

**LAPORAN TUGAS AKHIR**  
**ANALISIS SENTIMEN TWITTER MENGENAI**  
**PEMBELAJARAN ONLINE MENGGUNAKAN METODE KNN**

Diajukan untuk memenuhi salah satu syarat memperoleh gelar Sarjana Teknik  
Informatika



Disusun oleh :

Nama : Satria Pambayun  
Nim : A11.2018.11110  
Program Studi : Teknik Informatika

**FAKULTAS ILMU KOMPUTER**  
**UNIVERSITAS DIAN NUSWANTORO**  
**SEMARANG**

**2021**

## **PERSETUJUAN SKRIPSI**

Nama : Satria Pambayun.I.M  
NIM : A11.2018.11110  
Program Studi : Teknik Informatika  
Fakultas : Ilmu Komputer  
Judul Tugas Akhir : Analisis Sentimen Twiter Pembelajaran Online  
Menggunakan Metode KNN

Tugas Akhir ini telah diperiksa dan disetujui,

Semarang, 25 September 2021

Menyetujui :  
Pembimbing

Menyetujui :  
Dekan Fakultas Ilmu Komputer



**Dr. Catur Supriyanto, S. Kom, M. Cs    DR. Guruh Fajar Shidik, S. Kom, M.Cs**

## **PENGESAHAN DEWAN PENGUJI**

Nama : Satria Pambayun.I.M  
NIM : A11.2018.11110  
Program Studi : Teknik Informatika  
Fakultas : Ilmu Komputer  
Judul Tugas Akhir : Analisis Sentimen Twiter Pembelajaran Online  
Menggunakan Metode KNN

Tugas akhir ini telah diujikan dan dipertahankan dihadapan Dewan Penguji pada Sidang tugas akhir tanggal 19 September 2021. Menurut pandangan kami, tugas akhir ini memadai dari segi kualitas maupun kuantitas untuk tujuan penganugrahan gelar Sarjana Komputer (S.Kom.)

Semarang, 19 September 2021

Dewan Penguji:

**Dosen 1 :**

Erwin Yudi Hidayat. S.Kom. M.Cs

**Dosen 2 :**

Abas Setiawan. S.Kom.M.Cs

**Dosen Ketua :**

De Rosal Ignatius Moses Setiadi. M.Kom

## **PERNYATAAN KEASLIAN SRIPSI**

Sebagai mahasiswa Universitas Dian Nuswantoro, yang bertanda tangan di bawah ini, saya:

Nama : Satria Pambayun Ibrahim Maryana

NIM : A11.2018.11110

Menyatakan bahwa karya ilmiah saya yang berjudul:

“Analisis Sentimen Twitter Mengenai Pembelajaran Online Menggunakan Metode KNN”

Merupakan karya asli saya (kecuali cuplikan dan ringkasan yang masing-masing telah saya jelaskan sumbernya dan perangkat pendukung seperti web cam dll). Apabila di kemudian hari, karya saya disinyalir bukan merupakan karya asli saya, yang disertai dengan bukti-bukti yang cukup, maka saya bersedia untuk dibatalkan gelar saya beserta hak dan kewajiban yang melekat pada gelar tersebut. Demikian surat pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Semarang

Pada tanggal : 25 September 2021

Yang Menyatakan

SATRIA PAMBAYUN IBRAHIM MARYANA

## UCAPAN TERIMAKASIH

Dengan memanjatkan puji syukur kehadiran Allah SWT. Tuhan Yang Maha Pengasih dan Maha Penyayang yang telah melimpahkan segala rahmat, hidayah dan inayah-Nya kepada penulis sehingga laporan tugas akhir dengan judul “Analisis Sentimen Tweet Berbahasa Inggris dengan Machine Learning” dapat penulis selesaikan sesuai dengan rencana karena dukungan dari berbagai pihak yang tidak ternilai besarnya. Oleh karena itu penulis menyampaikan terimakasih kepada:

1. Prof. Dr. Ir. Edi Noersasongko, M.Kom., selaku Rektor Universitas Dian Nuswantoro.
2. DR. Guruh Fajar Shidik, S.Kom., M.Cs, selaku Dekan Fakultas Ilmu Komputer.
3. Dr. Muljono, S.Si, M.Kom, selaku Ketua Program Studi Teknik Informatika – S1.
4. Dr. Catur Supriyanto, S. Kom, M. Cs selaku dosen pembimbing tugas akhir yang memberikan ide penelitian, memberikan informasi referensi yang penulis butuhkan dan bimbingan yang berkaitan dengan penelitian penulis
5. Dosen-dosen pengampu di Fakultas Ilmu Komputer Teknik Informatika Universitas Dian Nuswantoro Semarang yang telah memberikan ilmu dan pengalamannya masing-masing, sehingga penulis dapat mengimplementasikan ilmu yang telah disampaikan.

Semoga Tuhan yang Maha Esa memberikan balasan yang lebih besar kepada beliau-beliau, dan pada akhirnya penulis berharap bahwa penulisan laporan tugas akhir ini dapat bermanfaat dan berguna sebagaimana fungsinya.

Semarang, 25 September 2021

SATRIA PAMBAYUN IBRAHIM MARYANA

## ABSTRAK

Wabah virus corona yang biasa disebut dengan COVID-19 ditetapkan secara resmi sebagai pandemi global oleh World Health Organization (WHO). Strategi yang diterapkan dalam upaya mengurangi penyebaran COVID-19 adalah dengan diberlakukannya pembatasan sosial yang memaksa semua sektor yang memungkinkan, melakukan kegiatan meraka secara daring. Salah satu sector yang diharuskan melakukan kegiatan secara daring adalah sektor pendidikan, dimana semua kegiatan belajar mengajar dari tingkat paling dasar sampai perguruan tinggi dilakukan secara *online*/daring. Kebijakan mengenai pembelajaran *online* ini menuai berbagai macam tanggapan dari masyarakat di media sosial Twitter, ada yang memberi tanggapan positif adapula yang memberi tanggapan negative. Tujuan dari tugas akhir ini adalah untuk mengetahui sentiment masyarakat terhadap pembelajaran online. Data yang digunakan pada tugas akhir ini sebanyak 500 *tweet*, yang dicari menggunakan kata kunci “pembelajaran *online*”, “kelas *online*”, “kuliah *online*”, “pembelajaran daring”, “sekolah daring”. Klasifikasi dilakukan menggunakan metode *K-Nearest Neighbor* (KNN) yang dibuat menggunakan bahasa pemrograman python yang memiliki beberapa tahapan yaitu pengambilan data yang dilakukan dengan menggunakan *library tweepy* yang telah disediakan oleh python, selanjutnya melakukan *data preprocessing* yang terdiri dari proses *cleansing* yang dilakukan untuk membersihkan data dari simbol-simbol seperti *user* (@), *hashtag* (#) dan lainnya, selanjutnya proses *case folding* yang dilakukan untuk mengubah semua huruf kapital yang ada pada data agar menjadi huruf kecil, selanjutnya adalah proses *Stopword Removal* yang dilakukan untuk mengilangkan semua kata yang tidak memiliki arti, kemudian yang terakhir proses *Stemming* yang dilakukan untuk menjadikan setiap kata yang terdapat pada data menjadi kata dasar. Kemudian setelah melalui data preprocessing, akan dilakukan pembobotan pada setiap kata yang ada pada data menggunakan metode *Term Frequency – Inverse Document Frequency* (TF-IDF) untuk mengubah dokumen menjadi vektor dengan elemen sebanyak term yg telah dapat dikenali pada tahapan ekstraksi dokumen. Kemudian yang terakhir akan dilakukan klasifikasi menggunakan metode *K-Nearest Neighbor* (KNN). Hasil yang didapatkan dengan menggunakan nilai  $k=1$  memperoleh nilai *accuracy*=0.675 , *f1 score* = 0.683 , *precision* = 0.667 , dan nilai *recall*=0.7 , menggunakan nilai  $k=3$  memperoleh nilai *accuracy* = 0.675 , *f1 score* = 0.683 , *precision* = 0.667 , dan nilai *recall* = 0.7 , menggunakan nilai  $k=5$  memperoleh nilai *accuracy* = 0.67 , *f1 score* = 0.667 , *precision* = 0.75 , dan nilai *recall* = 0.6 , menggunakan nilai  $k=7$  memperoleh nilai *accuracy* = 0.8 , *f1 score* = 0.778 , *precision* = 0.875 , dan nilai *recall* = 0.7 , dan dengan menggunakan nilai  $k=9$  memperoleh nilai *accuracy* = 0.75 , *f1 score*=0.722 , *precision* = 0.813 , dan nilai *recall* = 0.65.

Kata Kunci : COVID-19, Pembelajaran Online, Sentimen Masyarakat, KNN

## DAFTAR ISI

PERSETUJUAN SKRIPSI .....	i
PENGESAHAN DEWAN PENGUJI.....	ii
PERNYATAAN KEASLIAN SRIPSI.....	iii
UCAPAN TERIMAKASIH.....	iv
ABSTRAK.....	v
DAFTAR ISI.....	vi
DAFTAR GAMBAR .....	viii
DAFTAR GAMBAR .....	ix
BAB I PENDAHULUAN .....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	5
1.3 Batasan Masalah.....	5
1.4 Tujuan Penelitian.....	6
1.5 Manfaat Penelitian.....	6
BAB II LANDASAN TEORI .....	7
2.1 Tinjauan Studi .....	7
2.2 Data Mining .....	13
2.3 Text Mining.....	14
2.4 Analisis Sentimen .....	15
2.5 Twitter API .....	16
2.6 Pembobotan ( <i>Term Weightening</i> ) .....	16
2.6.1 Term Frequency – Inverse Document Frequency (TF-IDF).....	16
2.7 K-Nearest Neighbor (K-NN).....	17
2.8 Python .....	18

2.8.1 Library Re .....	19
2.8.2 Library Pandas .....	19
2.8.3 Library Matplotlib .....	19
2.8.4 Library Seaborn.....	20
2.8.5 Library Sastrawi .....	20
2.8.6 Library Scikit-learn .....	20
2.8.7 Library Matplotlib .....	20
2.9 Pembelajaran Online .....	21
BAB III METODE PENELITIAN.....	22
3.1 Pengumpulan Data .....	22
3.1.1 Twitter Developer .....	23
3.1.2 Proses Pengambilan Data <i>Twitter</i> .....	24
3.2 Analisis Data .....	26
3.3 Metode yang Diusulkan .....	26
3.4 Pengujian Metode.....	27
BAB IV HASIL PENELITIAN DAN PEMBAHASAN .....	28
4.1 Hasil Penelitian .....	28
4.1.1 Pengujian Program .....	28
4.1.2 Hasil Klasifikasi .....	35
4.2 Analisis .....	38
BAB V KESIMPULAN .....	40
5.1 Kesimpulan.....	40
5.2 Penelitian Selanjutnya .....	41
Daftar Pustaka .....	<b>Error! Bookmark not defined.</b>



## DAFTAR TABEL

Table 1 State of the Art Penelitian Terkait.....	9
Table 2 Hasil Klasifikasi dengan $K=1$ .....	36
Table 3 Hasil Klasifikasi dengan $K=3$ .....	36
Table 4 Hasil Klasifikasi dengan $K=5$ .....	37
Table 5 Hasil Klasifikasi dengan $K=7$ .....	37
Table 6 Hasil Klasifikasi dengan $K=9$ .....	38

## DAFTAR GAMBAR

Gambar 1 Grafik Tingkat Penggunaan Bahasa Pemrograman .....	3
Gambar 2 Proses Knowledge Discovery Database (KDD).....	14
Gambar 3 Tahapan data preprocessing (Wong & Mooney, 2018) .....	14
Gambar 4 Model Klasifikasi K-Nearest Neighbor (K-NN).....	18
Gambar 5 Tahapan Pengumpulan Data.....	23
Gambar 6 Tampilan Website Twitter Developer .....	23
Gambar 7 Tampilan Form Tujuan Penggunaan Application Programming Interface (API) Twitter .....	24
Gambar 8 Tampilan Pembuatan Project Baru.....	24
Gambar 9 API Key & Token Key .....	25
Gambar 10 Penggunaan API Key & Token Key .....	25
Gambar 11 Konversi Dataset Kedalam Format csv .....	25
Gambar 12 Tahapan Analisis Data .....	26
Gambar 13 Memasukkan Dataset .....	29
Gambar 14 Inisialisasi Label.....	29
Gambar 15 Jumlah Data Positif dan Negatif.....	30
Gambar 16 Proses Cleansing dan Case Folding .....	31
Gambar 17 Proses Stopword Removal .....	31
Gambar 18 Proses Stemming .....	32
Gambar 19 Membagi Data Menjadi Data Train dan Data Test .....	32
Gambar 20 Proses Pembobotan Dengan Metode Term Frequency – Inverse Document Frequency (TF-IDF). .....	33
Gambar 21 Proses Klasifikasi dengan K-Nearest Neighbor (K-NN). .....	33

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

*Coronavirus Disease 2019* atau yang sering disebut dengan COVID-19 merupakan virus yang menyerang sistem pernafasan manusia dan berpotensi menyebabkan kematian, yang melanda hamper di seluruh belahan dunia (Mahase, 2020). COVID-19 ini sendiri ditemukan pertama kali sekitar akhir desember 2019 di Wuhan, China. dan ditetapkan oleh *World Health Organization* (WHO) sebagai pandemi global pada tanggal 11 Maret 2020. Karena penularan COVID-19 yang mudah dan cepat, tidak butuh waktu lama untuk virus ini menyebar ke berbagai negara di seluruh dunia. Banyak negara telah melakukan isolasi untuk menghambat penyebaran COVID-19, isolasi sendiri merupakan pemisahan orang yang sedang terjangkit penyakit menular dari orang yang tidak terinfeksi untuk melindungi orang yang tidak terinfeksi (Smith & Fredman, 2020). Selain isolasi yang dilakukan untuk menghambat penyebaran COVID-19, berbagai negara juga menerapkan karantina untuk membatasi pergerakan orang yang diduga telah terkena penyakit menular namun tidak sakit (Smith & Fredman, 2020). Karantina ini sendiri juga dianggap sebagai salah satu cara yang cukup efektif untuk mencegah dan menghambat laju penyebaran penyakit menular di zaman modern ini (Goje, 2017).

COVID-19 ini mengakibatkan terganggunya kegiatan dari berbagai bidang seperti bidang ekonomi, sosal, dan pendidikan. Bidang pendidikan merupakan salah satu bidang yang terkena dampak cukup besar sejak hadirnya pandemi COVID-19 ini. *United Nations of Educational, Scientific, and Cultural Organization* (UNESCO) mengatakan bahwa ada sekitar 300 juta murid yang teraggu kegiatan belajar mengajarnya serta ada penutupan sekolah sementara dikarenakan pandemi COVID-19 (Satrianingrum & Prasetyo, 2021). Karena dilakukan penutupan sekolah untuk mengurangi dan menghambat laju penyebaran COVID-19, Kementria Pendidikan dan Kebudayaan memberlakukan system pendidikan jarak jauh untuk kegiatan belajar mengajar dari tingkat paling dasar hingga perguruan tinggi, yang dilakukan secara *online/daring*.

Pembelajaran *online*/daring telah menjadi tuntutan di dalam dunia pendidikan semenjak beberapa tahun terakhir (He, Xu, & Kruck, 2016). pembelajaran dilakukan karena dampak dari pandemi COVID-19 ini sendiri merupakan hal baru yang belum pernah dilakukan kebanyakan pelajar di Indonesia. Pembelajaran *online*/daring yang telah diberlakukan oleh Kemetrian Pendidikan dan Kebudayaan ini sendiri mengubah proses belajar mengajar yang biasa dilakukan dimana pelajar dituntut untuk duduk di kelas dan mendengarkan materi yang disampaikan oleh guru. Pembelajaran *online*/daring ini sendiri juga memiliki keuntungan tertentu, yaitu guru dan pelajar tidak perlu bertemu di sekolah untuk melakukan kegiatan belajar mengajar, serta guru dan pelajar tidak diharuskan untuk *online* diwaktu yang sama, sehingga dapat mengurangi waktu untuk kegiatan belajar mengajar sekitar sepertiga dari total waktu yang dibutuhkan (Oktavianto & Persada, 2020). Meski demikian, keuntungan tersebut tetap tidak ada artinya jika pelajar tidak memiliki minat belajar dan persepsi yang baik mengenai pembelajaran *online*. Dengan diberlakukannya pembelajaran *online* yang merubah proses belajar mengajar yang biasa dilakukan, tentunya banyak masyarakat yang mengungkapkan pendapat atau opini mereka mengenai pembelajaran *online* yang dilakukan sekarang ini terutama di media sosial.

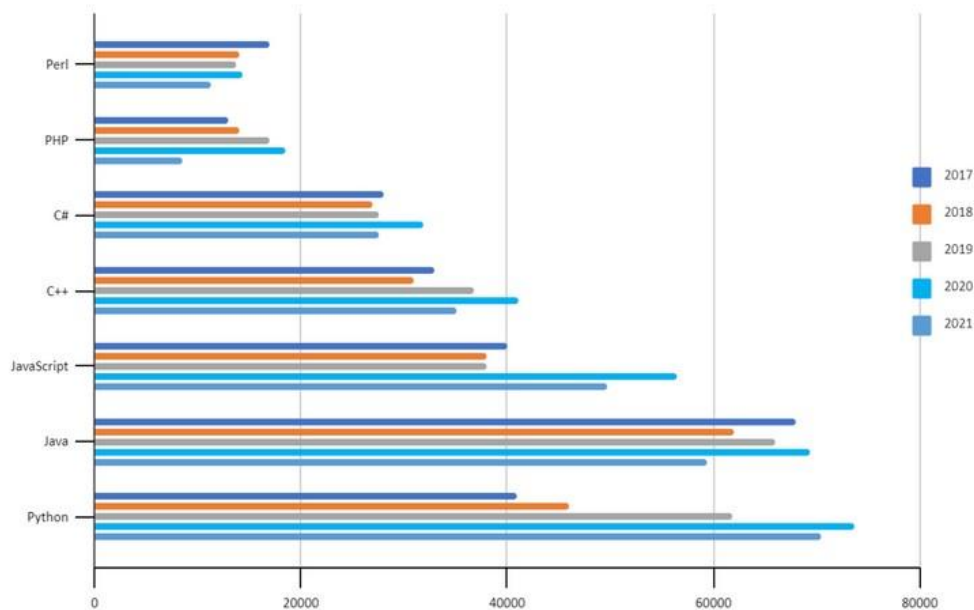
Media sosial merupakan salah satu media yang dapat digunakan untuk memperoleh suatu informasi atau bisa juga digunakan untuk mencurahkan opini keseharian yang sedang kita rasakan atau alami. Ada beberapa media sosial yang populer di masyarakat diantaranya adalah Facebook, Instagram, serta Twitter. Twitter sendiri merupakan salah satu media sosial yang memiliki lebih dari 19,5 juta pengguna di Indonesia serta 383 tweet per hari, dari 500 juta pengguna global. Indonesia sendiri juga diklaim sebagai salah satu negara dengan pertumbuhan pengguna aktif harian paling besar di twitter berdasarkan *financial report* dari twitter kuratal ke-3 tahun 2019 (Nurdiansyah, Bukhori, & Hidayat, 2018). Tweet atau postingan Twitter dapat dipergunakan sebagai sarana mengungkapkan perasaan pengguna, memberitahukan apa yang sedang dilakukan oleh pengguna, serta berbagai informasi lainnya. Dengan memanfaatkan data yang ada pada isi dari postingan masyarakat di Twitter, dapat dilakukan analisis sentiment masyarakat

mengenai pembelajaran *online*/daring yang mungkin selanjutnya dapat digunakan sebagai sarana evaluasi bagi pihak-pihak terkait.

*Sentiment Analysis* atau Analisis Sentimen adalah sebuah penilaian masyarakat yang menyangkut rasa suka atau tidak suka (Romadloni, Santoso, & Budilaksono, 2019). Analisis sentiment ini sendiri digunakan untuk menilai suatu komentar terhadap suatu permasalahan dan menentukan apakah komentar tersebut memiliki kecenderungan positif atau negative. Disisi lain, analisis sentiment juga dapat dimanfaatkan untuk mengevaluasi kinerja suatu pelayanan dan yang lainnya.

Secara sederhana penelitian ini memiliki dua tahap yaitu yang pertama pengambilan data, dan yang kedua adalah analisis data. Pada tahap pengambilan data, penulis menggunakan bahasa pemrograman yang populer serta banyak digunakan oleh masyarakat dunia menurut *Stack Overflow Trends* yaitu Python.

### The Top Programming Languages of Previous Years Compared to 2021



Gambar 1 Grafik Tingkat Penggunaan Bahasa Pemrograman

Dalam proses pengambilan data yang akan diambil dari Twitter, penulis memanfaatkan library yang disediakan oleh Python seperti *tweepy* dan *Textblob*, kemudian diolah pada tahap analisis data mulai dari proses *Data Preprocessing* yang terdiri dari proses menghilangkan *Twitter Handles* (@user), kemudian menghilangkan url dari teks, menghilangkan tanda baca serta karakter khusus, kemudian mengkonversi kata menjadi huruf kecil, *Stopword Removal*, *stemming* kata, dan *Tokenizing*. Kemudian setelah data dibersihkan akan dilakukan proses *weighting* atau pembobotan kata menggunakan metode *Term Frequency – Inverse Document Frequency* atau TF-IDF. Lalu selanjutnya akan dilakukan perhitungan akurasi.

Algoritma yang akan digunakan untuk penghitungan akurasi pada penelitian Tugas Akhir ini adalah metode *K-Nearest Neighbor* (K-NN). Algoritma *K-Nearest Neighbor* (K-NN) ini merupakan metode *lazy learning*, yang berarti pada metode ini tidak ada model yang dipelajari dari data pelatihan, jadi hanya mempelajari dari contoh uji yang harus diklasifikasikan. Tujuan dari algoritma *K-Nearest Neighbor* (K-NN) ini sendiri adalah untuk mengklasifikasikan objek berdasarkan atribut serta data training (Sari, 2020).

Alasan penulis menggunakan algoritma *K-Nearest Neighbor* (K-NN) ini sendiri adalah karena algoritma *K-Nearest Neighbor* (K-NN) merupakan algoritma klasifikasi yang mudah diimplementasikan serta di pahami, *dataset* yang digunakan pada algoritma *K-Nearest Neighbor* (K-NN) juga memiliki label sehingga dapat memudahkan dalam melakukan proses pengelompokan data ke dalam kelas yang paling sesuai yang telah disediakan. Algoritma *K-Nearest Neighbor* (K-NN) juga memiliki kelebihan dalam aspek kemudahan menerjemahkan hasil, waktu perhitungan, serta akurasi dari prediksi yang lebih baik dibandingkan dengan metode Logistic Regression, CART, dan Random Forest (Primartha, 2018).

Pada penelitian sebelumnya yang dilakukan oleh Fadholi Fat. H dan Bety Wulan. S yang melakukan analisis sentiment terhadap tanggapan masyarakat mengenai pelayanan Telkom dan Biznet, melakukan analisis sentimen dengan menggunakan data yang diambil dari Twitter sebanyak 500 data (Haranto & Sari, 2019). Pada penelitian yang dilakukan oleh A.M Pudjajana dan D. Manongga yang

melakukan analisis sentiment terhadap pronografi kaum homoseksual Indonesia juga menggunakan dataset yang diambil dari Twitter dengan menggunakan sebanyak 500 data tweet (Pudjajana & Manongga, 2018).

Berdasarkan latar belakang di atas, penulis akan melakukan penelitian yang berjudul “ANALISIS SENTIMEN TWITTER MENGENAI PEMBELAJARAN ONLINE MENGGUNAKAN METODE KNN” yang akan dilakukan dengan menggunakan data tweet yang diambil dari Twitter pada tanggal 18 November 2021, dengan menggunakan kata kunci yang berkaitan dengan pembelajaran online yang telah ditentukan oleh penulis yaitu “pembelajaran online”, “kelas online”, “kuliah online”, “pembelajaran daring”, serta “sekolah daring”. Penulis memilih kata kunci tersebut agar dapat memperoleh sentimen yang di keluarkan oleh siswa, mahasiswa, wali siswa, wali mahasiswa, serta orang yang tidak sedang menjalani pendidikan formal.

## **1.2 Rumusan Masalah**

Dari latar belakang masalah yang sudah ada, maka dapat dirumuskan masalah sebagai berikut :

1. Bagaimana data tanggapan masyarakat mengenai pembelajaran *online/daring*?
2. Bagaimana hasil *Sentiment Analysis* / Analisis Sentimen masyarakat mengenai pembelajaran *online/daring* dan implementasi metode *K-Nearest Neighbor* (K-NN) untuk melakukan kasifikasi *Sentiment Analysis* Twitter?
3. Bagaimana tingkat akurasi metode *K-Nearest Neihbor* (K-NN) dalam pengklasifikasian tanggapan masyarakat mengenai pemberlakuan pembelajaran online?

## **1.3 Batasan Masalah**

Agar penelitian pada tugas akhir ini tidak melenceng dari pokok permasalahan maka ruang lingkup dari pembahasan tugas akhir ini dibatasi dengan batasan masalah sebagai berikut:

1. Objek dalam penelitian tugas akhir ini adalah pendapat masyarakat di sosial media Twitter mengenai pemberlakuan pembelajaran *online/daring*.

2. Sumber pendapat / opini masyarakat didapatkan dari sosial media Twitter menggunakan library Python tweepy dengan kata kunci “pembelajaran *online*”, “kelas *online*”, “kuliah *online*”, “pembelajaran daring”, “sekolah daring”.
3. Analisis sentiment hanya mencakup positif dan negatif saja.
4. Analisis sentiment yang dilakukan pada penelitian tugas akhir ini menggunakan metode *K-Nearest Neighbor* (K-NN) menggunakan bahasa pemrograman *Python*.
5. *Tools* yang digunakan pada penelitian ini adalah *Jupyter Notebook* serta *Google Colabs*.

#### 1.4 Tujuan Penelitian

Tujuan dari penelitian Tugas Akhir ini adalah sebagai berikut:

1. Untuk mengetahui gambaran terhadap pendapat / opini masyarakat mengenai pemberlakuan pembelajaran *online*/daring.
2. Untuk mengetahui hasil klasifikasi *Sentiment Analysis* / Analisis Sentimen mengenai pemberlakuan pembelajaran *online* yang di terapkan dimasa pandemi COVID-19 dengan menggunakan metode *K-Nearest Neighbor* (K-NN).
3. Untuk mengetahui akurasi yang didapatkan dari sentiment analysis / analisis sentimen dengan menggunakan metode *K-Nearest Neighbor* (K-NN).

#### 1.5 Manfaat Penelitian

Manfaat dilakukannya penelitian Tugas Akhir ini adalah:

1. Untuk memberikan pengetahuan terkait klasifikasi *sentiment analysis* / analisis sentimen terhadap pemberlakuan pembelajaran *online* menggunakan metode *K-Nearest Neighbor* (K-NN).
2. Melakukan pengklasifikasian opini ke dalam kelas positif dan negative dalam jumlah besar dengan waktu singkat.
3. Hasil dari *sentiment analysis* /analisis sentimen menggunakan metode *K-Nearest Neighbor* (K-NN) ini dapat digunakan sebagai bahan evaluasi terhadap kebijakan pemerintah dalam memberlakukan pembelajaran *online*.



## **BAB II**

### **LANDASAN TEORI**

#### **2.1 Tinjauan Studi**

Pada laporan penelitian tugas akhir ini, penulis merujuk ke beberapa penelitian yang sudah pernah dilakukan sebelumnya mengenai analisis sentimen dengan menggunakan algoritma apriori. Adapun beberapa penelitian tersebut meliputi :

- a. Penelitian ini berjudul “Klasifikasi Sentimen Terhadap Tokoh Publik Pada Twitter Menggunakan Algoritma K-Nearest Neighbor” yang dibuat oleh Andres Kusumawardhana (2019). Dalam penelitian ini dijelaskan bahwa Indonesia merupakan negara yang menganut sistem demokrasi yang ditandai dengan adanya pemilihan umum yang dilakukan untuk menentukan presiden dan wakil presiden yang diadakan secara periodik, karena terdapat pasangan presiden dan wakil presiden yang ingin maju dalam pemilihan umum, maka akan menumbulkan opini-opini dari masyarakat di beberapa platform media sosial. Salah satunya adalah Twitter. Oleh karena itu penelitian ini mencoba untuk menganalisis twitter berbahasa Indonesia yang memiliki opini tentang tokoh-tokoh publik di Indonesia. Pada penelitian ini terdiri dari beberapa tahap. Tahap pertama adalah proses *data preprocessing* yang terdiri dari proses *case folding*, tokenisasi, *stemming*, normalisasi kata, dan *stopword removal*. Tahap kedua adalah proses pembobotan kata yang dilakukan dengan menggunakan metode TF-IDF. Selanjutnya tahap terkahir adalah melakukan klasifikasi dengan metode *K-Nearest Neighbor*. Hasil yang diperoleh dari analisis sentimen terhadap tokoh publik menggunakan algoritma *K-Nearest Neighbor* memperoleh akurasi tertinggi ketika menggunakan 20-fold dengan nilai k=1 sebesar 68,3%.
- b. Pada penelitian dengan judul “Penerapan Analisis Sentimen Pada Pengguna Twitter Menggunakan Metode *K-Nearest Neighbor*”. Akhmad Deviyanto (2018), menjelaskan bahwa media sosial kini menjadi tempat untuk mencurahkan opini dari penggunaanya terhadap berbagai hal salah satunya politik. Tema politik yang menjadi *trending* dalam beberapa waktu lalu saat

pembuatan penulisan ini adalah tentang Pemilihan Gubernur (Pilgub) DKI Jakarta tahun 2017. Maka, pada penelitian ini akan dilakukan analisis sentimen dari pengguna sosial media Twitter terhadap para kandidat Gubernur pada pilgub tersebut. Metode yang digunakan pada penelitian ini adalah algoritma *K-Nearest Neighbor* dengan menggunakan metode pembobotan kata TF-IDF dan fungsi *Cosine Similarity* untuk membangun sistem analisis sentiment. Lalu metode Confusion Matrix akan digunakan untuk mengukur kinerja algoritma. Berdasarkan penelitian ini diperoleh hasil akurasi terbesar adalah 67,2% ketika menggunakan  $k=5$ , nilai presisi terbesar adalah 56,94% ketika menggunakan  $k=5$ , dan recall terbesar adalah 78,24% ketika menggunakan  $k=15$ .

- c. Penelitian selanjutnya adalah penelitian yang dilakukan oleh Mahdi Muhammad Rizki (2019) yang berjudul “Analisis Sentimen Terhadap Produk Otomotif Dari Twitter Menggunakan Kombinasi Algoritma *K-Nearest Neighbor* dan pendekatan *Lexicon*”. Penelitian ini menjelaskan bahwa analisis sentiment merupakan cabang penelitian dari *text mining*. Dalam dunia bisnis, *opinion mining* sering digunakan untuk menganalisis secara otomatis opini dari pelanggan tentang produk dan pelayanannya. Penelitian ini bertujuan untuk mengetahui tingkat akurasi dari analisis sentiment masyarakat pada media sosial Twitter mengenai produk otomotif Toyota menggunakan algoritma *K-Nearest Neighbor*. Pengujian akan dilakukan dengan menggunakan model *confussion matrix*. Hasil dari penelitian ini memiliki akurasi terbaik ketika menggunakan nilai  $k=3$  dengan akurasi sebesar 83%.
- d. Penelitian ini memiliki judul “Analisis Sentimen Pengaruh COVID-19 Terhadap Kehidupan Bermasyarakat pada Media Sosial Twitter Menggunakan *K-Nearest Neighbor*” yang dilakukan oleh Pascal Aldwin Hernando (2021). Dalam penelitian ini dijelaskan bahwa dengan munculnya wabah COVID-19, membuat seluruh dunia termasuk Indonesia mengalami kemunduran dalam berbagai bidang kehidupan. Dengan adanya wabah ini, membuat pemerintah menetapkan lockdown pada berbagai daerah yang menyebabkan aktifitas setiap orang menjadi terbatas dari rumah saja. Untuk

mengungkapkan keadaan selama lockdown, maka orang-orang menggunakan media sosial untuk sebagai perantara, dimana salah satunya adalah Twitter. Dengan banyaknya penggunaan twitter, membuat data yang berhubungan tentang COVID-19 untuk dilakukan analisis sentimen menggunakan metode klasifikasi *K-Nearest Neighbor*. Hasil dari penelitian ini menampilkan bahwa pada data Jakarta mendapatkan akurasi terbaik pada 78.95% dengan nilai  $K=9$ , sementara pada data Surabaya mendapatkan akurasi terbaik pada 76.47% dengan nilai  $K=11,17,19,20$ .

- e. Pada penelitian yang berjudul “Klasifikasi Sentimen Menggunakan Data Twitter Mengenai Pandemi Korona di Indonesia” yang dilakukan oleh Dwi Retno Fibriyanti (2021), menjelaskan bahwa hadirnya kasus COVID-19 menimbulkan berbagai reaksi dan pandangan dari publik di media sosial. Penelitian ini akan melakukan analisis sentiment terhadap pandangan masyarakat mengenai pandemi COVID-19. Proses klasifikasi yang dilakukan pada ini menggunakan metode *K-Nearest Neighbor*, dan pengujian menggunakan *confussion matrix*. Hasil pengujian menggunakan Confusion Matrix diperoleh 0,673 pada sebuah Nilai akurasi 0,96. Tanggapan atas komentar dari pengguna Twitter di Indonesia terkait Pandemi Korona lebih positif jika dilihat dari hasil klasifikasi.

**Table 1 State of the Art Penelitian Terkait**

No.	Nama Peneliti dan Tahun	Judul Penelitian	Permasalahan	Hasil
1.	Andres Kusumawardhana (2019).	Klasifikasi Sentimen Terhadap Tokoh Publik Pada Twitter Menggunakan Algoritma K-Nearest Neighbor	Indonesia merupakan negara yang menganut sistem demokrasi yang ditandai dengan adanya pemilihan umum yang dilakukan untuk menentukan presiden dan wakil presiden yang diadakan secara	Hasil yang diperoleh dari analisis sentimen terhadap tokoh publik menggunakan algoritma <i>K-Nearest Neighbor</i> memperoleh akurasi tertinggi ketika menggunakan <i>20-fold</i>

			periodik, karena terdapat pasangan presiden dan wakil presiden yang ingin maju dalam pemilihan umum, maka akan menumbulkan opini-opini dari masyarakat di beberapa platform media sosial. Salah satunya adalah Twitter.	dengan nilai $k=1$ sebesar 68,3%.
2.	Akhmad Deviyanto (2018)	Penerapan Analisis Sentimen Pada Pengguna Twitter Menggunakan Metode <i>K-Nearest Neighbor</i>	Media sosial kini menjadi tempat untuk mencurahkan opini dari penggunaanya terhadap berbagai hal salah satunya politik. Tema politik yang menjadi <i>trending</i> dalam beberapa waktu lalu saat pembuatan penulisan ini adalah tentang Pemilihan Gubernur (Pilgub) DKI Jakarta tahun 2017. Maka, pada penelitian ini akan dilakukan analisis sentimen dari pengguna sosial media Twitter terhadap para kandidat Gubernur pada pilgub tersebut.	Berdasarkan penelitian ini diperoleh hasil akurasi terbesar adalah 67,2% ketika menggunakan $k=5$ , nilai presisi terbesar adalah 56,94% ketika menggunakan $k=5$ , dan recall terbesar adalah 78,24% ketika menggunakan $k=15$

3.	Mahdi Muhammad Rizki (2019)	Analisis Sentimen Terhadap Produk Otomotif Dari Twitter Menggunakan Kombinasi Algoritma <i>K-Nearest Neighbor</i> dan pendekatan <i>Lexicon</i>	analisis sentiment merupakan cabang penelitian dari <i>text mining</i> . Dalam dunia bisnis, <i>opinion mining</i> sering digunakan untuk menganalisis secara otomatis opini dari pelanggan tentang produk dan pelayanannya. Penelitian ini bertujuan untuk mengetahui tingkat akurasi dari analisis sentiment masyarakat pada media sosial Twitter mengenai produk otomotif Toyota menggunakan algoritma <i>K-Nearest Neighbor</i> .	Hasil dari penelitian ini memiliki akurasi terbaik ketika menggunakan nilai $k=3$ dengan akurasi sebesar 83%.
4.	Pascal Aldwin Hernando (2021)	Analisis Sentimen Pengaruh COVID- 19 Terhadap Kehidupan Bermasyarakat pada Media Sosial Twitter Menggunakan <i>K-Nearest Neighbor</i>	dengan munculnya wabah COVID-19, membuat seluruh dunia termasuk Indonesia mengalami kemunduran dalam berbagai bidang kehidupan. Dengan adanya wabah ini, membuat pemerintah menetapkan lockdown pada berbagai daerah yang menyebabkan	Hasil dari penelitian ini menampilkan bahwa pada data Jakarta mendapatkan akurasi terbaik pada 78.95% dengan nilai $K=9$ , sementara pada data Surabaya mendapatkan akurasi terbaik pada 76.47% dengan nilai $K=11,17,19,20$

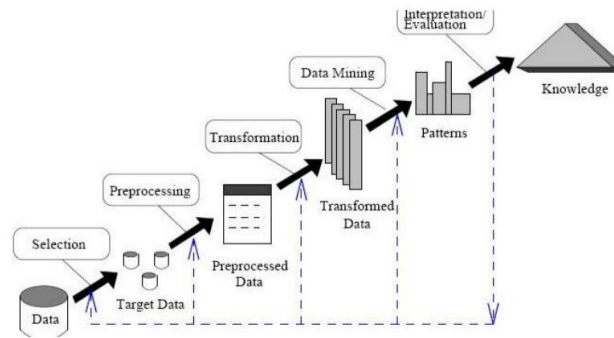
			<p>aktifitas setiap orang menjadi terbatas dari rumah saja. Untuk mengungkapkan keadaan selama lockdown, maka orang-orang menggunakan media sosial untuk sebagai perantara, dimana salah satunya adalah Twitter. Dengan banyaknya penggunaan twitter, membuat data yang berhubungan tentang COVID-19 untuk dilakukan analisis sentimen menggunakan metode klasifikasi <i>K-Nearest Neighbor</i></p>	
5.	Dwi Retno Fibriyanti (2021)	<p>Klasifikasi Sentimen Menggunakan Data Twitter Mengenai Pandemi Korona di Indonesia</p>	<p>hadirnya kasus COVID-19 menimbulkan berbagai reaksi dan pandangan dari publik di media sosial. Penelitian ini akan melakukan analisis sentiment terhadap pandangan masyarakat mengenai pandemi COVID-19.</p>	<p>Hasil pengujian menggunakan Confusion Matrix diperoleh 0,673 pada sebuah Nilai akurasi 0,96. Tanggapan atas komentar dari pengguna Twitter di Indonesia terkait Pandemi Korona lebih positif jika dilihat dari hasil klasifikasi</p>

## 2.2 Data Mining

Data mining merupakan kegiatan untuk menemukan suatu pola yang unik dari sebuah data dengan kapasitas yang besar, serta dapat disimpan di dalam database, *data warehouse*, maupun tempat penyimpanan informasi lainnya (Manning, Raghavan, & Schütze, 2017). Data mining merupakan salah satu bagian dari proses *Knowledge Discovery in Database* (KDD) yang bertujuan untuk mengekstrak suatu pola atau model dari data menggunakan suatu algoritma yang spesifik.

Tahapan proses dari *Knowledge Discovery Database* (KDD) adalah sebagai berikut:

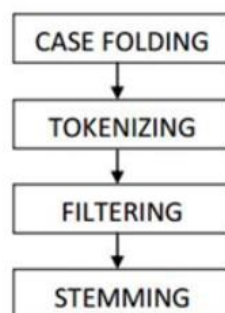
1. *Data selection* : merupakan tahapan pemilihan data dari sekumpulan data operasional
2. *Preprocessing* : di dalam *data mining* diperlukan proses cleaning yang bertujuan untuk membuang duplikasi data.
3. *Transformation* : merupakan tahapan yang dilakukan untuk mentransformasikan bentuk data yang belum memiliki entitas yang jelas menjadi data yang valid.
4. *Data mining* : pada tahapan ini dilakukan penerapan algoritma atau metode pencarian untuk mendapatkan suatu pola atau informasi menarik dalam data terpilih.
5. *Interoretation / Evaluation* : dalam tahapan ini dilakukan pemeriksaan apakah pola atau informasi yang didapatkan bertentangan dengan fakta atau hipotesa yang ada sebelumnya atau tidak .
6. *Knowledge* : merupakan tahapan yang melakukan teknik visualisasi serta representasi hasil dari pengolahan data.



Gambar 2 Proses Knowledge Discovery Database (KDD)

### 2.3 Text Mining

*Text mining* adalah suatu proses penggalian suatu informasi secara intensif yang dikerjakan menggunakan suatu alat dan metode tertentu untuk menganalisis suatu kumpulan dokumen. Tujuan serta proses dari *text mining* ini sendiri kurang lebih sama dengan *data mining*, hanya saja *text mining* memiliki *input* yang berbeda. *Input* dari *text mining* umumnya adalah data yang tidak atau kurang terstruktur, seperti kutipan teks, PDF, dan lain-lain, sedangkan *input* dari *data mining* merupakan data yang terstruktur. Karena *input* dari *text mining* merupakan data yang tidak atau kurang terstruktur, maka sebelum menentukan fitur yang mewakili, diperlukan tahapan *datapre-processing*. Tahapan data *pre-processing* yang umum dilakukan dalam *text processing* adalah *case folding*, *tokenizing*, *filtering*, serta *stemming* (Wong & Mooney, 2018).



Gambar 3 Tahapan data preprocessing (Wong & Mooney, 2018)



Berikut ini merupakan tahapan *data pre-processing*:

1. *Cleansing*: pada tahapan ini dilakukan penghapusan karakter *non-alfabetis* guna mengurangi *noise*. Karakter yang dihapus pada tahapan ini berupa tanda baca, serta simbol-simbol seperti hastag (#), @ pada nama pengguna, url dari suatu situs web, serta *emoticon*.
2. *Case folding*: merupakan proses pengubahan semua karakter yang sudah dibersihkan pada proses *cleansing* sebelumnya menjadi huruf kecil (*lower case*).
3. *Tokenizing*: adalah tahap yang memisahkan kata-kata dari sebuah kalimat.
4. *Normalization*: merupakan tahapan yang bertujuan untuk menormalkan kata-kata yang disingkat maupun diperpanjang menjadi kata-kata yang sesuai dengan yang ada di Kamus Besar Bahasa Indonesia (KBBI).
5. *Filtering*: merupakan proses dimana kata-kata yang sering muncul secara umum dan kurang relevan akan dihapus, serta menjadikan setiap kata menjadi kata dasar.
6. *Stemming*: merupakan proses pencarian *root* atau dasar suatu kata dari setiap kata dari hasil filtering.

## 2.4 Analisis Sentimen

Analisis Sentimen atau *Sentiment Analysis* merupakan proses memahami, mengekstrak serta mengolah data tekstual secara otomatis guna mendapatkan suatu informasi yang dibutuhkan. Analisis sentiment sendiri adalah salah satu bagian *opinion mining*, Analisis Sentiment umumnya dilakukan untuk melihat suatu pendapat masyarakat umum terhadap sebuah masalah, atau juga dapat digunakan untuk mengidentifikasi suatu kecenderungan perihal di pasar. Asalnya analisis sentiment atau *sentiment analysis* merupakan metode klasifikasi, namun pada penerapannya tidak semudah proses metode klasifikasi biasanya dikarenakan berkaitan dengan penggunaan bahasa. Karena dalam penggunaan bahasa terdapat ambiguitas didalam penggunaan suatu kata, tidak ada intonasi didalam suatu teks, serta dari perkembangan bahasa itu sendiri (Liu, 2016).

Analisis Sentimen atau *Sentiment Analysis* bisa juga digunakan untuk menjadi alat bantu dalam melihat respon konsumen ataupun masyarakat terhadap suatu produk. Sehingga dapat ditindak lanjuti dengan mengambil langkah strategis berikutnya. Analisis Sentimen juga dapat dipergunakan menjadi media untuk mengamati respon masyarakat terhadap suatu isu yang beredar, termasuk tanggapan masyarakat terhadap isu politik.

## 2.5 Twitter API

Twitter didirikan oleh Jack Dorsey sekitar bulan Maret pada tahun 2006, sedangkan situsnya sendiri diluncurkan pada bulan Juli 2006. Semenjak diluncurkan, Twitter menjadi satu diantara sepuluh situs yang sering dikunjungi oleh pengguna di internet. *Application Programming Interface* (API) merupakan suatu program atau aplikasi yang disediakan pihak developer tertentu agar pihak pengembang aplikasi lainnya dapat lebih mudah mengakses aplikasi tersebut.

Twitter juga menyediakan *Application Programming Interface* (API) bagi pengembang atau *developer* lain, sehingga *developer* dapat mempermudah pihak developer lain dalam mengakses informasi dari Twitter (Perdana, Pricillia, & Zulfachmi, 2021). Ketika mendapatkan API dari Twitter kita dapat memanggil *endpoint*. *Endpoint* sendiri merupakan alamat yang berkaitan dengan informasi tertentu yang disediakan oleh *developer*. Sedangkan *endpoint* yang dapat diakses dari Twitter adalah akun dan pengguna, tweet serta balasan, *Direct Message* dari pengguna yang telah memberi izin ke Twitter, serta iklan.

## 2.6 Pembobotan (*Term Weightening*)

Tujuan dilakukannya pembobotan atau *Term Weightening* adalah untuk memperoleh nilai dari suatu kata (Fitri, 2018)/*term* yang telah diekstrak. Dalam penelitian Tugas Akhir ini metode pembobotan yang akan digunakan adalah metode TF-IDF (*Term Frequency – Inverse Document Frequency*). Dalam tahapan pembobotan, setiap dokumen nantinya akan diubah menjadi vektor dengan elemen sebanyak term yg telah dapat dikenali pada tahapan ekstraksi dokumen. Anggota dari vektor tersebut adalah bobot dari setiap *term* atau kata yang dihitung berdasarkan metode TF-IDF.

### 2.6.1 Term Frequency – Inverse Document Frequency (TF-IDF)

*Term Frequency – Inverse Document Frequency* atau TF-IDF merupakan sebuah metode algoritma yang berfungsi untuk menghitung bobot dari setiap kata yang umum digunakan dalam suatu dokumen (Tirtana, Zulkarnain, & Listio, 2019). Metode ini akan melakukan perhitungan terhadap nilai *Term Frequency* (TF) serta *Inverse Document Frequency* (IDF) dalam setiap token atau kata pada setiap dokumen di

dalam korpus. Sederhananya metode Tf-IDF ini digunakan untuk melakukan perhitungan guna mengetahui seberapa sering suatu kata muncul didalam suatu dokumen. Rumus dari metode TF-IDF adalah sebagai berikut:

$$w(t, d) = tf(t, d) * idf$$

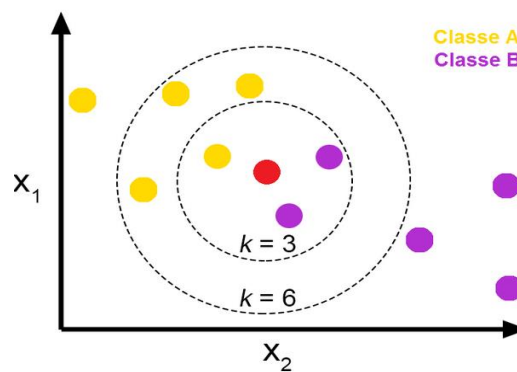
$$idf = \log\left(\frac{N}{df}\right)$$

Dengan  $tf(t,d)$  merupakan frekuensi kemunculan kata  $t$  dalam dokumen  $d$ ,  $N$  merupakan jumlah dari dokumen pada kumpulan dokumen, serta  $df$  merupakan jumlah dokumen yang mengandung *term*/kata  $t$ .

Kegunaan dari metode *Term Frequency – Inverse Document Frequency* (TF-IDF) ini sendiri adalah untuk mencari representasi dari nilai tiap dokumen dari kumpulan data *training* dan setelahnya akan dibuat suatu vektor dari dokumen dengan *term*/kata, lalu kesamaan antar dokumen dengan *cluster* akan di pilah oleh *prototype* vektor yang biasa disebut *cluster centroid* (Widyasanti, Putra, & Rusjyanthi, 2018).

## 2.7 K-Nearest Neighbor (K-NN)

*K -Nearest Neighbor* (K-NN) merupakan salah satu metode yang dapat dilakukan untuk melakukan pengklasifikasian, biasanya algoritma *K-Nearest Neighbor* (K-NN) ini digunakan untuk mengklasifikasi suatu teks atau data (Bonde, 2019). Algoritma *K-Nearest Neighbor* (K-NN) ini sendiri melakukan klasifikasi terhadap suatu objek berdasarkan kedekatan lokasi atau jarak suatu data dengan data lainnya.



Gambar 4 Model Klasifikasi K-Nearest Neighbor (K-NN)

Seperti yang ditunjukkan pada gambar 3, lingkaran putus putus kecil merupakan *K-Nearest Neighbor* (K-NN) dengan nilai K-tetangga 3, sedangkan pada lingkaran putus putus besar merupakan *K-Nearest Neighbor* (K-NN) dengan nilai K-tetangga 6. Dalam metode *K-Nearest Neighbor* (K-NN) ini sendiri menggunakan *Euclidean Distance* untuk metode perhitungan jarak terdekat. Tujuan dari *Euclidean Distance* adalah untuk menentukan kemiripan data yang muncul dalam suatu teks, setelah itu teks akan dibandingkan dengan tiap sampel data yang asli. Rumus dari *Euclidean Distance* adalah sebagai berikut :

$$d(A, B) = \sqrt{(A_1 - B_1)^2 + (A_2 - B_2)^2 + \dots + |A_i - B_i|^2}$$

Dimana:

$d(A, B)$  = Jarak antara dokumen A ke dokumen B

$A_i$  = Kata ke-i dalam dokumen A

$B_i$  = Kata ke-i dalam dokumen B

## 2.8 Python

Python merupakan bahasa pemrograman tingkat tinggi yang memiliki sifat interpretative, dimana dalam segi penulisannya python lebih simple dibandingkan dengan bahasa pemrograman lainnya (Kadarina, Dani, & Supegina, 2021). Python juga menggabungkan berbagai kemampuan, kapabilitas serta sintaks kode dan juga fungsi *library*/pustakanya yang memiliki kualitas yang tinggi. Python sering

digunakan untuk membuat suatu program yang digunakan masyarakat sehari-hari seperti program GUI (*desktop*), aplikasi *hand phone* atau *smartphone*, program CLI, game, *website*, serta masih banyak lagi.

Didalam bahasa pemrograman Python juga terdapat banyak library yang dapat digunakan untuk mempermudah pengguna dalam membuat sebuah algoritma. Dalam penelitian ini penulis juga menggunakan beberapa library yang telah disediakan oleh Python diantaranya.

### 2.8.1 Library Re

Library Re atau Regular expression (regex) adalah deretan karakter yang digunakan untuk pencarian string atau teks dengan menggunakan pola (pattern). Kalau kita perhatikan, email memiliki pola : huruf / angka, diikuti oleh tanda @, kemudian diikuti huruf, tanda titik, dan kemudian huruf lagi. Setelah mengetahui pola dari email, kita bisa mencari potongan teks mana saja yang termasuk email di dalam kumpulan teks yang banyak. Katakanlah kita ingin mencari email di dalam 10 lembar artikel yang tidak mungkin kita telusuri satu persatu secara manual.

### 2.8.2 Library Pandas

Dengan berdasarkan sistem dataframe, Library pandas atau biasa diinisialisasikan dengan pd ini dapat memuat sebuah file ke dalam tabel virtual menyerupai spreadsheet. Pandas juga berfungsi mengolah suatu data seperti teknik join, distinct, group by, agregasi, dan teknik lainnya seperti pada SQL. Bedanya, ini dilakukan pada tabel. Kelebihan dari library ini juga dapat membaca file dari berbagai format seperti .txt, .csv, dan .tsv.

### 2.8.3 Library Matplotlib

Library Matplotlib ini membantu dalam menampilkan hasil analisis berupa grafik berwarna dengan lebih rapi dan menarik. Ada dua jenis plot dalam menampilkan data yaitu secara 2D dan 3D. Sehingga data bisa ditampilkan sesuai dengan kebutuhan. Matplotlib ini merupakan library yang paling sering digunakan oleh data science dalam menyajikan datanya ke dalam visual yang lebih menarik.

#### 2.8.4 Library Seaborn

Seaborn adalah library untuk membuat grafik dan statistik dengan menggunakan Python. Library ini dibangun berdasarkan library matplotlib serta terintegrasi dengan struktur data pada pandas. Library Seaborn bertujuan agar visualisasi menjadi bagian penting dari proses eksplorasi dan pemahaman data. Seaborn berorientasi pada fungsi plotting berdasarkan dataset yang beroperasi pada dataframe dan array yang berisi seluruh dataset yang secara internal melakukan pemetaan semantik dan agregasi statistik yang diperlukan untuk menghasilkan grafik informatif.

#### 2.8.5 Library Sastrawi

Pada *natural language processing* (NLP), informasi yang akan digali berisi data-data tidak terstruktur. Oleh karena itu, diperlukan proses pengubahan bentuk menjadi data yang terstruktur agar proses analisis sentiment dapat dilakukan. Sastrawi adalah pengembangan dari proyek PHP Sastrawi. Python Sastrawi merupakan library sederhana yang dapat mengubah kata berimbuhan bahasa Indonesia menjadi bentuk dasarnya. Sastrawi juga dapat diinstal melalui “pip”. Jadi singkatnya library sastrawi adalah library yang dapat kita gunakan untuk melakukan tahapan *data preprocessing* dengan data yang menggunakan bahasa Indonesia.

#### 2.8.6 Library Scikit-learn

Scikit-learn atau sklearn adalah modul untuk bahasa pemrograman python yang dibangun diatas NumPy, SciPy, dan matplotlib, fungsi dari sklearn ini sendiri adalah untuk membantu melakukan processing data ataupun melakukan training data untuk kebutuhan machine-learning.

#### 2.8.7 Library Matplotlib

Matplotlib disusun oleh John Hunter di tahun 2002, dan didesain agar dapat digunakan selayaknya menggunakan MATLAB. Matplotlib dapat digunakan untuk memvisualisasikan data secara 2D maupun 3D dan menghasilkan gambar berkualitas yang bahkan dapat kamu simpan dalam berbagai format gambar, seperti JPEG dan PNG.

## 2.9 Pembelajaran Online

Perubahan pola pembelajaran saat ini terlihat masif dilakukan mulai dari tingkatan sekolah dasar, menengah bahkan sampai perguruan tinggi. Organisasi PBB yang mengurus pendidikan, keilmuan, dan kebudayaan UNESCO menyebutkan, lebih dari 1,5 miliar pelajar di dunia tidak bisa belajar di sekolah akibat virus asal Wuhan, China tersebut seperti yang dirilis oleh surat kabar *Antaraneews.com* pada hari Sabtu, 11 April 2020. Masalah ini tentunya menuntut instansi pendidikan dan pendidik (guru dan dosen) yang menjadi garda terdepan untuk menerapkan proses pembelajaran yang tepat. Regulasi yang diterapkan oleh pemerintah dengan belajar di rumah, bekerja di rumah dengan menerapkan *physical distancing* (jaga jarak) agar Covid-19 tidak menyebar secara cepat diharuskan untuk belajar secara online/daring.

Sayangnya ada beberapa permasalahan yang dihadapi saat melakukan pembelajaran online terutama akses internet (paket data) dan belum terbiasanya pengajar dan peserta didik menggunakan aplikasi pembelajaran online. Sebagaimana yang diungkapkan dari hasil survey yang dilakukan oleh Gunawan (Gunawan, Suranti, & Fathotoni, 2020). Kendala yang paling sering muncul selama pelaksanaan pembelajaran online yaitu paket internet yang tidak dimiliki mahasiswa, keterbatasan akses internet oleh dosen dan mahasiswa, dan belum terbiasanya dengan pembelajaran online. Sebagian besar pengajar mengadakan pembelajaran dan diskusi melalui aplikasi *socialmedia* seperti WhatsApp. Beberapa ada yang memanfaatkan LMS Moodle daring yang dikembangkan universitas atau secara mandiri maupun Google Classroom. Sebagian pembelajaran online juga dilaksanakan melalui video conference dengan memanfaatkan aplikasi Zoom Cloud meetings. Penyampaian materi paling banyak dilakukan yaitu dengan membagi file melalui pesan WhatsApp dan *socialmedia* lainnya.

## **BAB III**

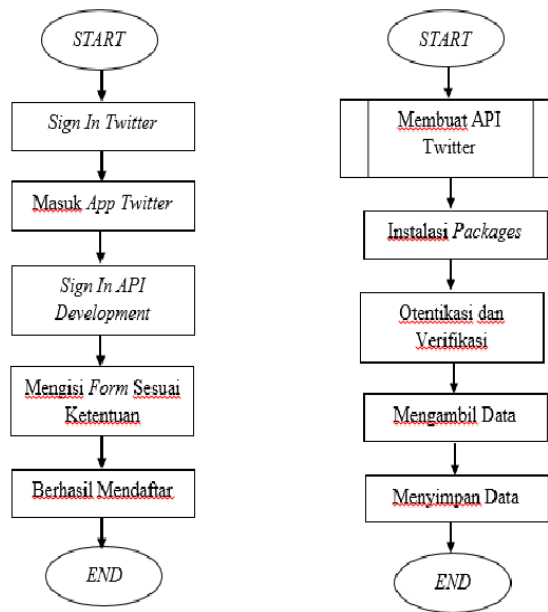
### **METODE PENELITIAN**

#### **3.1 Pengumpulan Data**

Dalam proses pengumpulan data penulis memilih data yang akan digunakan di dalam penelitian ini merupakan data yang didapatkan dari Twitter, data tersebut berupa *tweet* dari pengguna Twitter. Pada proses pengambilan data ini penulis memanfaatkan fasilitas *Application Programming Interface* (API) yang telah disediakan oleh *Twitter Developer*. Kemudian setelah mendapatkan *Application Programming Interface* (API) dari *Twitter Developer*, akan dilakukan *crawling* menggunakan bahasa pemrograman python dengan menggunakan *Application Programming Interface* (API) yang telah didapatkan tadi.

*Crawling* merupakan suatu teknik pengumpulan data otomatis guna mengindekskan suatu informasi dalam suatu halaman menggunakan *Uniform Resource Locator* (URL) dengan menyertakan *Application Programming Interface* (API). Dalam proses *crawling* penulis menggunakan kata kunci “pembelajaran *online*”, “kelas *online*”, “kuliah *online*”, “pembelajaran daring”, “sekolah daring”. Kemudian setelah memperoleh data-data tersebut, data-data tersebut akan disimpan dan kemudian akan diolah menggunakan fungsi serta library yang dibutuhkan. Berikut ini merupakan gambaran dalam proses pengumpulan data yang dilakukan:



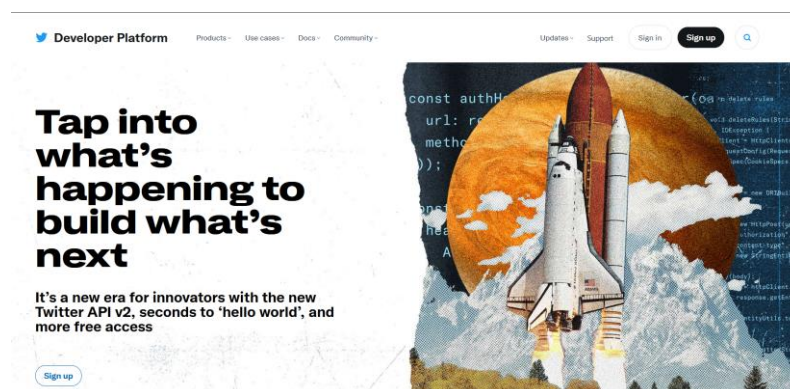


Gambar 5 Tahapan Pengumpulan Data

### 3.1.1 Twitter Developer

Untuk mendapatkan *Application Programming Interface* (API) dari twitter kita perlu menjadikan akun *Twitter* yang kita miliki menjadi akun *Developer*. Berikut ini merupakan tahapan untuk mendapatkan perizinan dari *Twitter* untuk menjadikan akun *Twitter* kita menjadi akun *Developer*:

1. Pertama kita perlu melakukan registrasi akun *Twitter Developer* dengan mengunjungi situs <https://developer.twitter.com> . Kemudian jika sudah memiliki akun *Twitter* klik *Sign in* dan jika belum memiliki akun *Twitter* klik *Sign Up*.



Gambar 6 Tampilan Website Twitter Developer

2. Selanjutnya ikuti alur perintah yang ada dan isikan apa tujuan kita menggunakan *Application Programming Interface (API) Twitter*.

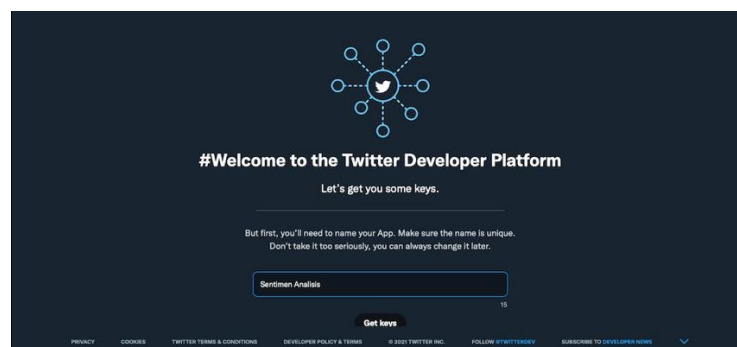
Gambar 7 Tampilan Form Tujuan Penggunaan *Application Programming Interface (API) Twitter*

3. Setelah konfirmasi *email* yang terkait dengan akun *Twitter Developer*, kita akan mendapatkan pemberitahuan serta *link* melalui email yang kita isikan jika pengajuan kita diterima.

### 3.1.2 Proses Pengambilan Data *Twitter*

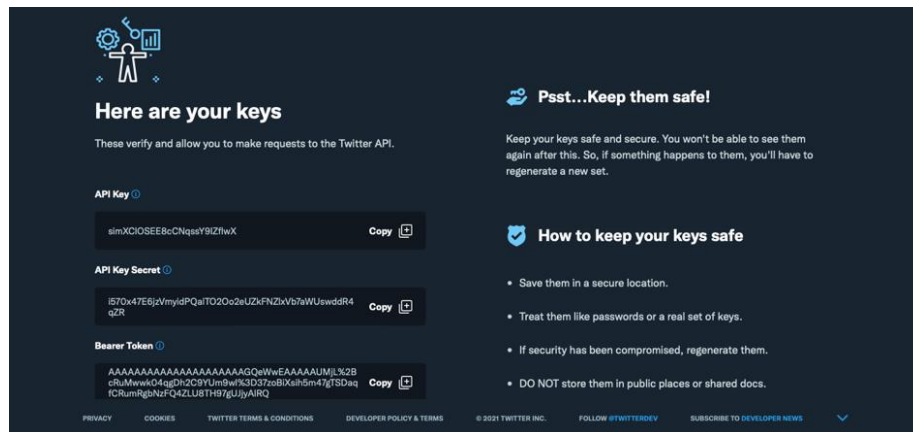
Setelah menjadikan akun *Twitter* kita menjadi akun *Developer* kita dapat memperoleh *Application Programming Interface (API) Twitter* yang dapat kita gunakan untuk mengambil data dari *Twitter* dengan langkah sebagai berikut:

1. Masuk ke website *Twitter Developer*, lalu pilih *project & apps* kemudian buat *project* baru.



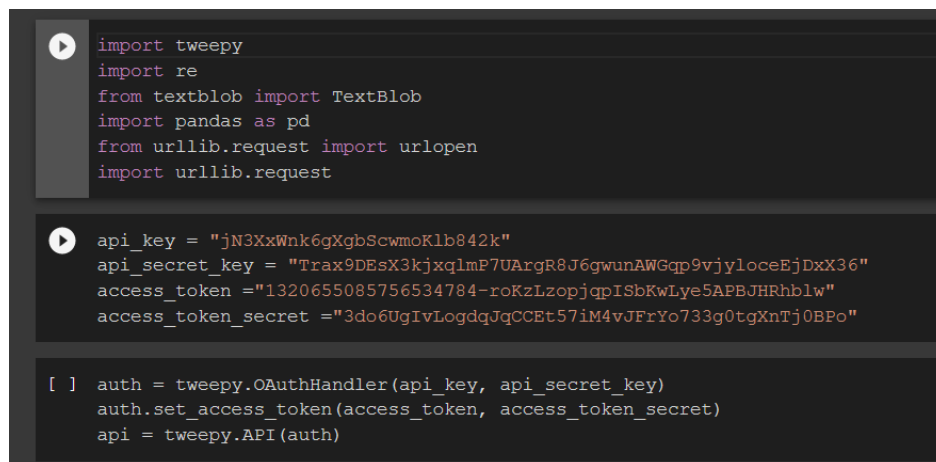
Gambar 8 Tampilan Pembuatan *Project Baru*

2. Setelah itu pilih *Get Keys*, dan kita akan mendapatkan *API* serta *Token* yang dibutuhkan dalam pengambilan data *Twitter*.



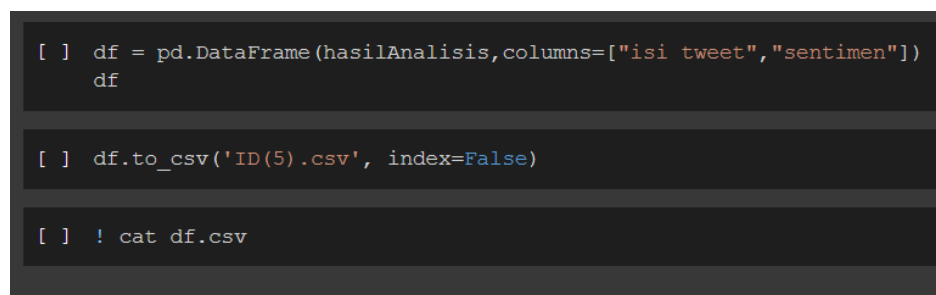
Gambar 9 API Key & Token Key

- Setelah itu salin API serta Token yang sudah kita peroleh lalu masukkan dalam deklarasi program pengambilan data.



Gambar 10 Penggunaan API Key & Token Key

- Kemudian setelah mendapatkan data yang dibutuhkan mengenai pembelajaran online, konversikan data yang kita peroleh kedalam format csv, lalu download dataset yang kita kumpulkan tadi.



Gambar 11 Konversi Dataset Kedalam Format csv

### 3.2 Analisis Data

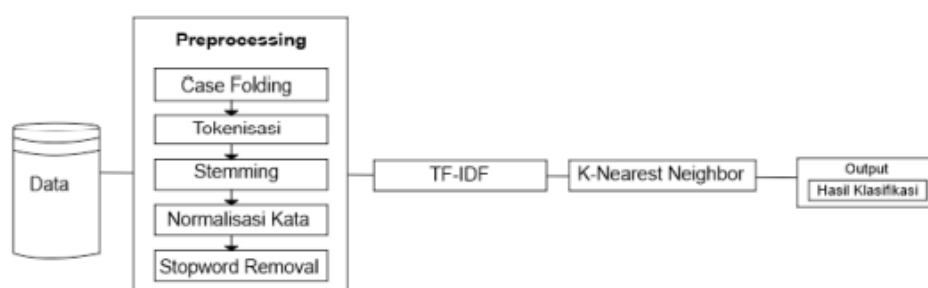
Setelah melakukan pengambilan data yang diambil dari Twitter, data yang diperoleh kemudian akan di analisis serta diberikan label apakah *tweet* yang diambil memiliki sentiment positif atau negatif. Proses pemberian label pada data yang telah diperoleh dilakukan dengan menggunakan algoritma dari *library* yang disediakan oleh *python* yang bernama *TextBlob*, serta dibantu dengan *library* lainnya yang diperlukan untuk menganalisis sentiment dari dataset yang diperoleh.

Berikut ini merupakan tahapan langkah yang dilakukan penulis untuk memberikan label pada data yang telah diperoleh:

1. Memasukkan semua *library* (*Python*) yang diperlukan dalam proses pemberian label pada data seperti *tweepy*, *re*, *textblob* dan lainnya.
2. Memasukan *Application Programming Interface* (API) yang telah diregistrasi yang didapatkan dengan meminta perizinan dari *twitter developer* sebelumnya.
3. Memasukkan kata kunci yang berkaitan dengan pembelajaran *online* yang telah ditentukan sebelumnya.
4. Memasukkan kode program yang digunakan untuk menentukan sentiment dari *tweet* yang telah diperoleh.

### 3.3 Metode yang Diusulkan

Setelah mendapatkan data yang telah diambil dari Twitter penulis kemudian melakukan *data preprocessing* pada data-data tersebut agar data-data tersebut dapat diolah. Setelah data tersebut dibersihkan melalui *data preprocessing* kemudian akan dilakukan pembobotan menggunakan metode TF-IDF. Baru setelah itu dapat dilakukan klasifikasi menggunakan metode *K-Nearest Neighbor* (K-NN).



Gambar 12 Tahapan Analisis Data

Metode yang akan dilakukan dalam klasifikasi analisis sentiment terhadap pemberlakuan pembelajaran online adalah seperti berikut ini:

1. Melakukan *data preprocessing* dari data yang sudah diperoleh guna membersihkan data yang dimiliki agar dapat di proses.
2. Melakukan pembobotan menggunakan metode *Term Frequency – Inverse Document Frequency* atau TF-IDF terhadap data yang dimiliki.
3. Kemudian baru bisa dilakukan klasifikasi analisis sentiment menggunakan metode *K-Nearest Neighbor* (K-NN).

### **3.4 Pengujian Metode**

Pada tahapan ini penulis akan menganalisa hasil kinerja dari program yang akan dilakukan yaitu mencari tau hasil akurasi dari hasil yang diprediksi oleh sistem kemudian akan dibandingkan oleh label aktual dari setiap *tweet* yang diperoleh.

## BAB IV

### HASIL PENELITIAN DAN PEMBAHASAN

#### 4.1 Hasil Penelitian

Dalam penelitian yang dilakukan oleh penulis ini, data yang akan digunakan adalah data berupa tweet sebanyak 500 tweet. Dari 500 dataset tersebut, data dengan sentiment positif ada sebanyak 291 data, dan data dengan sentiment negatif ada sebanyak 209 data. Kemudian selanjutnya data dengan label positif akan diinisialisasikan dengan 1 dan data negatif akan diinisialisasikan dengan 0. Sebelum diproses data harus dibersihkan dulu melalui proses *data preprocessing*, di dalam *data preprocessing* ini nanti data akan dibersihkan dari simbol-simbol seperti hastag (#), user (@), dan lainnya dalam proses *Cleansing*, lalu semua huruf yang ada pada data akan diubah ke dalam huruf kecil dalam proses *Case Folding*, kemudian kata-kata yang memiliki arti tidak penting juga akan dihilangkan dalam proses *Stopword Removal*, dan selanjutnya semua kata dari dataset akan diubah menjadi kata dasar (*root word*) dengan menghilangkan imbuhan di awal dan akhir kata dalam proses *Stemming*. Setelah melalui tahapan *data preprocessing*, akan dilakukan proses pembobotan pada data yang telah dibersihkan dengan menggunakan metode *Term Frequency – Inverse Document Frequency* (TF-IDF). Pembobotan ini sendiri memiliki tujuan untuk menghitung bobot dari setiap kata dengan cara menghitung frekuensi kemunculan kata dari setiap tweet dan dikalikan *idf* untuk mengetahui bobot dari masing-masing kata.

##### 4.1.1 Pengujian Program

Berikut ini merupakan urutan program pengujian dalam analisis sentimen beserta akurasi dengan menggunakan metode *K-Nearest Neighbor* (K-NN).

Membaca dataset & menampilkannya

```
[1] import pandas as pd

data = pd.read_csv("Dataset(2).csv")
data.tail(5)
```

	isi tweet	sentimen
495	@pengarang_sajak HahahaðŸ˜˜...ðŸ˜˜...ðŸ˜˜\nDasar dodoooo...	positif
496	RT @RyaWiedy: Bukti, pembelajaran online bikin...	negatif
497	Bukti, pembelajaran online bikin peserta didik...	negatif
498	RT @umairahbalqis_: Anda pelajar baru? Anda ak...	positif
499	@sevenamsun Bismillah mau saldo dana kak buat ...	negatif

Gambar 13 Memasukkan Dataset

Langkah pertama yang akan dilakukan adalah mengupload dataset yang telah diperoleh dan diubah dalam bentuk csv ke dalam program. Dataset yang digunakan dalam penelitian ini diberinama Dataset(2).csv yang memiliki 500 data tweet yang telah diambil dari Twitter dan diberikan label sentimen.

```
[31] def changetobiner(data):
      if data == 'positif':
          data = 1
      else:
          data = 0
      return data

[32] label = []
      for index, row in data.iterrows():
          label.append(changetobiner(row['sentimen']))

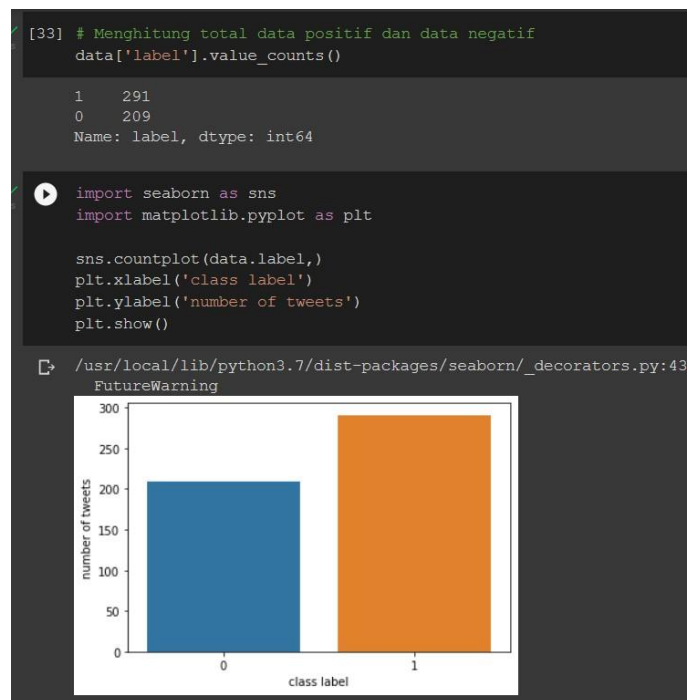
      data['label'] = label
      data.tail(20)
```

	isi tweet	sentimen	label
480	Yuk jadi sekolah kekinian bersama <a href="https://t.co...">https://t.co...</a>	negatif	0
481	Program ini bertujuan untuk memberikan ruang b...	positif	1
482	@Prambors Sama, setelah siswa hampir 2 tahunan...	negatif	0
483	Lucu betul tengok gelagat masing-masing bila k...	negatif	0
484	SUKSES TOEFL DARI NOL SELAMA SEBULANðŸ˜˜ðŸ˜˜ðŸ˜˜= Pro...	negatif	0

Gambar 14 Inisialisasi Label

Setelah kita memasukkan dataset yang kita miliki, untuk memudahkan proses pembuatan program akan dibuat kolom baru pada dataset, kolom tersebut akan diberikan nama label. Di dalam kolom label ini

data dengan sentiment positif akan diberikan label 1 dan data dengan sentiment negatif akan diberikan label 0.



Gambar 15 Jumlah Data Positif dan Negatif

Setelah menginisialisasikan sentiment positif dengan label 1, dan sentiment negatif dengan label 0, kita dapat melihat berapa jumlah data dengan label positif dan negative. Dalam kasus ini, dataset yang digunakan memiliki 291 data dengan sentiment positif, dan 209 data dengan sentimen negatif.

Kemudian agar data dapat diolah, perlu dilakukan *data preprocessing* guna membersihkan dataset yang digunakan. Pada penelitian ini tahapan data preprocessing diawali dengan proses *Cleansing* yang bertujuan untuk membersihkan data dari simbol-simbol seperti hastag (#), user (@), dan lainnya. Kemudian seluruh huruf kapital pada data akan diubah ke huruf kecil dalam proses *Case Folding*.



```
[119] import string, re

def cleansing(data):
    #lower text
    data = data.lower()

    #hapus punctuation
    remove = string.punctuation
    translator = str.maketrans(remove, ' '*len(remove))
    data = data.translate(translator)
    # remove stock market tickers like $GE
    data = re.sub(r'\$\w+', '', data)
    # remove old style retweet text "RT"
    data = re.sub(r'^RT[\s]+', '', data)
    #remove ASCII dan unicode
    data = data.encode('ascii', 'ignore').decode('utf-8')
    data = re.sub(r'^\x00-\x7f', r'', data)
    # remove hyperlinks
    data = re.sub(r'https?:\/\/\.*[\r\n]*', '', data)
    #remove newline
    data = data.replace('\n', ' ');

    return data

[120] comment = []
for index, row in data.iterrows():
    comment.append(cleansing(row['isi tweet']))

data['isi tweet'] = comment
data.head(200)
```

Gambar 16 Proses Cleansing dan Case Folding

Selanjutnya setelah data dibersihkan dan semua kata diganti menjadi huruf kecil, kemudian semua kata-kata yang memiliki arti tidak penting akan dihilangkan dalam proses *Stopword Removal*.

Proses StopWord Removal

```
[240] !pip install Sastrawi
Requirement already satisfied: Sastrawi in /usr/local/lib/python3.7/dist-packages (1.0.1)

[241] from Sastrawi.StopWordRemover.StopWordRemoverFactory import StopWordRemoverFactory

factory = StopWordRemoverFactory()
stopword = factory.create_stop_word_remover()

# implementasi pada data
comment = []
for index, row in data.iterrows():
    comment.append(stopword.remove(row["isi tweet"]))

data["isi tweet"] = comment
data.head()
```

	isi tweet	sentimen	label
320	aku baru selesai daring abis siapin bwt besok...	positif	1
63	rt gadgetnalna bestnya kalau orang hadiah al...	positif	1
221	pembelajaran daring matematika masa pandemi co...	positif	1
47	rt aanisafiqah fedtriyahya bila lecturer pu...	positif	1
69	pakej exclusive combo voicenote 1 basic i...	positif	1

Gambar 17 Proses Stopword Removal

Setelah semua kata yang tidak penting dihilangkan pada proses Stopword Removal sebelumnya, selanjutnya semua kata dari dataset akan diubah menjadi kata dasar (*root word*) dengan menghilangkan imbuhan di awal dan akhir kata dalam proses *Stemming*.

Proses Stemming

```

from Sastrawi.Stemmer.StemmerFactory import StemmerFactory

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

[244] # implementasi pada data
comment = []
for index, row in data.iterrows():
    comment.append(stemmer.stem(row['isi tweet']))

data['isi tweet'] = comment
data.head()

```

	isi tweet	sentimen	label
320	aku baru selesai daring abis siapin bwt besok ...	positif	1
63	rt gadgetnaina bestnya kalau orang hadiah airp...	positif	1
221	ajar daring matematika masa pandemi covid 19 h...	positif	1
47	rt aanisafiqah fedtriyahya bila lecturer puji ...	positif	1
69	pakej exclusive combo voicenote 1 basic iqra 2...	positif	1

Gambar 18 Proses Stemming

Setelah dataset dibersihkan melalui data preprocessing, selanjutnya kita akan membagi dataset yang kita miliki menjadi *data train* dan *data test*. *Data train* nantinya akan digunakan untuk membuat model dari algoritma yang akan digunakan, dimana dalam penelitian ini menggunakan algoritma *K-Nearest Neighbor* (K-NN). Sedangkan data test digunakan untuk menguji performa serta kebenaran dalam model yang bersangkutan.

Membagi data menjadi data training dan data testing

```

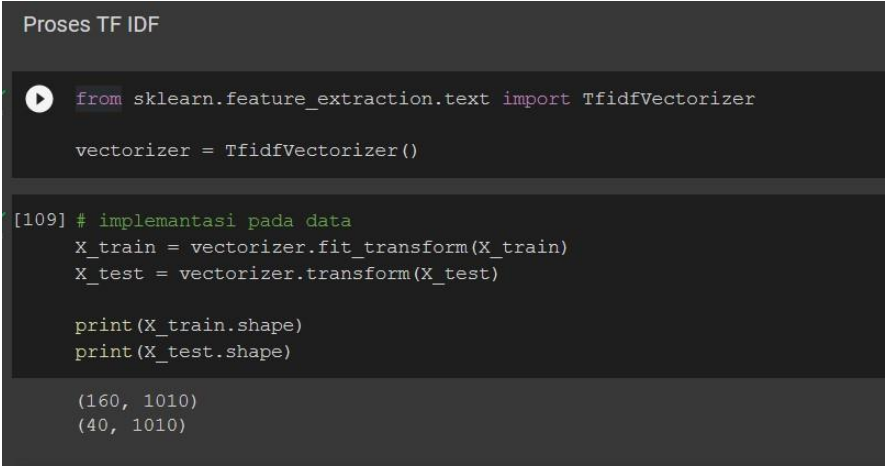
[ ] from sklearn.model_selection import train_test_split

X_train, X_test, Y_train, Y_test = train_test_split(data['komentar'], data['label'], test_size=0.2, stratify=data['label'], random_state=30)

```

Gambar 19 Membagi Data Menjadi Data Train dan Data Test

Setelah dataset yang kita miliki kita bagi menjadi *data train* dan *data test*, selanjutnya akan dilakukan proses pembobotan. Dalam tahapan pembobotan ini, setiap dokumen nantinya akan dirubah menjadi vektor dengan elemen sebanyak term yg telah dapat dikenali agar dapat dilakukan proses klasifikasi. Pada penelitian ini metode pembobotan yang digunakan adalah *Term Frequency – Inverse Document Frequency* (TF-IDF).



```

Proses TF IDF

from sklearn.feature_extraction.text import TfidfVectorizer

vectorizer = TfidfVectorizer()

[109] # implementasi pada data
X_train = vectorizer.fit_transform(X_train)
X_test = vectorizer.transform(X_test)

print(X_train.shape)
print(X_test.shape)

(160, 1010)
(40, 1010)

```

Gambar 20 Proses Pembobotan Dengan Metode Term Frequency – Inverse Document Frequency (TF-IDF).

Setelah dataset yang dimiliki melalui proses pembobotan, barulah dapat dilakukan klasifikasi menggunakan metode *K-Nearest Neighbor* (K-NN). Dalam penelitian kali ini nilai K yang akan digunakan dalam pengujian metode *K-Nearest Neighbor* (K-NN) ini adalah k=1, k=3, k=5, k=7, dan k=9.



```

Klasifikasi dengan KNN

from sklearn.neighbors import KNeighborsClassifier

# menentukan jumlah nilai K
k=1
knn = KNeighborsClassifier(n_neighbors=k)

[93] # melakukan prediksi pada data test
clf = knn.fit(X_train, Y_train)
predict = clf.predict(X_test)

```

Gambar 21 Proses Klasifikasi dengan K-Nearest Neighbor (K-NN).

Kemudian langkah yang akan dilakukan selanjutnya adalah menghitung hasil *accuracy*, *f1 score*, *precision*, serta *recall* menggunakan *confusion matrix* dari hasil pengklasifikasian yang sudah dilakukan dengan metode *K-Nearest Neighbor* (K-NN) sebelumnya.

*Accuracy* merupakan gambaran dari seberapa kuat model yang telah dibuat dalam mengklasifikasikan data dengan benar. Rumus untuk menentukan *accuracy* adalah sebagai berikut:

$$accuracy = \frac{(TP + TN)}{(TP + FP + TN + FN)}$$

Dimana:

TP (*True Positive*) = hasil aktual positif dan algoritma juga mengidentifikasikan positif

TN (*True Negative*) = hasil aktual negatif dan algoritma juga mengidentifikasikan negative

FP (*False Positive*) = hasil aktual negatif tetapi algoritma mengidentifikasikan positif

FN (*False Negative*) = hasil aktual positif tetapi algoritma mengidentifikasikan negatif

*Precision* merupakan gambaran dari hasil akurasi antara data yang diminta dengan hasil prediksi yang diberikan oleh model yang telah dibuat. Rumus untuk menentukan *precision* sendiri adalah sebagai berikut:

$$precision = \frac{(TP)}{(TP + FP)}$$

Dimana:

TP (*True Positive*) = hasil aktual positif dan algoritma juga mengidentifikasikan positif

FP (*False Positive*) = hasil aktual negatif tetapi algoritma mengidentifikasikan positif

*Recall* atau yang biasa disebut dengan *sensitivity* merupakan gambaran dari keberhasilan model dalam menemukan kembali sebuah informasi. Rumus untuk menentukan *recall* adalah sebagai berikut:

$$recall = \frac{(TP)}{(TP + FN)}$$

Dimana:

TP (*True Positive*) = hasil aktual positif dan algoritma juga mengidentifikasikan positif

FN (*False Negative*) = hasil aktual positif tetapi algoritma mengidentifikasikan negatif

*F1 Score* merupakan gambaran dari perbandingan rata-rata dari precision dan recall yang telah dibobotkan. Rumus untuk menentukan *f1 score* sendiri adalah sebagai berikut:

$$f1\ score = 2 \times \frac{(Recall \times Precision)}{(Recall + Precision)}$$

Dimana:

Recall = gambaran dari keberhasilan model dalam menemukan kembali sebuah informasi

Precision = gambaran dari hasil akurasi antara data yang diminta dengan hasil prediksi yang diberikan oleh model yang telah dibuat

#### 4.1.2 Hasil Klasifikasi

Dari proses pengklasifikasian yang dilakukan dengan metode *K-Nearest Neighbor* (K-NN) yang telah dilakukan dengan nilai K sebesar k=1, k=3, k=5, k=7, dan k=9, didapatkan hasil seperti berikut:

1. Klasifikasi Dengan  $k=1$ .Table 2 Hasil Klasifikasi dengan  $K=1$ 

	Hasil	Presentase
Accuracy	0,675	67,5%
F1 Score	0,683	68,3%
Precision	0,677	67,7%
Recall	0,7	70%

Dalam proses klasifikasi menggunakan metode *K-Nearest Neighbor* (K-NN) dengan nilai  $k=1$ , didapatkan hasil dari *accuracy*, *f1 score*, *precision*, dan *recall* seperti pada gambar 22. Dengan menggunakan nilai  $k=1$ , didapatkan nilai *accuracy* = 0.675 , *f1 score* = 0.683 , *precision* = 0.667 , dan nilai *recall* = 0.7 .

2. Klasifikasi Dengan  $k=3$ Table 3 Hasil Klasifikasi dengan  $K=3$ 

	Hasil	Presentase
Accuracy	0,675	67,5%
F1 Score	0,683	68,3%
Precision	0,677	67,7%
Recall	0,7	70%

Dalam proses klasifikasi menggunakan metode *K-Nearest Neighbor* (K-NN) dengan nilai  $k=3$ , didapatkan hasil dari *accuracy*, *f1 score*, *precision*, dan *recall* seperti pada gambar 23. Dengan menggunakan nilai  $k=3$ , didapatkan nilai *accuracy* = 0.675 , *f1 score* = 0.683 , *precision* = 0.667 , dan nilai *recall* = 0.7 .

### 3. Klasifikasi Dengan k=5

**Table 4 Hasil Klasifikasi dengan K=5**

	Hasil	Presentase
Accuracy	0,67	67%
F1 Score	0,667	66,7%
Precision	0,75	75%
Recall	0,6	60%

Dalam proses klasifikasi menggunakan metode *K-Nearest Neighbor* (K-NN) dengan nilai k=5, didapatkan hasil dari *accuracy*, *f1 score*, *precision*, dan *recall* seperti pada gambar 24. Dengan menggunakan nilai k=5, didapatkan nilai *accuracy* = 0.67 , *f1 score* = 0.667 , *precision* = 0.75 , dan nilai *recall* = 0.6 .

### 4. Klasifikasi Dengan k=7

**Table 5 Hasil Klasifikasi dengan K=7**

	Hasil	Presentase
Accuracy	0,8	80%
F1 Score	0,778	77,8%
Precision	0,875	87,5%
Recall	0,7	70%

Dalam proses klasifikasi menggunakan metode *K-Nearest Neighbor* (K-NN) dengan nilai k=7, didapatkan hasil dari *accuracy*, *f1 score*, *precision*, dan *recall* seperti pada gambar 25. Dengan menggunakan nilai k=7, didapatkan nilai *accuracy* = 0.8 , *f1 score* = 0.778 , *precision* = 0.875 , dan nilai *recall* = 0.7 .

## 5. Klasifikasi Dengan k=9

Table 6 Hasil Klasifikasi dengan K=9

	Hasil	Presentase
Accuracy	0,75	75%
F1 Score	0,722	72,2%
Precision	0,813	81,3%
Recall	0,65	65%

Dalam proses klasifikasi menggunakan metode *K-Nearest Neighbor* (K-NN) dengan nilai k=9, didapatkan hasil dari *accuracy*, *f1 score*, *precision*, dan *recall* seperti pada gambar 26. Dengan menggunakan nilai k=9, didapatkan nilai *accuracy* = 0.75 , *f1 score* = 0.722 , *precision* = 0.813 , dan nilai *recall* = 0.65 .

### 4.2 Analisis

Dari hasil uji akurasi program yang telah dibuat menggunakan metode *K-Nearest Neighbor* (K-NN) dengan metode pembobotan *Term Frequency – Inverse Document Frequency* (TF-IDF) yang dilakukan dengan menggunakan nilai k=1 didapatkan nilai *accuracy*=0.675 , *f1 score* = 0.683 , *precision* = 0.667 , dan nilai *recall*=0.7 , menggunakan nilai k=3 didapatkan nilai *accuracy* = 0.675 , *f1 score* =0.683 , *precision* = 0.667 , dan nilai *recall* = 0.7 , menggunakan nilai k=5 didapatkan nilai *accuracy* = 0.67 , *f1 score* = 0.667 , *precision* = 0.75 , dan nilai *recall* = 0.6 , menggunakan nilai k=7 didapatkan nilai *accuracy* = 0.8 , *f1 score* = 0.778 , *precision* = 0.875 , dan nilai *recall* = 0.7 , dan dengan menggunakan nilai k=9 didapatkan nilai *accuracy* = 0.75 , *f1 score*=0.722 , *precision* = 0.813 , dan nilai *recall* = 0.65.

Dari proses klasifikasi yang sudah dilakukan, nilai *accuracy*, *f1 score*, dan *precision* tertinggi didapatkan pada percobaan dengan menggunakan nilai k=7 dengan nilai *accuracy* sebesar 0.8, *f1 score* sebesar



0.778, dan nilai *precision* sebesar 0,875. Sedangkan nilai *recall* tertinggi didapatkan pada percobaan dengan menggunakan nilai  $k=1$ ,  $k=3$ , dan  $k=7$  dengan nilai *recall* sebesar 0,7. Dari data yang sudah didapatkan dari implementasi algoritma di atas, dapat diartikan bahwa besarnya nilai  $K$  yang digunakan untuk batas ketetanggaan algoritma *K-Nearest Neighbor* (K-NN) berpengaruh dalam menentukan nilai *accuracy*, *f1 score*, *precision*, serta nilai *recall*.

## BAB V

### KESIMPULAN

#### 5.1 Kesimpulan

Berdasarkan hasil pengujian algoritma *K-Nearest Neighbor* (K-NN) terhadap sentimen masyarakat mengenai pemberlakuan pembelajaran online pada *tweet* yang diambil dari Twitter dengan menggunakan 500 data, ada beberapa hal yang dapat disimpulkan, antara lain :

1. Dari 500 data *tweet* yang didapatkan mengenai tanggapan masyarakat terhadap pemberlakuan pembelajaran *online*, sebanyak 291 *tweet* masyarakat memberikan tanggapan positif, dan 209 *tweet* masyarakat memberikan tanggapan negatif. Jadi dapat disimpulkan bahwa opini masyarakat dengan jumlah data sebanyak 500 *tweet* memiliki tanggapan positif terhadap pemberlakuan pembelajaran online.
2. Hasil analisis sentimen twitter dengan menggunakan kata kunci pembelajaran *online*, kelas *online*, kuliah *online*, pembelajaran daring, dan sekolah daring pada program yang telah dibuat menggunakan metode *K-Nearest Neighbor* (K-NN) dengan metode pembobotan *Term Frequency – Inverse Document Frequency* (TF-IDF) yang dilakukan dengan menggunakan nilai  $k=1$  didapatkan nilai *accuracy* sebesar 0.675, nilai *f1 score* sebesar 0.683, nilai *precision* sebesar 0.667, dan nilai *recall* sebesar 0.7 ,lalu dengan menggunakan nilai  $k=3$  didapatkan nilai *accuracy* sebesar 0.675 , nilai *f1 score* sebesar 0.683, nilai *precision* sebesar 0.667 , dan nilai *recall* sebesar 0.7,lalu dengan menggunakan nilai  $k=5$  didapatkan nilai *accuracy* sebesar 0.67 , nilai *f1 score* sebesar 0.667, nilai *precision* sebesar 0.75, dan nilai *recall* sebesar 0.6 , sedangkan menggunakan nilai  $k=7$  didapatkan nilai *accuracy* sebesar 0.8, nilai *f1 score*

sebesar 0.778 , nilai *precision* sebesar 0.875 , dan nilai *recall* sebesar 0.7 , dan dengan menggunakan nilai  $k=9$  didapatkan nilai *accuracy* sebesar 0.75 , nilai *f1 score* sebesar 0.722, nilai *precision* sebesar 0.813 , dan nilai *recall* sebesar 0.65.

3. Dalam pengujian yang dilakukan pada penelitian ini nilai akurasi, *f1 score*, dan presisi tertinggi didapatkan pada nilai  $k=7$  dengan nilai akurasi sebesar 0.8 atau 80%, nilai *f1 score* sebesar 0,778 atau 77,8%, nilai presisi sebesar 0,875 atau 87,5%. Sedangkan nilai *recall* tertinggi didapatkan pada nilai  $k=1$ ,  $k=3$ , dan  $k=7$  dengan nilai *recall* sebesar 0,7 atau 70%.

## 5.2 Penelitian Selanjutnya

Penulis menyadari bahwa pada program analisis sentiment yang telah dibuat oleh penulis ini masih memiliki banyak kekurangan. Oleh karena itu, pada penelitian selanjutnya perlu ada beberapa hal yang harus dipertimbangkan serta diperbaiki, agar dapat diperoleh perkembangan yang lebih baik. Dalam pengembangan sistem selanjutnya, diharapkan sistem ini dapat diaplikasikan pada semua *platform*, baik itu *platform* media sosial maupun *platform* komersial, serta dapat dikombinasikan dengan algoritma pengklasifikasian yang lainnya seperti *Naïve Bayes*, *Support Vector Machine* (SVM), atau program dapat ditambahkan proses tambahan dalam melakukan klasifikasi yaitu dalam proses perhitungan nilai validitas guna mempertimbangkan validitas antar data latih yang digunakan serta perhitungan weighted voting yang digunakan untuk melakukan perhitungan bobota dari masing-masing terdekat dengan menggunakan metode *Modified K-Nearest Neighbor* (MKNN).

## Daftar Pustaka

- Bonde, A. (2019). K-Nearest Neighbor Dengan Feature Selection Menggunakan Backward Elimination Untuk Prediksi Harga Komoditi Kopi Arabika. *ILKOM Jurnal Ilmiah*, 188-195.
- Fitri, M. (2018). Perancangan Sistem Temu Balik Informasi Dengan Metode Pembobotan Kombinasi TF-IDF Untuk Pencarian Dokumen Berbahasa Indonesia. *Teknik Elektro Fakultas Teknik Universitas Tanjungpura*, 1-6.
- Goje, K. (2017). Preventative Prophetic Guidance in Infection and Quarantine. *Journal of Usuluddin*, 155-170.
- Gunawan, Suranti, N. Y., & Fathotoni. (2020). Variations of Models and Learning Platforms for Prospective Teachers During he COVID. *Indonesian*, 61-70.
- Haranto, F. F., & Sari, B. W. (2019). Implementasi Support Vector Machine Untuk Analisis Sentimen Pengguna Twitter Terhadap Pelayanan Telkom dan Biznet. *Jurnal PILAR Nusa Mandiri*, 171-176.
- He, W., Xu, G., & Kruck. (2016). Online IS Education for the 21st Century. *Journal of Information System Education*, 101-105.
- Kadarina, T. M., Dani, A., & Supegina, F. (2021). Peningkatan Kemampuan Turtle Library Untuk Menggambar Grafik Animasi Bagi Pelajar PKBM . *Universitas Mercu Buana*, 455-465.
- Liu, B. (2016). Sentiment Analysis and Subjectivity. *Handbook of Natural Language Processing*, 1-36.
- Mahase, E. (2020). Coronavirus: covid-19 has killed more people than SARS and MERS combined, despite lower case fatality rate. *NEWS*, 1 of 1.
- Manning, C. D., Raghavan, P., & Schütze, H. (2017). An Introduction to Information Retrieval. *Cambridge University Press*, 1-544.

Nurdiansyah, Y., Bukhori, S., & Hidayat, R. (2018). Sentiment analysis system for movie review in Bahasa Indonesia using naive bayes classifier method. *Journal of Physics*, 1-7.

Oktavianto, A., & Persada, S. F. (2020). Persepsi Publik Tentang Pembelajaran Daring di Indonesia: Studi Menggunakan ELK Stack dan Python untuk Analisis Sentimen di Twitter. *Jurnal Teknik ITS*, A170-A175.

Perdana, K., Pricillia, T., & Zulfachmi. (2021). Optimasi TextBlob Menggunakan Support Machine Untuk Analisis Sentimen (Studi Kasus Layanan Telkomsel). *Bangkit Indonesia*, 13-15.

Primartha, R. (2018). *Belajar Machine Learning Teori dan Praktik*. Bandung: Informatika Bandung.

Pudjajana, A. M., & Manongga, D. (2018). Sentimen Analisis Tweet Pornografi Kaum Homoseksual Indonesia di Twitter Dengan Naive Bayes. *Jurnal SIMETRIS*, 313-318.

R, P. (2018). Belajar Machine Learning Teori Dan Praktik. *Bandung : Informatika Bandung*.

Romadloni, N. T., Santoso, I., & Budilaksono, S. (2019). Perbandingan Metode Naive Bayes, K-NN dan Decision Tree Terhadap Analisis Sentimen Transportasi KRL Commuter Line. *Jurnal IKRA-ITH Informatika*, 1-9.

Sari, R. (2020). Analisis Sentimen Pada Review Objek Wisata Dunia Fantasi menggunakan Algoritma K-Nearest Neighbor (K-NN). *Jurnal Sains dan Manajemen*, 10-17.

Satrianingrum, A. P., & Prasetyo, I. (2021). Persepsi Guru Dampak Pandemi Covid-19 terhadap Pelaksanaan Pembelajaran Daring di PAUD. *Jurna Pendidikan Anak Usia Dini*, 633-640.

Smith, A., & Fredman. (2020). Isolation, quarantine, social distancing and community containment: pivotal role for old-style public health. *International Society of Travel Medicine*, 1-4.

Syarifudin, M. (2020). Analisis Sentimen Opini Publik Mengenai COVID-19 Pada Twitter Menggunakan Metode Naïve Bayes dan K-NN. *Inti Nusa Mandiri*, 23-28.

Tirtana, A., Zulkarnain, A., & Listio, Y. D. (2019). Pembuatan Sistem Pencarian Pekerjaan Menggunakan TF-IDF. *Jurnal Ilmiah Teknologi Informasi Asia*, 91-100.

Widyasanti, N. K., Putra, I. G., & Rusjyanthi, N. K. (2018). Seleksi Fitur Bobot Kata dengan Metode TFIDF untuk Ringkasan Bahasa Indonesia. *Merpati*, 119-126.

Wong , Y. W., & Mooney, R. J. (2018). Learning for Semantic Parsing with Statistical Machine Translation. *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, 439-446.