

**ANALISIS SENTIMEN TERHADAP BELAJAR DARING
MENGGUNAKAN OPTIMASI NAIVE BAYES
*CLASSIFIER DENGAN ADABOOST***

SKRIPSI



**UIN SUNAN AMPEL
S U R A B A Y A**

**Disusun Oleh:
TAZKIA SHABRINA AZ-ZAHRA
H06217021**

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL
SURABAYA
2021**

PERNYATAAN KEASLIAN

Saya yang bertanda tangan di bawah ini,

Nama : TAZKIA SHABRINA AZ-ZAHRA
NIM : H06217021
Program Studi : Sistem Informasi
Angkatan : 2017

Menyatakan bahwa saya tidak melakukan plagiat dalam penulisan skripsi daya yang berjudul “ANALISIS SENTIMEN TERHADAP BELAJAR DARING MENGGUNAKAN OPTIMASI NAIVE BAYES CLASSIFIER DENGAN ADABOOST”. Apabila suatu saat nanti terbukti saya melakukan tindakan plagiat, maka saya bersedia menerima sanksi yang telah ditetapkan.

Demikian pernyataan keaslian ini saya buat dengan sebenar-benarnya.

Surabaya, 12 Agustus 2021
Yang menyatakan,




The image shows a handwritten signature "Shabrina" written over a 20 Rupiah Indonesian postage stamp. The stamp features a portrait of Soekarno and is labeled "SPULUH RUPIAH", "TGL 20", "METERAI TEMPEL", and the serial number "4068AAJX014111699".

Tazkia Shabrina Az-zahra
NIM H06217021

PERNYATAAN KEASLIAN

Saya yang bertanda tangan di bawah ini,

Nama : TAZKIA SHABRINA AZ-ZAHRA
NIM : H06217021
Program Studi : Sistem Informasi
Angkatan : 2017

Menyatakan bahwa saya tidak melakukan plagiat dalam penulisan skripsi daya yang berjudul “ANALISIS SENTIMEN TERHADAP BELAJAR DARING MENGGUNAKAN OPTIMASI NAIVE BAYES CLASSIFIER DENGAN ADABOOST”. Apabila suatu saat nanti terbukti saya melakukan tindakan plagiat, maka saya bersedia menerima sanksi yang telah ditetapkan.

Demikian pernyataan keaslian ini saya buat dengan sebenar-benarnya.

Surabaya, 12 Agustus 2021
Yang Menyatakan,



Tazkia Shabrina Az-zahra
NIM.H06217021

LEMBAR PERSETUJUAN PEMBIMBING

Skripsi oleh

NAMA : TAZKIA SHABRINA AZ-ZAHRA

JUDUL : ANALISIS SENTIMEN TERHADAP BELAJAR
DARING MENGGUNAKAN OPTIMASI *NAIVE
BAYES CLASSIFIER DENGAN ADABOOST*

NIM : H06217021

Ini telah diperiksa dan disetujui untuk diujikan.

Surabaya, 12 Agustus 2021

Dosen Pembimbing 1



Dwi Roliawati, MT
NIP. 197909272014032001

Dosen Pembimbing 2



Ahmad Yusuf, M. Kom
NIP. 199001202014031003

PENGESAHAN TIM PENGUJI SKRIPSI

Skripsi Tazkia Shabrina Az-zahra ini telah dipertahankan
di depan tim penguji skripsi
di Surabaya, 12 Agustus 2021.

Mengesahkan,

Dewan Penguji

Dosen Penguji I



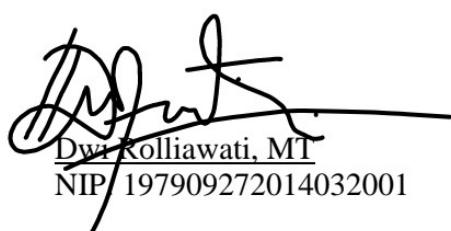
Mujib Ridwan, MT
NIP. 198604272014031004

Dosen Penguji II



Khalid, M.Kom
NIP. 197906092014031002

Dosen Penguji III



Dwi Kolliawati, MT
NIP. 197909272014032001

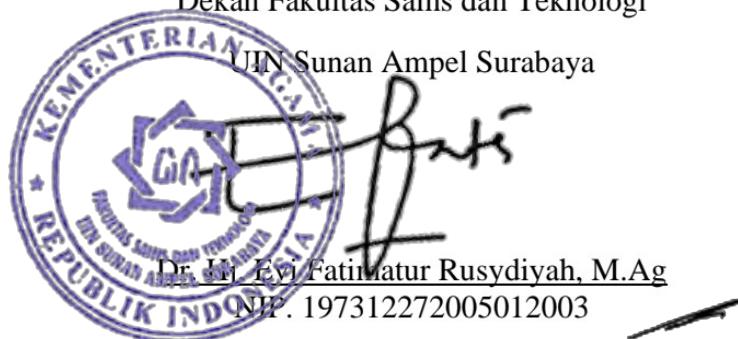
Dosen Penguji IV



Ahmad Yusuf, M. Kom
NIP. 199001202014031003

Mengetahui,

Dekan Fakultas Sains dan Teknologi





KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL SURABAYA
PERPUSTAKAAN

Jl. Jend. A. Yani 117 Surabaya 60237 Telp. 031-8431972 Fax.031-8413300
E-Mail: perpus@uinsby.ac.id

**LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademika UIN Sunan Ampel Surabaya, yang bertanda tangan di bawah ini, saya:

Nama : TAZKIA SHABRINA AZ-ZAHRA
NIM : H06217021
Fakultas/Jurusan : SAINS DAN TEKNOLOGI/ SISTEM INFORMASI
E-mail address : tazkiashabrina11@gmail.com

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Perpustakaan UIN Sunan Ampel Surabaya, Hak Bebas Royalti Non-Ekslusif atas karya ilmiah :

Sekripsi Tesis Desertasi Lain-lain (.....)
yang berjudul :

ANALISIS SENTIMEN TERHADAP BELAJAR DARING MENGGUNAKAN OPTIMASI

NAIVE BAYES CLASSIFIER DENGAN ADABOOST

beserta perangkat yang diperlukan (bila ada). Dengan Hak Bebas Royalti Non-Ekslusif ini Perpustakaan UIN Sunan Ampel Surabaya berhak menyimpan, mengalih-media/format-kan, mengelolanya dalam bentuk pangkalan data (database), mendistribusikannya, dan menampilkan/mempublikasikannya di Internet atau media lain secara **fulltext** untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan UIN Sunan Ampel Surabaya, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini yang saya buat dengan sebenarnya.

Surabaya, 15 Agustus 2021

Penulis

(TAZKIA SHABRINA AZ -ZAHRA)

ABSTRAK

ANALISIS SENTIMEN TERHADAP BELAJAR DARING MENGGUNAKAN OPTIMASI NAIVE BAYES *CLASSIFIER DENGAN ADABOOST*

Oleh:

Tazkia Shabrina Az-zahra

Kebijakan belajar daring yang berlaku sejak mewabahnya Covid-19 menuai beragam reaksi dari masyarakat. Twitter menjadi salah satu jejaring sosial yang digunakan masyarakat untuk mengungkapkan opini mengenai suatu isu. Kecenderungan sentimen masyarakat mengenai suatu isu dapat diketahui dengan melakukan analisis sentimen pada Twitter. Metode yang digunakan untuk analisis sentimen adalah *Naive Bayes Classifier* (NBC) dengan *Adaboost*. Penelitian ini bertujuan untuk mengoptimasi NBC dengan *Adaboost* untuk analisis sentimen, mengetahui performa NBC dengan *Adaboost*, dan mengetahui kecenderungan sentimen masyarakat mengenai belajar daring. Hasil penelitian menunjukkan *Adaboost* tidak dapat meningkatkan akurasi NBC, didapatkan akurasi NBC sebesar 78%, dan akurasi NBC dengan *Adaboost* sebesar 64%. Hal ini disebabkan karena NBC merupakan pengklasifikasi yang stabil sehingga sulit untuk *Adaboost* dapat meningkatkan akurasinya. Kombinasi antara data pada penelitian ini yang *imbalance* dengan NBC dan *Adaboost* juga menyebabkan penurunan akurasi. NBC dengan *Adaboost* melabelkan semua data dengan sentimen positif, hal tersebut menunjukkan bahwa NBC dengan *Adaboost* tidak dapat melakukan klasifikasi. Hasil analisis sentimen menggunakan NBC menunjukkan kecenderungan sentimen masyarakat terhadap belajar daring adalah positif dengan presentase 89,5% sedangkan sentimen negatif sebesar 10,5%.

Kata Kunci: Analisis Sentimen, Naive Bayes Classifier, Adaboost, Belajar Daring, Twitter

ABSTRACT

SENTIMENT ANALYSIS OF ONLINE LEARNING USING NAIVE BAYES CLASSIFIER OPTIMIZATION WITH ADABOOST

By:

Tazkia Shabrina Az-zahra

The online learning policy that has been in effect since the Covid-19 outbreak has drawn various reactions from the public. Twitter is one of the social networks used by the public to express opinions on an issue. The tendency of public sentiment regarding an issue can be known by conducting sentiment analysis on Twitter. The method used for sentiment analysis is the Naive Bayes Classifier (NBC) with Adaboost. This study aims to optimize NBC with Adaboost for sentiment analysis, determine NBC's performance with Adaboost, and determine the tendency of public sentiment regarding online learning. The results showed that Adaboost could not increase the accuracy of NBC, the accuracy of NBC was 78%, and the accuracy of NBC with Adaboost was 64%. This is because NBC is a stable classifier so it is difficult for Adaboost to improve its accuracy. The combination of the imbalanced data in this study with NBC and Adaboost also causes a decrease in accuracy. NBC with Adaboost labels all data with positive sentiment, it shows that NBC with Adaboost cannot do the classification. The results of sentiment analysis using NBC show that the tendency of public sentiment towards online learning is positive with a percentage of 89.5% while negative sentiment is 10.5%.

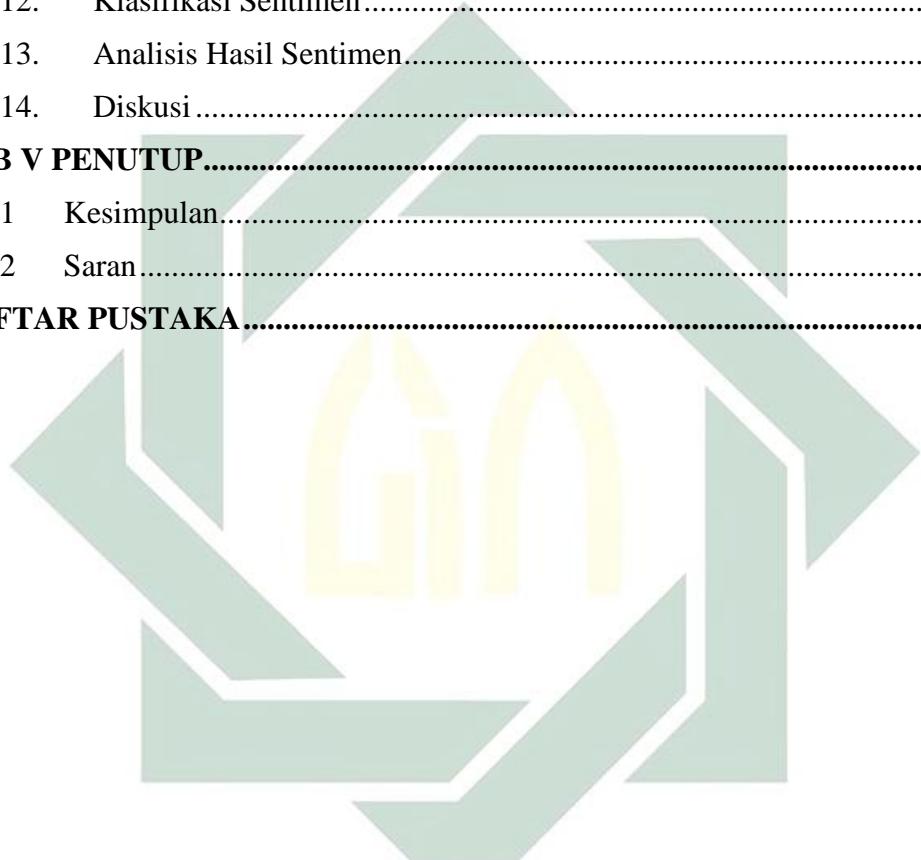
Keywords: Sentiment Analysis, Naive Bayes Classifier, Adaboost, Online Learning, Twitter

DAFTAR ISI

HALAMAN JUDUL	i
PERNYATAAN KEASLIAN	ii
LEMBAR PERSETUJUAN PEMBIMBING	iii
LEMBAR PENGESAHAN	iv
LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI	v
KATA PENGANTAR.....	vi
ABSTRAK	vii
ABSTRACT	viii
DAFTAR ISI.....	ix
DAFTAR TABEL	xii
DAFTAR GAMBAR.....	xiii
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang	1
1.2. Perumusan Masalah.....	4
1.3. Batasan Masalah.....	4
1.4. Tujuan Penelitian.....	4
1.5. Manfaat Penelitian.....	5
BAB II TINJAUAN PUSTAKA	6
2.1. Tinjauan Penelitian Terdahulu	6
2.2. Teori dasar yang digunakan	9
2.2.1. <i>Opinion Mining</i>	9
2.2.2. Analisis Sentimen	10
2.2.3. <i>Cross Validation</i>	11
2.2.4. <i>Data Preprocessing</i>	12
2.2.5. <i>TextBlob</i>	13
2.2.6. Validasi Data.....	14
2.2.7. <i>Judgement Sampling</i>	14
2.2.8. <i>Feature Extraction</i>	15
2.2.9. <i>Naïve Bayes Classifier</i> (NBC)	15
2.2.10. <i>Adaboost</i>	17
2.2.11. Evaluasi.....	18
2.2.12. <i>Confusion Matrix</i>	19

2.2.13. Visualisasi	19
2.3. Integrasi Keilmuan	20
BAB III METODOLOGI PENELITIAN	22
3.1. Jenis Penelitian	22
3.2. Sumber Data	23
3.3. Metode Penelitian	24
3.3.1. Studi Literatur	25
3.3.2. Pengumpulan Data	25
3.3.3. Penghapusan Data Duplikat	25
3.3.4. Penerjemahan Data	25
3.3.5. <i>Data Sampling</i>	26
3.3.6. <i>Data Preprocessing</i>	26
3.3.7. Pelabelan Data	30
3.3.8. Validasi Hasil Pelabelan	31
3.3.9. Visualisasi Hasil Pelabelan	31
3.3.10. <i>Feature Extraction</i>	32
3.3.11. <i>Cross Validation</i>	32
3.3.12. <i>Modeling</i>	32
3.3.13. Analisis Hasil Sentimen	33
BAB IV HASIL DAN PEMBAHASAN	34
4.1. Pengumpulan Data	34
4.2. Penghapusan Data Duplikat	35
4.3. Penerjemahan Data	35
4.4. <i>Data Sampling</i>	35
4.5. <i>Data Preprocessing</i>	36
4.5.1. <i>Data Cleaning</i>	36
4.5.2. <i>Case Folding</i>	36
4.5.3. <i>Tokenizing</i>	37
4.5.4. <i>Replace</i>	37
4.5.5. <i>Remove Punctuation</i>	38
4.5.6. <i>Lemmatization</i>	39
4.5.7. <i>Stopwords</i>	39
4.5.8. <i>Filtering</i>	40
4.6. Pelabelan Data	41

4.7.	Validasi Hasil Pelabelan.....	41
4.8.	Visualisasi Hasil Pelabelan	43
4.9.	<i>Feature Extraction</i>	45
4.10.	<i>Cross Validation</i>	46
4.11.	<i>Modeling</i>	47
4.11.1.	<i>Model Training</i>	47
4.11.2.	<i>Model Evaluation</i>	47
4.12.	Klasifikasi Sentimen.....	50
4.13.	Analisis Hasil Sentimen.....	51
4.14.	Diskusi	55
BAB V	PENUTUP.....	56
5.1	Kesimpulan.....	56
5.2	Saran.....	57
DAFTAR PUSTAKA		58

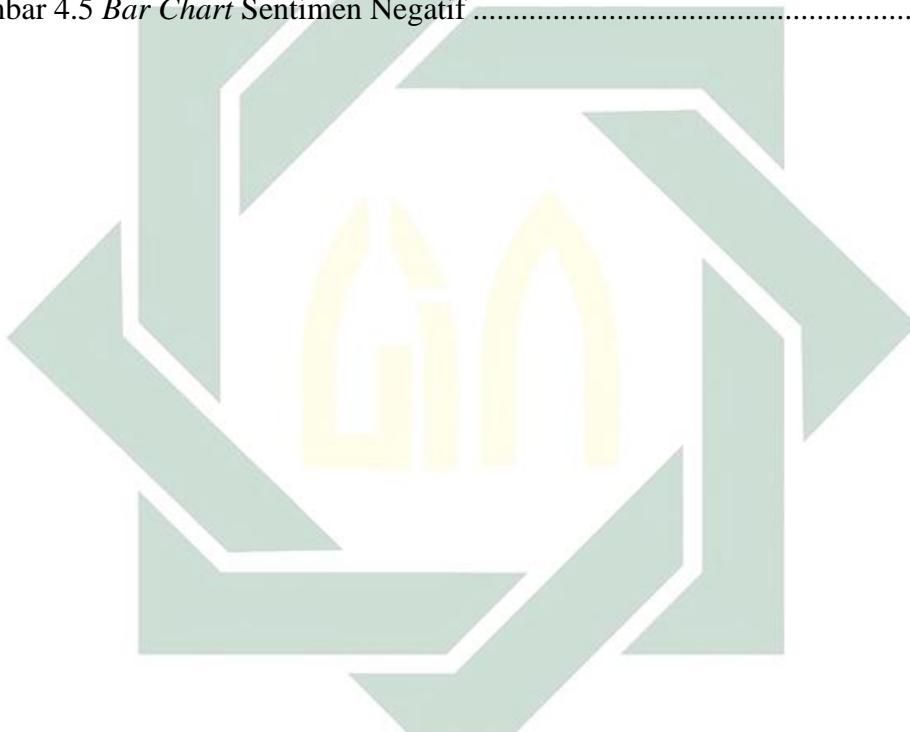


DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu	6
Tabel 2.2 <i>Confusion Matrix</i> (Sumber: Cahyani, 2020).....	19
Tabel 3.1 Hasil Pengumpulan Data	25
Tabel 3.2 Hasil Penerjemahan Data.....	26
Tabel 3.3 Hasil <i>Data Cleaning</i>	26
Tabel 3.4 Hasil <i>Case Folding</i>	27
Tabel 3.5 Hasil <i>Tokenizing</i>	27
Tabel 3.6 Hasil Proses <i>Replace</i>	27
Tabel 3.7 Hasil <i>Remove Punctuation</i>	28
Tabel 3.8 Hasil <i>Lemmatization</i>	28
Tabel 3.9 Hasil <i>Stopwords</i>	29
Tabel 3.10 Hasil <i>Filtering</i>	30
Tabel 3.11 Contoh Hasil Polaritas dan Subjektivitas.....	30
Tabel 3.12 Nilai Polaritas dan Label.....	31
Tabel 3.13 Contoh Hasil Pelabelan.....	31
Tabel 3.14 Contoh Hasil TF-IDF.....	32
Tabel 4.1 Jumlah <i>Tweets</i> Terkumpul	34
Tabel 4.2 Hasil Pengumpulan Data	34
Tabel 4.3 Hasil Penerjemahan Data.....	35
Tabel 4.4 Hasil <i>Data Cleaning</i>	36
Tabel 4.5 Hasil <i>Case Folding</i>	36
Tabel 4.6 Hasil <i>Tokenizing</i>	37
Tabel 4.7 Hasil <i>Replace</i>	38
Tabel 4.8 Hasil <i>Remove Punctuation</i>	38
Tabel 4.9 Hasil <i>Lemmatization</i>	39
Tabel 4.10 Hasil <i>Stopwords</i>	40
Tabel 4.11 Hasil <i>Filtering</i>	41
Tabel 4.12 Hasil Pelabelan Data	41
Tabel 4.13 Validasi Pelabelan.....	42
Tabel 4.14 Akurasi Pelabelan	43
Tabel 4.15 Hasil TF-IDF.....	45
Tabel 4.16 <i>Cross Validation Score</i>	46
Tabel 4.17 Hasil Akurasi dengan Berbagai Nilai Min_df	47
Tabel 4.18 <i>Confusion Matrix</i> NBC	48
Tabel 4.19 Hasil Evaluasi NBC	48
Tabel 4.20 <i>Confusion Matrix</i> NBC dengan <i>Adaboost</i>	49
Tabel 4.21 NBC dengan <i>Adaboost</i>	49
Tabel 4.22 Hasil Evaluasi <i>Decision Tree</i> dengan <i>Adaboost</i>	50
Tabel 4.23 Hasil Klasifikasi Data	51

DAFTAR GAMBAR

Gambar 2.1 <i>10-fold Cross Validation</i> (Berrar, 2018)	12
Gambar 2.2 <i>Word Cloud</i> (Castellà & Sutton, 2014)	20
Gambar 2.3 <i>Bar Chart</i> (Martinez, 2015)	20
Gambar 3.1 Grafik Belajar Daring (Google Trends, 2020)	22
Gambar 3.2 <i>Related Topics</i> (Google Trends, 2020)	23
Gambar 3.3 Metode Penelitian	24
Gambar 4.1 <i>Bar Chart</i> Hasil Sentimen.....	43
Gambar 4.2 <i>Word Cloud</i> Sentimen Positif	44
Gambar 4.3 <i>Word Cloud</i> Sentimen Negatif	44
Gambar 4.4 <i>Bar Chart</i> Sentimen Positif.....	52
Gambar 4.5 <i>Bar Chart</i> Sentimen Negatif	52



BAB I

PENDAHULUAN

1.1. Latar Belakang

Coronavirus disease (Covid-19) adalah wabah yang sangat berbahaya, berbagai penyakit dapat disebabkan oleh Covid-19 seperti Pneumonia, sindrom pernapasan akut, gagal ginjal, hingga kematian (Tosepu et al., 2020). Kasus pertama Covid-19 di Indonesia ditemukan pada tanggal 2 Maret 2020, hingga tanggal 23 Maret 2020 kasus terkonfirmasi positif Covid-19 terus bertambah menjadi 65 kasus (covid19.go.id, 2020). Pada tanggal 24 Maret 2020 pemerintah mengeluarkan Surat Edaran Nomor 4 Tahun 2020 Tentang Pelaksanaan Kebijakan Pendidikan dalam Masa Darurat Penyebaran *Coronavirus Disease* (Covid-19), diberlakukan proses belajar dari rumah yang dilaksanakan dengan beberapa ketentuan, salah satunya penerapan pembelajaran dalam jaringan (daring) atau jarak jauh. Kebijakan menutup sekolah bertujuan untuk mengurangi penyebaran Covid-19 (Sukendro et al., 2020).

Kebijakan belajar daring akan memberikan dampak besar pada para pelajar dan mahasiswa. Berdasarkan data pada Badan Pusat Statistik (BPS) pada tahun 2019, Indonesia memiliki sekitar 45.073.800 siswa dari jenjang pendidikan SD hingga SMA dan SMK. Jumlah mahasiswa di bawah Kementerian Riset, Teknologi dan Pendidikan Tinggi terdapat sekitar 7.339.164 orang yang terdiri dari mahasiswa Perguruan Tinggi Negeri (PTN) sebanyak 2.928.403 dan Perguruan Tinggi Swasta (PTS) 4.410.761 orang, sedangkan di bawah Kementerian Agama berjumlah 999.979 orang untuk PTN sebanyak 613.685 orang dan PTS 386.294 orang (Badan Pusat Statistik, 2020). Jutaan mahasiswa dan pelajar ini memiliki berbagai latar belakang dan tinggal di lokasi yang berbeda. Terdapat beberapa tantangan saat belajar daring, seperti lemahnya sinyal selular saat di tempat tinggal masing-masing (Sadikin & Hamidah, 2020), membutuhkan kuota internet lebih besar (Sujarwo et al., 2020), kurangnya komunikasi antara siswa dan guru (Mostafa, 2021), dan berkurangnya efektivitas pembelajaran dibandingkan dengan pembelajaran tatap muka (Dwi et al., 2020).

Peraturan belajar daring membuat penggunaan *social network* meningkat pesat. *Social network* dapat dimanfaatkan untuk belajar daring, dan berinteraksi dengan orang lain. *Social network* merupakan tempat berkumpulnya entitas sosial dapat saling berinteraksi dengan bertukar pendapat mengenai suatu hal (Panda et al., 2018). Twitter adalah salah satu *social network* populer di Indonesia, sebanyak 56% pengguna internet di Indonesia merupakan pengguna Twitter (Kemp, 2020). Setiap unggahan yang dibagikan di Twitter dapat dimanfaatkan untuk berbagai hal seperti, mengetahui produk favorit masyarakat, kecenderungan masyarakat mengenai suatu isu dan lain-lain. Hal-hal tersebut dapat dilakukan dengan *opinion mining*.

Opinion mining adalah bidang studi yang menganalisis opini, sentimen, penilaian, perilaku, dan emosi seseorang dari tulisan (Liu, 2012). Analisis sentimen termasuk dalam *opinion mining*, analisis sentimen berfungsi untuk menganalisis opini individu dari sebuah teks untuk ditentukan sentimennya (Fauzi, 2018). Analisis sentimen dapat menghasilkan informasi untuk membantu dalam pengambilan keputusan (Fitri et al., 2019). Pada penelitian yang dilakukan oleh Vidya, analisis sentimen dimanfaatkan untuk mengetahui reputasi *brand* perusahaan sehingga perusahaan dapat menetapkan rencana untuk meningkatkan persepsi *brand* dari sudut pandang masyarakat (Vidya, 2015). Analisis sentimen dapat diterapkan pada unggahan di *social network*, namun diperlukan Google Trends untuk mencari kata kunci yang tepat agar data yang diambil sesuai dengan topik analisis. Google Trends telah digunakan pada lebih dari 650 *paper* (Jun et al., 2018). Google Trends merupakan situs yang menampilkan data berdasarkan mesin pencari Google yang menunjukkan minat pada topik tertentu dan menampilkan kata yang berhubungan dengan topik tersebut (Google.com, 2020).

Banyak metode yang dapat digunakan untuk analisis sentimen, seperti metode pendekatan *machine learning* (ML), berbasis leksikon, dan pendekatan *hybrid* (Sharma et al., 2020). Pendekatan *machine learning* dibagi menjadi *supervised learning* dan *unsupervised learning* (Medhat et al., 2014). Pendekatan *supervised learning* memiliki kemampuan beradaptasi dan berlatih membuat model terhadap data untuk tujuan dan konteks yang spesifik, tetapi memerlukan data berlabel yang berpotensi mahal (D'Andrea et al., 2015). *Unsupervised learning*

dapat digunakan untuk pelabelan data. Metode berbasis leksikon tidak memerlukan data berlabel dan prosedur pembelajaran, namun metode ini memerlukan sumber daya linguistik yang kuat yang mana tidak selalu tersedia (Devika et al., 2016). Kelebihan metode *hybrid* adalah akurasi yang tinggi (Gupta & Joshi, 2020), tetapi ulasannya memiliki banyak *noise* (kata-kata yang tidak berhubungan dengan subjek yang diulas) (Auria & Moro, 2008).

Klasifikasi sentimen pada dasarnya merupakan masalah klasifikasi teks, oleh sebab itu metode *supervised learning* yang ada dapat langsung diimplementasikan, seperti klasifikasi *Support Vector Machines* (SVM) atau *Naïve Bayes Classifier* (NBC) (Liu, 2015). SVM dapat menyematkan data ke ruang berdimensi lebih tinggi sehingga lebih banyak data yang dapat dipisahkan (Russell & Norvig, 2010). Tetapi, biaya komputasi SVM mahal, dan memerlukan waktu *training* yang lama (Yue et al., 2019). NBC dapat menangani data yang memiliki banyak *noise* (Russell & Norvig, 2010), hal ini dibuktikan dengan penelitian yang menggunakan algortima NBC untuk klasifikasi SMS *spam* dan menghasilkan akurasi sebesar 0.94 (Setifani et al., 2020). Selain itu, NBC memiliki performa secara keseluruhan yang baik dan waktu pemrosesan jauh lebih cepat dibandingkan SVM (Colas & Brazdil, 2006). Kekurangannya, akurasi NBC termasuk rendah sehingga disebut *weak learner*, tetapi akurasi dapat ditingkatkan menggunakan algoritma *Adaboost* (Korada, 2012). *Adaboost* terbukti mampu meningkatkan akurasi NBC. Pada penelitian yang dilakukan Utami dan Wahono mengenai analisis sentimen *review* restoran, penerapan *Adaboost* dapat meningkatkan akurasi NBC dan *information gain* dari 73,00% menjadi 81,50% (Utami & Wahono, 2015).

Pada penelitian ini akan dilakukan analisis sentimen masyarakat terhadap belajar daring menggunakan algoritma *Naïve Bayes Classifier* (NBC) yang dioptimasi dengan *Adaboost*. Analisis sentimen terhadap belajar daring menjadi penting mengingat banyaknya jumlah pelajar dan mahasiswa yang terdampak dari kebijakan ini, selain itu analisis sentimen yang berasal dari Twitter dapat dimanfaatkan masyarakat terkhusus pemerintah untuk mengetahui pandangan masyarakat mengenai kebijakan yang diambil sehingga kedepannya dapat menjadi pertimbangan dalam pengambilan keputusan. Penentuan kata kunci untuk mendapatkan data Twitter didasari pada topik yang berkaitan dengan belajar daring

pada Google Trends. Kata yang digunakan adalah “belajar daring”, “daring rumah”, “belajar *online*”, “*online* siswa”, “*corona* sekolah”, “*corona* kuliah”, “kuliah *online*”, “sekolah *online*”, dan “*new normal*”. NBC dipilih untuk analisis sentimen, dikarenakan dapat bekerja dengan data yang memiliki banyak *noise* sehingga cocok untuk menangani data dari Twitter yang memiliki banyak *noise*. Tetapi, NBC tergolong *weak learner* sehingga proses pelatihan NBC perlu ditingkatkan dengan metode optimasi *Adaboost*. Pelatihan bertahap dengan metode NBC dan *Adaboost* bisa menghasilkan pemodelan dengan akurasi yang bagus dibandingkan NBC tanpa *Adaboost*.

1.2. Perumusan Masalah

1. Bagaimana implementasi *Naïve Bayes Classifier* dengan *Adaboost* untuk menganalisis sentimen terhadap belajar daring?
 2. Bagaimana performa *Naïve Bayes Classifier* dengan *Adaboost* untuk analisis sentimen?
 3. Bagaimana kecenderungan masyarakat Indonesia terhadap belajar daring berdasarkan hasil analisis sentimen?

1.3. Batasan Masalah

Batasan dalam penelitian ini sangat dibutuhkan agar pembahasannya jelas dan tidak menyimpang dari judul yang telah ditetapkan. Adapun batasan masalah dalam penelitian ini, yaitu:

1. Data diperoleh dari Twitter sejak bulan Juli 2020 hingga Desember 2020.
 2. Data yang dianalisis merupakan teks dalam Bahasa Indonesia yang diterjemahkan ke Bahasa Inggris.
 3. Tidak menangani teks sarkasme.
 4. *Output* klasifikasi sentimen dikelompokkan menjadi sentimen positif, dan negatif.

1.4. Tujuan Penelitian

1. Mengimplementasikan *Naïve Bayes Classifier* dengan *Adaboost* untuk menganalisis sentimen terhadap belajar daring.
 2. Mengetahui performa *Naïve Bayes Classifier* dengan *Adaboost*.
 3. Menganalisis kecenderungan sentimen masyarakat mengenai belajar daring.

1.5. Manfaat Penelitian

Berikut adalah manfaat yang dapat diambil dari penelitian ini:

1. Manfaat Akademis

Penelitian ini dapat memberikan wawasan mengenai *Adaboost* yang tidak dapat meningkatkan hasil *Naïve Bayes Classifier* dikarenakan NBC merupakan pengklasifikasi yang stabil dengan *high bias* dan *low variance* sehingga sulit untuk *Adaboost* meningkatkan akurasinya, juga kombinasi antara data pada penelitian ini yang *imbalance* dengan NBC dan *Adaboost* juga menyebabkan penurunan akurasi. Selain itu, penelitian ini dapat memberikan pandangan mengenai analisis sentimen masyarakat terhadap belajar daring. Pada proses *data preprocessing* juga ditambahkan *stopwords* untuk menghilangkan kata yang tidak memiliki efek pada sentimen.

2. Manfaat Praktis

Penelitian ini dapat menjadi referensi ilmiah yang dapat membantu pemerintah dalam pengambilan kebijakan saat pembelajaran daring di masa pandemi.

BAB II

TINJAUAN PUSTAKA

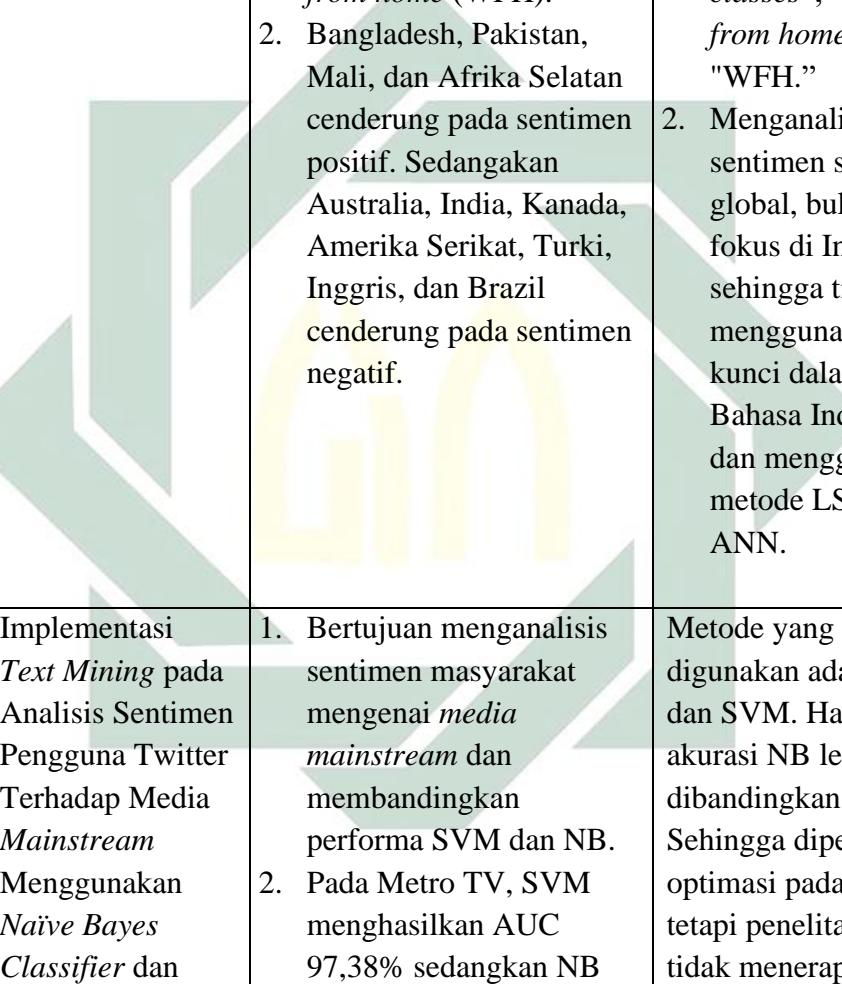
2.1. Tinjauan Penelitian Terdahulu

Tinjauan penelitian terdahulu dilakukan untuk mengetahui metode-metode yang telah dilakukan peneliti lain untuk memecahkan masalah yang sejenis serta mengetahui pembeda antara penelitian ini dengan penelitian sebelumnya. Pada Tabel 2.1 terdapat penelitian yang sudah dilakukan sebelumnya.

Tabel 2.1 Penelitian Terdahulu

No.	Judul	Hasil	Pembeda
1.	<i>Sentiment Analysis on Utilizing Online Transportation of Indonesian Customers using Tweets in the Normal Era and The Pandemic Covid-19 Era with Support Vector Machine</i> (Jaman et al., 2020)	<p>1. Bertujuan membandingkan analisis sentimen pengguna transportasi <i>online</i> pada sebelum dan saat pandemi <i>corona</i>.</p> <p>2. Sentimen pelanggan pada era normal dan saat era Covid-19 condong pada sentimen negatif.</p> <p>3. Hasil validasi sentimen pada era normal, Gojek mendapatkan sentimen negatif 363, dan positif 111, sedangkan Grab mendapat sentimen negatif 323, dan positif 140.</p> <p>4. Pada era Covid-19, Gojek mendapatkan sentimen negatif sebanyak 443, dan positif 127, sedangkan Grab mendapat sentimen negatif 429 dan positif 226.</p>	<p>1. Penelitian ini membahas topik terkait perilaku pada wabah Covid-19 yang berfokus pada layanan transportasi <i>online</i>, tidak membahas mengenai bidang pendidikan.</p> <p>2. Penelitian ini menggunakan metode SVM.</p>

Tabel 2.1 Penelitian Terdahulu lanjutan

No.	Judul	Hasil	Pembeda
2.	<p><i>Global Sentiment Analysis Of COVID-19 Tweets Over Time</i> (Mansoor et al., 2020)</p> 	<ol style="list-style-type: none"> Bertujuan menganalisis sentimen masyarakat di dunia berbasis data Twitter terkait pandemi virus <i>corona</i> khususnya terhadap pembelajaran <i>online</i> dan skenario <i>work from home</i> (WFH). Bangladesh, Pakistan, Mali, dan Afrika Selatan cenderung pada sentimen positif. Sedangkan Australia, India, Kanada, Amerika Serikat, Turki, Inggris, dan Brazil cenderung pada sentimen negatif. 	<ol style="list-style-type: none"> Fokus topik, pada pembelajaran <i>online</i> dan WFH, dengan kata kunci dalam bahasa inggris “<i>online learning</i>”, “<i>online classes</i>”, “<i>work from home</i>” dan “WFH.” Menganalisis sentimen secara global, bukan fokus di Indonesia sehingga tidak menggunakan kata kunci dalam Bahasa Indonesia, dan menggunakan metode LSTM dan ANN.
3.	<p>Implementasi <i>Text Mining</i> pada Analisis Sentimen Pengguna Twitter Terhadap Media <i>Mainstream</i> Menggunakan <i>Naïve Bayes Classifier</i> dan <i>Support Vector Machine</i> (Kurniawan, 2017)</p>	<ol style="list-style-type: none"> Bertujuan menganalisis sentimen masyarakat mengenai <i>media mainstream</i> dan membandingkan performa SVM dan NB. Pada Metro TV, SVM menghasilkan AUC 97,38% sedangkan NB menghasilkan 82,36%. 	<p>Metode yang digunakan adalah NB dan SVM. Hasil akurasi NB lebih kecil dibandingkan SVM. Sehingga diperlukan optimasi pada NB, tetapi penelitian ini tidak menerapkan optimasi.</p>

Tabel 2.1 Penelitian Terdahulu lanjutan

No.	Judul	Hasil	Pembeda
4.	<i>COVID-19: Social Media Sentiment Analysis on Reopening</i> (Samuel et al., 2020)	<ol style="list-style-type: none"> Bertujuan menganalisis sentimen masyarakat AS mengenai <i>reopening</i> setelah <i>lockdown</i> berbasis data Twitter. Hasil menunjukkan, <i>tweets</i> dengan sentimen netral sebesar 50,87%, sentimen positif sebesar 39,38% dan negatif sebesar 9,75%. 	COVID-19: <i>Social Media Sentiment Analysis on Reopening</i> menggunakan TextBlob untuk menganalisa sentimen masyarakat di AS.
5.	<i>Teknologi Opinion Mining untuk Mendukung Strategic Planning</i> (Rilliawati et al., 2020)	<ol style="list-style-type: none"> Bertujuan menganalisis tanggapan masyarakat mengenai UIN Sunan Ampel Surabaya sebagai bahan pertimbangan untuk menentukan <i>strategic planning</i>. Sentimen publik terhadap UIN Sunan Ampel Surabaya sebesar 97,54% cenderung netral, sedangkan sentimen positif sebesar 2,16%, dan sentimen negatif sebesar 0,34%. 	Topik bahasan analisis sentimen terhadap UIN Sunan Ampel Surabaya menerapkan pendekatan Vader untuk menganalisis sentimen positif, negatif dan netral.
6.	<i>An Analysis of University Students' Perspective on Online Learning in the Midst of Covid-19 Pandemic</i> (Sujarwo et al., 2020)	<ol style="list-style-type: none"> Bertujuan untuk mengetahui sudut pandang mahasiswa terhadap pembelajaran <i>online</i> saat pandemi. Pembelajaran <i>online</i> mendapatkan sentimen positif dari mahasiswa, karena dianggap cukup membantu terlepas dari tantangan yang dihadapi. 	Sudut pandang mahasiswa Universitas Megarezky menjadi topik pada <i>paper</i> ini, sentimen mahasiswa didapatkan dari hasil wawancara dan kuisioner.

Tabel 2.1 Penelitian Terdahulu lanjutan

No.	Judul	Hasil	Pembeda
7.	Twitter <i>Sentiment Analysis</i> Terhadap <i>Brand Reputation</i> : Studi Kasus PT.XL Axiata TBk (Vidya, 2015)	<p>1. Bertujuan untuk membandingkan algoritma SVM, NB, dan DT untuk menganalisis sentimen Twitter.</p> <p>2. Hasil ROC menunjukkan SVM memiliki performa terbaik dengan AUC 0,854, sedangkan NB mendapatkan 0,5, dan DT mendapatkan 0,5.</p>	SVM mendapatkan akurasi sebesar 0,854, sedangkan NB mendapatkan akurasi 0,5 sehingga diperlukan <i>boosting</i> untuk meningkatkan akurasinya.

Pada penelitian sebelumnya dilakukan analisis sentimen yang berkaitan dengan wabah Covid-19 dan juga yang berkaitan dengan penggunaan metode *Naïve Bayes Classifier* (NBC). Tetapi, penelitian-penelitian tersebut belum membahas mengenai topik belajar daring di Indonesia yang berdampak pada banyak orang, serta belum membahas mengenai optimasi metode, khususnya metode NBC dengan *Adaboost*. Sehingga hal tersebut menjadi pembeda antara penelitian ini dengan penelitian-penelitian sebelumnya.

2.2. Teori dasar yang digunakan

2.2.1. *Opinion Mining*

Menurut Benamara, opini adalah ungkapan bahasa subjektif dari entitas (orang, organisasi, dll) untuk menilai sesuatu (topik, objek, orang, dan sebagainya), memposisikannya pada skala terpolarisasi yang berkaitan dengan norma sosial (seperti penilaian keindahan) atau norma moral (seperti baik atau buruk) yang ada (Karoui et al., 2019). Opini seperti kalimat berisi faktor keyakinan yang kebenarannya belum jelas. Sentimen seseorang dapat diilustrasikan sebagai kejadian, entitas, dan properti, jika dinilai lebih jauh menggunakan pendekatan subjektif, dapat dianggap sebagai opini (S. Bhatia et al., 2020). *Opinion mining* adalah mengenai penentuan suatu teks berisi opini atau tidak (Giachanou & Crestani, 2016). Tujuan *opinion mining* untuk mengekstrak dan menangani opini mengenai suatu hal (Tsytsarau & Palpanas, 2012). Terdapat tiga tingkatan pada *opinion mining* (S. Bhatia et al., 2020):

1. Tingkat Dokumen

Pada tingkatan ini, sentimen dalam dokumen secara keseluruhan dibedakan menjadi positif atau negatif. Pendekatan secara umum dilakukan dengan mencari polaritas sentimen yang terdapat pada setiap kalimat atau kata lalu digabungkan untuk memperoleh polaritas dokumen (A. & Sonawane, 2016).

2. Tingkat Kalimat

Tingkatan kalimat lebih dalam daripada tingkatan dokumen. Pada tingkatan ini, setiap kalimat ditentukan sentimennya, termasuk positif, negatif atau netral, netral berarti tidak ada opini (Liu, 2012). Setiap kalimat dipandang sebagai unsur terpisah, dan polaritas dihitung dari satu kalimat (S. Bhatia et al., 2020).

3. Tingkat Aspek/*Feature*

Pada tingkat ini, dilakukan pelabelan pada setiap kata dengan sentimennya dan mengidentifikasi hal yang menjadi sasaran sentimen (A. & Sonawane, 2016). Opini langsung diperhatikan pada tingkat aspek, didasari konsep, bahwa opini berisi sentimen (positif atau negatif) dan target pendapat (Liu, 2012).

2.2.2. Analisis Sentimen

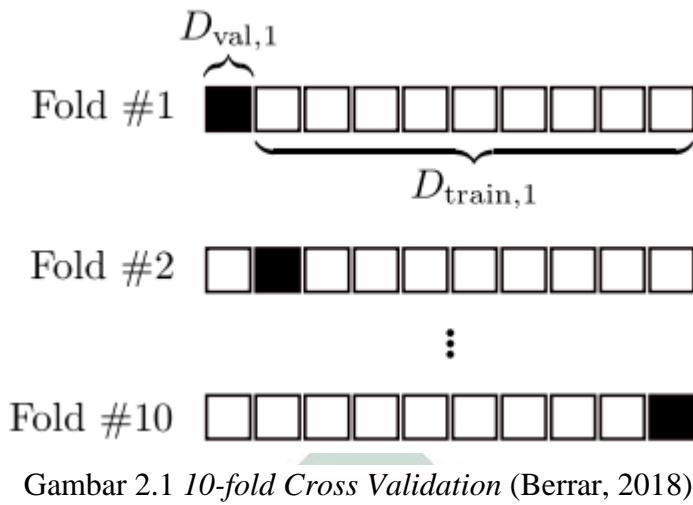
Tujuan dari *opinion mining* (OM) dan *sentiment analysis* (SA) adalah untuk mencari informasi opini dan polaritas sentimen, biasanya OM dan SA digunakan untuk mengungkapkan makna yang sama (Giachanou & Crestani, 2016). Terdapat perbedaan yang sangat tipis antara *opinion mining* dan *sentiment analysis*, *opinion mining* adalah mengenai penentuan suatu teks berisi opini atau tidak sedangkan *sentiment analysis* mencari polaritas sentimen dari opini sehingga dapat ditentukan opini tersebut termasuk dalam sentimen positif atau negatif (Tsytsarau & Palpanas, 2012). Selain OM dan SA, terdapat banyak nama lain dengan tugas yang sedikit berbeda seperti, *opinion extraction*, *subjectivity analysis*, *sentiment mining* dan sebagainya, namun sekarang semuanya ada dibawah naungan *sentiment analysis* dan *opinion mining* (Liu, 2012).

Analisis sentimen dapat dimanfaatkan untuk berbagai hal, dari sisi komersial analisis sentimen dapat memberikan bahan pertimbangan dan rekomendasi *online* untuk penjual dan pembeli, selain itu dapat dimanfaatkan oleh *platform e-commerce* untuk mengevaluasi produk atau layanan mereka (Yue et al., 2019). Selain, dalam hal bisnis, analisis sentimen dapat digunakan untuk menganalisis topik politik, sosial dan budaya. Analisis dari komentar-komentar diinternet akan sangat membantu pelaku politik, aktivis sosial dan budaya agar dapat mengambil keputusan secara tepat untuk menumbuhkan kehidupan sosial masyarakat (Hemmatian & Sohrabi, 2019).

2.2.3. *Cross Validation*

Cross validation merupakan teknik intensif komputer yang memanfaatkan semua contoh data yang tersedia sebagai contoh pelatihan dan pengujian (Bengio & Grandvalet, 2003). Banyak jurnal ilmiah mengenai klasifikasi sentimen data Twitter menggunakan *cross validation* untuk memperkirakan kinerja pengklasifikasi sentimen (Mozetić et al., 2018). *Cross validation* dapat digunakan untuk menghitung *average error rate* pada *dataset* latih dan uji (Russell & Norvig, 2010). Pada *cross validation* dihasilkan beberapa partisi, memungkinkan setiap sampel digunakan beberapa kali dengan tujuan untuk meningkatkan keandalan statistik dari hasil (Ramezan et al., 2019). *Cross validation* mempartisi data menjadi beberapa himpunan yang secara bergantian akan difungsikan sebagai data latih dan data uji. Prosedur pergantian data latih dan data uji akan terus berlanjut hingga semua bagian akan difungsikan sekali sebagai data uji (Kusrorong et al., 2019).

Cross validation memiliki beberapa macam, salah satunya adalah *k-fold cross validation*. *K-fold cross validation* membagi *dataset* menjadi k bagian, lalu menggunakan setiap bagian satu kali sebagai *test set* dan data lainnya sebagai *train set*. Kelebihan metode ini adalah semua data yang terkumpul dapat digunakan (Cielen et al., 2016). Pada penelitian analisis sentimen dengan menggunakan algoritma *Support Vector Machine*, *Decision Tree*, dan *Naïve Bayes* diterapkan *cross validation* dengan nilai $k=10$ menghasilkan akurasi paling optimal (Vidya, 2015). Pada Gambar 2.1 merupakan ilustrasi dari proses *10-fold cross validation*.



Gambar 2.1 *10-fold Cross Validation* (Berrar, 2018)

2.2.4. *Data Preprocessing*

Data preprocessing merupakan teknik *data mining* yang melibatkan pengubahan dari data mentah ke format yang dapat dipahami, karena data asli seringkali tidak lengkap, tidak konsisten, atau salah. *Data preprocessing* mengatasi masalah-masalah tersebut dan memastikan bahwa proses selanjutnya bebas dari kesalahan (P. Bhatia, 2019). Tahap *preprocessing* bertujuan untuk merepresentasikan setiap dokumen menjadi *feature vector* yang memisahkan kalimat menjadi kata individual (Srividhya & Anitha, 2010). Berikut adalah beberapa proses pada *data preprocessing*:

1. Data Cleaning

Data cleaning merupakan proses yang penting karena kesalahan yang terjadi saat proses pengumpulan data. Terdapat data yang hilang atau kesalahan yang mungkin terjadi saat proses mengumpulkan data (Aggarwal, 2015). *Data cleaning* berfungsi untuk menghilangkan *noise* dan data yang tidak konsisten (Han et al., 2012).

2. Case Folding

Case folding berfungsi untuk menyamakan huruf dalam dokumen. *Case folding* diperlukan agar pencarian dapat lebih mudah dilakukan dan seluruh teks pada dokumen dapat menjadi suatu bentuk standar (Vidya, 2015).

3. Tokenizing

Tokenizing berfungsi memecah kalimat menjadi bagian-bagian yang

memiliki makna dan mengidentifikasi entitas dalam kalimat (Ramachandran & Parvathi, 2019).

4. Remove Punctuation

Remove punctuation berfungsi untuk menghapus tanda baca seperti koma (,), titik koma (;), tanda hubung (-), dan lain sebagainya (Aggarwal, 2015).

5. Lemmatization

Lemmatization berfungsi untuk mengembalikan kata ke bentuk dasarnya dengan penggunaan kosa kata dan analisis morfologis kata (Manning et al., 2008).

6. Stopwords

Stopwords merupakan kata-kata yang sering muncul tetapi tidak terlalu diskriminatif untuk model seperti “*a*”, “*an*”, “*the*” dan kata-kata lain yang umum muncul pada dokumen tetapi memberikan sedikit informasi mengenai isi dari dokumen (Aggarwal, 2015).

7. Document Filtering

Document filtering adalah langkah untuk menghapus komponen dari dokumen yang tidak relevan untuk proses klasifikasi (Vidya, 2015).

2.2.5. TextBlob

TextBlob merupakan *library* Python untuk memproses data tekstual. Tujuan dari TextBlob adalah untuk menyediakan akses untuk operasi pemrosesan teks umum melalui *familiar interface*. Objek TextBlob dapat diperlakukan seperti Python *strings* yang memperlajari cara melakukan NLP (*Natural Language Processing*) (Loria, 2020a). TextBlob merupakan modul yang sangat efektif untuk pelabelan *tweets* dengan Python dengan memberi label sentimen positif, negatif, dan netral pada *tweets*. Analisis sentimen dapat dilakukan menggunakan polaritas dan subjektivitas pada TextBlob.

Polaritas mengacu pada penentuan sentimen cenderung pada positif, netral, atau negatif (Kaur & Sharma, 2020). Polaritas berkisar antara -1 hingga +1, dimana -1 hingga 0 adalah sentimen negatif, rentang 0 hingga +1 sebagai sentimen positif, dan 0 adalah netral. Polaritas kalimat dihitung sebagai jumlah dari polaritas semua kata dibagi dengan jumlah kata. TextBlob menangani kata negasi dengan cara

polaritas kata dikalikan dengan -0.5. Kata pengubah juga ditangani oleh TextBlob seperti kata *very* dengan memaksimalkan polaritas kata yang berkaitan dengannya. Objektivitas mengacu pada fakta, sedangkan subjektivitas mengacu pada opini atau emosi individu. Subjektivitas berkisar antara 0.0 hingga 1.0 dimana 0.0 sangat objektif dan 1.0 sangat subjektif (Loria, 2020b) (Schumacher, 2015).

2.2.6. Validasi Data

Validasi data hasil pelabelan dilakukan secara manual dengan melibatkan manusia. Validasi manusia sangat penting untuk penelitian ini karena sifat komunikasi yang rumit dan emotif tidak mudah dipahami oleh mesin (Lappeman et al., 2020). Sentimen analisis memiliki banyak tantangan, seperti ambiguitas, *slang*, dan konteks kalimat. Tantangan terbesar terhadap akurasi sentimen analisis adalah tingkat kesepakatan antara mesin dengan manusia (Kirilenko et al., 2018). Tingginya subjektivitas individu dalam menganalisis sentimen mengakibatkan sentimen analisis teks tidak akan pernah sempurna. Beberapa penelitian mengungkapkan tingkat kesesuaian yang realistik antara manusia dan mesin berkisar antara 70% hingga 80% (Ellis, 2014).

2.2.7. Judgement Sampling

Teknik *non-probability* dapat digunakan untuk penelitian sentimen analisis (Lappeman et al., 2020). Terdapat berbagai macam teknik *non-probability sampling*, salah satunya adalah *judgement sampling*. *Judgement sampling* didasari pilihan yang disengaja dan bukan termasuk proses acak. *Judgement sampling* digunakan pada *sample* kecil dari populasi yang dipahami dengan baik serta terdapat metode yang jelas untuk memilih *sampel* (National Audit Office, 2010). Kelebihan dari *judgement sampling* adalah tidak memerlukan banyak biaya, nyaman, cepat, dan ideal untuk desain penelitian eksplorasi. Sedangkan, kekurangannya adalah tidak memungkinkan generalisasi, dan subjektif (Taherdoost, 2016).

2.2.8. Feature Extraction

TF-IDF merupakan salah satu metode *feature extraction*. TF-IDF adalah singkatan dari *term frequency-inverse document frequency* untuk menghitung bobot kata berdasarkan seberapa sering kata tersebut muncul pada dokumen (VanderPlas, 2016). *Term frequency* (TF) adalah banyaknya kemunculan kata t dalam dokumen kata d , dinotasikan dengan $(tf_{t,d})$ yang mana t,d menyatakan kata dan dokumen dalam urutan (Manning et al., 2008). Berikut merupakan rumus IDF:

$$idf_t = \log \frac{N}{df_t} \quad (2.1)$$

Keterangan:

idf_t :IDF dari kata t

N :Jumlah dokumen dalam *corpus*

df_t : *Document frequency*, menunjukkan jumlah dokumen dalam koleksi yang mengandung kata t

Berikut adalah rumus untuk TF-IDF:

$$\text{tf-idf}_{t,d} = \text{tf}_{t,d} \times \text{idf}_t \quad (2.2)$$

TF-IDF berfungsi menghasilkan bobot gabungan untuk setiap kata pada tiap dokumen (Manning et al., 2008).

2.2.9. Naïve Bayes Classifier (NBC)

Naïve Bayes Classifier (NBC) adalah salah satu metode untuk membangun model klasifikasi. Klasifikasi merupakan proses menemukan fungsi (atau model) yang mendeskripsikan dan membedakan kelas data atau konsep. Model berasal dari analisis *set* data latih (objek data yang kelasnya diketahui). Model berfungsi untuk memprediksi label kelas dari objek yang tidak memiliki label pada kelasnya (Han et al., 2012).

Naïve Bayes Classifier (NBC) merupakan algoritma pengklasifikasi paling sederhana dan umum digunakan untuk analisis sentimen (Medhat et al., 2014). NBC termasuk ke dalam metode *supervised learning* yang berarti kinerjanya bertujuan untuk memprediksi secara akurat data uji yang masuk menggunakan label atau kelas dari data latih (Dey et al., 2018). NBC adalah metode klasifikasi dengan menerapkan metode probabilistik berdasarkan Teorema Bayes dengan asumsi

independensi yang kuat (*naïve*) (Group, 2014). Berikut merupakan rumus Teorema Bayes (Downey, 2013):

$$p(H|D) = \frac{p(H)p(D|H)}{p(D)} \quad (2.3)$$

Keterangan:

$p(H)$: *Prior probability* atau *prior*, merupakan probabilitas hipotesis sebelum memperhatikan data.

$p(H|D)$:Posterior, merupakan sesuatu yang ingin dihitung, probabilitas hipotesis setelah memperhatikan data

$p(D|H)$: *Likelihood*, merupakan probabilitas data di bawah hipotesis.

$p(D)$:Normalizing constant, merupakan probabilitas data di bawah hipotesis apapun.

Cara kerja NBC untuk mengklasifikasikan dokumen ke kelas yang tepat adalah dengan membandingkan konten dan daftar kata (A. & Sonawane, 2016). Rumus NBC ditampilkan di bawah ini. Misalkan d adalah *tweet* dan C^* adalah kelas yang menentukan d , yang mana:

$$P_{NB}(c|d) = \frac{(P(c)) \sum_{i=1}^m P(f|c)^{n_{i(d)}}}{P(d)} \quad (2.4)$$

$$C^* = \arg \max_c P_{NB}(c|d) \quad (2.5)$$

Berdasarkan rumus di atas, d adalah *tweet*, C^* adalah kelas yang menentukan d , ‘ f ’ adalah ‘fitur’, hitungan fitur (f_i) disimbolkan dengan $n_i(d)$ dan terdapat di d yang merepresentasikan *tweet*. Jumlah maksimal fitur dilambangkan dengan m . Parameter $P(c)$ dan $P(f/c)$ dihitung melalui prediksi kemungkinan maksimum (A. & Sonawane, 2016). Data yang *noisy* atau *missing data* tidak menjadi hal yang sulit untuk ditangani metode NBC dan jika diperlukan metode ini dapat memberikan prediksi probabilistik (Russell & Norvig, 2010). NBC secara komputasi efisien dan mudah digunakan, juga NBC memperlakukan fitur secara kondisional bebas satu dengan yang lain berdasarkan kelasnya (Agarwal & Mittal, 2016).

2.2.10. Adaboost

Boosting merujuk pada *ensemble learning* yang mengkombinasikan beberapa *weak learner* menjadi *strong learner*. Konsep dari metode *boosting* adalah melatih prediktor secara berurutan, yang setiap prediktor mengoreksi prediktor sebelumnya (Géron, 2017). *Boosting* melakukan pengulangan secara serial, yang mana kesalahan klasifikasi dari iterasi sebelumnya digunakan untuk memengaruhi pelatihan pengklasifikasi pada iterasi berikutnya dengan memberikan bobot lebih berat pada kesalahan klasifikasi tersebut pada langkah selanjutnya (Unpingco, 2016). *Adaboost* menjadi metode *boosting* yang banyak digunakan, karena sederhana dan adaptif (Freund, 2001). *Adaboost* dikatakan adaptif karena jumlah *update* ditentukan sebagai fungsi dari kesalahan bobot dari hipotesis yang dihasilkan oleh *weak learner* (Freund, 2001). *Adaboost* didefinisikan sebagai algoritma pengulangan yang menggunakan pengklasifikasi berbeda untuk set pelatihan yang sama, lalu menyatukannya untuk membuat pengklasifikasi terkuat diakhir (Feng, 2019). Algoritma *Adaboost* dibuat khusus untuk menangani masalah klasifikasi yang berfungsi untuk meningkatkan akurasi *weak learner* (Andreyestha & Subekti, 2020). *Adaboost* memiliki beberapa kelebihan, yaitu sederhana serta mudah diimplementasikan (Davidson-Pilon, 2015). Berikut adalah cara kerja algoritma *Adaboost* (Hu & Hu, 2005):

- a. Diberikan: $(x_1, y_1), \dots, (x_n, y_n)$ dimana $y_i \in \{+1, -1\}$ (2.6)

b. Inisialisasi bobot: $D_I(i) = 1/n \quad (i = 1, \dots, n)$ (2.7)

c. For $t=1, \dots, T$: (2.8)

1. Memilih *weak classifier* h_t yang meminimalkan *weighted error*:

$$h_t = \arg \min_{h_j \in H} \varepsilon_j = \sum_{i=1} D_t(i) I[y_i \neq h_j(x_i)] \quad (2.9)$$

2. Jika $c_t = \min_j c_j$ untuk $j = 1, 2, \dots, r$ dan tentukan λ_{opt} . (2.18)

- $$P(\{0\}) = \left(\alpha_0 + \beta_0 \right) \delta_{0,0} \quad (2.11)$$

- $$4. \text{ Memperbarui bobot. } D_{t+1}(i) \equiv \frac{\sum_{t \in S} P_t(i)}{z_t} \quad (2.12)$$

Dimana Z_t adalah faktor normalisasi yang memastikan D_{t+1} merupakan distribusi.

- d. Pengklasifikasi yang kuat adalah : $H(x) = \text{sign}(\sum_{t=1}^T \alpha_t h_t(x))$ (2.13)

2.2.11. Evaluasi

Pada proses evaluasi, hasil model dievaluasi untuk menentukan apakah model sesuai dengan tujuan bisnis yang dinyatakan semula atau tidak. Kemampuan pengklasifikasi dapat diukur dengan beberapa cara, beberapa diantaranya adalah *recall*, *precision*, *f-measure*, dan *accuracy*. *F-measure* disebut juga *F-1 score* adalah ukuran akurasi tes, memperhatikan *precision*, dan *recall* tes untuk menghitung nilai. *Precision* mengidentifikasi frekuensi model memprediksi kelas positif dengan benar. *Recall* mengidentifikasi dari semua kemungkinan label positif yang diklasifikasikan model dengan benar (P. Bhatia, 2019). *Precision*, *recall*, dan *f-measure* dapat dilakukan berdasarkan dua sudut pandang, yaitu sudut pandang label positif, dan negatif agar dapat diketahui performa pengklasifikasi dalam menemukan kembali suatu informasi berdasarkan kedua kelas tersebut (Mardiana et al., 2019)(Ankit & Saleena, 2018). *Accuracy* disebut juga sebagai *recognition rate* secara keseluruhan, menyatakan seberapa baik kemampuan pengklasifikasi dalam mengenali *tuple* dari berbagai kelas. Berikut adalah rumus *precision*, *recall*, *f-measure*, dan *accuracy* dari sudut pandang label positif (Han et al., 2012) :

$$Recall = \frac{TP}{TP + FN} = \frac{TP}{P} \quad (2.14)$$

$$Precision = \frac{TP}{TP + FP} \quad (2.15)$$

$$F = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (2.16)$$

$$Accuracy = \frac{TP + TN}{P + N} \quad (2.17)$$

Keterangan:

True Positives (TP) : Tuple positif yang dilabeli dengan benar

True Negatives (TN) : Tuple negatif yang dilabeli dengan benar

False Positives (FP) : Tuple negatif yang salah dilabeli sebagai positif

False Negatives (FN) : Tuple positif yang salah

Positive (P) : Jumlah tuple positif

2.2.12. Confusion Matrix

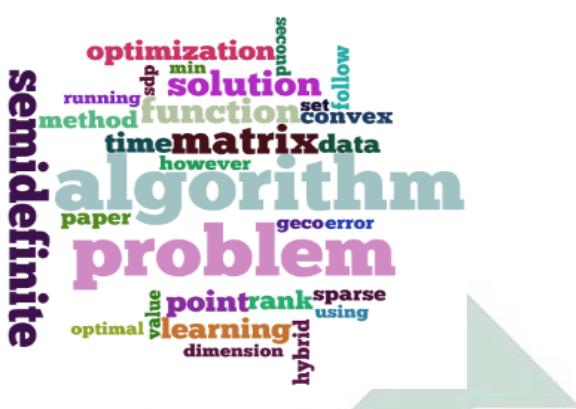
Confusion matrix berisi informasi yang dapat digunakan untuk menghitung berbagai *performance metrics*, seperti *precision*, *recall*, dan *f-measure*. *Confusion matrix* adalah tabel $N \times N$ yang merangkum akurasi dari prediksi model pengklasifikasi. N merepresentasikan jumlah kelas. *Confusion matrix* adalah hubungan antara label sebenarnya dengan label hasil prediksi model (P. Bhatia, 2019). *Confusion matrix* memeriksa kinerja model pada tingkat per-label. Gambar 2.2 merupakan contoh *confusion matrix*.

Tabel 2.2 *Confusion Matrix* (Sumber: Cahyani, 2020)

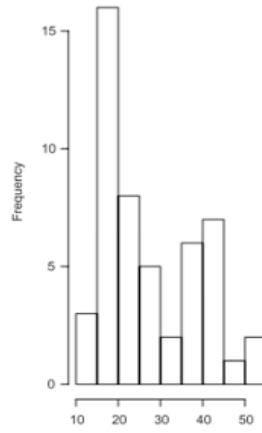
Aktual \ Prediksi	Positif	Negatif
Positif	<i>True Positive</i>	<i>False Negative</i>
Negatif	<i>False Positive</i>	<i>True Negative</i>

2.2.13. Visualisasi

Visualisasi data bertujuan untuk menggambarkan informasi dalam bentuk visual, dan dapat dianggap sebagai salah satu metode eksplorasi data yang paling kuat dan juga menarik (Gorunescu, 2011). Visualisasi data merupakan hal yang penting, karena informasi data akan disampaikan ke pembaca sehingga perlu memperhatikan akurasi informasi dan menggunakan metode yang tepat untuk mengevaluasi semua karakteristik penting dari distribusi data seperti jangkauan, variabilitas dan lain sebagainya (Martinez, 2015). *Bar chart* adalah metode visualisasi dasar yang populer dalam melaporkan analisis pemetaan data kategorikal ke tampilan visual dalam bentuk *x* dan *y* untuk dibandingkan secara numerik (Huang et al., 2009). Pada Gambar 2.3 ditampilkan contoh *bar chart*. *Word cloud* adalah metode visualisasi yang populer (Kurniawan, 2017). *Word cloud* merupakan presentasi grafis dari suatu dokumen, umumnya didapatkan dengan memplot kata-kata dalam dokumen yang paling umum dalam ruang dua dimensi, dengan ukuran font yang menunjukkan seberapa sering kata itu muncul (Castellà & Sutton, 2014). Contoh *word cloud* ditampilkan pada Gambar 2.2. *Word Cloud* (Castellà & Sutton, 2014).



Gambar 2.2 *Word Cloud* (Castellà & Sutton, 2014)



Gambar 2.3 *Bar Chart* (Martinez, 2015)

2.3. Integrasi Keilmuan

Integrasi antara penelitian ini dengan islam dapat diketahui dengan melakukan tanya jawab kepada ahli dibidang tafsir Al-Qur'an. Komunikasi secara daring telah dilakukan dengan Ustadzah Harista Hidayah, MA dan Ustadz Abdullah Khidhir. Ustadzah Harista Hidayah, MA adalah pengajar di AFI dan psikoterapi tasawuf UIN Sunan Ampel Surabaya, Ustadzah Harista menempuh pendidikan S1 jurusan akidah filsafat di UIN Sunan Ampel Surabaya dan S2 pemikiran islam di UIN Syarif Hidayatullah Jakarta. Ustadz Abdullah Khidhir adalah pengajar tafsir di Masjid Al-Mujahid Jrebeng dan Masjid Al-Abror Balong Bendo, Ustadz Abdullah pernah menjadi *staff* pengajar Mahad Al-Islam Karanganyar periode 2015-2017.

Penelitian ini membahas mengenai sentimen analisis dan belajar daring. Menurut Ustadzah Harista Hidayah, MA penelitian ini berkaitan dengan QS. Ash-Shu`ara' [26]:28.

قَالَ رَبُّ الْمَشْرِقِ وَالْمَغْرِبِ وَمَا بَيْنُهُمَا إِنْ كُنْتُمْ تَعْقِلُونَ

Artinya:

Dia (Musa) berkata, "Tuhan yang menguasai timur dan barat dan apa yang ada di antara keduanya; (Itulah Tuhanmu) jika kamu mengerti." (QS. Ash-Shu`ara' 26:28)

Dunia *cyber* atau daring merupakan ruang. Diri manusia harus memiliki wakil berupa akun/akun email/akun media sosial/nomor Hp. Dalam semesta alam

raya yang diartikan timur dan barat. Dunia *online* termasuk di dalamnya. Tetapi pasti pemahaman maksimal akan mengalami distorsi (target maksimal dalam kegiatan belajar mengajar bisa saja kurang maksimal) karena ada dilatas ruang dan waktu. Tetapi ada juga yang disebut ilmu *laduni*. Ilmu *laduni* adalah bagaimana seseorang Ulama/ahli ilmu mendapatkan pencerahan keilmuan setelah bermimpi didatangi *syaikh/master* ilmu bahkan bermimpi didatangi oleh Rasulullah SAW, banyak ulama yang menulis kitab setelah kejadian mimpi itu, alam mimpi juga tetap alam dan ruang. Sedangkan menurut Ustadz Abdullah Khidhir penelitian ini berhubungan dengan QS. An Naml [27]: 30.

إِنَّهُ مِنْ سُلَيْمَانَ وَإِنَّهُ بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Artinya:

Sesungguhnya (surat) itu dari Sulaiman yang isinya, Dengan nama Allah Yang Maha Pengasih, Maha Penyayang, (QS. An Naml 27: 30)

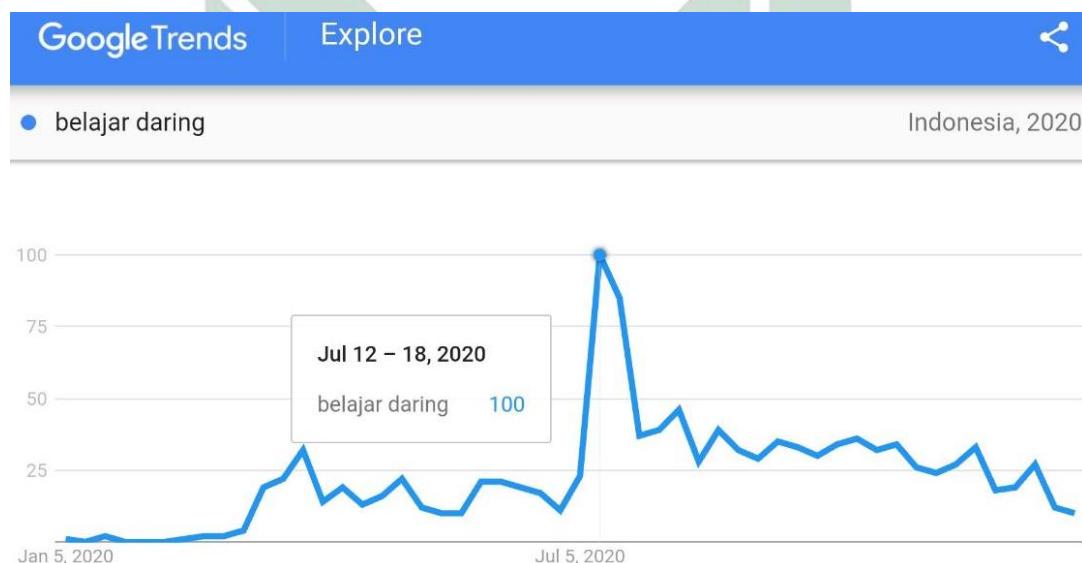
Ayat ini berisi kisah tentang surat menyurat antara Nabi Sulaiman dengan Ratu Balqis dengan media burung hud-hud, ini merupakan sebuah kecanggihan teknologi pada saat itu. Nabi Sulaiman mengirimkan surat yang berisi ajakan untuk beriman. Pada zaman sekarang, media burung hud-hud itu seperti media pembelajaran menggunakan kecanggihan teknologi daring.

BAB III

METODOLOGI PENELITIAN

3.1. Jenis Penelitian

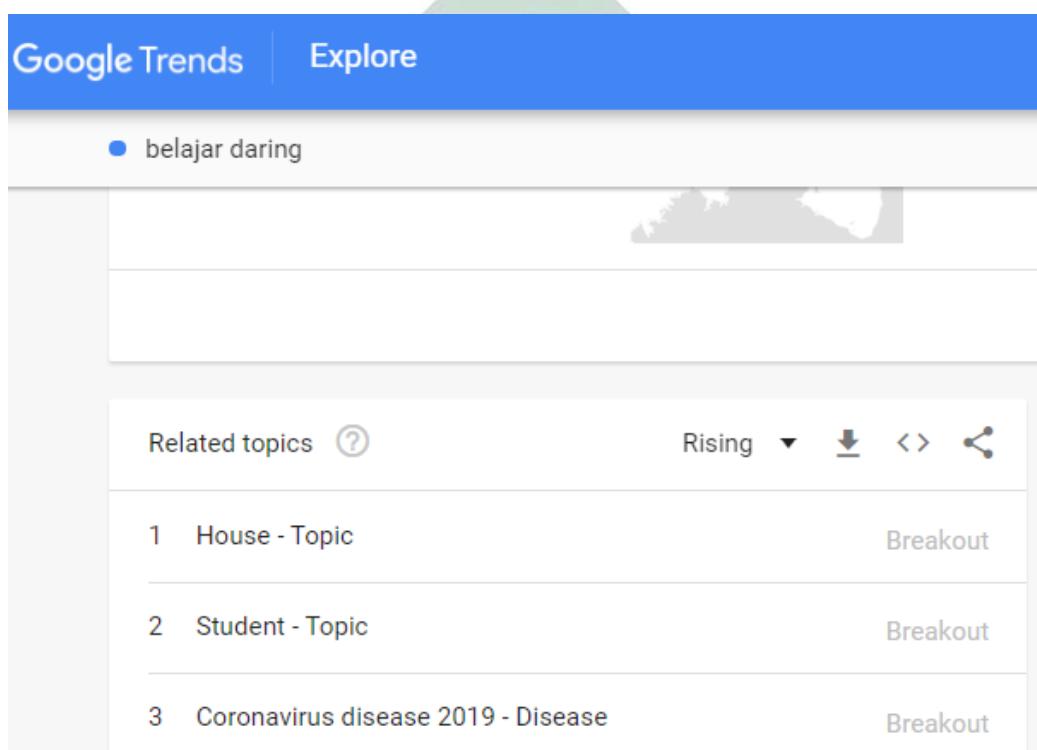
Penelitian ini menerapkan metode kuantitatif deskriptif. Penelitian metode kuantitatif menggunakan angka atau jumlah dengan cara mengumpulkan banyak data serta bertujuan untuk menampilkan hasil data (Sodik, Muhammad & Siyoto, 2015). Tujuan dari penelitian kuantitatif deskriptif adalah untuk menjelaskan atau mengukur secara jeli mengenai hal yang diteliti (Widhi Kurniawan & Puspitaningtyas, 2016). Penelitian ini mengambil data *tweets* mulai dari Juli hingga Desember 2020 dari Twitter melalui API Twitter. Bulan Juli 2020 dipilih karena berdasarkan Google Trends selama tahun 2020, pencarian mengenai belajar daring paling banyak pada bulan tersebut serta bertepatan dengan tahun ajaran baru. Desember 2020 dipilih karena Kemendikbud memberikan kewenangan penentuan kebijakan belajar daring atau luring (luar jaringan) kepada daerah mulai Januari 2021 dan Desember 2020 juga bertepatan dengan liburan semester. Grafik Google Trends dengan kata kunci belajar daring pada tahun 2020 ditampilkan pada Gambar 3.1.



Gambar 3.1 Grafik Belajar Daring (Google Trends, 2020)

3.2. Sumber Data

Penelitian ini menggunakan data yang bersumber dari Twitter. Data Twitter diambil menggunakan APIv2 Twitter dan Tweepy (*library* Python). Jenis sumber data ini adalah data sekunder. Data sekunder merupakan data yang didapatkan peneliti dari beragam sumber yang ada, sehingga peneliti tidak melakukan wawancara langsung untuk memperoleh data (Sodik, Muhammad & Siyoto, 2015). Kata kunci yang digunakan untuk mengumpulkan data adalah kata yang terkait dengan belajar daring dan ditentukan berdasarkan Google Trends, seperti yang ditampilkan pada Gambar 3.2.



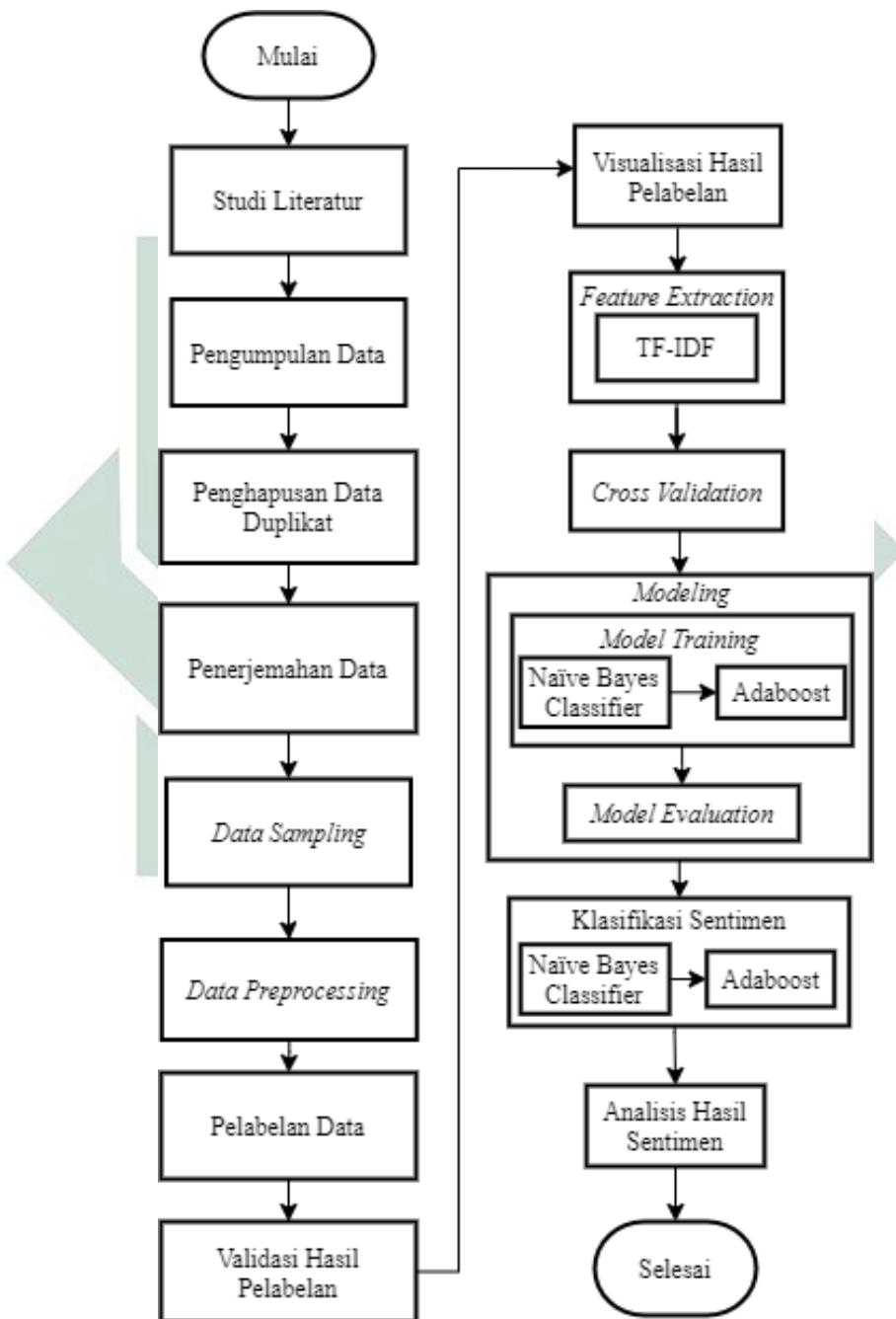
Gambar 3.2 *Related Topics* (Google Trends, 2020)

Pada Gambar 3.2 terdapat *related topics* atau topik yang sering dicari berkaitan dengan belajar daring, pada gambar tersebut ditunjukkan tiga kata teratas, yaitu: *house* yang artinya rumah, *student* yang artinya siswa, dan *coronavirus disease 2019* yang sering disebut *corona* (Google Trends, 2020). Ditambah dengan data tersebut, didapatkan beberapa kata kunci, yaitu: “belajar daring”, “daring rumah”, “belajar *online*”, “*online* siswa”, “*corona* sekolah”, “*corona* kuliah”, “kuliah *online*”, “sekolah *online*”, dan “*new normal*”. Kata yang ditangani pada penelitian ini merupakan kata yang dapat diterjemahkan oleh Google Translate.

Kata dengan penulisan tidak umum seperti alay, singkatan, gif dan emotikon tidak ditangani pada penelitian ini.

3.3. Metode Penelitian

Terdapat beberapa langkah dalam melakukan penelitian ini, langkah-langkah tersebut digambarkan pada Gambar 3.3.



Gambar 3.3 Metode Penelitian

3.3.1. Studi Literatur

Studi literatur dilakukan untuk memahami masalah penelitian, mencari dasar teori penelitian sebagai referensi penelitian serta mencari berbagai metode yang pernah digunakan untuk memecahkan masalah yang sejenis.

3.3.2. Pengumpulan Data

Pengumpulan data dilakukan dengan mengambil *tweets* yang berhubungan dengan belajar daring dari Twitter dengan beberapa kata kunci, yaitu: “belajar daring”, “daring rumah”, “belajar *online*”, “*online* siswa”, “*corona* sekolah”, “*corona* kuliah”, “kuliah *online*”, “sekolah *online*” dan “*new normal*”. Data diambil sejak Juli 2020 hingga Desember 2020. Beberapa contoh data yang telah terkumpul ditampilkan pada Tabel 3.1.

Tabel 3.1 Hasil Pengumpulan Data

Data
@collegemenfess Lagian egois gimana? Ga semua orang bisa belajar online juga. banyak noh yang kesusahan belajar daring
Pengalaman bagus nih. Akhirnya kualitas belajar daring makin bagus. https://t.co/yB21DQVzYr
nilaiku ada B 😣 🤢 astaga corona kapan selesai 🤢 aku gak suka belajar daring 🤢 🤢

3.3.3. Penghapusan Data Duplikat

Proses ini diperlukan karena pada jejaring sosial Twitter terdapat banyak *spam* yang memiliki isi *tweet* serupa. Data duplikat dihapus agar dapat mempercepat proses selanjutnya serta mengurangi data yang berulang, sehingga data yang diproses merupakan data yang unik.

3.3.4. Penerjemahan Data

Data yang telah terkumpul diterjemahkan ke Bahasa Inggris menggunakan Google Translate. Proses ini diperlukan karena pelabelan menggunakan subjektivitas dan polaritas dari *library* TextBlob hanya tersedia dalam Bahasa Inggris. Data diterjemahkan menggunakan Google Translate. Contoh penerjemahan data terlihat pada Tabel 3.2.

Tabel 3.2 Hasil Penerjemahan Data

Sebelum Penerjemahan Data	Setelah Penerjemahan Data
@collegemenfess Lagian egois gimananya? Ga semua orang bisa belajar online juga. banyak noh yang kesusahan belajar daring	@collegemenfess How selfish are you anyway? Not everyone can learn online either. many people have difficulty learning online
Pengalaman bagus nih. Akhirnya kualitas belajar daring makin bagus. https://t.co/yB21DQVzYr	It's a good experience. Finally, the quality of online learning is getting better. https://t.co/yB21DQVzYr
nilaiku ada B 😱 😱 astaga corona kapan selesai 😱 aku gak suka belajar daring 😱 😱	my score is B gosh when will corona finish I don't like online learning

3.3.5. Data Sampling

Sebanyak 50% data pada bulan-bulan awal digunakan untuk membuat model. Data awal dipilih agar model merepresentasikan kondisi nyata pada saat topik belajar daring sedang hangat dibicarakan banyak orang. Seluruh data digunakan untuk analisis sentimen.

3.3.6. *Data Preprocessing*

Tahap ini dilakukan untuk mengolah data mentah agar data terbebas dari kesalahan. Tahap ini terdiri dari beberapa proses, diantaranya:

1. Data Cleaning

Data dibersihkan dengan cara menghapus komponen yang tidak diperlukan dalam *tweets*, seperti, URL, *reserved words* (RT, dan FAV), *hashtags*, *mentions*, *username*, emotikon, *smileys*, dan angka. Perbedaan sebelum dan setelah dilakukan *data cleaning* ditampilkan pada Tabel 3.3.

Tabel 3.3 Hasil *Data Cleaning*

Sebelum <i>Data Cleaning</i>	Setelah <i>Data Cleaning</i>
@collegemenfess How selfish are you anyway? Not everyone can learn online either. many people have difficulty learning online	How selfish are you anyway? Not everyone can learn online either. many people have difficulty learning online
It's a good experience. Finally, the quality of online learning is getting better. https://t.co/yB21DQVzYr	It's a good experience. Finally, the quality of online learning is getting better.
my score is B gosh when will corona finish I don't like online learning	my score is B gosh when will corona finish i don't like online learning

2. Case Folding

Case folding berfungsi untuk mengubah seluruh huruf *tweets* menjadi huruf kecil agar semua teks diperlakukan sama, sehingga tidak menyebabkan gangguan pada proses selanjutnya. Tabel 3.4 menampilkan hasil proses *case folding*.

Tabel 3.4 Hasil *Case Folding*

Sebelum Case Folding	Setelah Case Folding
How selfish are you anyway? Not everyone can learn online either. many people have difficulty learning online	how selfish are you anyway? not everyone can learn online either. many people have difficulty learning online
It's a good experience. Finally, the quality of online learning is getting better.	it's a good experience. finally, the quality of online learning is getting better.
my score is B gosh when will corona finish i don't like online learning	my score is b gosh when will corona finish i don't like online learning

3. Tokenizing

Tokenizing digunakan untuk memotong teks menjadi perkata sehingga dapat mempermudah proses selanjutnya. Pada proses *tokenizing*, whitespace dianggap sebagai pemisah kata. Hasil *tokenizing* terlihat pada Tabel 3.5.

Tabel 3.5 Hasil *Tokenizing*

Sebelum Penerjemahan Data	Setelah Penerjemahan Data
how selfish are you anyway? not everyone can learn online either. many people have difficulty learning online	[how, selfish, are, you, anyway, ?, not, everyone, can, learn, online, either, ., many, people, have, difficulty, learning, online]
it's a good experience. finally, the quality of online learning is getting better.	[it, 's, a, good, experience, ., finally, ,, the, quality, of, online, learning, is, getting, better, .]
my score is b gosh when will corona finish i don't like online learning	[my, score, is, b, gosh, when, will, corona, finish, i, do, n't, like, online, learning]

4. Replace

Replace digunakan untuk mengubah kata. Tahap ini dilakukan agar kata yang mengandung sentimen tidak terhapus dan dapat dipahami oleh TextBlob. Pada Tabel merupakan contoh hasil dari *replace* kata *n't* menjadi *not*.

Tabel 3.6 Hasil Proses *Replace*

<i>Sebelum Replace</i>	<i>Setelah Replace</i>
[how, selfish, are, you, anyway, ?, not, everyone, can, learn, online, either, ., many, people, have, difficulty, learning, online]	[how, selfish, are, you, anyway, ?, not, everyone, can, learn, online, either, ., many, people, have, difficulty, learning, online]

[it, 's, a, good, experience, ., finally, ,, the, quality, of, online, learning, is, getting, better, .]	[it, 's, a, good, experience, ., finally, ,, the, quality, of, online, learning, is, getting, better, .]
[my, score, is, b, gosh, when, will, corona, finish, i, do, n't, like, online, learning]	[my, score, is, b, gosh, when, will, corona, finish, i, do, not, like, online, learning]

5. Remove Punctuation

Remove punctuation dilakukan agar tanda baca tidak ikut diproses pada langkah selanjutnya. Tabel 3.7 adalah contoh hasil *remove punctuation*.

Tabel 3.7 Hasil *Remove Punctuation*

Sebelum Remove Punctuation	Setelah Remove Punctuation
[how, selfish, are, you, anyway, ?, not, everyone, can, learn, online, either, .. many, people, have, difficulty, learning, online]	[how, selfish, are, you, anyway, , not, everyone, can, learn, online, either, , many, people, have, difficulty, learning, online]
[it, 's, a, good, experience, .., finally, ,, the, quality, of, online, learning, is, getting, better, .]	[it, s, a, good, experience, , finally, , the, quality, of, online, learning, is, getting, better,]
[my, score, is, b, gosh, when, will, corona, finish, i, do, not, like, online, learning]	[my, score, is, b, gosh, when, will, corona, finish, i, do, not, like, online, learning]

6. Lemmatization

Lemmatization dilakukan agar kata-kata yang terdapat didalam *tweets* dapat kembali dalam bentuk akar katanya. Dalam Bahasa Inggris suatu kata dapat memiliki banyak bentuk tergantung dari penulisan kalimatnya, seperti kata *write* jika ditulis dalam *past simple* akan menjadi *wrote*, dan jika ditulis dalam *past participle* menjadi *written*. Kata tersebut perlu untuk diubah ke akar kata agar kata yang ingin dibuang menggunakan *stopwords* dapat terhapus semua, dan untuk proses pelabelan agar kata dapat dikenali oleh TextBlob. Pada Tabel 3.8 terlihat hasil dari proses *lemmatization*.

Tabel 3.8 Hasil *Lemmatization*

<i>Sebelum Lemmatization</i>	<i>Setelah Lemmatization</i>
[how, selfish, are, you, anyway, , not, everyone, can, learn, online, either, , many, people, have, difficulty, learning, online]	[how, selfish, be, you, anyway, , not, everyone, can, learn, online, either, , many, people, have, difficulty, learn, online]
[it, s, a, good, experience, , finally, , the, quality, of, online, learning, is, getting, better,]	[it, s, a, good, experience, , finally, , the, quality, of, online, learn, be, get, good,]
[my, score, is, b, gosh, when, will, corona, finish, i, do, not, like, online, learning]	[my, score, be, b, gosh, when, will, corona, finish, i, do, not, like, online, learn]

7. Stopwords

Proses *stopwords* dilakukan agar kata yang sering muncul tetapi tidak berpengaruh pada sentimen dihapus. Hal ini dapat membuat proses lebih cepat dan efisien, serta kata yang diproses merupakan kata yang penting dan bermakna untuk analisis sentimen. *Library NLTK* digunakan pada tahap ini, kata yang terdapat dalam *library NLTK*, diantaranya: 'i', 'me', 'my', 'myself', 'we', 'our', 'ours', 'ourselves', 'you', 'your', 'yours', 'yourself', 'yourselves', 'he', 'him', 'his', 'himself', 'she', 'her', 'hers', 'herself', 'it', 'its', 'itself', 'they', 'them', 'their', 'theirs', 'themselves', 'what', 'which', 'who', 'whom', 'this', 'that', 'these', 'those', 'am', 'is', 'are', 'was', 'were', 'be', 'been', 'being', 'have', 'has', 'had', 'having', 'do', 'does', 'did', 'doing', 'a', 'an', 'the', 'and', 'but', 'if', 'or', 'because', 'as', 'until', 'while', 'of', 'at', 'by', 'for', 'with', 'about', 'against', 'between', 'into', 'through', 'during', 'before', 'after', 'above', 'below', 'to', 'from', 'up', 'down', 'in', 'out', 'on', 'off', 'over', 'under', 'again', 'further', 'then', 'once', 'here', 'there', 'when', 'where', 'why', 'how', 'all', 'any', 'both', 'each', 'few', 'more', 'most', 'other', 'some', 'such', 'no', 'nor', 'not', 'only', 'own', 'same', 'so', 'than', 'too', 'very', 's', 't', 'can', 'will', 'just', 'don', 'should', 'now'. Kata yang memiliki sentimen negatif dihapus dari daftar *stopwords* seperti *no* dan *not* agar sentimen negatif tidak hilang dari *tweets*. Pada Tabel 3.9 ditampilkan hasil dari *stopwords* menggunakan *library NLTK*.

Tabel 3.9 Hasil *Stopwords*

Sebelum Stopwords	Setelah Stopwords
[how, selfish, be, you, anyway, , not, everyone, can, learn, online, either, , many, people, have, difficulty, learn, online]	[selfish, anyway, , not, everyone, can, , difficult]
[it, s, a, good, experience, , finally, , the, quality, of, online, learn, be, get, good,]	[good, experience, , finally, , quality, get, good,]
[my, score, be, b, gosh, when, will, corona, finish, i, do, not, like, online, learn]	[score, b, gosh, corona, not, like]

8. *Filtering*

Filtering digunakan untuk menyaring kata pada dokumen, agar didapatkan data yang relevan untuk penelitian. *Tweets* yang memiliki huruf terlalu banyak dan terlalu sedikit akan dihapus. Jumlah kata kurang dari dua berisi kata yang tidak

terlalu penting seperti “*a*” dan lain sebagainya. Pada Tabel 3.10 ditampilkan hasil dari proses *filtering*.

Tabel 3.10 Hasil *Filtering*

Sebelum Filtering	Setelah Filtering
[selfish, anyway, , not, everyone, can, , difficult]	selfish anyway not everyone can difficult
[good, experience, , finally, , quality, get, good,]	good experience finally quality get good
[score, b, gosh, corona, not, like]	score gosh corona not like

9. Remove Duplicate

Menghapus *tweets* yang sama sehingga pada langkah selanjutnya hanya diproses *tweets* yang unik serta data akan diproses lebih cepat karena data sudah berkurang.

3.3.7. Pelabelan Data

Data untuk membuat model dilabelkan berdasarkan subjektivitas dan polaritas menggunakan *library* TextBlob. TextBlob menghitung rata-rata untuk memberikan nilai polaritas dan subjektivitas. TextBlob mengabaikan kata-kata yang tidak diketahui. Contoh hasil polaritas dan subjektivitas TextBlob ditampilkan pada Tabel 3.11.

Tabel 3.11 Contoh Hasil Polaritas dan Subjektivitas

Kata	Polaritas	Subjektivitas
selfish	-0.5	1
anyway	0	0
not	-0.5	0
everyone	0	0
can	0	0
difficult	-0.5	1

Nilai polaritas didapatkan dengan menghitung rata-rata kata dalam satu tweet, cara penghitungan sebagai berikut:

$$\text{Polaritas } (-0,5) + (-0,5) + (-0,5) = (-1,5)/3 = (-0,5)$$

Subjektivitas $1+1 = 2/2 = 1$

Berdasarkan penghitungan TextBlob, didapatkan nilai hasil polaritas dan subjektivitas. Nilai polaritas dengan angka tertentu akan diberikan label positif, netral, dan negatif seperti yang dituliskan pada Tabel 3.12.

Tabel 3.12 Nilai Polaritas dan Label

Nilai Polaritas	Label
Polaritas > 0	Positif
Polaritas = 0	Netral
Polaritas < 0	Negatif

Pada Tabel 3.13 ditampilkan contoh *tweets* dengan label, nilai polaritas, dan subjektivitas berdasarkan penghitungan dari TextBlob.

Tabel 3.13 Contoh Hasil Pelabelan

Label	Polaritas	Subjektivitas	Tweets
Negatif	-0.500000	1.000000	selfish anyway not everyone can difficult
Positif	0.466667	0.733333	good experience finally quality get good
Netral	0.000000	0.000000	score gosh corona not like

Data dengan label netral akan dihapus karena dapat menghambat pemrosesan selanjutnya, serta pada penelitian ini hanya menangani label positif, dan negatif. Berdasarkan contoh pada Tabel 3.13 dikarenakan data netral dihilangkan, maka pada proses selanjutnya hanya memproses dua *tweets*.

3.3.8. Validasi Hasil Pelabelan

Data dengan label dari TextBlob divalidasi manual dengan melibatkan 27 orang pakar untuk menilai sentimen pada *tweets*. Berdasarkan data yang telah dilabelkan, akan diambil *sample* lalu dinilai oleh pakar. *Sample* dipilih dengan menggunakan teknik *non-probability* yaitu *judgment sampling*. Berdasarkan penilaian dari pakar, akan didapatkan akurasi TextBlob dibandingkan dengan pakar.

3.3.9. Visualisasi Hasil Pelabelan

Hasil pelabelan divisualisasikan menggunakan *bar chart*, dan *word cloud*. *Bar chart* berfungsi untuk menampilkan perbandingan jumlah sentimen. *Word cloud* pada sentimen positif dan negatif untuk memvisualisasikan kata yang sering digunakan pada sentimen positif dan negatif.

3.3.10. *Feature Extraction*

Feature extraction menggunakan TF-IDF diterapkan pada data hasil pelabelan yang label netralnya telah dihapus. TF-IDF dihitung berdasarkan rumus (2.2). TF-IDF dilakukan agar data dapat berubah menjadi angka, dan dapat diproses pada tahap selanjutnya. Tabel 3.14 merupakan contoh hasil TF-IDF. Pada TF-IDF dimasukkan dua data dan urutan penghitungan dimulai dari 0, sehingga didapatkan hasil dari dokumen ke-0 dan dokumen ke-1.

Tabel 3.14 Contoh Hasil TF-IDF

Kata	Nilai Hasil TF-IDF	
	Dokumen ke-0	Dokumen ke-1
anyway	0.408248	0
can	0.408248	0
difficult	0.408248	0
everyone	0.408248	0
experience	0	0.353553
finally	0	0.353553
get	0	0.353553
good	0	0.707107
not	0.408248	0
quality	0	0.353553
selfish	0.408248	0

3.3.11. *Cross Validation*

Pada data diterapkan *cross validation* dengan nilai $k=10$. Proses ini dilakukan untuk mencegah terjadinya *overfitting*.

3.3.12. *Modeling*

Pada tahap ini, dilakukan pembangunan model atau *modeling*. Berikut merupakan penjelasan pada setiap tahapan pembangunan model seperti yang terlihat pada Gambar 3.3:

1. Model training diawali dengan melatih *Naïve Bayes Classifier*, pelatihan NBC dilakukan berdasarkan rumus (2.4) sebagai berikut:
 - a. Hitung jumlah maksimal fitur yang akan digunakan (m), berasal dari TF-IDF yang berupa *tweets*.
 - b. Mengalikan dengan prediksi kemungkinan maksimum ($P(f|c)^{n_{i(d)}}$)
 - c. Membagi hasil sebelumnya dengan parameter $P(d)$.

- d. Hitung kemungkinan tertinggi kelas C^* tergolong positif, negatif atau netral menggunakan rumus (2.5).
 - e. *Output* dari *Naïve Bayes Classifier* berupa hasil klasifikasi data.
 - f. Meningkatkan *weak learner Naïve Bayes Classifier* dengan *Adaboost*. Langkah-langkah *Adaboost* seperti yang terdapat pada Bab II, terdiri dari:
 1. Menentukan fitur dan kelas berdasarkan rumus (2.6).
 2. Menginisialisasi bobot *tweet* berdasarkan rumus (2.7).
 3. Masuk pada iterasi dengan nomor iterasi (t) dan jumlah iterasi (T) seperti pada rumus (2.8).
 4. Menentukan *weak classifier* berdasarkan rumus (2.9).
 5. Jika *error* lebih besar dari $\frac{1}{2}$ maka set $T=t-1$ dan hentikan *loop* berdasarkan rumus (2.10).
 6. Nilai α_t yang digunakan adalah 0,01.
 7. Bobot diperbarui berdasarkan rumus (2.12).
 8. *Output* diakhir akan dihasilkan pengklasifikasi yang kuat seperti berdasarkan rumus (2.13).
 - 2. Setelah model dilatih, model akan dievaluasi dengan dua sudut pandang, yaitu dari sudut pandang label positif, dan negatif agar dapat diketahui kemampuan pengklasifikasi dalam mengelompokkan masing-masing label. Evaluasi menggunakan *recall* dengan rumus (2.14), *precision* berdasarkan rumus (2.15), *f-measure* (2.16), dan *accuracy* (2.17). Hasil evaluasi dari NBC dan NBC dengan *Adaboost* akan dibandingkan.

3.3.13. Analisis Hasil Sentimen

Setelah model selesai dibuat, model digunakan untuk analisis sentimen. Seluruh data digunakan untuk analisis sentimen. Pada saat menganalisis hasil sentimen, dipaparkan jumlah sentimen positif, dan negatif.

BAB IV

HASIL DAN PEMBAHASAN

Pada bab ini, akan dipaparkan proses penelitian secara rinci dari awal hingga akhir.

4.1. Pengumpulan Data

Data dikumpulkan dengan melakukan pengambilan data dari Twitter berdasarkan kata kunci menggunakan APIv2 Twitter dan Tweepy (*library* Python). Kata kunci yang digunakan adalah “belajar daring”, “daring rumah”, “belajar *online*”, “*online* siswa”, “corona sekolah”, “corona kuliah”, “kuliah *online*”, “sekolah *online*”, dan “*new normal.*” Data diambil sejak Juli 2020 hingga Desember 2020. Jumlah *tweets* yang didapatkan dari proses pengumpulan data ditampilkan pada Tabel 4.1.

Tabel 4.1 Jumlah *Tweets* Terkumpul

Bulan	Jumlah Tweets
Juli	4.689
Agustus	3.927
September	4.350
Oktober	4.686
November	5.607
Desember	79.685
Total	102.944

Pada Tabel 4.2 ditampilkan beberapa data *tweets* yang telah didapatkan dari proses pengumpulan data.

Tabel 4.2 Hasil Pengumpulan Data

Hasil Pengumpulan Data

@tanyainrl mau offline si aku mah jujur karena udah cape banget sama online, apalagi aku udah kelas 9 semester 2 juga, tapi kalo harus ketemu temen-temen lagi takut ngerasa canggung banget 😱

pagi! awali pagimu dengan kelas online 🎉

Aku kuat, aku cinta diriku sendiri, kuliah online asik banget tugasnya dikit, dirumah suasannya harmonis, temen-temenku baik dukung aku terus, aku bahagia

dear my rapot, aku minta maaf ya semester ini belom bisa kasih yang terbaik. sekolah online gak masuk isi otak soalnya, maaf 😢 ❤️

knp aku terlanjur nyaman sekolah online aaaaakk.

4.2. Penghapusan Data Duplikat

Penghapusan data duplikat dilakukan menggunakan *library* Python. Penghapusan ini mengurangi jumlah data yang berulang, sehingga jumlah data yang dihasilkan dari proses ini sebanyak 77.919 *tweets*.

4.3. Penerjemahan Data

Setelah data duplikat dihapus, data diterjemahkan ke Bahasa Inggris menggunakan Google Translate. Saat melakukan proses penerjemahan, terdapat kemungkinan emotikon dapat dihilangkan. Hasil dari penerjemahan data ditampilkan pada Tabel 4.3.

Tabel 4.3 Hasil Penerjemahan Data

Sebelum Penerjemahan Data	Setelah Penerjemahan Data
@tanyainrl mau offline si aku mah jujur karena udah cape banget sama online, apalagi aku udah kelas 9 semester 2 juga, tapi kalo harus ketemu temen-temen lagi takut ngerasa canggung banget 😭	@tanyainrl I want to go offline, I'm honest because I'm really tired of being online, especially since I'm already in grade 9, semester 2 too, but if I have to meet friends again I'm afraid to feel really awkward 🤦
pagi! awali pagimu dengan kelas online ✨	morning! start your morning with an online class ✨
Aku kuat, aku cinta diriku sendiri, kuliah online asik banget tugasnya dikit, dirumah suasannya harmonis, temen-temenku baik dukung aku terus, aku bahagia	I'm strong, I love myself, online lectures are really fun to do a little work, at home the atmosphere is harmonious, my friends are kind to support me, I'm happy
dear my rapot, aku minta maaf ya semester ini belom bisa kasih yang terbaik. sekolah online gak masuk isi otak soalnya, maaf 😔 ❤️	dear my rapot, I'm sorry that this semester I can't give the best. online school doesn't enter the brain because, sorry
knp aku terlanjur nyaman sekolah online aaaaakk.	How come I'm already comfortable with online school aaaaakk.

4.4. *Data Sampling*

Data yang tersedia sebanyak 77.919 *tweets*, untuk membuat model digunakan 50% data awal sehingga jumlah data untuk membuat model sebanyak 38.959 *tweets*. Google Colab dengan RAM 25.46 GB dan *disk* sebesar 107.72 GB digunakan pada penelitian ini, mengingat besarnya data yang diperoleh dari proses pengumpulan data.

4.5. *Data Preprocessing*

Data preprocessing digunakan untuk mengolah data mentah menjadi data yang dapat dipahami agar didapatkan hasil yang maksimal. *Data preprocessing* yang dilakukan dalam penelitian ini meliputi:

4.5.1. Data Cleaning

Tahap ini dilakukan untuk proses pembersihan data awal. Hasil pada proses ini ditampilkan pada Tabel 4.4. *Tweet preprocessor* digunakan pada tahap ini. *Tweet preprocessor* berfungsi untuk menghapus URL, *reserved words* (RT, dan FAV), *hashtags*, *mentions*, *username*, emotikon, *smileys*. Angka juga dihapus pada tahap ini, karena tidak diperlukan pada penelitian ini.

Tabel 4.4 Hasil *Data Cleaning*

Sebelum Data Cleaning	Setelah Data Cleaning
@tanyainrl I want to go offline, I'm honest because I'm really tired of being online, especially since I'm already in grade 9, semester 2 too, but if I have to meet friends again I'm afraid to feel really awkward 	I want to go offline, I'm honest because I'm really tired of being online, especially since I'm already in grade , semester too, but if I have to meet friends again I'm afraid to feel really awkward
morning! start your morning with an online class 	morning! start your morning with an online class
I'm strong, I love myself, online lectures are really fun to do a little work, at home the atmosphere is harmonious, my friends are kind to support me, I'm happy	I'm strong, I love myself, online lectures are really fun to do a little work, at home the atmosphere is harmonious, my friends are kind to support me, I'm happy
dear my rapot, I'm sorry that this semester I can't give the best. online school doesn't enter the brain because, sorry	dear my rapot, I'm sorry that this semester I can't give the best. online school doesn't enter the brain because, sorry
How come I'm already comfortable with online school aaaaakkkk.	How come I'm already comfortable with online school aaaaakkkk.

4.5.2. Case Folding

Huruf pada data diubah semuanya menjadi huruf kecil, agar proses selanjutnya tidak terganggu dan semua teks diperlakukan sama. Hasil *case folding* ditampilkan pada Tabel 4.5.

Tabel 4.5 Hasil *Case Folding*

Sebelum Case Folding	Setelah Case Folding
I want to go offline, I'm honest because I'm really tired of being online, especially since I'm already in grade , semester too, but if I have to meet friends again I'm afraid to feel really awkward	i want to go offline, i'm honest because i'm really tired of being online, especially since i'm already in grade , semester too, but if i have to meet friends again i'm afraid to feel really awkward

morning! start your morning with an online class	morning! start your morning with an online class
I'm strong, I love myself, online lectures are really fun to do a little work, at home the atmosphere is harmonious, my friends are kind to support me, I'm happy	i'm strong, i love myself, online lectures are really fun to do a little work, at home the atmosphere is harmonious, my friends are kind to support me, i'm happy
dear my rapot, I'm sorry that this semester I can't give the best. online school doesn't enter the brain because, sorry	dear my rapot, i'm sorry that this semester i can't give the best. online school doesn't enter the brain because, sorry
How come I'm already comfortable with online school aaaaakkk.	how come i'm already comfortable with online school aaaaakkk.

4.5.3. Tokenizing

Tokenizing digunakan untuk memisah kalimat menjadi kata. *Tokenizing* yang digunakan dalam penelitian ini adalah *library* NLTK. Kata dipisahkan berdasarkan *whitespace*, tanda baca, dan juga NLTK dapat memisahkan kata negatif seperti kata *can't* menjadi *ca*, *n't* yang mana akan sangat berguna untuk penelitian ini. Hasil *tokenizing* terlihat pada Tabel 4.6.

Tabel 4.6 Hasil *Tokenizing*

Sebelum Tokenizing	Setelah Tokenizing
i want to go offline, i'm honest because i'm really tired of being online, especially since i'm already in grade , semester too, but if i have to meet friends again i'm afraid to feel really awkward	[i, want, to, go, offline, „, i, 'm, honest, because, i, 'm, really, tired, of, being, online, „, especially, since, i, 'm, already, in, grade, „, semester, too, „, but, if, i, have, to, meet, friends, again, i, 'm, afraid, to, feel, really, awkward]
morning! start your morning with an online class	[morning, !, start, your, morning, with, an, online, class]
i'm strong, i love myself, online lectures are really fun to do a little work, at home the atmosphere is harmonious, my friends are kind to support me, i'm happy	[i, 'm, strong, „, i, love, myself, „, online, lectures, are, really, fun, to, do, a, little, work, „, at, home, the, atmosphere, is, harmonious, „, my, friends, are, kind, to, support, me, „, i, 'm, happy]
dear my rapot, i'm sorry that this semester i can't give the best. online school doesn't enter the brain because, sorry	[dear, my, rapot, „, i, 'm, sorry, that, this, semester, i, ca, n't, give, the, best, „, online, school, does, n't, enter, the, brain, because, „, sorry]
how come i'm already comfortable with online school aaaaakkk.	[how, come, i, 'm, already, comfortable, with, online, school, aaaaakkk, .]

4.5.4. Replace

Replace mengubah suatu kata menjadi kata lain. Pada proses ini kata *n't* diubah menjadi *not* dan kata *difficulty* menjadi *difficult*. Hal ini dilakukan agar kata *n't* tidak hilang pada proses *stopwords* dan dapat dipahami oleh TextBlob pada saat

proses pelabelan, dikarenakan kata *n't* dan *difficulty* tidak terdapat pada kamus TextBlob. Hasil *replace* ditampilkan pada Tabel 4.7.

Tabel 4.7 Hasil *Replace*

Sebelum Replace	Setelah Replace
[i, want, to, go, offline, „ i, 'm, honest, because, i, 'm, really, tired, of, being, online, „ especially, since, i, 'm, already, in, grade, „ semester, too, „ but, if, i, have, to, meet, friends, again, i, 'm, afraid, to, feel, really, awkward]	[i, want, to, go, offline, „ i, 'm, honest, because, i, 'm, really, tired, of, being, online, „ especially, since, i, 'm, already, in, grade, „ semester, too, „ but, if, i, have, to, meet, friends, again, i, 'm, afraid, to, feel, really, awkward]
[morning, !, start, your, morning, with, an, online, class]	[morning, !, start, your, morning, with, an, online, class]
[i, 'm, strong, „ i, love, myself, „ online, lectures, are, really, fun, to, do, a, little, work, „ at, home, the, atmosphere, is, harmonious, „ my, friends, are, kind, to, support, me, „ i, 'm, happy]	[i, 'm, strong, „ i, love, myself, „ online, lectures, are, really, fun, to, do, a, little, work, „ at, home, the, atmosphere, is, harmonious, „ my, friends, are, kind, to, support, me, „ i, 'm, happy]
[dear, my, rapot, „ i, 'm, sorry, that, this, semester, i, ca, n't, give, the, best, .. online, school, does, n't, enter, the, brain, because, „ sorry]	[dear, my, rapot, „ i, 'm, sorry, that, this, semester, i, ca, not, give, the, best, .. online, school, does, not, enter, the, brain, because, „ sorry]
[how, come, i, 'm, already, comfortable, with, online, school, aaaaakkk, .]	[how, come, i, 'm, already, comfortable, with, online, school, aaaaakkk, .]

4.5.5. Remove Punctuation

Remove punctuation menghilangkan tanda baca yang terdapat pada *tweets*, dikarenakan pada proses sebelumnya tanda baca masih ada pada teks yang mana dapat menganggu proses selanjutnya. Tabel 4.8 menampilkan hasil dari proses *remove punctuation*.

Tabel 4.8 Hasil *Remove Punctuation*

Sebelum Remove Punctuation	Setelah Remove Punctuation
[I, want, to, go, offline, , I, ‘m, honest, because, I, ‘m, really, tired, of, being, online, ,, especially, since, I, ‘m, already, in, grade, ,, semester, too, ,, but, if, I, have, to, meet, friends, again, I, ‘m, afraid, to, feel, really, awkward]	[I, want, to, go, offline, , I, m, honest, because, I, m, really, tired, of, being, online, , especially, since, I, m, already, in, grade, , semester, too, , but, if, I, have, to, meet, friends, again, I, m, afraid, to, feel, really, awkward]
[morning, !, start, your, morning, with, an, online, class]	[morning, , start, your, morning, with, an, online, class]
[I, ‘m, strong, ,, I, love, myself, ,, online, lectures, are, really, fun, to, do, a, little, work, ,, at, home, the, atmosphere, is, harmonious, ,, my, friends, are, kind, to, support, me, ,, I, ‘m, happy]	[I, m, strong, , I, love, myself, , online, lectures, are, really, fun, to, do, a, little, work, , at, home, the, atmosphere, is, harmonious, , my, friends, are, kind, to, support, me, , I, m, happy]

[dear, my, rapot, ,, I, 'm, sorry, that, this, semester, I, ca, not, give, the, best, .. online, school, does, not, enter, the, brain, because, ,, sorry]	[dear, my, rapot, , I, m, sorry, that, this, semester, I, ca, not, give, the, best, , online, school, does, not, enter, the, brain, because, , sorry]
[how, come, I, 'm, already, comfortable, with, online, school, aaaaakkk, .]	[how, come, I, m, already, comfortable, with, online, school, aaaaakkk,]

4.5.6. Lemmatization

Lemmatization mengembalikan kata-kata dalam *tweets* ke bentuk akar kata, proses ini diperlukan agar kata dapat dikenali oleh TextBlob, dan juga kata yang ingin dihapus pada *stopwords* dapat terhapus secara keseluruhan. *Lemmatization* dapat mengembalikan kata kerja (*verb*) dan kata sifat (*adjective*) pada akar katanya, mengingat dalam Bahasa Inggris satu kata dapat menjadi banyak variasi tulisan tergantung pada konteksnya. Contohnya seperti kata *loose* jika ditulis menggunakan *past tense* menjadi *loosed*, kata *happy* yang ditulis dalam bentuk *comparative adjective* akan menjadi *happier*. *Lemmatization* dilakukan menggunakan *library NLTK*. Hasil *lemmatization* ditampilkan pada Tabel 4.9.

Tabel 4.9 Hasil Lemmatization

Sebelum Lemmatization	Setelah Lemmatization
[i, want, to, go, offline, , i, m, honest, because, i, m, really, tired, of, being, online, , especially, since, i, m, already, in, grade, , semester, too, , but, if, i, have, to, meet, friends, again, i, m, afraid, to, feel, really, awkward]	[i, want, to, go, offline, , i, m, honest, because, i, m, really, tire, of, be, online, , especially, since, i, m, already, in, grade, , semester, too, , but, if, i, have, to, meet, friends, again, i, m, afraid, to, feel, really, awkward]
[morning, , start, your, morning, with, an, online, class]	[morning, , start, your, morning, with, an, online, class]
[i, m, strong, , i, love, myself, , online, lectures, are, really, fun, to, do, a, little, work, , at, home, the, atmosphere, is, harmonious, , my, friends, are, kind, to, support, me, , i, m, happy]	[i, m, strong, , i, love, myself, , online, lecture, be, really, fun, to, do, a, little, work, , at, home, the, atmosphere, be, harmonious, , my, friends, be, kind, to, support, me, , i, m, happy]
[dear, my, rapot, , i, m, sorry, that, this, semester, i, ca, not, give, the, best, , online, school, does, not, enter, the, brain, because, , sorry]	[dear, my, rapot, , i, m, sorry, that, this, semester, i, ca, not, give, the, best, , online, school, do, not, enter, the, brain, because, , sorry]
[how, come, i, m, already, comfortable, with, online, school, aaaaakkk,]	[how, come, i, m, already, comfortable, with, online, school, aaaaakkk,]

4.5.7. Stopwords

Stopwords menggunakan *library* NLTK tetapi kata yang berpengaruh pada sentimen dihilangkan dari *stopwords*, seperti *no*, *not*, *very* dan sebagainya.

Stopwords ditambah dengan kata-kata lain yang sering muncul tetapi tidak mempengaruhi sentimen seperti “*online*”, “*school*”, “*how*”, dan lain-lain. Kata yang tidak mengandung sentimen, tidak bermakna, tidak berhasil diterjemahkan, dan bahasa daerah juga dimasukkan ke *stopwords* seperti “*aaaaakkk*”, “*zyuzyur*”, “*wkwkhwkwkhwk*”, “*arek*”, “*teros*”, “*jl*”, “*jo*” dan lain-lain. Tabel 4.10 adalah hasil dari proses *stopwords*.

Tabel 4.10 Hasil Stopwords

Sebelum Stopwords	Setelah Stopwords
[i, want, to, go, offline, , i, m, honest, because, i, m, really, tire, of, be, online, , especially, since, i, m, already, in, grade, , semester, too, , but, if, i, have, to, meet, friends, again, i, m, afraid, to, feel, really, awkward]	[want, offline, , honest, really, tire, , especially, since, grade, , semester, , meet, friends, afraid, feel, really, awkward]
[morning, , start, your, morning, with, an, online, class]	[morning, , morning]
[i, m, strong, , i, love, myself, , online, lecture, be, really, fun, to, do, a, little, work, , at, home, the, atmosphere, be, harmonious, , my, friends, be, kind, to, support, me, , i, m, happy]	[strong, , love, , really, fun, , atmosphere, harmonious, , friends, kind, support, , happy]
[dear, my, rapot, , i, m, sorry, that, this, semester, i, ca, not, give, the, best, , online, school, do, not, enter, the, brain, because, , sorry]	[dear, rapot, , sorry, semester, not, best, , not, enter, brain, , sorry]
[how, come, i, m, already, comfortable, with, online, school, aaaaakkk,]	[comfortable]

4.5.8. *Filtering*

Pada tahap ini, dilakukan *filtering* dengan menghapus kata yang mengandung huruf kurang dari dua dan lebih dari 21. Jumlah huruf lebih dari 21 dihapus, karena tidak relevan dengan penelitian. Pada Bahasa Inggris, kata terpanjang dan umum digunakan adalah *incomprehensibilities* yang memiliki 21 huruf, kata dengan huruf melebihi 21 merupakan istilah ilmiah yang jarang digunakan, maka dari itu kata dengan huruf lebih dari 21 dihapus. Setelah itu dilakukan *join* agar kata-kata dalam satu *tweets* bergabung menjadi kalimat dan dapat dilabelkan oleh TextBlob. Pada Tabel 4.11 ditampilkan hasil dari proses sebelum dan setelah *filtering*.

Tabel 4.11 Hasil *Filtering*

Sebelum Filtering	Setelah Filtering
[want, offline, , honest, really, tire, , especially, since, grade, , semester, , meet, friends, afraid, feel, really, awkward]	want offline honest really tire especially since grade semester meet friends afraid feel really awkward
[morning, , morning]	morning morning
[strong, , love, , lecture, really, fun, , atmosphere, harmonious, , friends, kind, support, , happy]	strong love really fun atmosphere harmonious friends kind support happy
[dear, rapot, , sorry, semester, not, best, , not, enter, brain, , sorry]	dear rapot sorry semester not best not enter brain sorry
[comfortable]	comfortable

4.6. Pelabelan Data

Data dilabelkan dengan subjektivitas dan polaritas menggunakan *library* TextBlob. Pada Tabel 4.12 ditampilkan beberapa hasil dari pelabelan. Data netral tidak ditangani pada penelitian ini, sehingga akan dihapus agar tidak menghambat proses selanjutnya. Data yang tersisa berjumlah 24.457 *tweets*.

Tabel 4.12 Hasil Pelabelan Data

Label	Polaritas	Subjektivitas	Tweets
Negatif	-0.080000	0.800000	want offline honest really tire especially since grade semester meet friends afraid feel really awkward
Netral	0.000000	0.000000	morning morning
Positif	0.526667	0.686667	strong love really fun atmosphere harmonious friends kind support happy
Negatif	-0.500000	0.766667	dear rapot sorry semester not best not enter brain sorry
Positif	0.400000	0.800000	comfortable

4.7. Validasi Hasil Pelabelan

Validasi hasil pelabelan dilakukan dengan cara divalidasi manual oleh pakar. Data yang divalidasi merupakan *sample* dari keseluruhan data yang telah dilabelkan oleh TextBlob. Pengambilan *sample* dilakukan dengan menggunakan teknik *non-probability* dengan *judgement sampling*. Data yang telah dilabelkan TextBlob berjumlah 24.457 dengan 15.593 label positif, dan 8.864 label negatif. *Sampel* yang digunakan untuk validasi pelabelan berjumlah 200 data dengan 91 label positif, dan 109 label negatif. Pada Tabel 4.13 ditampilkan beberapa hasil dari

validasi pelabelan pakar dengan 0 adalah label negatif, dan 1 adalah label positif. B adalah “Benar”, dimana hasil label dari TextBlob sama dengan pakar, dan S adalah “Salah” yang menyatakan hasil label dari TextBlob berbeda dengan pakar.

Tabel 4.13 Validasi Pelabelan

Tweets	Sentimen Pakar			Label TextBlob	B/S
	1	2	3		
ga siap senin besok mulai belajar daring aaaaaaa males bat	0	0		0	B
2020, i lost my father 🍀 sekolah online dpt nilai jelek td pagi motor ilang lgsg 3 ambil hikmah nya, semua uda diatur sm tuhan	0	1	0	0	B
yang sulit tuh ak masih anak kuliah, kuliah online = libur duit jajan 🤑 https://t.co/wmocfuof60	0			0	B
buat kalian yg mempersiapkan mau masuk kuliyah tahun depan banyak do'a aja ya soal nya masih banyak banget orang yg terkena covid19 semangatt yang daring 😊	0	1	1	1	B
uas gue open book, walaupun pas pertemuan daring doseinya killer dan kadang suka ngeselin, but i lop you buk makasih udah bikin mood aku mendingan menjelang akhir semester ❤️❤️❤️	1	1		1	B
walaupun sekolah online bikin sambat terus, tapi gapapa deh. seengganya aku bisa ngerasain sekolah sambil rebahan. yaa walaupun akhirnya otakku juga ikut rebahan sih... https://t.co/de0h10ncwx	0	1	0	1	S
blessed punya circle pertemanan yg kalo diajak bahas apapun ayo-ayo aja dan ga menggurui. temen sekolah di ngawi dulu, tapi belum kesempaan meet up lengkap. mereka off/online sama kocaknya dan ga jaim, dan bukan tipe pengen keliatan hits. ga ada beban kalo sama mereka ih 😊	1	1		1	B
ui kak magister. iyaa semoga cepat kembali normal. karena kuliah online kyk engga serius gituu :"" banyak jd problemnyaay	1			0	S
pagi-pagi sudah hujan, bikin tambah males kuliah online	1			0	S
jujur pingin balik kuliah offline. kuliah daring dari rumah beneran kurang dapet manfaat. mana kadang kuliah sambil rebahan, malah kadang sampek ketiduran	1	0	1	1	B

Hasil pelabelan menggunakan TextBlob dengan label dari pakar didapatkan akurasi sebesar 80% seperti yang ditampilkan pada Tabel 4.14.

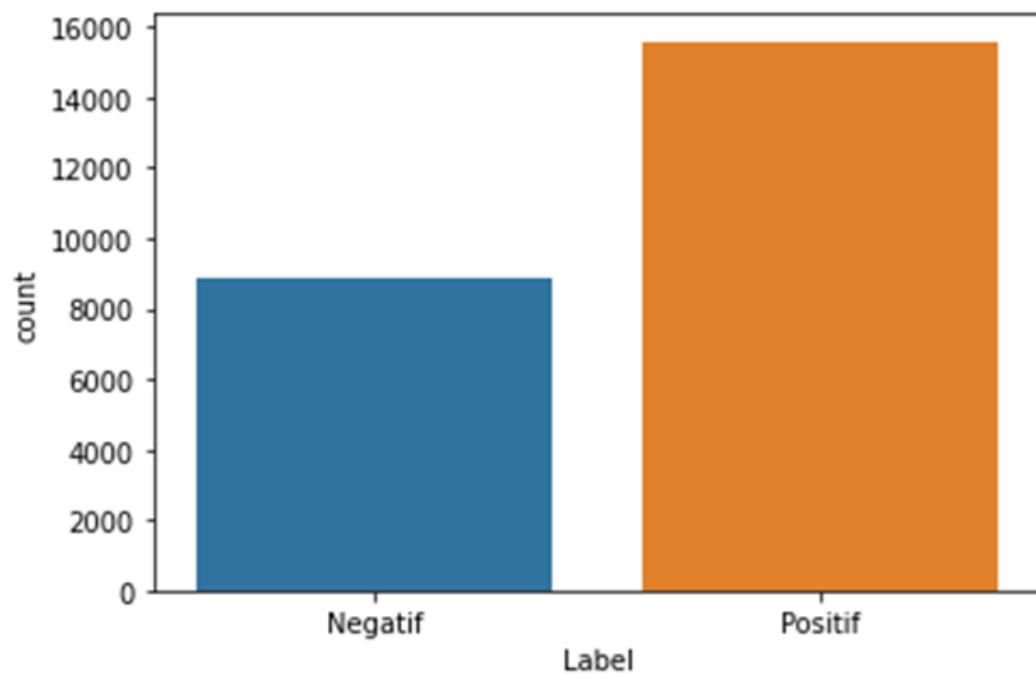
Tabel 4.14 Akurasi Pelabelan

Pakar \ TextBlob	Positif	Negatif
Positif	66	15
Negatif	25	94

$$Accuracy = \frac{66 + 94}{66 + 15 + 25 + 15} = 0.80$$

4.8. Visualisasi Hasil Pelabelan

Visualisasi hasil pelabelan ditampilkan menggunakan *bar chart* dan *word cloud*. *Bar chart* berfungsi untuk menampilkan jumlah label positif, dan negatif. Gambar 4.1 merupakan *bar chart* dengan *bar* berwarna jingga menunjukkan sentimen positif, dan *bar* berwarna biru menunjukkan sentimen negatif. Hasil pelabelan dengan TextBlob menunjukkan jumlah label positif lebih dominan, yaitu sebanyak 15.593 *tweets*, sedangkan sentimen negatif terdapat dalam 8.864 *tweets*.



Gambar 4.1 *Bar Chart* Hasil Sentimen

Hasil *word cloud* sentimen positif terlihat pada Gambar 4.2. Kata paling umum yang digunakan pada *tweets* dengan sentimen positif adalah “*friend*”, “*teacher*”, “*solution*”, dan “*material*”. *Word cloud* ini menunjukkan bahwa publik merasa senang hal yang berkaitan dengan “teman”, “guru”, “solusi”, dan “materi”.

Kata “bodoh” ini merepresentasikan bahwa banyak orang merasa bodoh saat kebijakan belajar daring ini dilakukan. Pengguna Twitter juga merasa tidak suka terhadap tugas yang diberikan oleh guru, dan dosen pada saat pelaksanaan belajar daring. Kata “tahun” juga sering digunakan pada sentimen negatif mengindikasikan bahwa tahun 2020 adalah tahun yang kurang disukai karena bertepatan dengan berlakunya penerapan belajar daring.

4.9. Feature Extraction

Pada tahap *feature extraction* digunakan TF-IDF agar didapatkan bobot gabungan untuk setiap kata pada tiap dokumen. TF-IDF memiliki parameter `min_df` yang berfungsi untuk memfilter kata, sebagai contoh jika ditulis `min_df=2` maka kata yang akan diproses adalah kata yang muncul minimal pada dua dokumen. Pada Tabel 4.15 ditampilkan contoh hasil TF-IDF dari empat dokumen atau empat *tweet* tanpa pengaturan nilai `min_df`, dikarenakan contoh hanya menggunakan empat *tweets*. Parameter `min_df` akan sangat berguna jika diterapkan pada data yang besar. Urutan penghitungan dimulai dari 0, sehingga dihasilkan urutan dokumen ke-0, dokumen ke-1, dokumen ke-2, dan dokumen ke-3. Hasil dari *feature extraction* berbeda-beda pada tiap *tweets* dikarenakan jumlah kemunculan tiap kata berbeda. Pada data diterapkan TF-IDF dengan berbagai parameter `min_df` agar didapatkan data yang relevan, dapat diproses lebih cepat, dan akurasi yang optimal. Perbedaan akurasi dari diterapkannya `min_df` akan terlihat pada proses evaluasi.

Tabel 4.15 Hasil TF-IDF

Kata	Nilai Hasil TF-IDF			
	Dokumen ke-0	Dokumen ke-1	Dokumen ke-2	Dokumen ke-3
afraid	0.260558	0	0	0
atmosphere	0	0.328919	0	0
awkward	0.260558	0	0	0
best	0	0	0.270948	0
brain	0	0	0.270948	0
comfortable	0	0	0	1
dear	0	0	0.270948	0
enter	0	0	0.270948	0
especially	0.260558	0	0	0
feel	0.260558	0	0	0
friends	0.205427	0.259324	0	0

fun	0	0.328919	0	0
grade	0.260558	0	0	0
happy	0	0.328919	0	0
harmonious	0	0.328919	0	0
honest	0.260558	0	0	0
kind	0	0.328919	0	0
love	0	0.328919	0	0
meet	0.260558	0	0	0
not	0	0	0.541896	0
offline	0.260558	0	0	0
rapot	0	0	0.270948	0
really	0.410855	0.259324	0	0
semester	0.205427	0	0.213619	0
since	0.260558	0	0	0
sorry	0	0	0.541896	0
strong	0	0.328919	0	0
support	0	0.328919	0	0
tire	0.260558	0	0	0
want	0.260558	0	0	0

4.10. *Cross Validation*

Cross validation dengan nilai $k=10$ diterapkan pada langkah ini. Nilai $k=10$ berarti membagi proporsi antara data latih 90% dan data uji 10%. *Cross validation* berfungsi untuk mencegah kebocoran data uji sehingga mencegah kemungkinan *overfitting*. Contoh hasil *cross validation score* pada tiap *fold* dengan $\text{min_df}=3$ ditampilkan pada Tabel 4.16.

Tabel 4.16 *Cross Validation Score*

Fold ke-	Cross Validation Score	
	NBC	NBC dengan Adaboost
1	0.77514309	0.6369583
2	0.7992641	0.63736713
3	0.80130826	0.63777596
4	0.77105478	0.6369583
5	0.78618152	0.63777596
6	0.78659035	0.63777596
7	0.7739166	0.63777596
8	0.77218814	0.63762781
9	0.800818	0.63762781
10	0.77750511	0.63762781
Rata-Rata	0.7843969955538611	0.6375271007128201

Berdasarkan Tabel 4.16, hasil rata-rata *cross validation score* menggunakan NBC sebesar 0.7843969955538611 sedangkan NBC dengan Adaboost mendapatkan nilai sebesar 0.6375271007128201.

4.11. *Modeling*

Pada tahap *modeling* atau pembangunan model, dilakukan proses *model training* dan *model evaluation*, untuk rincian prosesnya dipaparkan sebagai berikut:

4.11.1. *Model Training*

Library Scikit-learn digunakan pada proses *model training* atau pelatihan model. Pertama, dilakukan pelatihan pada model NBC. NBC dilatih menggunakan data hasil dari *cross validation*. Setelah selesai, dilanjutkan dengan proses pelatihan pada *Adaboost*, NBC menjadi *base_estimator* dari *Adaboost*. Detail alur proses *model training* tertulis pada Bab III bagian 3.3.12 mengenai *modeling*. Setelah semua model selesai dilatih, akan dilakukan proses evaluasi.

4.11.2. Model Evaluation

Pada tahap evaluasi, model akan diuji performanya menggunakan dua sudut pandang untuk *recall*, *precision*, *f-measure*, dan untuk mendapatkan hasil secara keseluruhan digunakan *accuracy*. Pada Tabel 4.17 ditampilkan hasil akurasi NBC dan NBC dengan Adaboost yang menerapkan nilai *min_df* yang berbeda. Jumlah baris merupakan jumlah *tweets*, sedangkan jumlah kolom merupakan jumlah kata atau *term* yang diproses.

Tabel 4.17 Hasil Akurasi dengan Berbagai Nilai Min_df

Min_df	Jumlah Baris dan Kolom	Rata-rata Akurasi	
		NBC	NBC dengan Adaboost
1	24457 rows × 17492 columns	0.74	0.64
2	24457 rows × 7999 columns	0.77	0.64
3	24457 rows × 5717 columns	0.78	0.64

Semakin besar nilai min_df, data yang diproses semakin kecil, sehingga dapat meningkatkan akurasi dari NBC. Hasil akurasi dari NBC dengan Adaboost tidak mengalami perubahan yang signifikan pada setiap perubahan jumlah data.

Model dengan min_df=3 mendapatkan hasil terbaik dibandingkan dengan yang lainnya, yaitu sebesar 0.78 sehingga model dengan min_df=3 akan digunakan untuk proses klasifikasi sentimen. Berikut merupakan hasil evaluasi dari model dengan min_df=3.

1. *Naïve Bayes Classifier* (NBC)

Data sebanyak 24.457 yang telah diterapkan *cross validation* digunakan untuk membuat model NBC. Tabel 4.18 adalah *confusion matrix* dari NBC.

Tabel 4.18 *Confusion Matrix NBC*

Aktual \ Prediksi	Positif	Negatif
Positif	14921	672
Negatif	4601	4263

Pada Tabel 4.19 berisi hasil *precision*, *recall*, dan *f1-score* dari label positif, dan negatif NBC.

Tabel 4.19 Hasil Evaluasi NBC

<i>Label</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
Negatif	0.86	0.48	0.62
Positif	0.76	0.96	0.85

Cara penghitungan untuk masing-masing evaluasi ditulis sebagai berikut:

$$Precision\ Negatif = \frac{4263}{4263 + 672} = 0.86$$

$$Precision\ Positif = \frac{14921}{14921 + 4601} = 0.764317181$$

$$Recall\ Negatif = \frac{4263}{4263 + 4601} = 0.48$$

$$Recall\ Positif = \frac{14921}{14921 + 672} = 0.96$$

$$F1 - Score Negatif = 2 \cdot \frac{0.86 \times 0.48}{0.86 + 0.48} = 0.62$$

$$F1 - Score Positif = 2 \cdot \frac{0.76 \times 0.96}{0.76 + 0.96} = 0.85$$

$$Accuracy = \frac{14921 + 4263}{14921 + 4263 + 1601 + 672} = 0.78$$

Hasil evaluasi dari model NBC didapatkan akurasi sebesar 0.78 atau 78% sedangkan pada *cross validation score* dihasilkan nilai rata-rata 0.7843969955538611, hal ini menunjukkan bahwa NBC mampu mengklasifikasikan dengan baik meskipun diberikan beragam data pada NBC. Selain itu, hasil *precision*, *recall*, dan *F1-Score* pada label positif

mendapatkan hasil yang cukup baik. Sedangkan, pada label negatif didapatkan hasil kurang baik dengan *recall* 0,48 dan *F1-Score* 0,62.

2. NBC dengan *Adaboost*

Adaboost dengan *base estimator* NBC. Data sebanyak 24.457 yang telah melalui proses *cross validation* digunakan untuk membangun model NBC dengan *Adaboost*. Tabel 4.20 adalah *confusion matrix* dari NBC dengan *Adaboost*.

Tabel 4.20 *Confusion Matrix* NBC dengan Adaboost

Aktual \ Prediksi	Positif	Negatif
Positif	15591	2
Negatif	8863	1

Pada Tabel 4.21 berisi hasil *precision*, *recall*, dan *f1-score* pada tiap label dari NBC dengan *Adaboost*.

Tabel 4.21 Hasil Evaluasi NBC dengan Adaboost

<i>Label</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
Negatif	0.33	0.00	0.00
Positif	0.64	1.00	0.78

$$Precision\ Negatif = \frac{1}{2+1} = 0.33$$

$$Precision\ Positif = \frac{15591}{15591 + 8863} = 0.64$$

$$Recall\ Negatif = \frac{1}{1 + 8863} = 0$$

$$Recall\ Positif = \frac{15591}{15591 + 2} = 1.0$$

$$F1 - Score Negatif = 2 \cdot \frac{0.33 \times 0}{0.33 + 0} = 0$$

$$F1 - Score \text{ Positif} = 2 \cdot \frac{0.64 \times 1.0}{0.64 + 1.0} = 0.78$$

$$Accuracy = \frac{15591+1}{15591+2+1+8862} = 0.64$$

Hasil evaluasi dari model NBC dengan *Adaboost* didapatkan akurasi sebesar 0.64 atau 64% sedangkan pada *cross validation score* dihasilkan nilai rata-rata 0.6375271007128201. Akurasi tersebut dapat dicapai karena jumlah data positif lebih banyak dari data negatif, meskipun NBC dengan

Adaboost tidak dapat mengklasifikasi label negatif. NBC merupakan pengklasifikasi yang stabil, hal ini akan menjadi sulit bagi *Adaboost* untuk dapat meningkatkan akurasi pengklasifikasi. NBC merupakan pengklasifikasi yang stabil dengan *high bias* dan *low variance* (Ming Ting & Zheng, 2003). Pada penelitian ini juga dilakukan percobaan dengan menerapkan *Adaboost* dengan *base estimator Decision Tree* yang mana merupakan *base estimator default* dari *Adaboost* di *library Scikit-learn*. Percobaan ini dilakukan agar dapat diketahui apakah terdapat ketidakcocokan antara data yang digunakan dengan *Adaboost*. Hasil evaluasi *Adaboost* dengan *Decision Tree* ditampilkan pada Tabel 4.22.

Tabel 4.22 Hasil Evaluasi *Decision Tree* dengan *Adaboost*

<i>Label</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
Negatif	0.83	0.64	0.72
Positif	0.82	0.93	0.87

Hasil menunjukkan bahwa data yang digunakan dapat diklasifikasikan dengan baik saat menggunakan *Decision Tree* dengan *Adaboost*. Kesimpulan yang dapat diambil dari percobaan tersebut, penurunan akurasi pada NBC dengan *Adaboost* bukan disebabkan adanya ketidakcocokan antara data dengan *Adaboost* ataupun data dengan NBC. Tetapi, penyebab turunnya akurasi NBC dengan *Adaboost* adalah gabungan antara data, NBC, dengan *Adaboost*. Data yang digunakan pada *modeling* merupakan data yang dominan pada label positif (*imbalance*) sehingga mempengaruhi performa NBC dengan *Adaboost* dalam mengklasifikasikan data.

4.12. Klasifikasi Sentimen

Setelah dilakukan *modeling*, selanjutnya dilakukan klasifikasi sentimen menggunakan model yang telah dibangun. Dari 77.919 *tweets* yang terkumpul, hasil klasifikasi sentimen dari NBC didapatkan 69.738 *tweets* berlabel positif, dan 8.181 *tweets* berlabel negatif. Sedangkan, *Adaboost* gagal untuk melakukan klasifikasi karena melabeli semua *tweets*, yaitu sebanyak 77.919 *tweets* dengan label positif. Tabel 4.23 menunjukkan beberapa hasil dari klasifikasi yang dilakukan oleh masing-masing model, dimana 0 menunjukkan label negatif, dan 1 menunjukkan label positif.

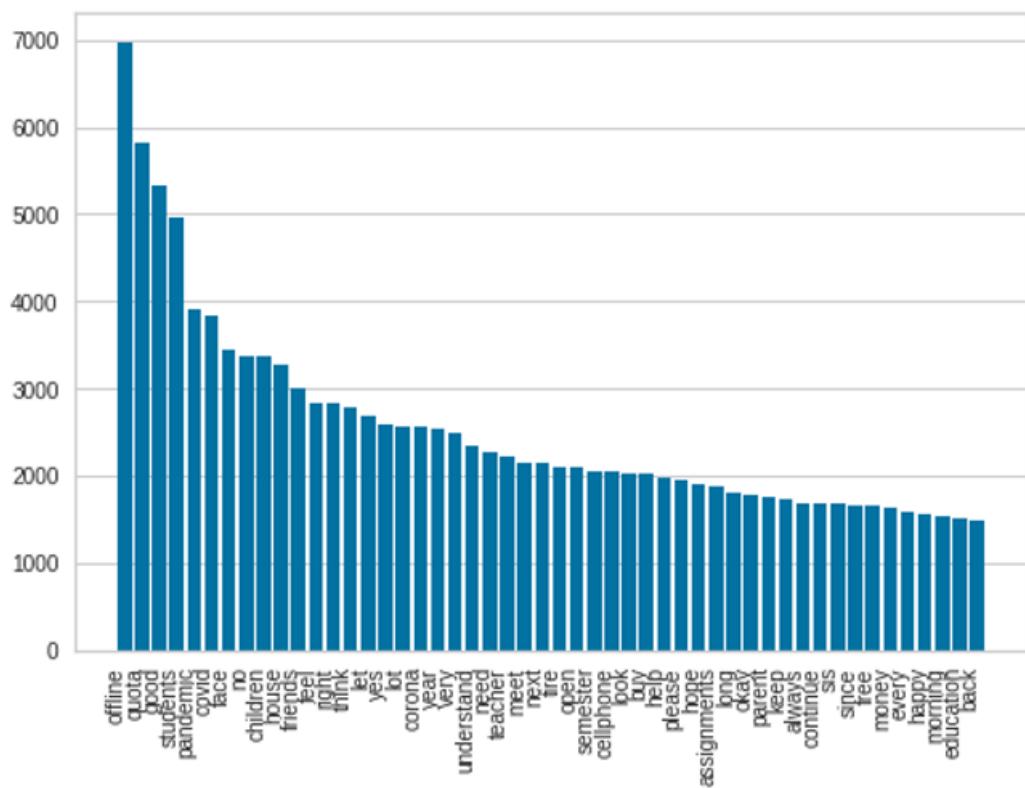
Tabel 4.23 Hasil Klasifikasi Data

Tweets	NBC	NBC dengan Adaboost
@tanyainrl I want to go offline, I'm honest because I'm really tired of being online, especially since I'm already in grade 9, semester 2 too, but if I have to meet friends again I'm afraid to feel really awkward 	0	1
morning! start your morning with an online class 	1	1
I'm strong, I love myself, online lectures are really fun to do a little work, at home the atmosphere is harmonious, my friends are kind to support me, I'm happy	1	1
dear my rapot, I'm sorry that this semester I can't give the best. online school doesn't enter the brain because, sorry	0	1
How come I'm already comfortable with online school aaaaakkk,	1	1

4.13. Analisis Hasil Sentimen

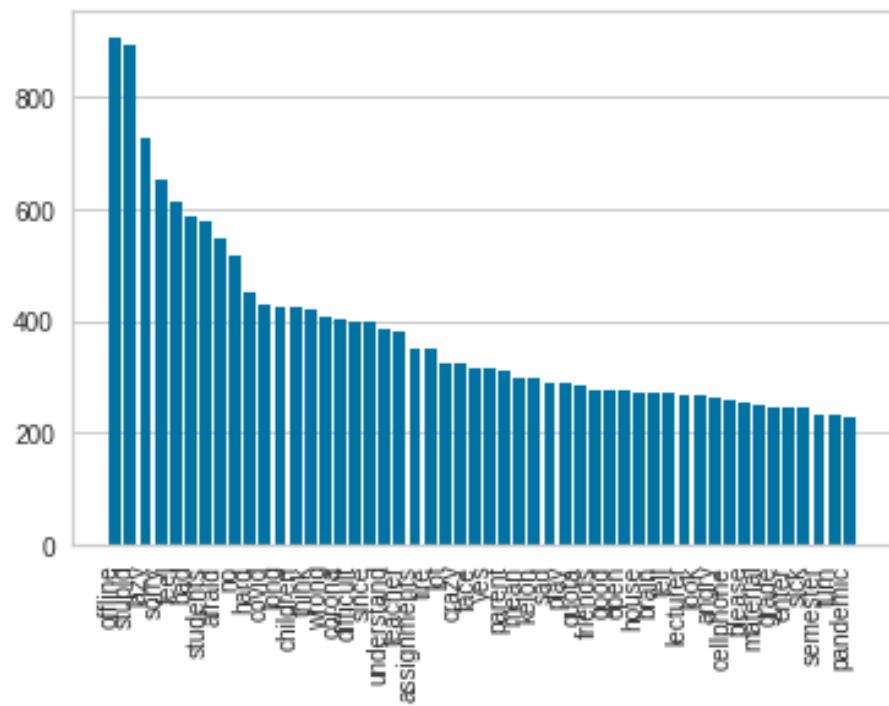
Berdasarkan hasil klasifikasi sentimen yang didapatkan dari NBC, dari 77.919 *tweets* dihasilkan 89,5% data atau 69.738 *tweets* berlabel positif, dan 10,5% data atau 8.181 *tweets* berlabel negatif. Sedangkan, NBC dengan *Adaboost* melabeli 77.919 *tweets* dengan label positif. Berdasarkan kedua hasil klasifikasi sentimen tersebut, dapat disimpulkan bahwa klasifikasi sentimen terhadap belajar daring menunjukkan kecenderungan pada sentimen positif.

Pada Gambar 4.4 dan Gambar 4.5 ditampilkan hasil visualisasi menggunakan *bar chart*. *Bar chart* tersebut menunjukkan 50 kata teratas yang sering muncul pada tiap sentimen. Jumlah kemunculan kata terdapat pada sumbu y sedangkan kata yang sering muncul terdapat pada sumbu x. Kata terbanyak yang digunakan pada *tweets* dengan sentimen positif adalah “*offline*”, “*quota*”, “*good*”, “*students*”, “*pandemic*”, “*covid*”, “*face*”, “*no*”, “*children*”, dan “*house*” seperti yang ditampilkan pada Gambar 4.4.



Gambar 4.4 *Bar Chart* Sentimen Positif

Pada Gambar 4.5 ditampilkan hasil sentimen negatif yang menunjukkan 50 kata yang paling sering digunakan pada *tweets* negatif.



Gambar 4.5 *Bar Chart* Sentimen Negatif

Kata terbanyak yang digunakan pada *tweets* negatif adalah “*offline*”, “*stupid*”, “*lazy*”, “*sorry*”, “*feel*”, “*bad*”, “*students*”, “*afraid*”, “*no*”, “*hard*”, “*covid*”, “*long*”, “*children*”, dan “*understand*”.

Berdasarkan hasil yang didapatkan dari *bar chart* sentimen positif pada Gambar 4.4, kata “*quota*” dominan digunakan. Hal ini disebabkan pada tahun 2020 pemerintah memberikan bantuan kuota untuk pelajar yang mendapatkan respon positif dari masyarakat. Kata “*good*” bermakna luas, yaitu “*good*” karena belajar dilakukan dari rumah (dikaitkan dengan kata “*house*”) sehingga suasana belajar lebih santai, dapat juga dikarenakan pembagian bantuan kuota yang diberikan pemerintah. Kata “*face*” dapat berkaitan dengan kata “pembelajaran tatap muka” yang mana dapat menjadi hal yang cenderung disukai masyarakat.

Hasil dari *bar chart* sentimen negatif seperti pada Gambar 4.5 kata “*stupid*”, “*lazy*”, “*feel*” dominan digunakan, hal tersebut berkaitan dengan banyaknya pelajar yang merasa bodoh dikarenakan belajar daring menyebabkan rasa malas menjadi meningkat, materi pelajaran sulit dipahami (berkaitan dengan kata “*hard*”, “*no*”, dan, “*understand*”). Kata “*sorry*” dapat berkaitan dengan hal negatif, seperti kesedihan ataupun penyesalan, hal ini dapat dikaitkan dengan banyak masyarakat merasa sedih dengan penerapan belajar daring karena belajar daring juga memerlukan *gadget* yang mendukung, padahal tidak semua orang memiliki *gadget*. Kata “*bad*” dapat dikaitkan dengan masyarakat yang menganggap belajar daring merupakan sesuatu yang buruk, hal ini dapat disebabkan banyak masyarakat yang masih kesulitan dalam penerapan belajar daring, karena belum meratanya fasilitas jaringan internet yang memadai. Adanya pandemi Covid-19 juga berlangsung lama, dan hal ini menimbulkan ketakutan sehingga kegiatan belajar menjadi kurang nyaman. Hal ini berkaitan dengan kata “*covid*”, “*long*”, dan “*afraid*”.

Kata yang sering muncul pada sentimen positif, dan negatif adalah “*offline*”, “*students*”, “*covid*”, dan “*children*”. Hal ini menunjukkan bahwa terdapat beragam pandangan mengenai belajar secara *offline*. Kata “*students*”, “*covid*”, “*children*” juga merupakan kata yang sering muncul di kedua sentimen, dikarenakan kata tersebut adalah kata netral yang memiliki beragam konteks.

Berdasarkan penelitian yang dilakukan (Mutaqinah & Hidayatullah, 2020) faktor pendorong siswa yang menikmati belajar dari rumah daripada pembelajaran tatap muka, dikarenakan banyak waktu untuk bersama dengan keluarga (antara orang tua dan anak), suasana rumah yang mendukung kegiatan belajar daring, serta *gadget* dan jaringan internet yang mendukung. Kelebihan belajar daring adalah materi yang sudah diajarkan dapat diberikan kembali (Nengrum et al., 2021). Kelebihan lainnya adalah belajar daring lebih fleksibel terkait dengan waktu belajar sehingga orang tua dapat mendampingi anak-anaknya pada saat belajar daring, materi yang mudah diakses, dan pembelajaran lebih santai (Anugrahana, 2020).

Dampak negatif mengenai belajar daring yang diteliti oleh (Jatira & S, 2021) menghasilkan bahwa murid mengalami *stress*, merasa bosan dan menjadi pemalas dikarenakan hambatan yang terjadi saat belajar daring. Kondisi rumah yang kurang kondusif untuk belajar membuat perasaan atau emosi anak menjadi tidak stabil. Saat dilaksanakan belajar atau ujian secara daring, terkadang orang tua menuntut sampai memarahi anak agar membantu mengerjakan pekerjaan rumah sehingga menyebabkan anak tidak dapat konsentrasi dengan proses pembelajaran.

Siswa juga merasa adanya kesulitan untuk berdiskusi dengan guru, dan juga merasa kesulitan jika hanya dapat materi pembelajaran dari buku (Puspaningtyas & Dewi, 2020). Terbatasnya kuota internet, dan jaringan juga menjadi kendala untuk masyarakat dalam melaksanakan belajar daring. Disarankan bagi pemerintah, khususnya pemerintah daerah untuk menyediakan infrastruktur yang mendukung kegiatan belajar daring, seperti jaringan internet yang merata di seluruh daerah (Mutaqinah & Hidayatullah, 2020). Pada mahasiswa, kesulitan untuk memahami materi juga dialami selama melakukan belajar daring dikarenakan terdapat dosen yang hanya memberikan banyak tugas tanpa memberi penjelasan materi, dan cara penyampaian materi terbatas hanya melalui teks atau pesan suara. Belajar daring juga menimbulkan kemalasan dikarenakan tidak bertemu secara langsung sehingga daya saing antar mahasiswa menurun.(Hariyanti et al., 2020)

Berdasarkan hasil analisis penelitian ini dan penelitian sebelumnya, dapat disarankan untuk pemerintah agar dapat terus memberikan bantuan kuota, memperbaiki sarana dan prasarana untuk belajar daring seperti jaringan internet, perangkat yang memadai, dan memberikan pelatihan, serta fasilitas kepada

pengajar, agar pengajar dapat menjalankan kegiatan mengajar secara daring dengan baik. Pengajar diharapkan dapat menyampaikan materi secara kreatif agar mudah dipahami, dan tidak memberi beban tugas terlalu banyak. Wali murid disarankan agar dapat memahami situasi dan kondisi anak saat melakukan belajar daring, sehingga diharapkan dapat memberikan dukungan serta tidak terlalu memberikan terlalu banyak tuntutan pada anak. Untuk siswa dan mahasiswa dapat meningkatkan motivasi internal agar tujuan pembelajaran tercapai maksimal. Belajar daring adalah hal yang tidak dapat dihindari sehingga diharapkan terdapat kolaborasi antara pemerintah, pengajar, dan peserta didik agar tujuan pembelajaran dapat dicapai.

4.14. Diskusi

Berdasarkan latar belakang yang terdapat pada Bab I yang menyatakan bahwa “*Adaboost* dapat meningkatkan hasil akurasi NBC” dengan hasil penelitian ini terdapat kontradiksi, dimana pada penlitian ini didapatkan hasil bahwa akurasi NBC dengan *Adaboost* menurun. Menurut analisa peneliti, penurunan akurasi NBC setelah dioptimasi dengan *Adaboost* dapat disebabkan berbagai hal. Data yang *imbalance* dengan banyak *noise* serta *outliers* dapat menjadi penyebabnya. Selain itu, *Adaboost* menentukan klasifikasi berdasarkan jumlah *vote* dan mengkombinasikan banyak *weak learners* untuk melakukan klasifikasi, sehingga data akan condong pada label positif dikarenakan data pada penelitian ini merupakan data yang *imbalance* dan cenderung pada label positif. Gabungan antara data, NBC, dan *Adaboost* pada penelitian ini juga menyebabkan akurasi dari pengklasifikasi menurun.

BAB V

PENUTUP

Bab ini berisi kesimpulan dan saran dari penelitian analisis sentimen terhadap belajar daring menggunakan optimasi *Naïve Bayes Classifier* dengan *Adaboost*.

5.1 Kesimpulan

Berdasarkan penelitian yang telah dilakukan, berikut merupakan kesimpulan yang dapat ditarik dari penelitian ini:

1. NBC dengan *Adaboost* dapat diimplementasikan untuk analisis sentimen, menggunakan bahasa pemrograman Python dengan *library* Scikit-learn, Pandas, Matplotlib, NLTK, dan TextBlob. Proses analisis sentimen dimulai dengan studi literatur, pengumpulan data, penghapusan data duplikat, penerjemahan data, *data sampling*, *data preprocessing*, pelabelan data, validasi hasil pelabelan, visualisasi hasil pelabelan, *feature extraction*, *cross validation*, *modeling*, klasifikasi sentimen, dan analisis hasil sentimen.
 2. *Adaboost* bekerja kurang baik dengan *base estimator* NBC dengan menghasilkan akurasi sebesar 64%. Sedangkan, *Naïve Bayes Classifier* menghasilkan akurasi yang baik, yaitu sebesar 78% untuk klasifikasi sentimen.
 - a. Perbedaan nilai *min_df* dapat mempengaruhi akurasi NBC, semakin besar nilai *min_df* maka akurasi dari NBC dapat bertambah. Penerapan nilai *min_df* yang berbeda pada NBC dengan *Adaboost*, tidak menghasilkan perbedaan signifikan pada nilai akurasinya.
 - b. Kombinasi antara data pada penelitian ini yang *imbalance* dengan NBC dan *Adaboost* menyebabkan penurunan akurasi. Selain itu, NBC merupakan pengklasifikasi yang stabil sehingga sulit untuk *Adaboost* dapat meningkatkan akurasinya.
 3. Hasil dari klasifikasi sentimen menggunakan NBC menunjukkan bahwa sentimen masyarakat mengenai belajar daring cenderung pada sentimen positif dengan jumlah 15.593 data atau sebesar 89,5%, sedangkan sentimen

negatif didapatkan 8.864 data atau sebesar 10,5%. Klasifikasi NBC dengan *Adaboost* melabelkan semua data dengan positif.

5.2 Saran

Penelitian yang telah dilakukan tidak luput dari kekurangan, diperlukan perbaikan pada penelitian ini. Berikut beberapa saran yang dapat dilakukan untuk mengembangkan penelitian ini:

1. NBC dan NBC dengan *Adaboost* hanya memperhatikan pola dari kalimat, dan tidak mengetahui konteks dari kalimat. Terlebih lagi, pada masa pandemi Covid-19 banyak istilah yang terkesan mengandung sentimen positif tetapi bermakna negatif, seperti kalimat “positif Covid-19”. Pada penelitian selanjutnya diharapkan hal ini dapat ditangani.
 2. Variabel yang diteliti pada *tweets* dapat ditambahkan variabel lain, seperti lokasi, dan waktu. Hal ini dilakukan agar dapat terlihat daerah mana saja yang menyukai, dan kurang menyukai belajar daring.
 3. Emotikon pada *tweets* dapat ikut diproses, mengingat banyaknya penggunaan emotikon di Twitter.
 4. Label untuk sentimen dapat diperluas lagi bukan hanya positif dan negatif, tetapi dispesifikkan seperti marah, sedih, senang, dan sebagainya.
 5. Metode *machine learning* lain dapat diterapkan pada penelitian ini agar memungkinkan didapat akurasi yang lebih tinggi dan didapatkan pengklasifikasi sentimen yang lebih baik.
 6. NBC dapat dikombinasikan dengan metode lain yang cocok agar didapatkan pengklasifikasi dengan performa yang lebih baik.

DAFTAR PUSTAKA

- A., V., & Sonawane, S. S. (2016). Sentiment Analysis of Twitter Data: A Survey of Techniques. *International Journal of Computer Applications*, 139(11), 5–15. <https://doi.org/10.5120/ijca2016908625>

Agarwal, B., & Mittal, N. (2016). Prominent Feature Extraction for Sentiment Analysis. In *Prominent Feature Extraction for Sentiment Analysis* (Issue i). <https://doi.org/10.1007/978-3-319-25343-5>

Aggarwal, C. C. (2015). Data Mining. In *Cancer Letters* (Vol. 14, Issue 3). Springer International Publishing. <https://doi.org/10.1007/978-3-319-14142-8>

Andreyestha, A., & Subekti, A. (2020). Analisa Sentiment Pada Ulasan Film Dengan Optimasi Ensemble Learning. *Jurnal Informatika*, 7(1), 15–23. <https://doi.org/10.31311/ji.v7i1.6171>

Ankit, & Saleena, N. (2018). An Ensemble Classification System for Twitter Sentiment Analysis. *Procedia Computer Science*, 132(Iccids), 937–946. <https://doi.org/10.1016/j.procs.2018.05.109>

Anugrahana, A. (2020). Hambatan, Solusi dan Harapan: Pembelajaran Daring Selama Masa Pandemi Covid-19 Oleh Guru Sekolah Dasar. *Scholaria: Jurnal Pendidikan Dan Kebudayaan*, 10(3), 282–289. <https://doi.org/10.24246/j.js.2020.v10.i3.p282-289>

Auria, L., & Moro, R. A. (2008). *Support Vector Machines (SVM) as a technique for solvency analysis* (Issue 811). Deutsches Institut für Wirtschaftsforschung (DIW). <http://hdl.handle.net/10419/27334>

Badan Pusat Statistik. (2020). *Statistik Indonesia 2020 Penyediaan Data untuk Perencanaan Pembangunan*. <https://doi.org/1101005>

Bengio, Y., & Grandvalet, Y. (2003). *No unbiased Estimator of the Variance of K-Fold Cross-Validation*. <https://econpapers.repec.org/RePEc:cir:cirwor:2003s-22>

- Berrar, D. (2018). Cross-validation. *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, 1–3(January 2018), 542–545. <https://doi.org/10.1016/B978-0-12-809633-8.20349-X>

Bhatia, P. (2019). Data Mining and Data Warehousing. In *Studies in Computational Intelligence* (Vol. 47). Cambridge University Press. <https://doi.org/10.1017/9781108635592>

Bhatia, S., Chaudhary, P., & Dey, N. (2020). *Opinion Mining in Information Retrieval*. Springer Singapore. <https://doi.org/10.1007/978-981-15-5043-0>

Cahyani, R. (2020). Analisis Sentimen Pada Media Sosial Twitter Terhadap Tokoh Publik Peserta Pilpres 2019. In *Digilib* (Vol. 12, Issue 1). <https://doi.org/10.18860/mat.v12i1.8356>

Castellà, Q., & Sutton, C. (2014). Word storms: Multiples of word clouds for visual comparison of documents. *WWW 2014 - Proceedings of the 23rd International Conference on World Wide Web*, 665–675. <https://doi.org/10.1145/2566486.2567977>

Cielen, D., Meysman, A. D. B., & Ali, M. (2016). *Davy Cielen Arno D . B . Meysman*.

Colas, F., & Brazdil, P. (2006). On the behavior of SVM and some older algorithms in binary text classification tasks. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 4188 LNCS, 45–52. https://doi.org/10.1007/11846406_6

covid19.go.id. (2020). *Peta Sebaran COVID-19*. <https://covid19.go.id/peta-sebaran-covid19>

D’Andrea, A., Ferri, F., Grifoni, P., & Guzzo, T. (2015). Approaches, Tools and Applications for Sentiment Analysis Implementation. *International Journal of Computer Applications*, 125(3), 26–33. <https://doi.org/10.5120/ijca2015905866>

Davidson-Pilon, C. (2015). *Bayesian Methods for Hackers: Probabilistic*

Programming and Bayesian Inference (1st ed.). Addison-Wesley Professional.

Devika, M. D., Sunitha, C., & Ganesh, A. (2016). Sentiment Analysis: A Comparative Study on Different Approaches. *Procedia Computer Science*, 87, 44–49. <https://doi.org/10.1016/j.procs.2016.05.124>

Dey, N., Borah, S., Babo, R., & Ashour, S. A. (2018). Social Network Analytics: Computational Research Methods and Techniques. In *Academic Press*.
https://www.cambridge.org/core/product/identifier/CBO9781107415324A009/type/book_part

Downey, A. B. (2013). Think Bayes. In M. L. Spencer & A. Spencer (Eds.), *O'Reilly Media* (1st ed.).

Dwi, B., Amelia, A., Hasanah, U., & Putra, A. M. (2020). Analisis Keefektifan Pembelajaran Online di Masa Pandemi Covid-19. *Jurnal Pendidikan Guru Sekolah Dasar*, 2(1), 3.

Ellis, B. (2014). *Sentiment Analysis: Why It's Never 100% Accurate*.
<https://brnrd.me/sentiment-analysis-never-accurate>

Feng, X. (2019). Research of sentiment analysis based on adaboost algorithm. *Proceedings - 2019 International Conference on Machine Learning, Big Data and Business Intelligence, MLDBI 2019*, 279–282. <https://doi.org/10.1109/MLDBI48998.2019.00062>

Fitri, V. A., Andreswari, R., & Hasibuan, M. A. (2019). Sentiment analysis of social media Twitter with case of Anti-LGBT campaign in Indonesia using Naïve Bayes, decision tree, and random forest algorithm. *Procedia Computer Science*, 161, 765–772. <https://doi.org/10.1016/j.procs.2019.11.181>

Freund, Y. (2001). An adaptive version of the boost by majority algorithm. *Machine Learning*, 43(3), 293–318.
<https://doi.org/10.1023/A:1010852229904>

Géron, A. (2017). *Hands-on machine learning with Scikit-Learn and TensorFlow : concepts, tools, and techniques to build intelligent systems.*

O'Reilly Media.

- Giachanou, A., & Crestani, F. (2016). Like it or not: A survey of Twitter sentiment analysis methods. *ACM Computing Surveys*, 49(2).
<https://doi.org/10.1145/2938640>

Google.com. (2020). *How can we help you?*
<https://support.google.com/trends/?hl=ko#topic=6248052>

Google Trends. (2020). *Google Trends Belajar Daring.*

Gorunescu, F. (2011). *Data Mining* (Vol. 12). Springer Berlin Heidelberg.
<https://doi.org/10.1007/978-3-642-19721-5>

Group, F. (2014). *Data Algorithms: Recipes for scaling up with Hadoop and Spark.*

Gupta, I., & Joshi, N. (2020). Enhanced twitter sentiment analysis using hybrid approach and by accounting local contextual semantic. *Journal of Intelligent Systems*, 29(1), 1611–1625. <https://doi.org/10.1515/jisys-2019-0106>

Han, J., Kamber, M., & Pei, J. (2012). Getting to Know Your Data. In *Data Mining* (Vol. 2, Issue January 2013, pp. 39–82). Elsevier.
<https://doi.org/10.1016/B978-0-12-381479-1.00002-2>

Hariyanti, D., Haq, A., & Hidayat, N. (2020). Identifikasi Hambatan Mahasiswa Dalam Pelaksanaan Pembelajaran Biologi Secara Daring Selama Pandemi Covid-19 Di Kabupaten Jember. *Jurnal Pendidikan Biologi*, 1(1), 11–21.

Hemmatian, F., & Sohrabi, M. K. (2019). A survey on classification techniques for opinion mining and sentiment analysis. *Artificial Intelligence Review*, 52(3), 1495–1545. <https://doi.org/10.1007/s10462-017-9599-6>

Hu, W., & Hu, W. (2005). *Network-based intrusion detection using adaboost algorithm*. Proceedings - 2005 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2005. <https://doi.org/10.1109/WI.2005.107>

Huang, M. L., Huang, T. H., & Zhang, J. (2009). TreemapBar: Visualizing additional dimensions of data in bar chart. *Proceedings of the International*

Conference on Information Visualisation, December 2016, 98–103.
<https://doi.org/10.1109/IV.2009.22>

Jaman, J. H., Abdulrohman, R., Suharso, A., Sulistiowati, N., & Dewi, I. P. (2020). Sentiment analysis on utilizing online transportation of indonesian customers using tweets in the normal era and the pandemic covid-19 era with support vector machine. *Advances in Science, Technology and Engineering Systems*, 5(5), 389–394. <https://doi.org/10.25046/AJ050549>

Jatira, Y., & S, N. (2021). Fenomena Stress dan Pembiasaan Belajar Daring dimasa Pandemi Covid-19. *Edukatif: Jurnal Ilmu Pendidikan*, 3(1), 35–43.
<https://doi.org/10.31004/edukatif.v3i1.187>

Jun, S.-P., Yoo, H. S., & Choi, S. (2018). Ten years of research change using Google Trends: From the perspective of big data utilizations and applications. *Technological Forecasting and Social Change*, 130(November), 69–87. <https://doi.org/10.1016/j.techfore.2017.11.009>

Karoui, J., Benamara, F., & Moriceau, V. (2019). *Automatic Detection of Irony Opinion Mining in Microblogs and Social Media*.

Kaur, C., & Sharma, A. (2020). Sentiment analysis of tweets on social issues using machine learning approach. *International Journal of Advanced Trends in Computer Science and Engineering*, 9(4), 6303–6311.
<https://doi.org/10.30534/ijatcse/2020/310942020>

Kemp, S. (2020). *More Than Half of The People on Earth Now Use Social Media.*
<https://wearesocial.com/blog/2020/07/more-than-half-of-the-people-on-earth-now-use-social-media>

Kirilenko, A. P., Stepchenkova, S. O., Kim, H., & Li, X. (Robert). (2018). Automated Sentiment Analysis in Tourism: Comparison of Approaches. *Journal of Travel Research*, 57(8), 1012–1025.
<https://doi.org/10.1177/0047287517729757>

Korada, N. K. (2012). *Implementation of Naive Bayesian Classifier and Ada-Boost Algorithm Using Maize Expert System.*

- Kurniawan, T. (2017). *Implementasi Text Mining Pada Analisis Sentimen Pengguna Twitter Terhadap Media Mainstream Menggunakan Naïve Bayes Classifier Dan Support Vector Machine Media Mainstream Menggunakan Naïve Machine*. 1.
- Kusrorong, N. S. B., Sina, D. R., Rumlaklak, N. D., Komputer, J. I., & Cendana, U. N. (2019). *Kajian Machine Learning Dengan Komparasi Klasifikasi Prediksi Dataset Tenaga Kerja Non-Aktif*. 7(1), 37–49.
- Lappeman, J., Clark, R., Evans, J., Sierra-Rubia, L., & Gordon, P. (2020). Studying social media sentiment using human validated analysis. *MethodsX*, 7, 100867. <https://doi.org/10.1016/j.mex.2020.100867>
- Liu, B. (2012). Sentiment Analysis and Opinion Mining. *Synthesis Lectures on Human Language Technologies*, 5(1), 1–167.
<https://doi.org/10.2200/S00416ED1V01Y201204HLT016>
- Liu, B. (2015). Sentiment Analysis: Mining Opinions, Sentiments, and Emotions. In *Cambridge: Cambridge University Press*.
<https://doi.org/10.1017/CBO9781139084789>
- Loria, S. (2020a). *TextBlob: Simplified Text Processing*.
<https://textblob.readthedocs.io/en/dev/>
- Loria, S. (2020b). TextBlob Documentation. *TextBlob*, 69.
<https://media.readthedocs.org/pdf/textblob/latest/textblob.pdf>
- Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press. <https://doi.org/10.5555/1394399>
- Mansoor, M., Gurumurthy, K., U, A. R., & Prasad, V. R. B. (2020). *Global Sentiment Analysis Of COVID-19 Tweets Over Time*.
<http://arxiv.org/abs/2010.14234>
- Mardiana, T., Syahreva, H., & Tuslaela, T. (2019). Komparasi Metode Klasifikasi Pada Analisis Sentimen Usaha Waralaba Berdasarkan Data Twitter. *Jurnal Pilar Nusa Mandiri*, 15(2), 267–274.
<https://doi.org/10.33480/pilar.v15i2.752>

- Martinez, E. Z. (2015). Description of continuous data using bar graphs: A misleading approach. *Revista Da Sociedade Brasileira de Medicina Tropical*, 48(4), 494–497. <https://doi.org/10.1590/0037-8682-0013-2015>

Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 5(4), 1093–1113. <https://doi.org/10.1016/j.asej.2014.04.011>

Ming Ting, K., & Zheng, Z. (2003). A study of AdaBoost with naive Bayesian classifiers: Weakness and improvement. *Computational Intelligence*, 19(2), 186–200. <https://doi.org/10.1111/1467-8640.00219>

Mostafa, L. (2021). Egyptian Student Sentiment Analysis Using Word2vec During the Coronavirus (Covid-19) Pandemic. In *Advances in Intelligent Systems and Computing: Vol. 1261 AISC* (Issue September). Springer International Publishing. https://doi.org/10.1007/978-3-030-58669-0_18

Mozetič, I., Torgo, L., Cerqueira, V., & Smailović, J. (2018). How to evaluate sentiment classifiers for Twitter time-ordered data? *PLoS ONE*, 13(3), 1–20. <https://doi.org/10.1371/journal.pone.0194317>

Mutaqinah, R., & Hidayatullah, T. (2020). Implementasi Pembelajaran Daring (Program BDR) Selama Pandemi Covid-19 di Provinsi Jawa Barat. *Jurnal Petik*, 6(2), 86–95. <https://doi.org/10.31980/jpetik.v6i2.869>

National Audit Office. (2010). A Practical Guide to Sampling. *A Practical Guide to Sampling*, 1–19. <http://www.nao.org.uk/wp-content/uploads/2001/06/SamplingGuide.pdf>

Nengrum, T. A., Pettasolong, N., & Nuriman, M. (2021). Kelebihan dan Kekurangan Pembelajaran Luring dan Daring dalam Pencapaian Kompetensi Dasar Kurikulum Bahasa Arab di Madrasah Ibtidaiyah 2 Kabupaten Gorontalo. *Jurnal Pendidikan*, 30(1), 1–12. <http://journal.univetbantara.ac.id/index.php/jp/article/view/1190>

Panda, M., Hassanien, A.-E., & Abraham, A. (Eds.). (2018). *Big Data Analytics*. CRC Press. <https://doi.org/10.1201/9781315112626>

- Puspaningtyas, N. D., & Dewi, P. S. (2020). Persepsi Peserta Didik Terhadap Pembelajaran Berbasis Daring. *Jurnal Pembelajaran Matematika Inovatif (JPMI)*, 3(6), 703–712. <https://doi.org/10.22460/jpmi.v3i6.703-712>

Ramachandran, D., & Parvathi, R. (2019). Analysis of Twitter Specific Preprocessing Technique for Tweets. *Procedia Computer Science*, 165, 245–251. <https://doi.org/10.1016/j.procs.2020.01.083>

Ramezan, C. A., Warner, T. A., & Maxwell, A. E. (2019). Evaluation of sampling and cross-validation tuning strategies for regional-scale machine learning classification. *Remote Sensing*, 11(2). <https://doi.org/10.3390/rs11020185>

Rolliawati, D., Khalid, K., & Rozas, I. S. (2020). Teknologi Opinion Mining untuk Mendukung Strategic Planning. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 7(2), 293. <https://doi.org/10.25126/jtiik.2020721685>

Russell, S., & Norvig, P. (2010). Artificial Intelligence: A Modern Approach. In *Upper Saddle River, NJ: Prentice Hall* (3rd ed). Upper Saddle River, NJ : Prentice Hall.

Sadikin, A., & Hamidah, A. (2020). Pembelajaran Daring di Tengah Wabah Covid-19. *BIODIK*, 6(2), 109–119. <https://doi.org/10.22437/bio.v6i2.9759>

Samuel, J., Rahman, M. M., Nawaz Ali, G. G. M., Samuel, Y., & Pelaez, A. (2020). Feeling Like It is Time to Reopen Now? COVID-19 New Normal Scenarios based on Reopening Sentiment Analytics. *ArXiv*. <https://doi.org/10.2139/ssrn.3604802>

Schumacher, A. (2015). *TextBlob Sentiment : Calculating Polarity and Subjectivity*. https://planspace.org/20150607-textblob_sentiment/

Setifani, N. A., Fitriana, D. N., & Yusuf, A. (2020). Perbandingan Algoritma Naïve Bayes, SVM, dan Decision Tree untuk Klasifikasi SMS Spam. *JUSIM (JURNAL SISTEM INFORMASI MUSIRAWAS)*, 05(02), 153–160. <https://doi.org/10.32767/JUSIM.V5I02.956>

Sharma, D., Sabharwal, M., Goyal, V., & Vij, M. (2020). Sentiment analysis techniques for social media data: A review. In *Advances in Intelligent*

Systems and Computing (Vol. 1045). Springer Singapore.
https://doi.org/10.1007/978-981-15-0029-9_7

Sodik, Muhammad & Siyoto, S. (2015). *DASAR METODOLOGI PENELITIAN* (Ayup (Ed.)). Literasi Media Publishing.

Sridhya, V., & Anitha, R. (2010). Evaluating Preprocessing Techniques in Text Categorization. *International Journal of Computer Science and Application*, 49–51.

Sujarwo, S., Sukmawati, S., Akhiruddin, A., Ridwan, R., & Suharti Siradjuddin, S. S. (2020). An Analysis of University Students' Perspective On Online Learning in The Midst of Covid-19 Pandemic. *Jurnal Pendidikan Dan Pengajaran*, 53(2), 125. <https://doi.org/10.23887/jpp.v53i2.24964>

Sukendro, S., Habibi, A., Khaeruddin, K., Indrayana, B., Syahruddin, S., Makadada, F. A., & Hakim, H. (2020). Using an extended Technology Acceptance Model to understand students' use of e-learning during Covid-19: Indonesian sport science education context. *Heliyon*, 6(11), e05410. <https://doi.org/10.1016/j.heliyon.2020.e05410>

Taherdoost, H. (2016). Sampling Methods in Research Methodology ; How to Choose a Sampling Technique for Research Hamed Taherdoost To cite this version : HAL Id : hal-02546796 Sampling Methods in Research Methodology ; How to Choose a Sampling Technique for. *International Journal of Academic Research in Management (IJARM)*, 5(2), 18–27.

Tosepu, R., Gunawan, J., Effendy, D. S., Ahmad, L. O. A. I., Lestari, H., Bahar, H., & Asfian, P. (2020). Correlation between weather and Covid-19 pandemic in Jakarta, Indonesia. *Science of the Total Environment*, 725. <https://doi.org/10.1016/j.scitotenv.2020.138436>

Tsytsarau, M., & Palpanas, T. (2012). Survey on mining subjective data on the web. *Data Mining and Knowledge Discovery*, 24(3), 478–514.
<https://doi.org/10.1007/s10618-011-0238-6>

Unpingco, J. (2016). Python for probability, statistics, and machine learning. In

Python for Probability, Statistics, and Machine Learning.

<https://doi.org/10.1007/978-3-319-30717-6>

Utami, L. D., & Wahono, R. S. (2015). Integrasi Metode Information Gain untuk Seleksi Fitur dan AdaBoost untuk Mengurangi Bias pada Analisis Sentimen Review Restoran Menggunakan Algoritma Naive Bayes. *Journal of Intelligent Systems*, 1(2), 120–126.

VanderPlas, J. (2016). Python Data Science Handbook: Essential Tools for Working with Data. In *O'Reilly*.

<http://shop.oreilly.com/product/0636920034919.do> <https://jakevdp.github.io/PythonDataScienceHandbook/05.01-what-is-machine-learning.html>

Vidya, N. A. (2015). *Twitter Sentiment Analysis Terhadap Brand Reputation: Studi Kasus PT.XL Axiata TBk*. Universitas Indonesia.

Widhi Kurniawan, A., & Puspitaningtyas, Z. (2016). *METODE PENELITIAN KUANTITATIF*.

Yue, L., Chen, W., Li, X., Zuo, W., & Yin, M. (2019). A survey of sentiment analysis in social media. *Knowledge and Information Systems*, 60(2), 617–663. <https://doi.org/10.1007/s10115-018-1236-4>