4.1 Identifikasi Masalah

#### 1. Studi Literatur

Studi literatur dilakukan oleh peneliti yaitu mengumpulkan literatur berupa jurnal, buku dan revisi lainnya untuk memperkuat pengetahuan penelitian yang dilakukan. Tujuan penggunaan jurnal dan buku pada referensi 65 penulisan tinjauan pustaka dan landasan teori serta memperkuat pengetahuan peneliti ketika penelitian ini berlangsung dan tidak sebagai referensi penulisan. Pada revisi lainnya seperti website yang digunakan sebagai untuk mengetahui tentang analisis sentimen, data mining, *natural language processing*.

#### 2. Mencari Permasalahan Penelitian

Pada peneliti saat ini yang dilakukan yaitu mencari pokok permasalahan pada penelitian yang harus diselesaikan dari penelitian seperti *datamining*, *text mining*, *natural language processing* dan lain-lainnya.

#### 3. Menentukan Objek Penelitian

Setelah menemukan permasalahan pada penelitian, selanjutnya dilakukan pencarian objek yang akan digunakan sebagai penelitian ini.

### 3.4.2 Pengumpulan Data

#### 1. Crawling Data Tweet

Ditahap *crawling data tweet* dibutuhkan aplikasi anaconda sebagai tempat proses pengolahan data serta menggunakan salah satu *library* untuk penarikan data *tweet* skala besar dengan ketentuan dalam penarikan data yaitu pengguna kata kunci dan tanggal awal dan akhir yang ditentukan.

Kata kunci dalam penarikan terdiri dari kata dan hashtag seperti “#belajardaricovid19”, “#belajardaring”, “#belajaronline”, “#daring”, “#kuliahonline”, “#merdekabelajar”, “#pembelajaranjarakjauh”, “#pjj”, “#rumahbelajar”, “#sekolahonline”, “Daring”, “Kuliah online”, “Sekolah online”, dan penarikan data dimulai dari tanggal 16 maret 2020 sampai 10 april 2021.

## 2. Membersihkan Data *Tweet*

Setelah proses pengumpulan data yang dilakukan, Selanjutnya menggabungkan semua data yang telah ditarik menjadi 1 data pusat dengan menggunakan salah satu *library* anaconda dan membuang data yang sama atau duplikat dari hasil dari pencarian berdasarkan kata kunci yang digunakan.

## 3. Klasifikasi Sentimen Responden

Pada proses labelling sentimen manual dengan cara melakukan wawancara dengan 4 responden yang bukan dari pakar ahli bahasa, tetapi dari golongan terdekat. waktu proses klasifikasi dibutuhkan rentang waktu 2 minggu untuk menghasilkan sentimen.

### 3.4.3 Preprocessing

*Preprocessing* merupakan proses mengubah data yang tidak baik menjadi baik untuk memperoleh sumber keaslian informasi yang akan dilakukan proses selanjutnya. Dithap *preprocessing* awal menampilkan data informasi yang dilakukan 3 tahapan yaitu membersihkan, memperbaiki dan menggabungkan data-data tersebut. Maka penelitian ini yang akan di lakukan di *preprocessing* data

dengan memecah kalimat jadi perkata (*Tokenization*), lalu dilakukan proses perbaikan kata memperbaiki kata typo menjadi benar (*Spell Checker*) dan perbaikan kata yang tidak baku menjadi baku (*Normalization*), dan membuang kata yang banyak kemunculannya (*Stopword*) serta membuang kata imbuhan dari kata yang memiliki imbuhan (*Stemming*). Berikut langkah-langkah dan penjelasan dari proses *preprocessing*:

1. Case Folding

Berfungsi untuk mengubah huruf yang tidak *lowercase* atau huruf kapital menjadi huruf kecil semua atau disebut dengan *lowercase*.

2. Tokenization

Berfungsi untuk memecah perkata pada kalimat yang telah di proses pembersihan kata yaitu membuang angka, tanda baca dan simbol pada kalimat.

3. Spell Checker

Berfungsi untuk memperbaiki kata yang *typo* dengan mencari nilai probabilitas dengan semua kadidat yang paling benar menggunakan Norvig dan SymSpell.

4. Stopword

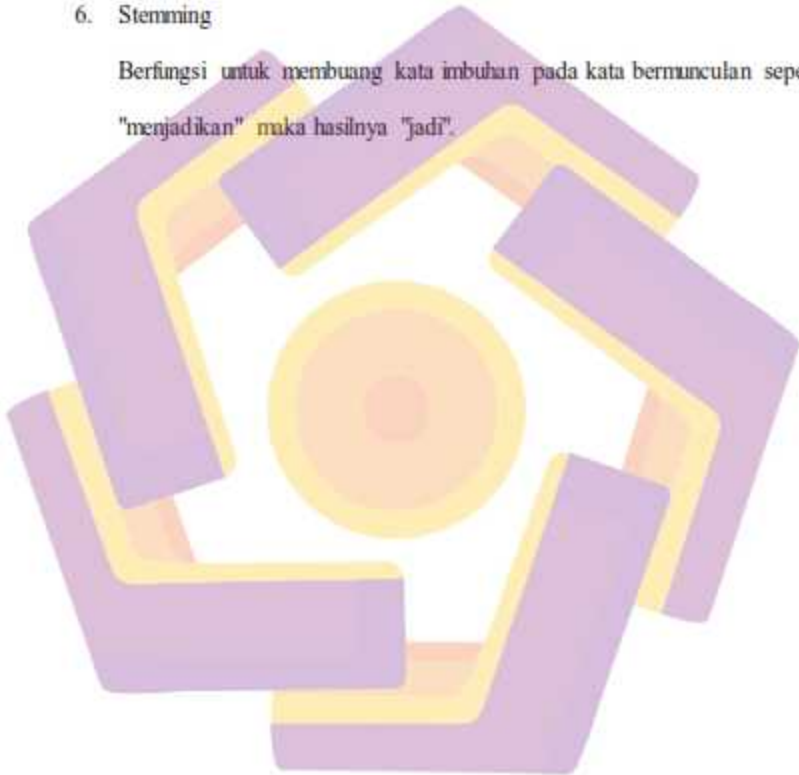
Berfungsi membuang kata dasar yang memiliki kemunculan sangat banyak serta tidak bermakna seperti "yang", "dan", "di", "dari", dan lainnya.

5. Normalization

Berfungsi untuk memperbaiki kata yang tidak baku menjadi baku, seperti kata jelas yang tidak baku menjadi kata baku sesuai dengan Kamus Besar Bahasa Indonesia (KBBI).

6. Stemming

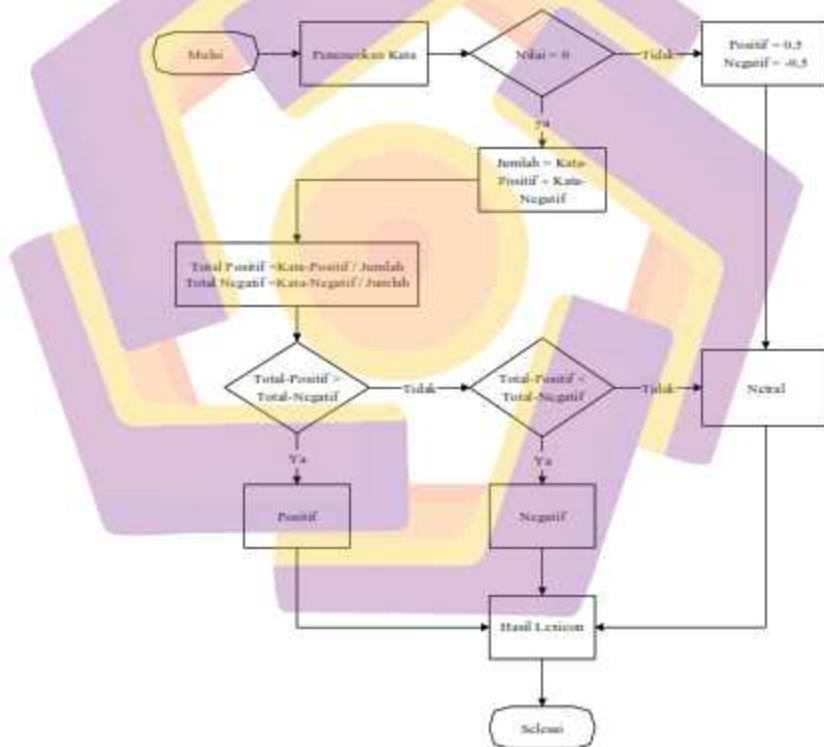
Berfungsi untuk membuang kata imbuhan pada kata bermunculan seperti "menjadikan" maka hasilnya "jadi".



### 3.4.4 Lexicon

#### 1. Klasifikasi Sentimen Lexicon

Berfungsi untuk mengidentifikasi sebuah kata dalam kalimat yang memiliki sifat positif, negatif atau netral dengan menghitung nilai *polarity*. Cara melakukan untuk mengidentifikasi kata untuk mengetahui sentiment pada kata, bisa mengunakan dataset lexicon berbahasa Indonesia dan berikut alur perhitungan lexicon.



Gambar 3. 2. Lexicon

## 2. Klasifikasi Opini

Setelah dilakukan proses klasifikasi dengan lexicon serta menghasilkan nilai *polarity*, kemudian dilakukan proses label sentimen sifat dengan menghitung nilai *polarity* pada kondisi tertentu. Berikut kondisi perhitungan sentimen pada tabel 3.1.

Tabel 3. 1. Perhitungan Klasifikasi Opini

Negatif	Positif	Netral
Nilai kurang sama dengan -8 dan nilai besar sama dengan -99 - <b>Menolak</b>	Nilai besarsama dengan -99 dan nilai kurang sama dengan 8 - <b>Menerima</b>	Nilai 0 - <b>biasa saja</b>
Nilai kurang sama dengan -4 dan nilai besar sama dengan -7 - <b>Sangat tidak Setuju</b>	Nilai besarsama dengan -7 dan nilai kurang sama dengan 4- <b>Sangat Setuju</b>	
Nilai kurang sama dengan -3 dan nilai besar sama dengan -3 - <b>cukup setuju</b>		

### 3.4.5 Pembobotan TF-IDF

Pembobotan kata TF-IDF yang bertujuan untuk menghitung nilai bobot masing-masing kata setiap dokumen. Pada tahapan ini dibagi menjadi 2 proses yaitu TF dan IDF. TF (*Term Frequency*) hitung jumlah kemunculan setiap kata pada dokumen serta dengan kemunculan kata tersebut paling banyak maka nilai kata itu paling besar. IDF (*Inverse Document Frequency*) menghitung jumlah dokumen pada setiap kata yang jarang kemunculan pada sebuah dokumen yang dianggap nilai paling besar. Jika kata tersebut banyak kemunculan kata pada dokumen maka hasil nilainya kecil.

### 3.4.6 Klasifikasi Sentimen dan Hasil Pengujian

Pada klasifikasi sentimen dilakukan pada dataset responden dan lexicon dengan pembobotan TF-IDF lalu di proses dari masing-masing metode *k-nearest neighbor* dan *naïve bayes*. Hasil dari melakukan klasifikasi sentimen dengan *k-*



*nearest neighbor* dan *naïve bayes* akan dibandingkan kemiripan sentimen yang sudah ada dengan dataset sentimen sebelumnya, berapa presentase kemiripan data telah diklasifikasi.

Hasil pengujian dari metode *k-nearest neighbor* dan *naïve bayes* pada dataset telah di klasifikasi sebelumnya dengan membagi data menjadi data training 80% dan data testing 20% untuk mengetahui tingkat akurasi yang diperoleh. Berikut penjelasan pada metode *k-nearest neighbor* dan *naïve bayes*:

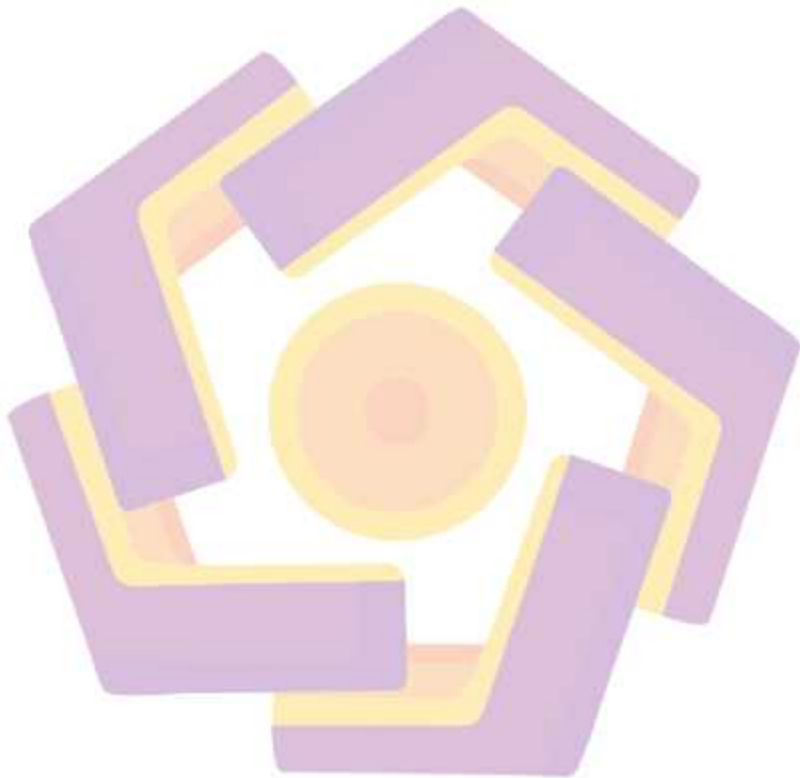
1. **Naïve bayes**

Proses klasifikasi dengan metode *naive bayes* dengan menggunakan 2 data yaitu data training dan data testing. Data training digunakan untuk mencari nilai probabilitas dari masing-masing kata yang ada pada dataset. Selanjutnya data testing ini digunakan untuk membandingkan kata dari data testing dengan diketahuinya nilai probabilitas pada data training tersebut.

2. **K-Nearest Neighbor**

Proses klasifikasi selanjutnya menggunakan metode *k-nearest neighbor* untuk klasifikasi berdasarkan objek pada data yang mempunyai jarak paling dekat dari objek untuk menghitung seberapa jauh atau dekat dari objek masing-masing kelompok. Tahapan dalam metode *k-nearest neighbor* yaitu, Menentukan jumlah data tetangga K dan menghitung jarak objek dengan menggunakan *euclidean distance* yang dimana nilai D adalah jarak, nilai X adalah data training serta nilai Y adalah data testing.

Diproses terakhir menentukan jarak tetangga yang terdekat dari minimum jarak ke K.



## BAB IV

### HASIL PENELITIAN DAN PEMBAHASAN

#### 4.1. Pengumpulan Data

Pada pengumpulan data menggunakan *tools* jupyter notebook (anaconda) untuk memperoleh data berdasarkan *keyword* yang sudah ditentukan. Kemudian dilakukan penggabungan data dan dilakukan pembuangan kata yang sama dari pengguna twitter yang sama. Supaya data lebih baik dan dilakukan proses klasifikasi sentimen dengan 4 responden.

##### 4.1.1 Crawling Data Tweet

Pada proses *crawling* data *tweet* dibutuhkan sebuah *library python* yang bernama *snsrcape* yang memiliki keunggulan dalam pengambilan data lebih banyak dan cepat. Data *tweet* yang akan diambil dari tanggal 16 Maret 2020 sampai 10 April 2021 dengan *keyword* yang telah ditentukan. Berikut hasil *crawling* data pada tabel 4.1.

Tabel 4. 1. Crawling Data Tweet

No	Keyword	Jumlah Tweet	Waktu Penarikan
1	#belajardaricovid19	88	1 menit
2	#belajardaring	37	3 detik
3	#belajaronline	103	1 menit
4	#daring	166	1 menit
5	#kuliahonline	208	1.30 menit
7	#merdekabelajar	384	1.30 menit
8	#pembelajaranjarakjauh	18	3 detik

Tabel 4.1. Crawling Data Tweet (lanjutan)

No	Keyword	Jumlah Tweet	Waktu Penarikan
9	#pjj	114	1 menit
10	#rumahbelajar	45	3 detik
11	#sekolahonline	48	3 detik
12	Darng	9432	5 jam
13	Kuliah <i>online</i>	6216	3 jam
14	Sekolah <i>online</i>	515	4 menit
Jumlah		17.373	8 Jam 11 menit 9 detik

Berikut *source code* untuk *crawling data tweet* berdasarkan *keyword* yang digunakan pada gambar 4.1.

```
import requests, module.twttr as twitttr
import pandas as pd
import itertools

search_term = 'kuliah online'

df = pd.DataFrame(twitttr.TwitterSearchScrapor(search_term=' #pjj', since='2020-03-10', until='2021-04-10', lang='Indonesia', lang_id='').get_items())['data', 'content']
data = pd.DataFrame({'tweet_id': df['data'].astype(str).str[-10], 'text': df['content']})
data.to_excel('dataset/'+search_term+'.xlsx', index=False)
```

Gambar 4. 1. Kode Crawling Data Tweet

#### 4.1.2 Membersihkan Data

Hasil *crawling data tweet* dari *keyword* yang digunakan akan dilakukan proses penggabungan dataset menjadi satu dataset terpusat. Untuk mempermudah proses cek data yang kembar dengan nama pengguna yang sama agar dataset lebih baik. Berikut hasil pembersihan data yang kembar yang telah dibersihkan pada tabel 4.2.

Tabel 4. 2. Pembersihan Data

No	Dataset	Jumlah Tweet
1	Dataset tidak bersih	17.373
2	Dataset bersih	15.759

Berikut *source code* untuk menggabungkan semua data dari *keyword* dipilih dan membersihkan data yang kembar digunakan dengan bantuan *library panda* dan *glob* pada gambar 4.2.

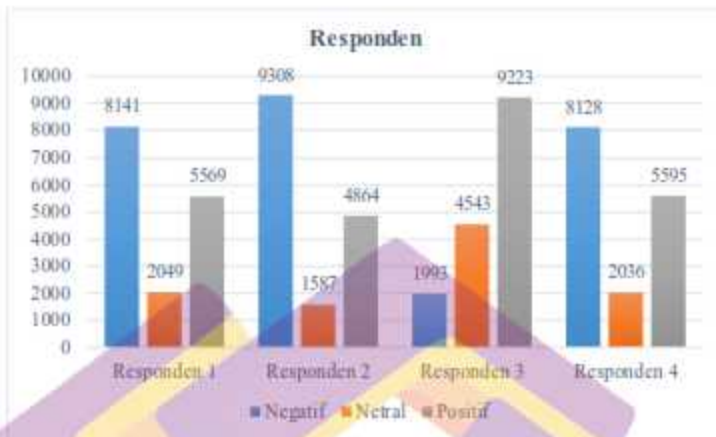
```
import pandas as pd
import glob

all_data = pd.DataFrame()
for f in glob.glob("dataset/*.xlsx"):
    df = pd.read_excel(f)
    all_data = all_data.append(df, ignore_index=True)
all_data.drop_duplicates(subset=["text"], keep=False)
all_data.to_excel("dataset.xlsx", index=False)
```

Gambar 4. 2. Kode Menggabungkan data dan Membersih Data Kembar

#### 4.1.3 Klasifikasi Sentimen Responden

Pada proses *labelling* sentimen manual dengan melakukan wawancara dengan 4 responden yang bukan dari pakar ahli bahasa, Akan tetapi dari golongan terdekat. Waktu proses klasifikasi dibutuhkan rentang waktu selama 2 minggu untuk memperoleh hasil sentimen. Berikut hasil yang telah dikumpul ke dalam bentuk grafik perbandingan hasil responden 1 sampai 4.



Gambar 4. 3. Klasifikasi sentimen Manual

Pada Gambar 4.3 merupakan hasil perbandingan dari 4 responden yang telah dilakukan proses klasifikasi sentimen manual. Hasil dari kemiripan data yang memiliki tingkat presentase tinggi di negatif yaitu responden 1, 2, dan 4 dari pada positif dan netral. Akan tetapi di responden 3 lebih cenderung netral dari pada positif dan netral

Pada Tabel 4.3. merupakan hasil dari klasifikasi sentimen manual terhadap masing-masing responden dengan *tweet* para pengguna twitter terhadap pembelajaran secara *online* di masa pandemi covid-19 yang dilakukan proses wawancara.

Tabel 4. 3. Klasifikasi Sentimen Responden

No	Tweet	R 1	R 2	R 3	R 4	Kesimpulan
1	@bdgfess Kuliah <i>online</i> , mahar , rebahan	negatif	negatif	netral	negatif	negatif
2	@biaseact_ aktifitas bias dipagi hari, kuliah <i>online</i> , rung adus tapi dandan. wkwk	positif	positif	netral	positif	positif

Tabel 4. 3. Klasifikasi Sentimen Responden (lanjutan)

No	Tweet	R 1	R 2	R 3	R 4	Kesimpulan
3	Ngga kuliah online ataupun offline, ngantuk tetep menjadi hal yang mengikutinya	negatif	negatif	netral	negatif	negatif
4	Katanya mo simulasi kuliah online jam 7 pagi, udh jam segini ga mulai mulai. Dosennya telat bangun apa gimana	negatif	negatif	netral	negatif	negatif
5	Sempat sekarang masih ku tunggu kabar baik kuliah online diperpanjang sampai setelah lebaran ^_^	negatif	negatif	netral	negatif	negatif

Di tabel 4.3. merupakan klasifikasi sentimen pada data *tweet* oleh responden untuk mengetahui tanggap atau makra dari kalimat tersebut untuk dilakukan kesimpulan terhadap *tweet* dari 4 responden dengan sentimen paling terbanyak yang dipilih.

#### 4.2. Preprocessing

Di *preprocessing* memiliki banyak tahapan dalam proses perbaikan kata, membuang kata, dan pengabungan data yang akan nantinya dilakukan proses selanjutnya supaya *tweet* baik. Maka pada dataset *tweet* sebanyak 15.759 akan dilakukan proses *preprocessing* dengan bantuan *library* *pandas* dan sebagainya untuk dilakukan proses menampilkan data dan mengolah data dengan 6 tahapan tersebut.

Berikut *source code* digunakan untuk menampilkan data excel yang telah diproses sebelum nya serta akan dilakukan proses *preprocessing* data.

```
import pandas as pd
import nltk
nltk.download('stopwords')
nltk.download('punkt')

TWEET_DATA = pd.read_excel("data_bersih.xlsx")
```

Gambar 4. 4. Kode Tampil Data Mentah

### 4.2.1 Case Folding

Pada tahapan *case folding* dilakukan untuk mengubah huruf besar menjadi huruf kecil dan proses tersebut hanya pengaruh pada huruf “a” sampai “z”. Jika tidak menggunakan proses *case folding* pada tahap awal, maka menimbulkan kesalahan dari proses selanjutnya dalam perbaikan kata dengan bantuan dari *tools* yang digunakan. Berikut proses *case folding* pada data *tweet* di tabel 4.4.

Tabel 4. 4. Case Fokling

No	Tweet	Case Fokling
1	@bdgfess Kuliah <i>online</i> , mubar , rebahan	@bdgfess kuliah <i>online</i> , mubar , rebahan
2	@biasexact_ aktifitas bias dipagi hari, kuliah <i>online</i> , rung adus tapi dandan. wkwk	@biasexact_ aktifitas bias dipagi hari, kuliah <i>online</i> , rung adus tapi dandan. wkwk
3	Ngga kuliah <i>online</i> ataupun offline, ngantuk tetep menjadi hal yang mengikutinya	ngga kuliah <i>online</i> ataupun offline, ngantuk tetep menjadi hal yang mengikutinya
4	Katanya mo simulasi kuliah <i>online</i> jam 7 pagi, udh jam segini ga mulai mulai. Dosenya telat bangun apa gimana	katanya mo simulasi kuliah <i>online</i> jam 7 pagi, udh jam segini ga mulai mulai. dosenya telat bangun apa gimana
5	Sampai sekarang masih ku tunggu kabar baik kuliah <i>online</i> diperpanjang sampai setelah lebaran ^ ^	sampai sekarang masih ku tunggu kabar baik kuliah <i>online</i> diperpanjang sampai setelah lebaran ^ ^

Berikut *source code* digunakan untuk mengubah kata yang besar menjadi kecil atau lowercase pada gambar

```
# ----- Case Folding -----
# gunakan fungsi Series.str.lower() pada Pandas
TWEET_DATA['tweet_cf'] = TWEET_DATA['text'].str.lower()

print('Case Folding Result : \n')
print(TWEET_DATA['tweet_cf'].head(5))
print('\n\n')
```

Gambar 4. 5. Kode Case Folding



#### 4.2.2 Tokenization

Pada *tokenization* merupakan proses pemotongan kalimat menjadi perkata dari hasil proses *case folding*. Akan tetapi dalam proses *tokenization* harus yang melakukan penbuangan kata yang bermakna simbol, angka, dan url agar di proses selanjutnya lebih baik. Jika tidak dilakukan pembersihan terhadap simbol, angka dan url, maka dataset yang digunakan sebagai klasifikasi sentimen lebih cenderung tidak sempurna. Serta akan menghasilkan sentimen netral yang disebabkan susah memahami dari kata tersebut dan menurun dari tingkat akurasi data. Berikut hasil dari proses *tokenization* di tabel 4.5.

Tabel 4. 5. Tokenization

No	Case Folding	Tokenization
1	kuliah <i>online</i> mabar rebahan	[kuliah', 'online', 'mabar', 'rebahan']
2	aktivitas bias dipagi hari kuliah <i>online</i> rung adus tapi dandan wkwk	[aktivitas', 'bias', 'di', 'pagi', 'hari', 'kuliah', 'online', 'rung', 'adus', 'tapi', 'dandan', 'wkwk']
3	ngga kuliah <i>online</i> ataupun offline ngantuk tetep menjadi hal yang mengikutinya	[tidak', 'kuliah', 'online', 'ataupun', 'offline', 'mengantuk', 'tetap', 'menjadi', 'hal', 'yang', 'mengikutinya']
4	katanya mo simulasi kuliah <i>online</i> jam pagi udh jam segini ga mulai mulai dosennya telat bangun apa gimana	[katanya', 'mo', 'simulasi', 'kuliah', 'online', 'jam', 'pagi', 'sudah', 'jam', 'segini', 'tidak', 'mulai', 'mulai', 'dosennya', 'telat', 'bangun', 'apa', 'bagaimana']
5	sampai sekarang masih ku tunggu kabar baik kuliah <i>online</i> diperpanjang sampai setelah lebaran	[sampai', 'sekarang', 'masih', 'ku', 'tunggu', 'kabar', 'baik', 'kuliah', 'online', 'diperpanjang', 'sampai', 'setelah', 'lebaran']

Berikut *source code* untuk membersihkan kalimat dari simbol, *emotion*, url website.

```
import string
import re
import sys
import re
import preprocessor as p

# ----- Tolentring -----

def remove_tweet_special(text):
    # remove tab, new line, and back slice
    text = text.replace("\t", " ").replace("\n", " ").replace("\r", " ")
    # remove non ASCII (emoticon, chinese word, etc)
    text = text.encode('ascii', 'replace').decode('ascii')
    # remove mention, link, hashtag
    text = " ".join(re.sub("([@][A-Z0-9-]+)|(#)[\w/]+", "", text).split())
    # remove incomplete url
    return text.replace("http://", "").replace("https://", "")

TWEET_DATA['tweet_cf'] = TWEET_DATA['tweet_cf'].apply(remove_tweet_special)

# remove number
def remove_number(text):
    return re.sub("^\d+", "", text)

TWEET_DATA['tweet_clean'] = TWEET_DATA['tweet_cf'].apply(remove_number)

# remove punctuation
def remove_punctuation(text):
    return text.translate(str.maketrans("", "", string.punctuation))

TWEET_DATA['tweet_clean'] = TWEET_DATA['tweet_clean'].apply(remove_punctuation)

# remove whitespace leading & trailing
def remove_whitespace_lt(text):
    return text.strip()

TWEET_DATA['tweet_clean'] = TWEET_DATA['tweet_clean'].apply(remove_whitespace_lt)

# remove multiple whitespace into single whitespace
def remove_whitespace_multiple(text):
    return re.sub(" +", " ", text)

TWEET_DATA['tweet_clean'] = TWEET_DATA['tweet_clean'].apply(remove_whitespace_multiple)

# remove single char
def remove_singl_char(text):
    return re.sub("^\b[0-9-]+", "", text)

TWEET_DATA['tweet_clean'] = TWEET_DATA['tweet_clean'].apply(remove_singl_char)

def preprocessing_data(x):
    return p.clean(x)

TWEET_DATA['tweet_clean'] = TWEET_DATA['tweet_clean'].apply(preprocessing_data)

def tokenize_data(x):
    return p.tokenize(x)

TWEET_DATA['tweet_clean'] = TWEET_DATA['tweet_clean'].apply(tokenize_data)
```

Gambar 4. 6. Kode Bersihkan url, simbol dan lain-lainnya

Berikut *source code* untuk memecah kalimat menjadi perkata dalam proses

*Tokenization*.

```
import nltk
from nltk.tokenize import word_tokenize

# nltk word tokenize
def word_tokenize_wrapper(text):
    return word_tokenize(text)

TWEET_DATA['tweet_token'] = TWEET_DATA['tweet_clean'].apply(word_tokenize_wrapper)
```

Gambar 4. 7. Kode Tokenization

### 4.2.3 Spell Checker

*Spell checker* merupakan proses pencarian nilai probabilitas yang tinggi dari sebagai kandidat yang paling benar. *Spell checker* membutuhkan data *corpus* yang berbahasa Indonesia sebagai referensi perbaikan kata yang tepat untuk data *tweet* yang *typo*. *Spell checker* dan *normalization* proses kerjanya sama dengan memperbaiki kata yang salah. Akan tetapi *spell checker* ini lebih cocok dalam perbaikan kata *typo* walaupun itu tidak baku tetap akan dilakukan proses kebenaran kata tersebut, supaya di proses *normalization* lebih baik dijadikan kata yang baku.

Berikut hasil proses *spell checker* di tabel 4.6.

Tabel 4. 6. Spell Checker

No	Tokenization	Spell Checker
1	[kuliah', 'online', 'min', 'bareng', 'rebahan']	[kuliah', 'online', 'min', 'bareng', 'rebahan']
2	[aktivitas', 'bus', 'di', 'pagi', 'hari', 'kuliah', 'online', 'rung', 'adus', 'tap?', 'dandan', 'wkwk']	[aktivitas', 'bus', 'di', 'pagi', 'hari', 'kuliah', 'online', 'rung', 'adus', 'tap?', 'dandan', 'wkwk']
3	[tidak', 'kuliah', 'online', 'ataupun', 'offline', 'mengantuk', 'tetap', 'menjadi', 'hal', 'yang', 'mengikutinya']	[tidak', 'kuliah', 'online', 'ataupun', 'offline', 'mengantuk', 'tetap', 'menjadi', 'hal', 'yang', 'mengikutinya']
4	[katanya', 'mau', 'simulasi', 'kuliah', 'online', 'jam', 'pagi', 'sudah', 'jam', 'segi?', 'tidak', 'mulai', 'mulai', 'dosennya', 'telat', 'bangun', 'apa', 'bagaimana']	[katanya', 'mau', 'simulasi', 'kuliah', 'online', 'jam', 'pagi', 'sudah', 'jam', 'begini', 'tidak', 'mulai', 'mulai', 'dosennya', 'telat', 'bangun', 'apa', 'bagaimana']
5	[sampai', 'sekarang', 'masih', 'ku', 'tunggu', 'kabar', 'baik', 'kuliah', 'online', 'diperpanjang', 'sampai', 'setelah', 'lebaran']	[sampai', 'sekarang', 'masih', 'ku', 'tunggu', 'kabar', 'baik', 'kuliah', 'online', 'diperpanjang', 'sampai', 'setelah', 'lebaran']

Berikut *source code* untuk memperbaiki kata yang *typo* dengan mencari nilai probabilitas dengan semua kandidat yang paling benar.

```
import re
from collections import Counter
import time
from spellcheck import Spellcheck, VerbsOnly

path_corpus = "/content/gymnasium_online_bimbingan/ukki.txt"
spn_spell = Spellcheck()
spn_spell.create_dictionary(path_corpus)

def spell_checker(kata):
    content = []
    for u in kata:
        suggestions = spn_spell.lookup(u, VerbsOnly, CLASSIFY, use_orth_distance=True, include_unknown=True)
        content.append([u, term for u in suggestions][0])
    hasil = content
    return hasil

TEST_DATA['spell_checker'] = TEST_DATA['spell_checker'].apply(spell_checker)
```

Gambar 4. 8. Kode Spell Checker

#### 4.2.4 Stopword

*Stopword* merupakan proses pengolahan kata dengan membuang kata yang relevan seperti kata hubung seperti “di”, “yang”, “ke” dan lain-lainnya. Ketika proses *stopword* tidak dilakukan pada dataset ini, Maka nantinya pengaruh terhadap tingkat presentase sentimen dan akurasi data. Akan tetapi dari riset sebelum dari kata yang tidak dilakukan *stopword* akan lebih cenderung netral hasilnya. Berikut hasil proses *stopword* di tabel 4.7

Tabel 4. 7. Stopword

No	Spell Checker	Stopword
1	['kuliah', 'online', 'main', 'bareng', 'rebahan']	['kuliah', 'online', 'main', 'bareng', 'rebahan']
2	['aktivitas', 'bus', 'di', 'pagi', 'hari', 'kuliah', 'online', 'rung', 'adus', 'tapi', 'dandan', 'wkwk']	['aktivitas', 'bias', 'pagi', 'kuliah', 'online', 'rung', 'adus', 'dandan', 'wkwk']
3	['tidak', 'kuliah', 'online', 'ataupun', 'offline', 'mengantuk', 'tetap', 'menjadi', 'hal', 'yang', 'mengikutinya']	['kuliah', 'online', 'offline', 'mengantuk', 'mengikutinya']

Tabel 4. 7. Stopword (lanjutan)

No	Spell Checker	Stopword
4	[katanya, 'mau', 'simulasi', 'kuliah', 'online', 'jam', 'pagi', 'sudah', 'jam', 'segitu', 'tidak', 'mula', 'mula', 'dosennya', 'telat', 'bangun', 'apa', 'bagaimana']	[simulasi, 'kuliah', 'online', 'jam', 'pagi', 'jam', 'dosennya', 'telat', 'bangun']
5	[sampai, 'sekarang', 'masih', 'ku', 'tunggu', 'kabar', 'baik', 'kuliah', 'online', 'diperpanjang', 'sampai', 'setelah', 'lebaran']	[ku, 'tunggu', 'kabar', 'kuliah', 'online', 'diperpanjang', 'lebaran']

Berikut *source code* digunakan untuk membuang kata yang relevan di *stopword*.

```
import sys
ALTA_Aplikasi('stopword')
from nltk.corpus import stopwords

# ..... get stopwords from nltk stopwords
# get stopwords manually
list_stopwords = stopwords.words('indonesian')
print(len(list_stopwords))

# ..... manually add stopwords
# tambah substitusi_stopword
list_stopwords.extend(['ig', 'dg', 'rt', 'lgn', 'ep', 'di', 'ku',
                        'aku', 'mau', 'sudah', 'kapan', 'telat',
                        'apa', 'bagaimana', 'sudah', 'jam',
                        'pagi', 'sudah', 'jam', 'segitu', 'tidak',
                        'mula', 'mula', 'dosennya', 'telat',
                        'bangun', 'apa', 'bagaimana', 'jam',
                        'pagi', 'sudah', 'jam', 'segitu',
                        'tidak', 'mula', 'mula', 'dosennya', 'telat',
                        'bangun', 'apa', 'bagaimana'])

len(list_stopwords)

# ..... get stopwords from list file
# get list stopwords using regex
list_stopword = re.findall('stopword[0-9].txt', os.listdir('.'))
# convert stopwords string to list & append substitusi_stopword
list_stopwords.extend(list_stopword[0].split('\n'))
len(list_stopwords)

# .....
# convert list to dictionary
list_stopwords = set(list_stopwords)

# .....
# fungsi stopword pada list token
def stopword_remove(words):
    return [word for word in words if word not in list_stopwords]

tweet_data['tweet_stopword'] = tweet_data['tweet_token'].apply(stopword_remove)

print(tweet_data['tweet_stopword'])
```

Gambar 4. 9. Kode Stopword

#### 4.2.5 Normalization

*Normalization* untuk memperbaiki kata yang tidak baku menjadi baku sesuai Kamus Besar Bahasa Indonesia (KBBI) dengan mencocokkan dataset yang dimiliki. Proses *normalization* berbeda dengan cara *spell checker* untuk memperbaiki kata dengan mencari nilai probabilitas yang paling benar. Kelamahan dari

normalization terhadap kamus bahasa yang terbatas dan gagal dalam memperbaiki kata yang identik bahasa daerah. Berikut hasil proses normalization di tabel 4.8.

Tabel 4. 8. Normalization

No	Stopword	Normalization
1	[kuliah', 'online', 'main', 'bareng', 'rebahan']	['kuliah', 'daring', 'main', 'bareng', 'rebahan']
2	['aktivitas', 'bias', 'pagi', 'kuliah', 'online', 'rung', 'adus', 'dandan', 'wkwk']	['aktivitas', 'bias', 'pagi', 'kuliah', 'daring', 'rung', 'adus', 'dandan', 'wkwk']
3	['kuliah', 'online', 'offline', 'mengantuk', 'mengikutinya']	['kuliah', 'daring', 'offline', 'mengantuk', 'mengikutinya']
4	['simulasi', 'kuliah', 'online', 'jam', 'pagi', 'jam', 'dosennya', 'telat', 'bangun']	['simulasi', 'kuliah', 'daring', 'jam', 'pagi', 'jam', 'dosennya', 'terlambat', 'bangun']
5	['kn', 'tunggu', 'kabar', 'kuliah', 'online', 'diperpanjang', 'lebaran']	['ku', 'tunggu', 'kabar', 'kuliah', 'daring', 'diperpanjang', 'lebaran']

Berikut *source code* yang digunakan untuk memperbaiki kata yang tidak baku menjadi baku sesuai dengan data KBBI.

```
def normalisasi(tweet):
    kamus_slangword = eval(open("combined_slang_words.txt").read())
    pattern = re.compile(r'\b( ' + '|'.join(kamus_slangword.keys())+r')\b')
    content = {}
    for kata in tweet:
        filteredSlang = pattern.sub(lambda x: kamus_slangword[x.group()],kata)
        content.append(filteredSlang.lower())
    tweet = content
    return tweet
TWEET_DATA['tweet_normalisasi'] = TWEET_DATA['tweet_stemming'].apply(normalisasi)
print(TWEET_DATA['tweet_normalisasi'])
```

Gambar 4. 10. Kode Normalization

#### 4.2.6 Stemming

*Stemming* merupakan proses akhir dari memperbaiki kata yang tidak baku menjadi baku sesuai KBBI serta dilakukan proses *stemming* untuk menghilangkan kata inbuan yang ada pada data *tweet*. Dari hasil *stemming* juga memiliki kelemahan dengan kata inbuan yang huruf *double* atau salah penulisan kata tersebut. Contoh kata "Dkecewakkn" yang gagal dalam proses *stemming* di bagian kata "kn" dan "D", karena disebabkan dalam penulisan menggunakan angka atau

simbol yang menyebabkan di proses awal *tokenization* dibuang. Berikut hasil proses *stemming* di tabel 4.9.

Tabel 4.9. Stemming

No	Normalisasi	Stemming
1	['kuliah', 'daring', 'main', 'bareng', 'rebahan']	['kuliah', 'daring', 'main', 'bareng', 'rebah']
2	['aktivitas', 'bias', 'pagi', 'kuliah', 'daring', 'tungg', 'adus', 'dandan', 'kwkw']	['aktivitas', 'bias', 'pagi', 'kuliah', 'daring', 'tungg', 'adus', 'dandan', 'kwkw']
3	['kuliah', 'daring', 'offline', 'mengantuk', 'mengikutinya']	['kuliah', 'daring', 'offline', 'antuk', 'ikut']
4	['sumulas', 'kuliah', 'daring', 'jam', 'pagi', 'jam', 'dosenya', 'lambat', 'bangun']	['sumulas', 'kuliah', 'daring', 'jam', 'pagi', 'jam', 'dosen', 'lambat', 'bangun']
5	['ku', 'tunggu', 'kabar', 'kuliah', 'daring', 'diperpanjang', 'lebaran']	['ku', 'tunggu', 'kabar', 'kuliah', 'daring', 'panjang', 'lebaran']

Berikut *source code* digunakan untuk membuang kata yang memiliki imbuhan dengan *stemming*.

```

# Import nltk and package
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
import swifter

# Create stemmer
factory = StemmerFactory()
stemmer = factory.create_stemmer()

# stemmed
def stemmed_wrapper(terms):
    return stemmer.stem(terms)

terms_dict = {}

for document in TWEET_DATA['tweet_stopword']:
    for term in document:
        if term not in terms_dict:
            terms_dict[term] = ""

print(len(terms_dict))
print("-----")

for term in terms_dict:
    terms_dict[term] = stemmed_wrapper(terms)
    print(term, " ", terms_dict[term])

print(terms_dict)
print("-----")

# apply stemmed term to dataframe
def get_stemmed_term(document):
    return [terms_dict[term] for term in document]

TWEET_DATA['tweet_stemming'] = TWEET_DATA['tweet_stopword'].swifter.apply(get_stemmed_term)
print(TWEET_DATA['tweet_stemming'])

```

Gambar 4.11. Kode Stemming





Kesimpulan hasil akhir dari *preprocessing* dengan 6 tahapan masih memiliki kekurangan dari dataset dalam perbaikan bahasa gaul atau alay di *tweet* para pengguna twitter terhadap opini pembelajaran secara *online*. Maka dibutuhkan library yang canggih dalam proses perbaikan kata gaul atau alay serta dengan bahasa daerah seluruh Indonesia.

### 4.3. Lexicon

Lexicon merupakan proses klasifikasi sentimen yang sistematis dengan mencocokkan kata pada dataset lexicon berbahasa Indonesia dan memiliki nilai *polarity* dari masing-masing kata. Proses klasifikasi sudah dijelaskan pada Gambar 3.2 sebagai alur proses untuk mengetahui sentiment positif, negatif, dan netral kata tersebut.

#### 4.3.1 Klasifikasi Sentimen Lexicon

Setelah dilakukan *preprocessing* di *stemming*, kemudian dilakukan proses klasifikasi sentimen dengan mencocokkan kata pada dataset lexicon berbahasa Indonesia serta menghasilkan klasifikasi sentimen positif, negatif, dan netral. Berikut hasil dari klasifikasi lexicon di tabel 4.10.

Tabel 4. 10. Klasifikasi Lexicon

No	Tweet Preprocessing	Sentimen	Polarity
1	kuliah daring main bareng rebah	negatif	-1
2	aktivitas bias pagi kuliah daring rung udus dandan wkwk	positif	2
3	kuliah daring offline antuk ikut	negatif	-7
4	simulasi kuliah daring jam pagi jam dosen lambat bangun	positif	1
5	ku tunggu kabar kuliah daring panjang lebaran	netral	0

Pada tabel 4.10 merupakan hasil dari klasifikasi sentimen lexicon dengan mencocokkan kata dan menghitung nilai *polarity* yang dimiliki dari kalimat tersebut serta menentukan sentimen apa di dalamnya. Hasil lexicon sendiri memiliki kelamhan dan kekurangan dari dataset yang digunakan dalam terjemahan bahasa yang memiliki makna sentimen di dalamnya.

Pada gambar 4.14 merupakan perbandingan hasil klasifikasi sentimen yang dilakukan manual (responden) dan sistematis (lexicon). Klasifikasi manual merupakan hasil dari kesimpulan responden 1 sampai 4 dengan mencari predikat paling banyak sentimen pada satu data *tweet*. Maka berikut hasil dari perbandingan klasifikasi responden dan lexicon tersebut.



Gambar 4. 14. Klasifikasi Responden dan Klasifikasi Lexicon

Hasil Gambar 4.14 merupakan hasil perbandingan terhadap klasifikasi sentimen dengan tingkat presentase yang tinggi bagian sentimen negatif pada responden sebanyak 8148 *tweet* dan disusul lexicon 7671 *tweet*. Akan tetapi proses klasifikasi positif dan netral cenderung lebih tinggi lexicon dari pada responden

sebanyak 5698 *tweet* positif dan netral sebanyak 2390 *tweet* serta responden positif sebanyak 5556 *tweet* dan netral sebanyak 2055 *tweet*.

Berikut *source code* untuk mengubah kata dari *stemming* menjadi nilai *polarity* dan memiliki sentimen dalam kata dengan mencocokkan dataset lexicon yang berbahasa Indonesia dimiliki.

```
lexicon_positive = dict()
import csv
with open('lexicon_positive.csv', 'w') as csvfile:
    reader = csv.reader(csvfile, delimiter=',')
    for row in reader:
        lexicon_positive[row[0]] = int(row[1])

lexicon_negative = dict()
import csv
with open('lexicon_negative.csv', 'w') as csvfile:
    reader = csv.reader(csvfile, delimiter=',')
    for row in reader:
        lexicon_negative[row[0]] = int(row[1])

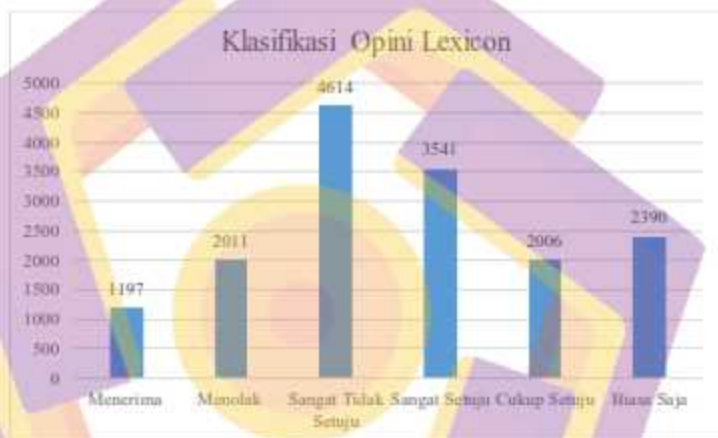
# function to determine sentiment polarity of tweets
def sentiment_analysis_lexicon_indonesia(text):
    score = 0
    for word in text:
        if word in lexicon_positive:
            score = score + lexicon_positive[word]
    for word in text:
        if word in lexicon_negative:
            score = score - lexicon_negative[word]
    polarity = ''
    if (score > 0):
        polarity = 'positif'
    elif (score < 0):
        polarity = 'negatif'
    else:
        polarity = 'netral'
    return score, polarity

results = TWEET_DATA['tweet_stemming'].apply(sentiment_analysis_lexicon_indonesia)
results = list(zip(*results))
TWEET_DATA['jumlah_polarity'] = results[0]
TWEET_DATA['sentimen_lexicon'] = results[1]
print(TWEET_DATA['sentimen_lexicon'].value_counts())
```

Gambar 4. 15. Kode Klasifikasi Lexicon

#### 4.3.2 Klasifikasi Opini

Setelah hasil klasifikasi sentimen *lexicon* dan memiliki nilai *polarity* yang akan dilakukan proses untuk klasifikasi opini terhadap pembelajaran di masa pandemi covid-19. Cara proses klasifikasi opini yang akan dilakukan sudah dijelaskan pada tabel 3.1. Berikut hasil presentase klasifikasi opini dari nilai *polarity lexicon* di setiap data *tweet* pada gambar 4.16.



Gambar 4. 16. Klasifikasi Opini Lexicon

Pada gambar 4.16 menjelaskan kondisi terhadap pembelajaran secara daring dari hasil proses klasifikasi opini dengan menghitung nilai *polarity* yang disesuaikan aturan pada tabel 3.1. Hasil yang diperoleh mengatakan bahwa opini masyarakat terhadap pembelajaran secara *online* cenderung "**Sangat Tidak Setuju**" sebanyak 4614 *tweet* dan "**Sangat Setuju**" sebanyak 3541 *tweet* kepada kebijakan di masa pandemi covid-19.

#### 4.4. Pembobotan TF-IDF

Selanjutnya dilakukan pembobotan kata dengan TF-IDF untuk mengetahui kemunculan *term* pada setiap dokumen tersebut. *Term* menghitung nilai probabilitas dari kemunculan *term* pada setiap dokumen yang seperti di dokumen D1 sampai D5. DF menghitung jumlah kemunculan *term* dari keseluruhan dokumen D1 sampai D5 serta perhitungan IDF menggunakan metode log dari total DF dengan membagi total dokumen  $=\text{LOG}(\text{jumlah dokumen} / \text{jumlah DF})$ . TF.IDF menghitung untuk pembobotan kata pada masing-masing dokumen dengan mengalikan hasil IDF dengan jumlah setiap *term* pada dokumen. Berikut hasil dari pembobotan kata TF-IDF pada tabel 4.11.

Tabel 4. 11. Pembobotan Kata TF-IDF

Term	Dokumen					DF	IDF	TF.IDF				
	D1	D2	D3	D4	D5			D1	D2	D3	D4	D5
kuliah	1	1	1	0	1	4	0,097	0,097	0,097	0,097	0	0,097
daring	1	1	1	1	1	5	0	0	0	0	0	0
main	1	0	0	0	0	1	0,699	0,699	0	0	0	0
bareng	1	0	0	0	0	1	0,699	0,699	0	0	0	0
rebah	1	0	0	0	0	1	0,699	0,699	0	0	0	0
aktivitas	0	1	0	0	0	1	0,699	0	0,699	0	0	0
bias	0	1	0	0	0	1	0,699	0	0,699	0	0	0
pagi	0	1	0	1	0	2	0,398	0	0,398	0	0,398	0
rung	0	1	0	0	0	1	0,699	0	0,699	0	0	0
adus	0	1	0	0	0	1	0,699	0	0,699	0	0	0
dandan	0	1	0	0	0	1	0,699	0	0,699	0	0	0
offline	0	0	1	0	0	1	0,699	0	0	0,699	0	0

Tabel 4. 11. Pembobotan Kata TF-IDF (lanjutan)

Term	Dokumen					DF	IDF	TF.IDF				
	D1	D2	D3	D4	D5			D1	D2	D3	D4	D5
antuk	0	0	1	0	0	1	0,699	0	0	0,699	0	0
ikat	0	0	1	0	0	1	0,699	0	0	0,699	0	0
simulasi	0	0	0	1	0	1	0,699	0	0	0	0,699	0
jam	0	0	0	2	0	2	0,398	0	0	0	0,796	0
dosen	0	0	0	1	0	1	0,699	0	0	0	0,699	0
lambat	0	0	0	1	0	1	0,699	0	0	0	0,699	0
bangun	0	0	0	1	0	1	0,699	0	0	0	0,699	0
tunggu	0	0	0	0	1	1	0,699	0	0	0	0	0,699
kabar	0	0	0	0	1	1	0,699	0	0	0	0	0,699
panjang	0	0	0	0	1	1	0,699	0	0	0	0	0,699
lebaran	0	0	0	0	1	1	0,699	0	0	0	0	0,699

#### 4.5. Klasifikasi Sentimen

Setelah dilakukan proses pembobotan kata TF-IDF dan dilanjutkan dengan proses klasifikasi sentimen dengan *naïve bayes* dan *k-nearest neighbor* untuk mengetahui hasil prediksi sentimen dari data sentimen responden dan *lexicon*. Tidak hanya dengan mengetahui hasil sentimen saja, akan tetapi akan dilakukan perbandingan hasil sebelum dan sesudah di klasifikasi dari data responden dan *lexicon*. Berikut tabel 4.12 data sentimen yang akan dilakukan proses klasifikasi dengan *naïve bayes* dan *k-nearest neighbor*.

Tabel 4. 12. Data Sentimen Responden dan Lexicon

Klasifikasi	Negatif	Netral	Positif
Lexicon	7671	2390	5698
Responden	8148	2055	5556

#### 4.5.1 Naïve Bayes

Ditahap klasifikasi *naïve bayes* dibantu dengan *library Natural Language Toolkit* di proses *NaiveBayesClassifier* untuk memperoleh hasil klasifikasi dari *naïve bayes* di tabel 4.12. Waktu dalam proses awal sampai akhir klasifikasi *naïve bayes*, membutuhkan waktu 3 jam terhadap data responden dan *lexicon* yang akan diklasifikasi sentimennya. Berikut hasil dari klasifikasi sentimen pada responden dan *lexicon* di tabel 4.13.

Tabel 4. 13. Hasil Klasifikasi Naive Bayes

No	Tweet Preprocessing	Responden	Lexicon	NB_Responden	NB_Lexicon	Kesimpulan_Responden	Kesimpulan_Lexicon
1	kuliah daring main bareng rebah	negatif	negatif	negatif	negatif	1	1
2	aktivitas bias pagi kuliah daring rung adus dandan wkwk	positif	positif	positif	positif	1	1
3	kuliah daring offline antuk ikut	negatif	negatif	negatif	negatif	1	1
4	simulasi kuliah daring jam pagi jam dosen lambat bangun	negatif	positif	negatif	negatif	1	0
5	ku tunggu kabar kuliah daring panjang lebaran	negatif	netral	negatif	negatif	1	0

Pada tabel 4.13 merupakan hasil perbandingan yang telah diklasifikasi dengan *naïve bayes* serta memiliki makna dalam kesimpulan pada data tersebut. Jika nilai 1 adalah hasil kesamaan dari hasil klasifikasi sebelum dan sesudah serta

nilai 0 hasil yang terbalik. Berikut perbandingan hasil kesamaan pada data sentimen sebelum dan sesudah diklasifikasikan dengan *naïve bayes* di gambar 4.17.



Gambar 4. 17. Perbandingan Klasifikasi Sentimen Naive Bayes

Pada gambar 4.17 merupakan hasil presentase perbandingan setelah diklasifikasi dari data sebelumnya dengan hitung nilai 1 adalah sama dan 0 adalah tidak sama. Bahwa hasil dari responden dan *lexicon* memiliki tingkat presentase kemiripan hasil sentimen sebanyak 84 % dan 16% tidak sama. Di tabel 4.14 akan memaparkan atau menjelaskan hasil data yang "Sama" dan "Tidak sama" setelah di klasifikasi dengan *naïve bayes*.

Tabel 4. 14. Total Perbandingan Lexicon dan Responden di Naïve Bayes

Klasifikasi	Sama	Tidak Sama
NB_Responden	13273	2486
NB_Lexicon	13307	2452



Berikut hasil dari keseheruhan data sentimen responden dan *lexicon* sebelum dan sesudah diklasifikasi dengan *naïve bayes* pada gambar 4.18.



Gambar 4. 18. Hasil Klasifikasi Sentimen Naïve Bayes

Kesimpulan dari gambar 4.18 menyatakan bahwa hasil dari sentimen negatif dan netral cenderung meningkat serta bagian positif cenderung menurun setelah diklasifikasikan dengan *naïve bayes*. Hasil pada *nb\_responden* negatif sebanyak 8998 *tweet* dan *nb\_lexicon* negatif sebanyak 8413 *tweet* dari pada hasil sebelumnya responden negatif sebanyak 8148 *tweet* dan *lexicon* negatif sebanyak 7671 *tweet*.

Hasil pada *nb\_responden* netral sebanyak 2291 *tweet* dan *nb\_lexicon* netral sebanyak 2487 *tweet* dari pada hasil sebelumnya responden netral sebanyak 2055 *tweet* dan *lexicon* netral sebanyak 2390 *tweet*.

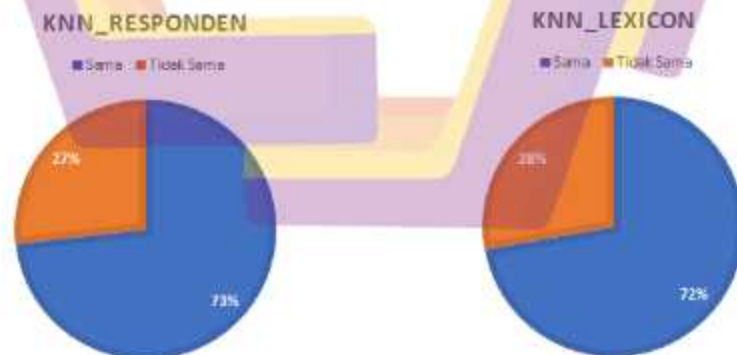
Hasil pada *nb\_responden* positif sebanyak 4470 *tweet* dan *nb\_lexicon* positif sebanyak 4859 *tweet* dari pada hasil sebelumnya responden positif sebanyak 5556 *tweet* dan *lexicon* positif sebanyak 5698 *tweet*.



Tabel 4. 15. Hasil Klasifikasi K-Nearest Neighbor

No	Tweet Preprocessing	Responden	Lexicon	KNN_Responden	KNN_Lexicon	Kesimpulan_Responden	Kesimpulan_Lexicon
1	kuliah daring mau bareng rebah	negatif	negatif	negatif	negatif	1	1
2	aktivitas bias pagi kuliah daring rung adus dandan wkwk	positif	positif	positif	positif	1	1
3	kuliah daring offline antuk ikut	negatif	negatif	negatif	negatif	1	1
4	simulasi kuliah daring jam pagi jam dosen lambat bangun	negatif	positif	negatif	netral	1	0
5	ku tunggu kabar kuliah daring panjang lebaran	negatif	netral	netral	netral	0	0

Pada tabel 4.15 merupakan hasil perbandingan yang telah diklasifikasikan dengan *k-nearest neighbor* serta memiliki makna dalam kesimpulan pada data tersebut. Berikut perbandingan hasil kesamaan pada data sentimen sebelum dan sesudah diklasifikasikan dengan *k-nearest neighbor* di gambar 4.19.



Gambar 4. 20. Perbandingan Klasifikasi Sentimen K-Nearest Neighbor

Pada gambar 4.20 merupakan hasil presentase perbandingan setelah di klasifikasi dari data sebelumnya dengan hitung nilai 1 adalah sama dan 0 adalah tidak sama. Bahwa hasil *lexicon* memiliki tingkat presentase kemiripan 72% sama dan 28% tidak sama dan responden juga miliki tingkat presentase kemiripan 73% sama dan 27 % tidak sama. Dari kesimpulan responden dan *lexicon* memiliki tingkat presentase selisih 1 % dari kemiripan dan perbedaan data yang di peroleh. Di tabel 4.16 akan memaparkan atau menjelaskan hasil data yang “Sama” dan “Tidak Sama” setelah di klasifikasi dengan *k-nearest neighbor*.

Tabel 4. 16. Total Perbandingan Lexicon dan Responden di K-Nearest Neighbor

Klasifikasi	Sama	Tidak Sama
KNN_Responden	11541	4218
KNN_Lexicon	11382	4377

Berikut hasil dari keseluruhan data sentimen responden dan *lexicon* sebelum dan sesudah di klasifikasi dengan *k-nearest neighbor* pada gambar 4.20.



Gambar 4. 21. Hasil Klasifikasi K-Neareset Neighbor

Kesimpulan dari gambar 4.21 menyatakan hasil dari sentimen negatif dan netral cenderung meningkat dan juga bagian positif cenderung menurun setelah diklasifikasikan dengan *k-nearest neighbor*. Hasil pada *knn\_responden* negatif sebanyak 8651 *tweet* dan *knn\_lexicon* negatif sebanyak 8047 *tweet* dari pada hasil sebelumnya responden negatif sebanyak 8148 *tweet* dan *lexicon* negatif sebanyak 7671 *tweet*.

Hasil pada *knn\_responden* netral sebanyak 3021 *tweet* dan *knn\_lexicon* netral sebanyak 3234 *tweet* dari pada hasil sebelumnya responden netral sebanyak 2055 *tweet* dan *lexicon* netral sebanyak 2390 *tweet*.

Hasil pada *knn\_responden* positif sebanyak 4087 *tweet* dan *knn\_lexicon* positif sebanyak 4478 *tweet* dari pada hasil sebelumnya responden positif sebanyak 5556 *tweet* dan *lexicon* positif sebanyak 5698 *tweet*.

Berikut source code klasifikasi sentimen menggunakan *k-nearest neighbor*.

```
testing_kalimat = TWEET_DATA['tweet_test']
bersihkan_kalimat = []
for test in testing_kalimat:
    if test not in list_stopswords:
        lowercase_test = test.translate(str.maketrans("", "", string.punctuation)).lower()
        hapus_angka_test = re.sub("\d+", "", lowercase_test)
        hapus_space_test = hapus_angka_test.strip()
        bersihkan_kalimat.append(hapus_space_test)

print("Preprocessing Komentar Test : %d" % len(bersihkan_kalimat))

x_test = vectorizer.transform(bersihkan_kalimat)
y_pred = classifier.predict(x_test)
y_test = [0,1] * len(TWEET_DATA['sentimen_lexicon'])
rows = len(TWEET_DATA)
sentimen_KNN = []

for xu in range(rows):
    a = y_pred[xu]
    sentimen_KNN.append(a)

TWEET_DATA['hasil_KNN_lexicon'] = sentimen_KNN
```

Gambar 4. 22. Kode Klasifikasi Sentimen K-Nearest Neighbor

#### 4.6. Hasil Pengujian

Setelah klasifikasi sentimen dengan *naïve bayes* dan *k-nearest neighbor* dengan pembobotan kata TF-IDF untuk mengetahui sentimen yang didapatkan

setelah klasifikasi serta dibandingkan dengan data sentimen sebelumnya. Pada akhir proses pengujian ini, akan dilakukan pengujian dengan metode *naïve bayes* dan *k-nearest neighbor* untuk mengetahui hasil tingkat akurasi, presisi, dan recall dari dataset sentimen.

Ditahap awal proses pengujian dilakukan pembagian data *tweet* sebanyak 15.759 menjadi data training 80% dan data testing 20% di tabel 4.18 serta dilakukan proses penghitungan akurasi, presisi, dan recall dengan *confusion matrix*. Berikut penjelasan *confusion matrix* dan perhitungannya di tabel 4.17.

Tabel 4. 17. Confusion Matrix

Data Aktual	Data Prediksi		
	Negatif	Positif	Netral
Negatif	TN	NPF	NNNF
Positif	PNF	TP	PNNF
Netral	XNNF	XNPF	XNNT

- TN : Data negatif yang diprediksi hasil benar negatif.
- NPF : Data negatif yang diprediksi hasil tidak sesuai menjadi positif.
- NNNF : Data negatif yang diprediksi hasil tidak sesuai menjadi netral.
- PNF : Data positif yang diprediksi hasil tidak sesuai menjadi negatif.
- TP : Data positif yang diprediksi hasil benar positif.
- PNNF : Data positif yang diprediksi hasil tidak sesuai menjadi netral.
- XNNF : Data netral yang diprediksi hasil tidak sesuai menjadi negatif.
- XNPF : Data netral yang diprediksi hasil tidak sesuai menjadi positif.
- XNNT : Data netral yang diprediksi hasil benar netral.

Berikut rumus perhitungan pada akurasi, persisi, dan recall:

$$\text{Persisi} = \frac{TN}{(TN+PNF+XNNF)} * 100 = ?$$

$$\text{Recall} = \frac{TN}{(TN+NPF+NNNF)} * 100 = ?$$

$$\text{Akurasi} = \frac{TN+TP+XNNT}{(TN+NPF+NNNF+PNF+TP+PNNE+XNNF+XNPF+XNNT)} * 100 = ?$$

#### 4.6.1 Data Training dan Data Testing

Pada proses pembagian data training dan data testing dilakukan sebagai tahapan awal untuk pengujian klasifikasi menggunakan metode *k-nearest neighbor* dan *naïve bayes* untuk mengetahui hasil yang diperoleh. Berikut pembagian data training dan data testing di tabel 4.18.

Tabel 4.18. Pembagian Data Training dan Data Testing

Pembagian data		Jumlah data	
Data training	Data testing	Data training	Data testing
80%	20%	12608	3152

Pada tabel 4.18 merupakan hasil pembagian data *training* dan data *testing* sebagai pengujian metode *naïve bayes* dan *k-nearest neighbor* yang akan dilakukan. Dalam proses pengujian dengan metode tersebut, data *training* digunakan sebagai melatih metode *naïve bayes* dan *k-nearest neighbor* serta data *testing* untuk mengetahui performa metode yang digunakan dan menghasilkan akurasi, persisi dan recall.

Berikut *source code* untuk pembagian data training 80% dan data testing

20%.

```
from sklearn.model_selection import train_test_split
X=TWEET_DATA.tweet_join
y=TWEET_DATA.sentimen
x_train, x_test, y_train, y_test = train_test_split(X,y,test_size = 0.2)
print('Banyak data train :',len(x_train))
print('Banyak data test :',len(x_test))
```

Gambar 4. 23. Kode Pembagian Data Training Dan Data Testing

#### 4.6.2 Naïve Bayes

Pada proses pengujian *naïve bayes* untuk mengetahui akurasi persisi dan recall pada data sentimen. *nb\_responden* dan *nb\_lexicon* yang akan dibandingkan dengan sub *naïve bayes* untuk mengetahui tingkat akurasi terbaik dari *naïve bayes multinomial*, *naïve bayes complement* dan *naïve bayes bernoulli*. Berikut hasil pengujian dari 3 metode *naïve bayes* di tabel 4.19.

Tabel 4. 19. Pengujian Naïve Bayes

Naïve Bayes	Sentimen	Akurasi	Persisi	Recall
Multinomial	Responden	72,46	76,46	58,33
Complement		74,08	69,72	64,51
Bernoulli		67,83	73,03	50,94
Multinomial	Lexicon	69,8	72,13	57,62
Complement		71,89	68,16	63,73
Bernoulli		64,59	65,94	50,34

Pada tabel 4.19 merupakan hasil pengujian dari sub *naïve bayes* untuk mengetahui tingkat akurasi yang terbaik kepada *naïve bayes complement* di data *nb\_responden* dengan tingkat akurasi sebanyak 74,08%, presisi sebanyak 69,72%, dan recall sebanyak 64,51% serta *nb\_lexicon* dengan tingkat akurasi sebanyak



71,89%, presisi sebanyak 68,16%, dan recall sebanyak 63,73%. Jika dibandingkan hasil akurasi mana yang terbaik. Maka data nb\_responden paling terbaik dari tingkat akurasi yang di peroleh dengan selisih 3% dari nb\_lexicon. Berikut proses *confusion matrix* dan perhitungan untuk mendapatkan hasil akurasi presisi dan recall di nb\_responden pada tabel 4.20 dan nb\_lexicon pada tabel 4.21.

Tabel 4. 20. Naïve Bayes Complement Responden

Aktual	Predikat			Total
	Negatif	Netral	Positif	
Negatif	1509	53	214	1776
Netral	240	170	54	464
Positif	199	57	656	912
Total	1948	280	924	3152

Pada tabel 4.20 akan dilakukan perhitungan presisi dan recall pada data nb\_responden yang dimiliki untuk mengetahui bukti ketepatan antara informasi dan tingkat keberhasilan sistem dalam memperoleh sebuah informasi. Hasil dari keseluruhan presisi dan recall pada nb\_responden, akan dilakukan proses perhitungan rata-rata dari presisi dan recall untuk mendapatkan nilai yang relevan terhadap informasi.

#### Sentimen negatif

$$Precision : \frac{1509}{(1509+240+199)} = \frac{1509}{1948} * 100 = 77,46$$

$$Recall : \frac{1509}{(1509+53+214)} = \frac{1509}{1776} * 100 = 84,97$$

#### Sentimen positif

$$Precision : \frac{656}{(656+54+214)} = \frac{656}{924} * 100 = 71,00$$

$$\text{Recall} : \frac{656}{(656+57+199)} = \frac{656}{912} * 100 = 71,93$$

#### Sentimen netral

$$\text{Precision} : \frac{170}{(170+53+57)} = \frac{170}{280} * 100 = 60,71$$

$$\text{Recall} : \frac{170}{(170+54+240)} = \frac{170}{464} * 100 = 36,64$$

Selanjutnya dilakukan proses perhitungan akurasi untuk mengetahui tingkat keberhasilan sebuah data yang kelola dengan hitung hasil aktual dan hasil predikat.

$$\text{Accuracy} : \frac{1509+170+656}{(1509+240+199+53+170+57+214+54+656)} = \frac{2335}{3152} * 100 = 74,08$$

Hasil akhir dari nb responden dengan tingkat akurasi sebanyak 74,08%, presisi sebanyak 69,72% dan recall sebanyak 64,51%.

Tabel 4. 21. Naive Bayes Complement Lexicon

	Predikat			Total	
	Aktual	Negatif	Netral		Positif
Negatif		1380	67	214	1661
Netral		253	189	71	513
Positif		216	65	697	978
Total		1849	321	982	3152

Pada tabel 4.21 akan dilakukan perhitungan presisi dan recall pada data nb\_lexicon yang dimiliki untuk mengetahui bukti ketepatan antara informasi dan tingkat keberhasilan sistem dalam memperoleh sebuah informasi. Hasil dari keseluruhan presisi dan recall pada nb\_lexicon, akan dilakukan proses perhitungan rata-rata dari presisi dan recall untuk mendapatkan nilai yang relevan terhadap informasi.

**Sentimen negatif**

$$\textit{Precision} : \frac{1380}{(1380+253+216)} = \frac{1380}{1849} * 100 = 74,63$$

$$\textit{Recall} : \frac{1380}{(1380+67+214)} = \frac{1380}{1661} * 100 = 83,08$$

**Sentimen positif**

$$\textit{Precision} : \frac{697}{(697+71+214)} = \frac{697}{982} * 100 = 70,98$$

$$\textit{Recall} : \frac{697}{(697+65+216)} = \frac{697}{978} * 100 = 71,27$$

**Sentimen netral**

$$\textit{Precision} : \frac{189}{(189+67+65)} = \frac{189}{321} * 100 = 58,88$$

$$\textit{Recall} : \frac{189}{(189+71+253)} = \frac{189}{513} * 100 = 36,84$$

Selanjutnya dilakukan proses perhitungan akurasi untuk mengetahui tingkat keberhasilan sebuah data yang dikelola dengan hitung hasil aktual dan hasil prediksi.

$$\textit{Accuracy} : \frac{1380+189+697}{(1380+253+216+67+189+65+214+71+697)} = \frac{2266}{3152} * 100 = 71,89$$

Hasil akhir dari nb\_lexicon dengan tingkat akurasi sebanyak 71,89%, presisi sebanyak 68,16% dan recall sebanyak 63,73%.

### 4.6.3 K-Nearest Neighbor

Pada proses pengujian menggunakan *k-nearest neighbor* untuk mengetahui tingkat akurasi, presisi, dan recall yang di peroleh dari data sentimen *knn\_responden* dan *knn\_lexicon* dengan menguji nilai K ganjil yaitu K-3, K-5, K-7, K-9, K-11, K-13 dan K-15. Berikut hasil pengujian dari nilai K metode *k-nearest neighbor* di tabel 4.22.

Tabel 4. 22. Pengujian K pada K-Nearest Neighbor

Nilai K	Akurasi	Presisi	Recall	sentimen
K-3	78,46	76,43	76,64	Responden
K-5	76,74	75,04	74,96	
K-7	74,81	72,53	73,09	
K-9	73,67	71,42	71,54	
K-11	73,06	70,80	70,85	
K-13	73,41	71,08	71,24	
K-15	72,78	70,68	69,87	
K-3	77,09	75,52	75,37	Lexicon
K-5	76,40	75,07	75,09	
K-7	74,84	73,09	73,61	
K-9	73,95	72,34	72,05	
K-11	73,26	71,67	71,56	
K-13	73,32	71,54	71,43	
K-15	72,94	71,20	70,72	

Pada tabel 4. 22 merupakan hasil pengujian nilai K dari *k-nearest neighbor* untuk mengetahui peringkat yang terbaik jatuh kepala nilai K-3 di data *knn\_responden* dengan tingkat akurasi sebanyak 78,46%, presisi sebanyak 76,43%,

dan recall sebanyak 76,64% serta *knn\_lexicon* dengan tingkat akurasi sebanyak 77,09%, presisi sebanyak 75,52%, dan recall sebanyak 75,37%. Jika dibandingkan hasil akurasi mana yang terbaik, maka data *knn\_responden* paling terbaik tingkat akurasi dengan selisih 1% dari *knn\_lexicon*. Berikut hasil *confusion matrix* dan perhitungan untuk memperoleh akurasi presisi dan recall di *knn\_responden* tabel 4.23 dan *knn\_lexicon* tabel 4.24.

Tabel 4. 23. K-Nearest Neighbor Niki K-3 Responden

Aktual	Predikat			Total
	Negatif	Netral	Positif	
Negatif	1419	172	115	1706
Netral	119	453	23	595
Positif	184	66	601	851
Total	1722	691	739	3152

Pada tabel 4.23 akan dilakukan perhitungan presisi dan recall pada data *knn\_responden* yang dimiliki untuk mengetahui bukti ketepatan antara informasi dan tingkat keberhasilan sistem dalam memperoleh sebuah informasi. Hasil dari keseluruhan presisi dan recall pada *knn\_responden*, akan dilakukan proses perhitungan rata-rata dari presisi dan recall untuk mendapatkan nilai yang relevan terhadap informasi.

#### Sentimen negatif

$$Precision : \frac{1419}{(1419+119+184)} = \frac{1419}{1722} * 100 = 82,40$$

$$Recall : \frac{1419}{(1419+172+115)} = \frac{1419}{1706} * 100 = 83,18$$

**Sentimen positif**

$$\textit{Precision} : \frac{601}{(601+23+115)} = \frac{601}{739} * 100 = 81,33$$

$$\textit{Recall} : \frac{601}{(601+66+184)} = \frac{601}{851} * 100 = 70,62$$

**Sentimen netral**

$$\textit{Precision} : \frac{453}{(453+172+66)} = \frac{453}{691} * 100 = 65,56$$

$$\textit{Recall} : \frac{453}{(453+23+119)} = \frac{453}{595} * 100 = 76,13$$

Selanjutnya dilakukan proses perhitungan akurasi untuk mengetahui tingkat keberhasilan sebuah data yang kebela dengan hitung hasil aktual dan hasil predikat.

$$\textit{Accuracy} : \frac{1419+453+601}{(1419+119+184+172+453+66+115+23+601)} = \frac{2473}{3152} * 100 = 78,46$$

Hasil akhir dari knn\_responden dengan tingkat akurasi sebanyak 78,46%, presisi sebanyak 76,43% dan recall sebanyak 76,64%.

Tabel 4. 24. K-Nearest Neighbor Nilai K-3 Lexicon

	Predikat			Total
	Aktual	Negatif	Netral	
Negatif	1315	175	109	1599
Netral	134	457	38	629
Positif	191	75	658	924
Total	1640	707	805	3152

Pada tabel 4.24 akan dilakukan perhitungan presisi dan recall pada data knn\_lexicon yang dimiliki untuk mengetahui bukti ketepatan antara informasi dan tingkat keberhasilan sistem dalam memperoleh sebuah informasi. Hasil dari keseluruhan presisi dan recall pada knn\_lexicon, akan dilakukan proses perhitungan

rata-rata dari presisi dan recall untuk mendapatkan nilai yang relevan terhadap informasi.

#### Sentimen negatif

$$\textit{Precision} : \frac{1315}{(1315+134+191)} = \frac{1315}{1640} * 100 = 80,18$$

$$\textit{Recall} : \frac{1315}{(1315+175+109)} = \frac{1315}{1599} * 100 = 82,24$$

#### Sentimen positif

$$\textit{Precision} : \frac{658}{(658+38+109)} = \frac{658}{805} * 100 = 81,74$$

$$\textit{Recall} : \frac{658}{(658+75+191)} = \frac{658}{924} * 100 = 71,21$$

#### Sentimen netral

$$\textit{Precision} : \frac{457}{(457+175+75)} = \frac{457}{707} * 100 = 64,64$$

$$\textit{Recall} : \frac{457}{(457+134+38)} = \frac{457}{629} * 100 = 72,66$$

Selanjutnya dilakukan proses perhitungan akurasi untuk mengetahui tingkat keberhasilan sebuah data yang kelola dengan hitung hasil aktual dan hasil predikat.

$$\textit{Accuracy} : \frac{1315+457+658}{(1315+134+191+175+457+75+109+38+658)} = \frac{2430}{3152} * 100 = 77,09$$

Hasil akhir dari *knn\_lexicon* dengan tingkat akurasi sebanyak 77,09%, presisi sebanyak 75,52% dan recall sebanyak 75,37%.

Kesimpulan hasil akhir dari pengujian pada data *tweet* 15.759 dengan metode *naïve bayes* dan *k-nearest neighbor* lebih baik dari pada penelitian sebelumnya yang menggunakan metode *naïve bayes* dan *k-nearest neighbor* terhadap pembelajaran secara *online*. Bisa dikata lebih baik dari di penelitian sebelumnya, karena data yang dimiliki dari 16 maret 2020 sampai 10 april 2021

sebanyak 17.373 *tweet* dan di tahapan preprocessing memiliki 6 proses dalam perbaikan data. Di pengujian *naïve bayes* dan *k-nearest neighbor* dilakukan analisis terhadap masing-masing metode *naïve bayes* yaitu *naïve bayes multinomial*, *naïve bayes complement* dan *naïve bayes bernoulli* mana yang terbaik hasil terkait dengan akurasinya serta *k-nearest neighbor* dengan nilai K-Ganjil yaitu K-3, K-5, K-7, K-9, K-11, K-13, K-15 mana yang terbaik hasil dengan akurasinya. Maka hasil dari penelitian ini yang peroleh hasil terbaik yaitu *nb\_respoden naïve bayes complement* dengan akurasi 74,08%, presisi 69,72% dan recall 64,51% serta di pengujian *k-neareast neighbor* nilai k-3 pada *knn\_respoden* dengan akurasi 78,46%, presisi 76,47% dan recall 76,64%. Dengan tingkat akurasi yang di peroleh cukup baik dari penelitian sebelumnya dari segi data topik pembelajaran secara *online*, *preprocessing* dan pengujian dari masing-masing *naïve bayes* dan *k-nearest neighbor*.



## BAB V

### PENUTUP

#### 5.1. Kesimpulan

Setelah melakukan pengujian berbagai macam-macam menggunakan *Naïve Bayes* dan *K-Nearest Neighbor* menggunakan pembobotan TF-IDF dapat diambil beberapa kesimpulan hasil yaitu:

1. Hasil yang diperoleh mengatakan bahwa opini masyarakat terhadap pembelajaran secara *online* cenderung "**Sangat Tidak Setuju**" sebanyak 4614 *tweet* dan "**Sangat Setuju**" sebanyak 3541 *tweet* kepada kebijakan di masa pandemi covid-19.
2. Hasil sentimen yang telah diklasifikasi menggunakan *naive bayes* dan *k-nearest neighbor* dengan pembobotan TF-IDF yang memperoleh hasil di responden dan *lexicon* sebanyak 84% yang sama hasil sentimen serta 16% yang tidak sama di *naïve bayes*. *K-nearest neighbor* pada hasil responden yang sama 73% dan tidak sama 27% serta *lexicon* yang sama 72% dan tidak sama 28% dengan masing-masing selisih 1% yang berbeda.
3. Hasil akurasi yang terbaik dari klasifikasi *naive bayes* dan *k-nearest neighbor* dengan menghasilkan *nb\_respoden* di pengujian *naïve bayes complement* dengan akurasi 74,08%, presisi 69,72% dan recall 64,51% serta di pengujian *k-nearvast neighbor* nilai K-3 pada *knm\_respoden* dengan akurasi 78,46%, presisi 76,47% dan recall 76,64%.

## 5.2. Saran

Dari hasil kesimpulan penelitian yang telah dilakukan masih banyak kekurangan dan kendala, maka dari itu untuk penelitian selanjutnya di masa depan dapat melanjutkan penelitian ini dari hasil saran sebagai berikut:

1. Penambahkan kata gaul pada dataset *preprocessing*, karena data media sosial rata-rata tidak menggunakan kata baku.
2. Dibutuhkan pakar bahasa dibidangnya untuk menjelaskan opini-opini pada dataset dengan labeling manual sentimen.
3. Dibutuhkan kajian ulang pada proses *preprocessing* tanpa *stopword* dan *stemming* untuk mengetahui tingkat akurasi dan sentimen pada dataset.
4. Dibutuhkan optimalisasi tingkat akurasi dari metode yang digunakan penelitian ini.
5. Masih memiliki kekurangan dalam kamus bahasa Indonesia yang digunakan pada proses *normalization*.

## DAFTAR PUSTAKA

### PUSTAKA BUKU

- Goleman, D. (2015). Emotional Intelligence, Kecerdasan Emosional Mengapa EI Lebih Penting daripada IQ. *Jakarta: PT Gramedia Pustaka Utama.*
- Hienstra, D. (2009). Information Retrieval Models. In *Information Retrieval: Searching in the 21st Century*. <https://doi.org/10.1002/9780470033647.ch1>
- Olson, D. L., & Delen, D. (2008). *Advanced data mining techniques. Advanced Data Mining Techniques*. <https://doi.org/10.1007/978-3-540-76917-0>
- Pawar, A. B., Jawale, M. A., & Kyataravar, D. N. (2016). Fundamentals of sentiment analysis: Concepts and methodology. In *Studies in Computational Intelligence* (Vol. 639). [https://doi.org/10.1007/978-3-319-30319-2\\_2](https://doi.org/10.1007/978-3-319-30319-2_2)
- Worth, D. (2010). Introduction to Modern Information Retrieval, 3rd Edition. *Australian Academic and Research Libraries*, 41(4). <https://doi.org/10.1080/00048623.2010.10721488>

### PUSTAKA MAJALAH, JURNAL ILMIAH ATAU PROSIDING

- Alizah, M. D., Nugroho, A., Radayah, U., & Gata, W. (2020). Sentimen Analisis Terkait Lockdown pada Sosial Media Twitter. *Indonesian Journal on Software Engineering (IJSE)*, 6(2). <https://doi.org/10.31294/ijse.v6i2.8991>
- Andika, L. A., Azizah, P. A. N., & Respatiwiulan, R. (2019). Analisis Sentimen Masyarakat terhadap Hasil Quick Count Pemilihan Presiden Indonesia 2019 pada Media Sosial Twitter Menggunakan Metode Naive Bayes Classifier. *Indonesian Journal of Applied Statistics*, 2(1).

<https://doi.org/10.13057/ijas.v2i1.29998>

Aprilia, W., Kurniawan, I., Baydhowi, M., & Haryati, T. (2021). Prediksi Kemungkinan Diabetes pada Tahap Awal Menggunakan Algoritma Klasifikasi Random Forest. *SISTEMASI*, 10(1).

<https://doi.org/10.32520/stmsi.v10i1.1129>

Berkovic, D., Ackerman, I., Briggs, A. M., & Ayton, D. (2020). "Coronavirus could kill me #StayHomeSaveLives": A content and sentiment analysis of tweets by people with arthritis during COVID-19. *Journal of Medical Internet Research*.

<https://doi.org/10.2196/24550>

Chalida, M. (2019). Analisis Sentimen Ujaran Kebencian Pemilihan Presiden 2019 Menggunakan Algoritma Naïve Bayes (Studi Kasus: Tweet #Pilpres2019 Di Kota Jakarta, Bandung, Semarang, Surabaya Dan Yogyakarta). *Jnanaloka (Jurnal Open Access Yayasan Lentera Dua Indonesia)*, (2001).

Fanny, N. S., Bambang, A., & Hapsari, D. T. (2020). FENOMENA CAPTION "TWITTER, PLEASE DO YOUR MAGIC!" MEMBANGKITKAN NETIZEN UNTUK MENOLONG. *JOURNAL OF DIGITAL EDUCATION, COMMUNICATION, AND ARTS (DECA)*, 3(01).

<https://doi.org/10.30871/deca.v3i01.1961>

Fitriyah, S. N. J., Safriadi, N., & Pratama, E. E. (2019). Analisis Sentimen Calon Presiden Indonesia 2019 dari Media Sosial Twitter Menggunakan Metode Naïve Bayes. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, 5(3).

<https://doi.org/10.26418/jp.v5i3.34368>

Giovani, A. P., Ardiansyah, A., Haryanti, T., Kurniawati, L., & Gata, W. (2020).

ANALISIS SENTIMEN APLIKASI RUANG GURU DI TWITTER MENGGUNAKAN ALGORITMA KLASIFIKASI. *Jurnal Teknoinfo*, 14(2). <https://doi.org/10.33365/jti.v14i2.679>

- H. Manguri, K., N. Ramadhan, R., & R. Mohammed Amin, P. (2020). Twitter Sentiment Analysis on Worldwide COVID-19 Outbreaks. *Kurdistan Journal of Applied Research*. <https://doi.org/10.24017/covid.8>
- Hananto, B. K., Pinandito, A., & Kharisma, A. P. (2018). Penerapan Maximum TF-IDF Normalization Terhadap Metode KNN Untuk Klasifikasi Dataset Multiclass Panichella Pada Review Aplikasi Mobile. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIIK)*, 2(12), 6812–6823.
- Hermanto, H., Mustopa, A., & Kuntoro, A. Y. (2020). ALGORITMA KLASIFIKASI NAIVE BAYES DAN SUPPORT VECTOR MACHINE DALAM LAYANAN KOMPLAIN MAHASISWA. *JITK (Jurnal Ilmu Pengetahuan Dan Teknologi Komputer)*, 5(2). <https://doi.org/10.33480/jitk.v5i2.1181>
- Lazuardi, D. R., Munandar, T. A., Harsiti, H., Mutaqin, Z., & Hays, R. N. (2020). Sentiment analysis of public opinions on the welfare of honorary educators using Naive Bayes. In *IOP Conference Series: Materials Science and Engineering* (Vol. 830). <https://doi.org/10.1088/1757-899X/830/3/032018>
- Musfiroh, D., Khaira, U., Eko, P., Utomo, P., & Suratno, T. (2021). Sentiment Analysis of Online Lectures in Indonesia from Twitter Dataset Using InSet Lexicon Analisis Sentimen terhadap Perkuliahan Daring di Indonesia dari Twitter Dataset Menggunakan InSet Lexicon. *MALCOM: Indonesian Journal*

*of Machine Learning and Computer Science, 1*(April).

- Natasuwarna, A. P. (2020). Seleksi Fitur Support Vector Machine pada Analisis Sentimen Keberlanjutan Pembelajaran Daring. *Techno.Com, 19*(4). <https://doi.org/10.33633/te.v19i4.4044>
- Oktavianto, A., & Persada, S. F. (2021). Persepsi Publik Tentang Pembelajaran Daring di Indonesia: Studi Menggunakan ELK Stack dan Python untuk Analisis Sentimen di Twitter. *Jurnal Teknik ITS, 9*(2).
- Ratnawati, F. (2018). Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter. *INOVTEK Polbeng - Seri Informatika, 3*(1). <https://doi.org/10.35314/isi.v3il.335>
- Samsir, Ambiyar, Verawardina, U., Edi, F., & Watrianthos, R. (2021). Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naive Bayes. *Jurnal Media Informatika Budidarma, 5*.
- Samuel, J., Ali, G. G. M. N., Rahman, M. M., Esawi, E., & Samuel, Y. (2020). COVID-19 public sentiment insights and machine learning for tweets classification. *Information (Switzerland), 11*(6). <https://doi.org/10.3390/info11060314>
- Saputri, M. S., Mahendra, R., & Adriani, M. (2019). Emotion Classification on Indonesian Twitter Dataset. In *Proceedings of the 2018 International Conference on Asian Language Processing, IALP 2018*. <https://doi.org/10.1109/IALP.2018.8629262>
- Sari, R. (2020). ANALISIS SENTIMEN PADA REVIEW OBJEK WISATA DUNIA FANTASI MENGGUNAKAN ALGORITMA K-NEAREST

NEIGHBOR (K-NN). *EVOLUSI: Jurnal Sains Dan Manajemen*, 8(1).  
<https://doi.org/10.31294/evolusi.v8il.7371>

Septian, J. A., Fahrudin, T. M., & Nugroho, A. (2019). Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF - IDF dan K - Nearest Neighbor. *Journal of Intelligent Systems and Computation*, (September).

Setyawati, K. S., Handoyo, A., & Palit, H. N. (2021). Aplikasi Sentiment Analysis Terhadap Pelaksanaan Pembelajaran Jarak Jauh Universitas Kristen Petra Dengan Metode Naive Bayes Classifier. *Jurnal Infra*, 1.

Suryono, S., & Taufiq Luthfi, E. (2021). Analisis sentimen pada Twitter dengan menggunakan metode Naive Bayes Classifier. *JNANALOKA*.  
<https://doi.org/10.36802/jnamaloka.2020.v1-no2-81-86>

Syariffuddin, M. (2020). Analisis Sentimen Opini Publik Mengenai Covid-19 Pada Twitter Menggunakan Metode Naive Bayes Dan Knn. *Inti Nusa Mandiri*, 15(1).

Wenando, F. A., Hayami, R., & Anggrawan, A. J. (2020). ANALISIS SENTIMEN PADA PEMERINTAHAN TERPILIH PADA PILPRES 2019 DITWITTER MENGGUNAKAN ALGORITME NAIVEBAYES. *JURTEKSI (Jurnal Teknologi Dan Sistem Informasi)*, 7(1).  
<https://doi.org/10.33330/jurteksi.v7il.851>

#### **PUSTAKA ELEKTRONIK**

Kemdikbud. (2020a). Kementerian Pendidikan dan Kebudayaan» Republik Indonesia. Retrieved June 9, 2021, from

<https://www.kemdikbud.go.id/main/blog/2020/03/surat-edaran-pencegahan-covid-19-pada-satuan-pendidikan>

Kemdikbud. (2020b). Pembelajaran Tatap Muka Daring | Blog Rumah Belajar. Retrieved July 8, 2021, from <http://pena.belajar.kemdikbud.go.id/2020/12/pembelajaran-tatap-muka-daring/>

Kompas. (2020). UPDATE 2 Desember: Ada 71.074 Suspek Terkait Covid-19 Halaman all - Kompas.com. Retrieved November 6, 2021, from <https://nasional.kompas.com/read/2020/12/02/16300291/update-2-desember-ada-71074-suspek-terkait-covid-19?page=all>

Monkeylearn. (2019). Go-to Guide for Text Classification with Machine Learning. Retrieved July 14, 2021, from <https://monkeylearn.com/blog/text-classification-machine-learning/>