

SKRIPSI

**ANALISIS SENTIMEN SEKOLAH TATAP MUKA DI
INDONESIA**



YALVI HIDAYAT

NPM: 6181801044

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INFORMASI DAN SAINS
UNIVERSITAS KATOLIK PARAHYANGAN
2022**

UNDERGRADUATE THESIS

**SENTIMENT ANALYSIS OF FACE-TO-FACE SCHOOL IN
INDONESIA**



YALVI HIDAYAT

NPM: 6181801044

**DEPARTMENT OF INFORMATICS
FACULTY OF INFORMATION TECHNOLOGY AND SCIENCES
PARAHYANGAN CATHOLIC UNIVERSITY
2022**

LEMBAR PENGESAHAN

ANALISIS SENTIMEN SEKOLAH TATAP MUKA DI INDONESIA

YALVI HIDAYAT

NPM: 6181801044

Bandung, 23 Juni 2022

Menyetujui,

Pembimbing

**Digitally signed
by Mariskha Tri
Adithia**

Mariskha Tri Adithia, P.D.Eng

Ketua Tim Penguji

**Digitally signed
by Luciana
Abednego**

Luciana Abednego, M.T.

Anggota Tim Penguji

**Digitally signed
by Vania Natali**

Vania Natali, M.T.

Mengetahui,

Ketua Program Studi

**Digitally signed
by Mariskha Tri
Adithia**

Mariskha Tri Adithia, P.D.Eng

PERNYATAAN

Dengan ini saya yang bertandatangan di bawah ini menyatakan bahwa skripsi dengan judul:

ANALISIS SENTIMEN SEKOLAH TATAP MUKA DI INDONESIA

adalah benar-benar karya saya sendiri, dan saya tidak melakukan penjiplakan atau pengutipan dengan cara-cara yang tidak sesuai dengan etika keilmuan yang berlaku dalam masyarakat keilmuan.

Atas pernyataan ini, saya siap menanggung segala risiko dan sanksi yang dijatuhkan kepada saya, apabila di kemudian hari ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya, atau jika ada tuntutan formal atau non-formal dari pihak lain berkaitan dengan keaslian karya saya ini.

Dinyatakan di Bandung,
Tanggal 23 Juni 2022



YALVI HIDAYAT
NPM: 6181801044

ABSTRAK

Pandemi *Covid-19* yang sedang berlangsung di seluruh dunia membuat banyak aktivitas menjadi tidak dapat bekerja dengan baik. Salah satu aktivitas tersebut adalah kegiatan belajar mengajar. Seiring berjalannya waktu dan jumlah kasus positif *Covid-19* yang terus menurun, sudah mulai dibuka kembali kegiatan ajar-mengajar sekolah tatap muka. Pembukaan kembali sekolah tatap muka menuai kontroversi tersendiri. Banyak dari mereka khawatir dengan kembalinya sekolah tatap muka akan terjadi suatu kluster penularan *Covid-19* secara masif. Ada juga yang sudah rindu dengan suasana sekolah tatap muka. Berdasarkan hal ini, penelitian ini muncul untuk menganalisis sentimen pada sekolah tatap muka di Indonesia.

Pada penelitian ini, penelitian akan membuat perangkat lunak untuk melakukan analisis sentimen dengan menggunakan alur tahapan *data mining*. Analisis sentimen merupakan bagian dari *data mining* yang di mana untuk menganalisis ukuran emosi positif, negatif, atau netral dari suatu opini. Analisis sentimen termasuk dalam konsep *data mining*. *Data mining* merupakan proses terdiri dari tahapan seleksi, eksplorasi, dan analisis untuk mencapai pengetahuan. Proses diawali dengan pengumpulan data dengan menggunakan media sosial *Twitter* sebagai sumber data. Data tersebut kemudian diproses kembali untuk memudahkan proses analisis data yang akan dilakukan. Analisis data yang dilakukan akan melihat kategori sentimen apa yang paling dominan untuk topik sekolah tatap muka. Selain melakukan analisis data, penelitian ini akan membangun model yang akan digunakan untuk proses prediksi sentimen. Pembangunan model ini akan menggunakan teknik klasifikasi *Decision Tree* sebagai klasifikator dan kemudian mengevaluasi hasil pembangunan model untuk mengukur seberapa baik akurasi model yang berhasil dibangun. Pembangunan model akan kemudian dioptimasi dengan menggunakan parameter `max_depth`. Parameter `max_depth` merupakan parameter untuk mengatur kedalaman *tree* yang sedang dibangun.

Berdasarkan eksperimen yang telah dilakukan dalam penelitian ini, didapatkan nilai akurasi model yang dihasilkan oleh teknik klasifikasi *Decision Tree* sebesar 86.4%. Selain itu juga, terdapat proses optimasi yang dilakukan dengan mengatur `max_depth`. Nilai `max_depth` yang paling optimal didapatkan 135 dengan akurasi model sebesar 86.6%. Model ini kemudian digunakan untuk melakukan prediksi sentimen yang diberikan oleh pengguna yang kemudian akan mendapatkan keluaran berupa label sentimen. Hasil yang didapatkan dalam prediksi sentimen didapatkan secara umum model sudah dapat melakukan prediksi sentimen meskipun terdapat beberapa kasus kategori sentimen tidak tepat dikategorikan. Untuk analisis data, didapatkan bahwa sentimen negatif menjadi sentimen yang paling dominan untuk topik sekolah tatap muka dengan persentase sebesar 37%. Adapun kata bersentimen yang paling sering muncul untuk kategori sentimen negatif adalah kata 'lama'.

Kesimpulan yang didapatkan dalam penelitian ini adalah analisis sentimen untuk topik sekolah tatap muka dapat diselesaikan dengan konsep *data mining* dan konsep analisis data untuk mencapai tujuan penelitian ini. Selain itu juga, pembangunan model untuk melakukan prediksi sentimen berhasil dibangun dengan menggunakan teknik klasifikasi *Decision Tree*.

Kata-kata kunci: Analisis Sentimen, Sekolah Tatap Muka, Pohon Keputusan, Data Mining, Twitter, Analisis Data

ABSTRACT

The ongoing Covid-19 pandemic around the world has brought many activities to a standstill. One of these activities is teaching and learning activities. As time goes by and the number of positive cases of Covid-19 continues to decline, face-to-face school teaching activities have begun to reopen. The reopening of face-to-face schools has been controversial. Many of them are worried that with the return of face-to-face schools there will be a massive cluster of Covid-19 transmissions and some already miss the face-to-face school atmosphere. Based on this, this study appears to analyze sentiment in face-to-face schools in Indonesia.

In this study, the research will create software to perform sentiment analysis using the flow of data mining stages. Sentiment analysis is part of data mining which is to analyze the measure of positive, negative, or neutral emotions from an opinion. Sentiment analysis is a subset concept of data mining. Data mining is a sequence of process from selection, exploration, and analysis to gather knowledge. The process begins with data collection by using social media Twitter as a data source. The data is then reprocessed to facilitate the data analysis process to be carried out. The data analysis carried out will see which sentiment category is the most dominant for face-to-face school topics. In addition to conducting data analysis, this research will build a model that will be used for the sentiment prediction process. The construction of this model will use the Decision Tree classification technique as a classifier and then evaluate the results of model development to measure how well the accuracy of the model has been built. The model development will then be optimized using the `max_depth` parameter. The `max_depth` parameter is a parameter to set the depth of the tree being built.

Based on the experiments that have been carried out in this study, the accuracy value of the model generated by the Decision Tree classification technique was 86.4%. In addition, there is an optimization process that is carried out by setting `max_depth`. The optimal `max_depth` value is 135 with an accuracy of 86.6%. This model then used to predict the sentiment given by the user who then will get an output in the form of a sentiment label. The results obtained in sentiment prediction are generally found that the model is able to predict sentiment even though there are some cases where the sentiment category is not properly categorized. For data analysis, it was found that negative sentiment became the most dominant sentiment for face-to-face school topics with a percentage of 37%. The word with sentiment that most often appears for the category of negative sentiment is 'lama'.

The conclusion obtained in this study is that sentiment analysis for face-to-face school topics can be completed with the concept of data mining and the concept of data analysis to achieve the objectives of this study. In addition, the construction of a model to predict sentiment was successfully built using the Decision Tree classification technique.

Keywords: Sentiment Analysis, Face-to-face Schools, Decision Tree, Data Mining, Twitter, Data Analysis

*Skripsi ini ditujukan untuk diri sendiri, keluarga, dan teman-teman
yang terbaik yang pernah saya temui*

KATA PENGANTAR

Puji dan syukur yang penulis panjatkan ke hadirat Tuhan Yang Maha Esa karena penulis telah menyelesaikan penelitian yang berjudul “Analisis Sentimen Sekolah Tatap Muka Di Indonesia”. Penelitian ini tidak akan berhasil dilakukan tanpa adanya dukungan dari banyak pihak. Oleh karena itu, penulis ingin mengucapkan terima kasih kepada:

- Ibu Marishka Tri Adhithia, P.D.Eng yang telah memberikan bimbingan, motivasi hidup, dan arahan selama proses bimbingan skripsi berlangsung.
- Ibu Luciana Abednego, M.T. dan Ci Vania Natali, M.T yang telah melakukan pengujian skripsi dan memberikan masukan yang membangun terkait penelitian yang sudah dilaksanakan.
- Ko Kris yang telah sabar dalam memberikan masukan tambahan dan ilmu baru yang didapatkan untuk menunjang penelitian ini.
- Keluarga baik dari mamah, papap, teh Yulia, hingga Nuna dan Zavier yang mendukung penulis dalam membuat penelitian skripsi ini.
- Ko Indra sebagai *peer-reviewer* tambahan sekaligus juga memberikan ilmu baru yang didapatkan oleh penulis.
- Sahabat dekat penulis semasa kuliah Vino, Arthur, Gian, Devon, Mario, Bryan, Gio, Jiang Han, Afi, dan Alfred yang telah mendukung penulis dalam penelitian ini.
- Sahabat dekat penulis semasa SMA Bram, Mikha, Fari, dan Famian yang telah mendukung penulis dalam penelitian ini.
- Kawan-kawan seluruh angkatan 18 jurusan Informatika.

Penulis menyadari bahwa penelitian ini tidaklah sempurna. Oleh karena itu, penulis memohon maaf jika terdapat kekurangan dalam hasil penyusunan dokumen penelitian ini. Penulis juga berharap penelitian ini dapat berguna bagi banyak pihak yang berkepentingan.

Bandung, Juni 2022

Penulis

DAFTAR ISI

KATA PENGANTAR	xv
DAFTAR ISI	xvii
DAFTAR GAMBAR	xxi
DAFTAR TABEL	xxv
DAFTAR KODE PROGRAM	xxvii
1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan	4
1.4 Batasan Masalah	4
1.5 Metodologi	5
1.6 Sistematika Penulisan	5
2 LANDASAN TEORI	7
2.1 Twitter [1]	7
2.2 Machine Learning [2]	8
2.2.1 Tipe Machine Learning [2]	8
2.2.2 Teknik Klasifikasi [3]	9
2.2.3 Evaluasi Teknik Klasifikasi [3, 4]	11
2.3 Data Mining [3]	13
2.3.1 Teknik Ekstraksi Fitur [5]	15
2.4 Teknik Klasifikasi Decision Tree [3]	16
2.4.1 Pembangunan Decision Tree [3]	18
2.4.2 Entropy [6]	18
2.4.3 Pruning Tree [3]	19
2.5 Text Mining [7]	20
2.5.1 Definisi <i>Text Mining</i>	20
2.5.2 Tujuan <i>Text Mining</i>	21
2.5.3 Proses <i>Text Mining</i> [4]	21
2.6 Analisis Sentimen [1]	21
2.7 Natural Language Processing [8]	21
2.7.1 Lexicon [8]	22
2.7.2 Case Folding [4]	22
2.7.3 Tokenization [5]	23
2.7.4 Stemming [5]	23
2.7.5 Algoritma Stemming Nazief-Adriani [9, 10, 11]	24
2.7.6 Filtering [5]	25
2.7.7 N-Gram [5]	26

2.8	Morfologi dalam Bahasa Indonesia [12]	26
3	ANALISIS MASALAH DAN EKSPLORASI TOOLS	29
3.1	Analisis Masalah	29
3.1.1	<i>Tools</i> dan <i>Library</i> yang Digunakan	30
3.2	Tahap Penelitian	31
3.2.1	Analisis Sentimen Data Twitter Utama	31
3.2.2	Analisis Sentimen Data Twitter Lanjut	31
3.3	Eksplorasi Tools	32
3.3.1	Eksplorasi Pengumpulan Data	32
3.3.2	Eksplorasi Penggabungan Data	35
3.3.3	Eksplorasi Data Cleaning	36
3.3.4	Eksplorasi Tokenisasi Data	38
3.3.5	Eksplorasi Normalisasi Data	39
3.3.6	Eksplorasi Ekstraksi Fitur untuk Pemberian Label	44
3.3.7	Eksplorasi Penentuan Label dari Skor Sentimen	47
3.3.8	Eksplorasi Visualisasi Data	48
3.3.9	Eksplorasi Ekstraksi Fitur untuk Klasifikasi Data	51
3.3.10	Eksplorasi Teknik Klasifikasi <i>Decision Tree</i>	52
3.3.11	Eksplorasi Evaluasi Klasifikasi	54
3.3.12	Eksplorasi Proses Penyimpanan Model	56
3.3.13	Eksplorasi Tools Pembuatan GUI PAGE	57
4	PENYIAPAN DATASET DAN PEMBANGUNAN MODEL	63
4.1	Deskripsi Dataset	63
4.2	Analisis Sentimen Data Twitter Utama	65
4.2.1	Proses Data Cleaning	65
4.2.2	Proses Normalisasi Data	66
4.2.3	Proses Ekstraksi Fitur untuk Pelabelan Data	69
4.2.4	Proses Pelabelan Data	70
4.2.5	Proses Visualisasi Data	73
4.2.6	Proses Ekstraksi Fitur	80
4.2.7	Proses Klasifikasi dan Evaluasi Klasifikasi	82
4.2.8	Proses Prediksi Sentimen dengan Model yang Sudah Disimpan	85
4.2.9	Kesimpulan Analisis Sentimen Data Twitter Utama	87
4.3	Analisis Sentimen Data Twitter Lanjut	88
4.3.1	Proses Penarikan Data	88
4.3.2	Proses Pembersihan Data	89
4.3.3	Proses Normalisasi Data	89
4.3.4	Proses Pelabelan Data	93
4.3.5	Proses Visualisasi Data	94
4.3.6	Proses Ekstraksi Fitur	112
4.3.7	Proses Klasifikasi dan Evaluasi Klasifikasi	114
4.3.8	Proses Prediksi Sentimen dengan Model yang Sudah Disimpan	119
4.3.9	Kesimpulan Analisis Sentimen Data Twitter Lanjut	120
5	IMPLEMENTASI PERANGKAT LUNAK	123
5.1	Pengembangan Prototipe Perangkat Lunak	123
5.1.1	Pengembangan <i>Mockup</i> Perangkat Lunak	123
5.1.2	Pengembangan Perangkat Lunak	132
6	KESIMPULAN DAN SARAN	145

6.1 Kesimpulan	145
6.2 Saran	145
DAFTAR REFERENSI	147
A KODE PROGRAM	149
B HASIL EKSPERIMEN	185

DAFTAR GAMBAR

1.1	Contoh <i>Tweet</i> pada Platform <i>Twitter</i>	2
1.2	Contoh Ulasan Positif	3
1.3	Contoh Ulasan Negatif	3
2.1	Tahap Klasifikasi <i>Learning</i>	10
2.2	Tahap Klasifikasi <i>Classification</i>	10
2.3	Contoh <i>Confusion Matrix</i> dengan Dimensi 2×2 Beserta Rumus <i>Precision</i> , <i>Recall</i> , dan <i>Accuracy</i>	11
2.4	Contoh Teknik <i>Cross-Validation</i> untuk <i>4-Fold Cross-Validation</i>	12
2.5	Alur Proses dalam <i>Data Mining</i>	14
2.6	Anatomi <i>Decision Tree</i>	16
2.7	Contoh <i>Decision Tree</i> Sederhana dengan Kasus Pembelian Komputer	17
2.8	Contoh <i>Tree</i> Sebelum Dilakukan Pemotongan	19
2.9	Contoh <i>Tree</i> Setelah Dilakukan Pemotongan	20
2.10	Contoh Aplikasi SimSimi sebagai Implementasi dari NLP	22
3.1	Diagram Alir Penelitian yang Berhasil Dibuat	30
3.2	Contoh Tampilan Halaman Developer Platform <i>Twitter</i>	32
3.3	Diagram Alir Proses Normalisasi Data Analisis Data Sebelum Optimasi	39
3.4	Diagram Alir Proses Normalisasi Data untuk Proses Optimasi Data	39
3.5	Hasil <i>Word Cloud</i> Sentimen Positif	49
3.6	Hasil <i>Word Cloud</i> Sentimen Negatif	49
3.7	Hasil <i>Word Cloud</i> Sentimen Netral	49
3.8	Hasil <i>Word Cloud</i> untuk Seluruh Data	49
3.9	Hasil <i>Bar Plot</i> Sentimen Positif	50
3.10	Hasil <i>Bar Plot</i> Sentimen Negatif	50
3.11	Hasil <i>Barplot</i> Sentimen Netral	50
3.12	Hasil <i>Bar Plot</i> untuk Seluruh Data	50
3.13	Hasil Visualisasi <i>Pie Chart</i> Sebaran Sentimen	51
3.14	<i>Tree</i> yang Berhasil Dibuat	54
3.15	Visualisasi <i>Bar Plot</i> Evaluasi Model Klasifikasi	55
3.16	Visualisasi <i>Line Plot</i> Evaluasi Model Klasifikasi	56
3.17	Aplikasi PAGE yang Berhasil Dijalankan	57
3.18	Aplikasi PAGE Bagian Aplikasi Utama	58
3.19	Aplikasi PAGE Bagian <i>Widget Toolbar</i>	58
3.20	Aplikasi PAGE Bagian <i>Widget Tree</i>	59
3.21	Aplikasi PAGE Bagian <i>Preview Window</i>	60
3.22	Aplikasi yang Berhasil Dibuat	61
3.23	Aplikasi Sederhana yang Berhasil Dibuat	61
4.1	Struktur dalam <i>Tweet</i>	64
4.2	Hasil Visualisasi <i>Bar Plot</i> Tanpa Penghilangan <i>Stopword</i>	68
4.3	Hasil Visualisasi <i>Bar Plot</i> dengan Penghilangan <i>Stopword</i>	68

4.4	Hasil Visualisasi <i>Pie Chart</i> Sebaran Sentimen	74
4.5	Hasil <i>Bar Plot</i> Sebaran Kata untuk Sentimen Positif	75
4.6	Hasil <i>Bar Plot</i> Sebaran Kata untuk Sentimen Positif yang Telah Dipotong	75
4.7	Hasil <i>Bar Plot</i> Kata Bersentimen Positif pada Sentimen Positif	76
4.8	Hasil <i>Bar Plot</i> untuk Kategori Sentimen Negatif	76
4.9	Hasil <i>Bar Plot</i> Kata Bersentimen pada Sentimen Negatif	77
4.10	Hasil <i>Bar Plot</i> Sentimen Netral	77
4.11	Hasil <i>Bar Plot</i> Kata Bersentimen pada Sentimen Netral	78
4.12	Hasil <i>Bar Plot</i> Jumlah Kata Bersentimen pada Seluruh Kategori Sentimen	79
4.13	Hasil <i>Word Cloud</i> dengan Kata Bersentimen untuk Sentimen Positif	79
4.14	Hasil <i>Word Cloud</i> dengan Kata Bersentimen untuk Sentimen Negatif	79
4.15	Hasil <i>Word Cloud</i> dengan Kata Bersentimen untuk Sentimen Netral	80
4.16	Hasil <i>Decision Tree</i> yang Berhasil Dibentuk	82
4.17	Visualisasi <i>Line Plot</i> Evaluasi Model Klasifikasi dengan Teknik <i>Cross-Validation</i>	84
4.18	Visualisasi <i>Bar Plot</i> Evaluasi Model Klasifikasi dengan Teknik Perhitungan Metrik	85
4.19	Hasil <i>Bar Plot</i> Kata yang Memiliki Panjang 2 atau Kurang	91
4.20	Hasil <i>Bar Plot</i> Kata Slang atau Tidak Baku	92
4.21	Hasil <i>Pie Chart</i> Sebaran Sentimen untuk Seluruh Bulan Setelah Dilakukan Optimasi	94
4.22	Hasil <i>Pie Chart</i> Sebaran Sentimen untuk Seluruh Bulan Sebelum Dilakukan Optimasi	94
4.23	Hasil Sebaran Kata yang Berhasil Divisualisasikan dengan <i>Bar Plot</i> dengan Dataset yang Sudah Dioptimasi	95
4.24	Hasil Sebaran Kata yang Berhasil Divisualisasikan dengan <i>Bar Plot</i> dengan Dataset yang Belum Dioptimasi	95
4.25	Hasil Sebaran Kata Bersentimen yang Berhasil Divisualisasikan dengan <i>Bar Plot</i> dengan Dataset yang Sudah Dioptimasi	96
4.26	Hasil Sebaran Kata Bersentimen Positif yang Berhasil Divisualisasikan dengan <i>Bar Plot</i> dengan Dataset yang Sudah Dioptimasi	97
4.27	Hasil Sebaran Kata Bersentimen Negatif yang Berhasil Divisualisasikan dengan <i>Bar Plot</i> dengan Dataset yang Sudah Dioptimasi	97
4.28	Hasil Sebaran Kata Bersentimen Netral yang Berhasil Divisualisasikan dengan <i>Bar Plot</i> dengan Dataset yang Sudah Dioptimasi	98
4.29	Hasil <i>Pie Chart</i> Sebaran Sentimen untuk Bulan Juni	98
4.30	Hasil Sebaran Kata Bersentimen yang Berhasil Divisualisasikan dengan <i>Bar Plot</i> untuk Bulan Juni	99
4.31	Hasil Sebaran Kata Bersentimen yang Berhasil Divisualisasikan dengan <i>Bar Plot</i> untuk Bulan Juni Kategori Sentimen Positif	100
4.32	Hasil Sebaran Kata Bersentimen yang Berhasil Divisualisasikan dengan <i>Bar Plot</i> untuk Bulan Juni Kategori Sentimen Negatif	100
4.33	Hasil <i>Pie Chart</i> Sebaran Sentimen untuk Bulan Juli	101
4.34	Hasil Sebaran Kata Bersentimen yang Berhasil Divisualisasikan dengan <i>Bar Plot</i> untuk Bulan Juli	101
4.35	Hasil Sebaran Kata Bersentimen yang Berhasil Divisualisasikan dengan <i>Bar Plot</i> untuk Bulan Juli Kategori Sentimen Positif	102
4.36	Hasil Sebaran Kata Bersentimen yang Berhasil Divisualisasikan dengan <i>Bar Plot</i> untuk Bulan Juli Kategori Sentimen Negatif	103
4.37	Hasil <i>Pie Chart</i> Sebaran Sentimen untuk Bulan Agustus	103
4.38	Hasil Sebaran Kata Bersentimen yang Berhasil Divisualisasikan dengan <i>Bar Plot</i> untuk Bulan Agustus	104
4.39	Hasil Sebaran Kata Bersentimen yang Berhasil Divisualisasikan dengan <i>Bar Plot</i> untuk Bulan Agustus Kategori Sentimen Positif	105

4.40	Hasil Sebaran Kata Bersentimen yang Berhasil Divisualisasikan dengan <i>Bar Plot</i> untuk Bulan Agustus Kategori Sentimen Negatif	105
4.41	Hasil <i>Pie Chart</i> Sebaran Sentimen untuk Bulan September	106
4.42	Hasil Sebaran Kata Bersentimen yang Berhasil Divisualisasikan dengan <i>Bar Plot</i> untuk Bulan September	106
4.43	Hasil Sebaran Kata Bersentimen yang Berhasil Divisualisasikan dengan <i>Bar Plot</i> untuk Bulan September Kategori Sentimen Positif	107
4.44	Hasil Sebaran Kata Bersentimen yang Berhasil Divisualisasikan dengan <i>Bar Plot</i> untuk Bulan September Kategori Sentimen Negatif	108
4.45	Hasil <i>Pie Chart</i> Sebaran Sentimen untuk Bulan Oktober	108
4.46	Hasil Sebaran Kata Bersentimen yang Berhasil Divisualisasikan dengan <i>Bar Plot</i> untuk Bulan Oktober	109
4.47	Hasil Sebaran Kata Bersentimen yang Berhasil Divisualisasikan dengan <i>Bar Plot</i> untuk Bulan Oktober Kategori Sentimen Positif	110
4.48	Hasil Sebaran Kata Bersentimen yang Berhasil Divisualisasikan dengan <i>Bar Plot</i> untuk Bulan Oktober Kategori Sentimen Positif Untuk Data Selain Kata ‘Utama’	110
4.49	Hasil Sebaran Kata Bersentimen yang Berhasil Divisualisasikan dengan <i>Bar Plot</i> untuk Bulan Oktober Kategori Sentimen Negatif	111
4.50	Hasil Sebaran Kata Bersentimen yang Berhasil Divisualisasikan dengan <i>Bar Plot</i> untuk Bulan Oktober Kategori Sentimen Negatif Untuk Data Selain Kata ‘batas’	111
4.51	Hasil <i>Line Chart</i> untuk Seluruh Sentimen	112
4.52	Visualisasi <i>Bar Plot</i> untuk Jumlah Fitur Sebelum Optimasi dan Sesudah Optimasi	113
4.53	Hasil <i>Decision Tree</i> yang Berhasil Dibentuk	114
4.54	Visualisasi <i>Line Plot</i> Evaluasi Model Dengan Teknik <i>Cross-Validation</i>	115
4.55	Visualisasi <i>Bar Plot</i> Evaluasi Model Klasifikasi	116
4.56	Visualisasi <i>Line Plot</i> Evaluasi Model Klasifikasi dengan Teknik <i>Cross-Validation</i>	117
4.57	Visualisasi <i>Bar Plot</i> Evaluasi Model Klasifikasi dengan Teknik Perhitungan Metrik	118
5.1	Hasil Perancangan <i>Mockup</i> Aplikasi Halaman Utama yang Berhasil Dibuat	124
5.2	Alternatif Aplikasi Halaman Utama jika Dipilih <i>Radio Button CSV</i>	125
5.3	Rancangan Aplikasi Ketika Menghadapi Kasus Eksepsi	126
5.4	Rancangan Aplikasi Ketika Menghadapi Kasus Eksepsi untuk CSV	127
5.5	Rancangan Aplikasi Ketika Aplikasi Berhasil Melakukan Prediksi Sentimen untuk Data Satu Kalimat	128
5.6	Rancangan Aplikasi Ketika Aplikasi Mengeluarkan Hasil Prediksi Sentimen untuk Data CSV	128
5.7	Rancangan Aplikasi Ketika Aplikasi Berhasil Melakukan Prediksi Sentimen untuk Data Satu Kalimat	129
5.8	Rancangan Aplikasi Ketika Aplikasi Mengeluarkan Hasil Prediksi Sentimen untuk Data CSV	130
5.9	Rancangan Aplikasi Halaman Tentang <i>Dataset</i>	131
5.10	Rancangan Aplikasi Halaman Cara Menggunakan Aplikasi	132
5.11	Hasil Perangkat Lunak yang Sudah Dibuat	133
5.12	Perangkat Lunak dengan Memilih Satu Kalimat Sebagai Pilihan	134
5.13	Perangkat Lunak dengan Memilih Satu Kalimat Sebagai Pilihan	134
5.14	Perangkat Lunak dengan Masukkan Kalimat yang Valid	135
5.15	Perangkat Lunak Berhasil Menampilkan Prediksi dari Masukkan Pengguna	136
5.16	Penanganan Eksepsi pada Data Kosong	137
5.17	Penanganan Eksepsi pada Data yang Tidak Valid	137
5.18	Perangkat Lunak dengan Memilih CSV sebagai Pilihan	138
5.19	Aplikasi akan Menampilkan <i>Window</i> Baru untuk Membuka suatu <i>File</i>	139

5.20 Aplikasi akan Menampilkan Notifikasi apabila Pengguna Sudah Berhasil Melakukan Unggah CSV	139
5.21 Aplikasi Berhasil Melakukan Prediksi Sentimen	140
5.22 Aplikasi dalam Penanganan Kasus Eksepsi Data Belum Diunggah	141
5.23 Aplikasi dalam Penanganan Kasus Eksepsi Data Tidak Sesuai Format yang Ditentukan	141
5.24 Hasil Perangkat Lunak pada Tentang <i>Dataset</i>	142
5.25 Hasil Perangkat Lunak pada Cara Menggunakan Aplikasi	143

DAFTAR TABEL

2.1	Contoh Data untuk Prediksi Pembelian Komputer	17
2.2	Hasil Data Prediksi Pembelian Komputer	18
2.3	Tabel Sebelum dan Sesudah dalam Proses <i>Case Folding</i>	23
2.4	Tabel Sebelum dan Sesudah dalam Proses <i>Tokenization</i>	23
2.5	Tabel Aturan <i>Stemming</i>	24
2.6	Tabel Sebelum dan Sesudah dalam Proses <i>Stemming</i>	24
2.7	Tabel Kombinasi Awalan dan Akhiran yang Tidak Diizinkan	25
3.1	Contoh Data <i>Twitter</i> yang Berhasil Ditarik	35
3.2	Contoh Data <i>Twitter</i> Pertama	35
3.3	Contoh Data <i>Twitter</i> Kedua	36
3.4	Contoh Hasil Penggabungan <i>Dataset</i> dari 2 <i>Dataset</i>	36
3.5	Contoh Data <i>Twitter</i> yang Telah Dilakukan Proses <i>Data Cleaning</i>	38
3.6	Contoh Data <i>Twitter</i> yang Telah Dilakukan Proses Tokenisasi	38
3.7	Contoh Data <i>Twitter</i> yang Telah Dilakukan Proses <i>Case Folding</i>	40
3.8	Contoh Daftar Kata Slang yang Digunakan	40
3.9	Contoh Data <i>Twitter</i> yang Telah Dilakukan Proses Perubahan Kata Slang	41
3.10	Contoh Data <i>Twitter</i> yang Telah Dilakukan Penghilangan <i>Stopwords</i>	42
3.11	Contoh Data <i>Twitter</i> yang Telah Dilakukan Penghilangan Kata dengan panjang 2 Huruf atau Kurang	43
3.12	Contoh Data <i>Twitter</i> yang Telah Dilakukan Proses <i>Stemming</i>	44
3.13	Contoh Perhitungan Ekstraksi Fitur untuk Pemberian Label	46
3.14	Contoh <i>Tweet</i> yang Telah Diberi Skor Sentimen	47
3.15	Kategori Sentimen	47
3.16	Contoh <i>Tweet</i> yang Telah Diberi Kategori Sentimen	47
3.17	<i>Document-Term Matrix</i> yang Berhasil Dibuat	52
3.18	<i>Dataset</i> yang Digunakan untuk Membangun Model <i>Decision Tree</i>	53
3.19	Hasil Evaluasi Klasifikasi <i>Decision Tree</i>	55
3.20	Hasil Evaluasi Klasifikasi <i>Decision Tree</i> dengan <i>Cross-Validation</i>	56
4.1	Contoh <i>Dataset Twitter</i> yang Berhasil Ditarik	63
4.2	Contoh Data <i>Twitter</i> yang Telah Dilakukan Proses <i>Data Cleaning</i>	66
4.3	Contoh Data <i>Twitter</i> yang Telah Dilakukan Proses Normalisasi	67
4.4	Contoh Potongan Kata Sentimen yang Digunakan untuk Proses Pelabelan	69
4.5	Contoh Data <i>Twitter</i> yang Telah Dilakukan Proses Ekstraksi Data untuk Pelabelan Data	70
4.6	Kategori Sentimen	70
4.7	Contoh Data <i>Twitter</i> yang Telah Dilakukan Proses Pelabelan	70
4.8	Contoh <i>Dataset</i> Murni Netral atau Tidak Memiliki Sentimen	71
4.9	Contoh <i>Dataset</i> Sentimen Netral dengan Pelabelan yang Tidak Tepat	71
4.10	Contoh <i>Dataset</i> Sentimen Netral yang Mengandung Kata Bersentimen Positif dan Negatif	72

4.11	Contoh <i>Dataset</i> Sentimen Netral yang Mengandung Kata Sentimen Positif atau Negatif Berbahasa Asing	73
4.12	Perubahan Jumlah Data Sentimen Sebelum dan Sesudah Dilakukan Penelusuran Kembali Sentimen Netral	73
4.13	Contoh Hasil Ekstraksi Fitur yang Sudah Dilakukan	81
4.14	Contoh Pelabelan Data untuk Tiap Baris Data	81
4.15	Rincian <i>Dataset</i> untuk Proses Klasifikasi	82
4.16	Hasil Perhitungan dengan <i>Cross-Validation</i>	83
4.17	Hasil Perhitungan dengan Pengukuran Metrik	84
4.18	Rincian <i>Dataset</i> yang Digunakan untuk Melakukan Prediksi Sentimen	86
4.19	Hasil Prediksi yang Berhasil Dilakukan	87
4.20	Contoh <i>Dataset</i> yang Berhasil Dipisah untuk Bulan Juni	88
4.21	Contoh <i>Dataset</i> yang Berhasil Dipisah untuk Bulan Oktober	89
4.22	Contoh <i>Dataset</i> yang Berhasil Dibersihkan	89
4.23	Contoh <i>Dataset</i> yang Berhasil Dinormalisasi	90
4.24	Contoh <i>Dataset Twitter</i> yang Telah Dilakukan Proses Normalisasi pada Tahap Analisis Sentimen Utama	90
4.25	Contoh <i>Dictionary</i> untuk Merubah Kata dengan 2 Huruf atau Kurang Menjadi Kata Rujukan Asli	92
4.26	Contoh <i>Dictionary</i> untuk Merubah Kata Slang atau Tidak Baku Menjadi Kata Baku	93
4.27	Perbandingan Proses Normalisasi Sebelum Optimasi dengan Proses Normalisasi Setelah Optimasi	93
4.28	Contoh Data <i>Twitter</i> yang Telah Dilakukan Proses Pelabelan	94
4.29	Contoh Hasil Ekstraksi Fitur yang Sudah Dilakukan Menggunakan Teknik Normalisasi yang Dioptimasi	113
4.30	Perbandingan Banyaknya Data Ekstraksi Fitur Sebelum Optimasi dan Sesudah Optimasi	113
4.31	Rincian <i>Dataset</i> untuk Proses Klasifikasi	114
4.32	Hasil Evaluasi Model <i>Cross-Validation</i> Sebelum Dilakukan Optimasi dan Setelah Dilakukan Optimasi	115
4.33	Hasil Perhitungan dengan Pengukuran Metrik	116
4.34	Hasil <i>Grid Search</i> yang Berhasil Dilakukan	117
4.35	Hasil Evaluasi Model Dengan Parameter <code>max_depth=135</code> Dengan Teknik <i>Cross-Validation</i>	118
4.36	Hasil Perhitungan dengan Pengukuran Metrik dengan Menggunakan <code>Max_Depth</code>	118
4.37	Rincian <i>Dataset</i> yang Digunakan untuk Memprediksi Sentimen	119
4.38	Hasil Prediksi yang Berhasil Dilakukan	120

DAFTAR KODE PROGRAM

3.1	Kode Otentikasi <i>Key</i> dan <i>Token</i>	33
3.2	Kode untuk Mendapatkan <i>Tweet</i> dari <i>Twitter</i> dan Menyimpan <i>Dataframe</i> menjadi CSV	34
3.3	Kode untuk Menggabungkan Seluruh <i>Dataset</i> menjadi Sebuah <i>Dataset</i>	36
3.4	Kode untuk Menggabungkan Seluruh <i>Dataset</i> menjadi Sebuah <i>Dataset</i>	37
3.5	Kode untuk Melakukan Tokenisasi	38
3.6	Kode untuk Melakukan Perubahan Kata Slang	41
3.7	Kode untuk Melakukan Penghilangan <i>Stop Words</i>	42
3.8	Kode untuk Melakukan Penghilangan Kata dengan 2 Huruf Kurang dari 3 Huruf	43
3.9	Kode untuk Melakukan <i>Stemming</i>	44
3.10	Kode untuk Melakukan Perhitungan Skor Sentimen	46
3.11	Kode untuk Memisahkan <i>tweet</i> berdasarkan sentimen <i>tweet</i>	48
3.12	Kode untuk Membuat Visualisasi <i>Word Cloud</i>	48
3.13	Kode untuk Membuat visualisasi <i>bar plot</i>	49
3.14	Kode untuk Membuat Visualisasi <i>Pie Chart</i>	50
3.15	Kode untuk Melakukan Ekstraksi Fitur	51
3.16	Kode untuk Melakukan Klasifikasi dengan <i>Decision Tree</i> dan Konstruksi <i>Decision Tree</i>	53
3.17	Kode untuk Melakukan Evaluasi Klasifikasi	55
3.18	Kode untuk Melakukan Penyimpanan Model	56
A.1	testProgram_support.py	149
A.2	testProgram.py	149
A.3	twitter_retrieval.py	150
A.4	preprocessing.py	151
A.5	merge.py	156
A.6	wordCount.py	157
A.7	tfidf.py	168
A.8	predict.py	173
A.9	application.py	175
A.10	application_support.py	181
B.1	result_allData_17K.txt	185
B.2	result_DFAugust.txt	189
B.3	result_DFJuly.txt	189
B.4	result_DFJune.txt	189
B.5	result_DFOctober.txt	190
B.6	result_DFSeptember.txt	190

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Pandemi *Covid-19* adalah suatu wabah penyakit yang disebabkan oleh virus *Covid-19* yang sedang terjadi saat ini meliputi daerah geografis yang luas dan terjadi serentak. *Covid-19* tersendiri adalah suatu jenis virus menular pada manusia yang berasal dari *SARS-CoV-2*. Virus ini bersifat menular dan virus ini menular melalui kontak langsung, udara, atau benda-benda sekitar. Pandemi *Covid-19* yang saat ini sedang terjadi, belakangan ini, sudah mengalami penurunan kasus yang cukup signifikan. Berdasarkan data dari pemerintah melalui *website* resmi pemerintah mengenai *Covid-19*¹, kasus harian *Covid-19* hanya terkonfirmasi sebanyak 997 kasus per 16 Oktober 2021. Penurunan kasus inilah yang menjadi latar belakang pemerintah untuk membuka berbagai tempat. Salah satunya adalah membuka kembali sekolah atau perguruan tinggi agar dapat melakukan pembelajaran tatap muka (PTM). Istilah ini juga biasa disebut sebagai sekolah tatap muka.

Sekolah tatap muka adalah sekolah yang dilakukan dengan kontak langsung dengan pelajar maupun pengajar. Sekolah tatap muka merupakan sekolah yang lazim ditemui dalam dunia pendidikan. Akan tetapi karena pandemi *Covid-19* yang sedang terjadi, sekolah tatap muka ditiadakan dan beralih ke sekolah *online* atau dikenal sebagai sekolah tatap maya. Sekolah *online* pada dasarnya sama dengan tatap muka, namun yang membedakannya mereka tidak perlu melakukan kontak langsung dan datang ke sekolah secara langsung. Sekolah *online* membutuhkan komputer atau perangkat yang mampu menunjang kebutuhan sekolah *online* dan koneksi internet.²

Program pembukaan sekolah tatap muka ternyata menuai kontroversi. Ada yang mengatakan bahwa tindakan pembukaan sekolah tatap muka dapat memicu kluster positif *Covid-19*. Namun, ada juga yang berpendapat bahwa PTM menjadi hal yang harus dijalankan demi memaksimalkan performa pengajar maupun pelajar dalam proses kegiatan ajar-mengajar. Argumen tersebut banyak bersebaran di internet. Media berita mengemukakan pendapatnya mengenai kebijakan pemerintah ini. Hal ini juga terdapat pada media sosial yang sering kita pakai di kehidupan sehari-hari.

Media sosial merupakan media yang digunakan oleh banyak pengguna yang bersifat *online* atau terhubung dengan internet di mana pengguna bisa mengekspresikan opini, mengeluh kesah, melakukan interaksi antar pengguna, dan membagikan informasi dalam *platform* sosial³. Media sosial sudah menjadi kebutuhan oleh banyak pengguna di seluruh dunia karena media sosial merangkum apa saja yang sedang terjadi di dunia ini dan juga menjadi media berkomunikasi antar pengguna tanpa harus melakukan kontak langsung. Pengguna media sosial memiliki rentang umur yang bervariasi, mulai dari umur yang masih sangat muda, remaja, dewasa, maupun lanjut usia. Platform sosial dalam hal ini meliputi *Instagram*, *Twitter*, *Facebook*, *Whatsapp*, maupun *Line*.

Twitter adalah sebuah platform media sosial yang menawarkan jasa berupa *mikroblog* dan fitur berbagi pesan. Hal ini memungkinkan pengguna untuk melihat pesan pengguna lain. Pesan atau *mikroblog* yang dibuat oleh pengguna dikenal juga sebagai *tweet*. *Tweet* adalah sebuah teks yang ada dalam *platform Twitter* yang memiliki panjang maksimum 280 karakter. *Tweet* dapat

¹<https://covid19.go.id/peta-sebaran>

²<https://gig.id/stories/lifestyle/pembelajaran-daring-dan-luring>

³<https://www.sekawanmedia.co.id/media-sosial-untuk-bisnis/>

melakukan *mentioning*—perujukan ke pengguna *Twitter* lainnya, *hashtag*—perujukan pada suatu topik tertentu, mengirim berupa gambar, bahkan melakukan *retweet*—melakukan *tweet* dengan menggunakan *tweet* dari pengguna lain. Untuk melihat contoh *tweet* dalam platform *Twitter*, dapat dilihat pada Gambar 1.1.



Gambar 1.1: Contoh *Tweet* pada Platform *Twitter*

Berdasarkan Gambar 1.1, *tweet* ditunjukkan pada tulisan *happy holidays*. Selain itu juga, gambar tersebut menampilkan nama pengguna yang terletak pada bagian atas dari *tweet*. Pengguna juga dapat mengirimkan *tweet* dengan menggunakan gambar yang tertera juga pada gambar tersebut. Sebuah *tweet* juga terdapat waktu publikasi *tweet* tersebut termasuk tanggal dan platform apa yang dipakai untuk melakukan publikasi *tweet* tersebut.

Perkembangan teknologi media sosial yang sangat pesat juga melahirkan banyak sekali opini yang diekspresikan pengguna dalam platform media sosial terhadap suatu hal. Opini, secara umum, adalah suatu ekspresi atau tanggapan seseorang dalam menilai suatu hal atau fenomena yang sedang terjadi. Misalnya, pada kasus pengguna melakukan suatu ulasan suatu produk. Pengguna secara bebas dapat mengekspresikan tentang kualitas produk, kegunaan produk, dan rekomendasi produk tersebut. Opini yang dihasilkan dari suatu ulasan produk pun beragam. Mulai dari setuju bahwa produk tersebut bagus, netral, hingga mengatakan produk sangat buruk.

Pengguna dapat membedakan opini yang bersifat negatif, positif, maupun netral secara mudah karena manusia sudah terbiasa dengan kata-kata yang mengacu apakah kalimat tersebut negatif atau positif. Opini negatif adalah sebuah ekspresi yang dikemukakan oleh pengguna yang berisikan kekecewaan, keburukan, maupun penolakan dari suatu jasa atau produk. Opini positif adalah sebuah ekspresi yang dikemukakan oleh pengguna yang berisikan kegembiraan, kesukaan, dan kepuasan dari suatu jasa atau produk. Suatu opini yang bersifat positif atau negatif akan memiliki kata yang mengandung sentimen. Opini netral adalah opini yang tidak mengandung unsur ekspresi atau sentimen apapun dalam sebuah kalimat. Sentimen yang positif dalam bahasa Indonesia dapat dicontohkan seperti kata 'bagus', 'cepat', 'cantik', dan 'hebat'. Kata sentimen yang negatif dalam bahasa Indonesia seperti kata 'jelek', 'lama', 'buruk', dan 'payah'. Untuk contoh opini positif

ditunjukkan pada Gambar 1.2 dan opini negatif ditunjukkan pada Gambar 1.3⁴:



Great product, excellent service.
Awesome



Barang sudah diterima.. Tp ketika saya scan QR code undiannya. ehh ada peringatan dari ***** bahwa barang saya palsu.. dan disuruh menghubungi penjual untuk refund. Karena katanya kualitas gak sama dengan yang asli performanya padahal saya kira asli.. hehee ya udah lah biar aja gak saya kembalikan. tp juga gak saya pakai.. Males repot repot menukar. Mau beli baru aja di toko offline aja.

Gambar 1.3: Contoh Ulasan Negatif

Gambar 1.2: Contoh Ulasan Positif

Berdasarkan Gambar 1.2, pengguna memberikan ulasan yang positif yaitu dengan mengatakan produk yang dibeli sebagai produk yang sangat baik dengan pelayanan yang sangat baik. Hal ini ditandai dengan kemunculan kata “great”, “excellent”, dan “awesome”, di mana kedua kata tersebut memiliki arti sentimen yang positif. Pada Gambar 1.3, pengguna memberikan ulasan yang negatif yaitu dengan mengatakan produk tidak bagus karena merupakan barang palsu dengan beberapa bukti yang menjelaskan alasan benda tersebut tidak asli. Kemudian, pengguna juga menginginkan produk yang dibeli secara *offline* atau datang ke toko langsung. Pada gambar tersebut memiliki sentimen yang negatif yang ditandai oleh kata “palsu”. Kedua gambar tersebut mengulas produk yang sama di *website* yang sama.

Perlu diperhatikan bahwa gambar yang dicontohkan sesuai dengan Gambar 1.2 dan Gambar 1.3, hanya menampilkan 2 buah contoh *review* saja. Permasalahan akan muncul jika ada banyak sekali *review* yang dihasilkan dalam suatu platform. Sebagai contoh, sebuah produk memiliki jumlah *review* sebanyak 100000 buah *review*. Jika manusia membaca ulasan satu per satu, waktu yang digunakan akan banyak digunakan. Hal ini tentu tidak efektif untuk menyelesaikan masalah ini. Oleh karena itu, perlu ada bantuan yang efektif dan efisien dengan menggunakan komputer untuk menyelesaikan permasalahan dengan menggunakan analisis sentimen.

Analisis sentimen adalah salah satu bagian dari teknik *text mining* dan *data mining* yang merupakan ukuran pengukuran emosi dari berbagai opini yang dikemukakan oleh pengguna dalam suatu topik permasalahan [1]. *Text mining* adalah suatu cabang dari *data mining* yang memproses data berupa dokumen teks yang bersifat tidak terstruktur maupun semi terstruktur [7]. Analisis sentimen akan mengukur kecenderungan sentimen suatu masukan yang positif, negatif, atau netral. Analisis sentimen juga merupakan salah satu teknik klasifikasi. Proses analisis sentimen dibantu dengan teknik pemrosesan bahasa alami atau dikenal sebagai *Natural Language Processing (NLP)*.

NLP merupakan cabang bidang ilmu yang mempelajari komunikasi antara mesin komputer dengan manusia dengan menggunakan bahasa alami. Hal ini bertujuan untuk komputer memahami arti dari ucapan maupun tulisan manusia. Teknik NLP memiliki beberapa macam jenis seperti *lexicon* atau kamus, *case folding* atau normalisasi kapitalisasi huruf, tokenisasi atau pemecahan kalimat menjadi kata, dan *stemming* atau proses penyederhanaan imbuhan. Teknik NLP penting dilakukan dalam penelitian ini agar komputer dapat membedakan apakah *tweet* yang digunakan termasuk sentimen positif, negatif, ataupun netral.

Teknik *Decision tree* adalah suatu teknik klasifikasi yang dapat menentukan label data pada suatu *test set* yang memiliki struktur seperti pohon berisikan *flowchart* dalam yang bertujuan untuk menentukan label kelas data yang akan diklasifikasikan [3]. Hasil dari klasifikasi dalam penelitian ini akan dipakai dalam analisis sentimen sebagai label apakah sebuah opini merupakan opini yang positif, negatif, atau netral. Selain itu juga, dapat juga diketahui tingkat akurasi dari klasifikasi tersebut untuk menunjukkan kualitas hasil klasifikasi yang sudah dilakukan.

Hal yang dibuat dalam penelitian ini adalah suatu model yang dapat melakukan prediksi sentimen dengan aplikasi berbasis *Graphical User Interface (GUI)* untuk pengguna dapat memasukkan data

⁴<https://bit.ly/3iLVuU3>

yang ingin diprediksi dan sebuah analisis data yang berkaitan tentang sentimen sekolah tatap muka untuk didapatkan *insight-insight* yang menarik. Untuk analisis data, keluaran yang didapatkan adalah sebaran sentimen positif, negatif, maupun netral beserta temuan-temuan lainnya. Untuk langkah-langkah yang dilakukan adalah pengumpulan data *Twitter* yang berkaitan dengan sekolah tatap muka di Indonesia sebagai sumber data yang akan dianalisis. Data tersebut kemudian akan masuk ke dalam tahap pemrosesan bahasa alami untuk menyiapkan data-data tersebut. Setelah menyiapkan data, dilakukan pemberian skor sentimen untuk melakukan pelabelan data yang digunakan, kemudian dilakukan ekstraksi fitur untuk mengubah fitur berupa teks menjadi fitur yang dapat diolah oleh algoritma klasifikasi, dan diklasifikasikan dengan menggunakan teknik klasifikasi *decision tree* untuk mengetahui kelas-kelas dari teks tersebut. Teknik *Decision Tree* digunakan karena hasil keluaran dari teknik tersebut mudah dibaca oleh manusia. Library yang akan digunakan dalam penelitian ini akan digunakan antara lain `tweepy`, `matplotlib.pyplot`, `nlTK`, `pySastrawi`, dan `sklearn`. Hasil klasifikasi *Decision Tree* kemudian akan dijadikan suatu model untuk memprediksi data *Twitter* yang lain.

1.2 Rumusan Masalah

Berdasarkan uraian yang telah dipaparkan dalam bagian Deskripsi, masalah-masalah dapat dirumuskan sebagai berikut:

1. Bagaimana cara mengumpulkan data *Twitter* berbahasa Indonesia tentang sekolah tatap muka di Indonesia?
2. Bagaimana cara menyiapkan pesan *Twitter* agar dapat digunakan analisis lebih lanjut?
3. Bagaimana cara untuk melakukan analisis sentimen sekolah tatap muka di Indonesia?
4. Bagaimana membangun perangkat lunak yang dapat menampilkan hasil sentimen sekolah tatap muka di Indonesia?

1.3 Tujuan

Berdasarkan rumusan masalah yang sudah dirumuskan, tujuan dari penelitian ini adalah sebagai berikut:

1. Mempelajari dan mengimplementasikan teknik pengumpulan data *Twitter* berbahasa Indonesia mengenai sekolah tatap muka di Indonesia dengan menggunakan API *Twitter*, `tweepy`, dengan menggunakan bahasa *Python*.
2. Mempelajari teknik-teknik dalam *Natural Language Processing* (NLP) yang digunakan sebagai proses *pre-processing* dan termasuk juga mempelajari morfologi dalam bahasa Indonesia. Pada proses ini juga dilanjutkan oleh proses implementasi teknik-teknik NLP yang sesuai dengan tujuan dan hasil penelitian ini.
3. Mempelajari algoritma teknik klasifikasi *decision tree*, melakukan analisis data, dan membangun model dengan klasifikasi *decision tree* yang kemudian akan digunakan sebagai prediksi model.
4. Mempelajari, merancang, dan mengimplementasikan teknik-teknik dalam pembangunan perangkat lunak dengan menggunakan *Python*. Hal ini termasuk melakukan implementasi teknik *pre-processing data*, teknik ekstraksi fitur, teknik NLP, teknik klasifikasi *decision tree*, dan perancangan GUI.

1.4 Batasan Masalah

Pelaksanaan penelitian ini memiliki batasan masalah berupa:

1. *Dataset* yang digunakan memiliki rentang bulan Juni 2021 hingga Oktober 2021. Hal ini dibatasi karena dalam rentang waktu tersebut topik sekolah tatap muka sedang mengalami

trend yang tinggi pada rentang waktu tersebut. Di luar rentang tersebut sudah tidak banyak topik sekolah tatap muka yang sedang dibicarakan. Selain itu, jumlah data yang terkumpul pada rentang waktu tersebut sudah cukup banyak untuk menunjang penelitian ini.

2. Bahasa pada data *tweet* adalah Bahasa Indonesia.

1.5 Metodologi

Berikut metodologi penelitian yang dilakukan dalam skripsi ini:

1. Melakukan studi literatur materi yang akan dilakukan pada penelitian ini. Berikut merupakan materi yang akan dipelajari pada penelitian ini sebagai acuan dasar penelitian:
 - (a) Studi literatur tentang *Twitter* dan *API Twitter*
 - (b) Studi literatur tentang sentimen analisis
 - (c) Studi literatur tentang *machine learning*
 - (d) Studi literatur tentang konsep *data mining* secara sederhana dan *text mining* secara sederhana
 - (e) Studi literatur tentang konsep-konsep pada NLP atau pemrosesan bahasa alami seperti *lexicon*, *case folding*, *token*, *stemming*, dan *filtering*.
 - (f) Studi literatur tentang teknik klasifikasi dan teknik klasifikasi *Decision Tree*
 - (g) Studi literatur tentang teknik evaluasi klasifikasi seperti mempelajari tentang *accuracy*, *precision*, *recall*, *f-score*, dan teknik *cross-validation*
 - (h) Studi literatur tentang morfologi kata dalam Bahasa Indonesia
 - (i) Studi literatur mengenai konsep dasar dari data visualisasi dan jenis data visualisasi yang digunakan
 - (j) Studi literatur mengenai teknik ekstraksi fitur yang akan digunakan seperti konsep *term frequency* dan *inverse document frequency*
 - (k) Studi literatur mengenai teknik optimasi *Grid Search*
2. Melakukan pengumpulan data yang relevan dari platform *Twitter*.
3. Melakukan eksplorasi data yang telah dikumpulkan untuk melihat karakteristik data yang akan diproses.
4. Melakukan eksplorasi teknologi yang akan digunakan seperti mempelajari *library* pada *Python* seperti **pandas**, **sciPy**, **tweepy**, teknik-teknik NLP sederhana dengan menggunakan *Python*, dan teknik klasifikasi *decision tree* pada kasus sederhana. Pada proses ini juga akan dipilih teknologi yang cocok digunakan dalam penelitian ini.
5. Melakukan persiapan data seperti melakukan *data cleaning*, *tokenization*, *feature extracting*, dan *feature engineering*.
6. Melakukan perancangan model dan penentuan cara mengevaluasi model.
7. Melakukan implementasi perangkat lunak.
8. Melakukan eksperimen pengujian dalam suatu model.
9. Melakukan peluncuran model pada aplikasi/prototipe sederhana.
10. Menulis dokumen skripsi.

1.6 Sistematika Penulisan

Berikut merupakan sistematika penulisan yang terdapat dalam dokumen skripsi ini:

1. **Bab 1 Pendahuluan**

Bab ini berisikan latar belakang penelitian, tujuan penelitian, rumusan masalah, batasan-batasan masalah, metodologi penelitian, dan sistematika penulisan dokumen skripsi.

2. **Bab 2 Landasan Teori**

Bab ini membahas mengenai teori-teori yang akan digunakan dalam penelitian ini. Teori yang akan digunakan dalam penelitian ini adalah *Twitter*, *Data Mining*, Teknik Klasifikasi *Decision Tree*, *Text Mining*, Analisis Sentimen, NLP, dan Morfologi dalam Bahasa Indonesia.

3. **Bab 3 Analisis Masalah dan Eksplorasi Tools**

Bab ini membahas mengenai permasalahan yang akan diselesaikan. Selain itu, pada bab ini akan melakukan eksplorasi *tools* yang akan digunakan. Eksplorasi *tools* juga terdiri dari penggambaran proses yang akan dilakukan untuk bagian pembangunan model yang akan dilakukan pada Bab 4.

4. **Bab 4 Penyiapan *Dataset* dan Pembangunan Model**

Bab ini membahas mengenai profil dari *dataset* yang akan digunakan. Selain itu juga dibahas mengenai proses penyiapan data. Setelah proses penyiapan data, akan dilakukan juga pembangunan model yang *dataset*-nya berasal dari *dataset* yang sudah disiapkan.

5. **Bab 5 Implementasi Perangkat Lunak**

Bab ini membahas mengenai implementasi perangkat lunak untuk peluncuran prototipe sederhana dalam bentuk GUI. Selain itu juga, terdapat bagian pengujian model sederhana untuk mengecek fungsionalitas pada prototipe sederhana.

6. **Bab 6 Kesimpulan dan Saran**

Bab ini membahas mengenai kesimpulan yang didapatkan dalam penelitian ini. Selain itu juga, terdapat bagian saran untuk melanjutkan penelitian ini.