

Analisis Sentimen Terhadap Pembelajaran Daring dengan Algoritma *Naïve Bayes Classifier*

Fajar Sidik^{#1}, Ibnu Suhada^{#2}, Azhar Haikal Anwar^{#3}, Firman Noor Hasan^{#4}

[#]Teknik Informatika, Universitas Muhammadiyah Prof. Dr. Hamka
Jl. Tanah Merdeka No.6, Kec. Pasar Rebo, DKI Jakarta 13830, Indonesia

¹fjrsk88@gmail.com

²ibnu.suhada10@gmail.com

³azharlolypop@gmail.com

⁴firman.noorhasan@uhamka.ac.id

Abstrak— Pandemi *COVID-19* merupakan penyakit yang skala penyebarannya terjadi secara global di seluruh dunia, termasuk Indonesia. Banyak bidang yang terkena dampak pandemi ini termasuk pendidikan. Indonesia saat ini sedang menjalankan strategi pembelajaran daring yang menimbulkan banyak opini masyarakat. Analisis sentiment pada cabang *Text Mining* digunakan untuk mengklasifikasi suatu entitas pada dokumen teks yang terdiri dari dua kelas yaitu positif dan negatif, kelas tersebut diperoleh dengan mengklasifikasikan *dataset* headline dan substansi berita terkait pembelajaran daring. Tujuan dari artikel ini adalah untuk melakukan prediksi pendapat orangtua terhadap pembelajaran daring serta mengetahui nilai akurasi dari pendapat tersebut dengan algoritma *Naïve Bayes Classifier*. Artikel ini menggunakan metode *TF-IDF* untuk melakukan pembobotan teks. Hasil dari artikel ini terlihat bahwa algoritma *Naïve Bayes Classifier* mampu melakukan prediksi kalimat positif dan negatif dengan nilai 62,5% positif dan 37,5% negatif dari 40 data pendapat orangtua dan nilai akurasi sebesar 65% berdasarkan 100 berita positif dan 100 berita negative.

Kata kunci— analisis sentimen, naïve bayes classifier, pembelajaran daring.

I. PENDAHULUAN

Covid-19 merupakan penyakit yang penyebarannya disebabkan oleh *Sars-coV2* yang awalnya muncul di Wuhan, Tiongkok pada 31 desember 2019. *Covid-19* ini dapat menyebabkan masalah pernapasan ringan, infeksi paru-paru serius, dan bahkan kematian. Sampai sekarang, ada lima jenis *covid-19* yang telah diidentifikasi pada manusia [1]. Sebagai dampak pandemi *covid-19* ini, banyak pendekatan dilakukan untuk menghentikan rantai penyebaran *covid-19* termasuk Indonesia [2]. Kebijakan ini dilakukan agar pemerintah bisa mengurangi mobilitas

pelajaran mahasiswa sehingga dapat menahan penyebaran *covid-19* [3].

Pandemi *covid-19* menyebabkan adanya perubahan sosial yang tidak direncanakan [4]. Perubahan utama terkait model cara belajar [5]. Guru harus melakukan pembelajaran jarak jauh (online) [6]. Dengan metode belajar daring seperti ini sangat membutuhkan pendampingan dari orang tua [7]. Peran orang tua tidak hanya memberikan dasar pendidikan, sikap serta keterampilan dasar, tetapi kini meluas sebagai pendamping belajar [8]. Dengan sistem pembelajaran yang baru siswa sangat membutuhkan peran guru dan orang tua agar bisa beradaptasi dengan sistem yang baru[9].

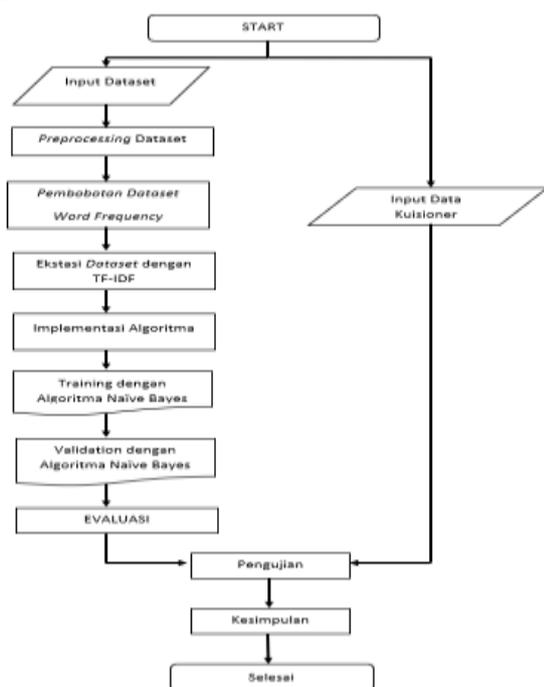
Analisis sentimen merupakan suatu proses memahami, mengekstrak dan memproses data berbasis teks secara otomatis, proses ini dilakukan untuk memperoleh sebuah informasi sentimen yang tersirat dari suatu opini [10]. Dilakukannya analisis sentimen untuk mengetahui kecenderungan penilaian terhadap permasalahan dari sudut pandang orang lain, umumnya apakah mereka mempunyai opini negatif atau positif terhadap suatu permasalahan [11]. Dalam beberapa penelitian, sentiment dapat dikategorikan sebagai *big data* dimana ukuran dari data teks tersebut semakin banyak dan makna konteksnya semakin beragam [12]. Proses analisis meliputi *tweet*, teks ulasan, *blog*, forum, dan data *preprocessing* mencakup *stopword*, *stemming*, penghapusan, identifikasi sentimen, proses *tokenization*, dan klasifikasi sentimen [13]. *Text Mining* merupakan bidang ilmu baru yang berkembang untuk mengekstrak informasi yang bermakna dari teks alami [14], hal tersebut dapat dicirikan sebagai proses analisis teks ke ekstrak informasi yang berguna untuk tujuan tertentu [15]. Pengolahan teks atau *text mining* dapat dilakukan dengan beberapa tahap dalam *preprocessing* atau pembersihan data teks untuk diselesaikan ke tahap pengelompokkan (klasifikasi)

berikutnya [16]. *Text mining* mempunyai tugas khusus yaitu pengelompokan teks [17].

Oleh karena itu, analisis sentimen dilakukan dengan tujuan untuk melakukan analisis dari sebuah opini terkait metode pembelajaran online yang sedang dilakukan, penelitian ini menggunakan *googleform* untuk pengambilan data dan diproses menggunakan perhitungan *naïve bayes* untuk mendapatkan sebuah hasil analisis sentimen terhadap pembelajaran online. Penelitian ini memberikan gambaran bagaimana cara agar tetap dapat melakukan kegiatan belajar dari rumah selama pandemi *covid-19*, dan diharapkan dapat menjadi motivasi belajar dalam melaksanakan pembelajaran online.

II. METODOLOGI PENELITIAN

Gambar.1 menunjukkan langkah-langkah penelitian yang dilakukan oleh peneliti.



Gambar.1 Diagram alir penelitian

A. Input Data

Setiap penelitian tentukan membutuhkan data sebagai bahan olah atau sebagai sumber dalam melakukan penelitian. Pada penelitian ini, peneliti menggunakan 2 sumber data.

B. Perancangan Model *Preprocessing Data*

Pada tahapan ini, data yang sudah terkumpul langsung diolah, awal atau biasa disebut *preprocessing*. Tahap ini bertujuan untuk membuat teks kalimat menjadi kalimat yang diinginkan sebelum dilakukannya proses selanjutnya yaitu implementasi. Adapun langkah yang dilakukan pada proses ini seperti:

1. *Case Folding*

Tahap ini melakukan perubahan huruf kapital menjadi huru kecil.

2. *Cleansing*

Tahap ini dilakukan untuk menghilangkan kata yang tidak dibutuhkan. Contoh atribut yang tidak dibutuhkan yaitu angka (1), tanda baca (!) dan link ('http') dan symbol lainnya.

3. *Stemming*

Tahap ini dilakukan untuk menghilangkan imbuhan pada kata di dalam sebuah kalimat laporan atau melakukan perubahan kata kerja menjadi kata benda. Contoh kata yang dihilangkan setelah imbuhan di- dan -kan berubah menjadi "hilang".

C. *Word Frequency*

Pada tahap ini dilakukan pengambilan *Word Frequency*, yaitu menghitung banyak kata muncul dalam sebuah *dataset* yang telah melewati tahap *preprocessing*. Selanjutnya menggunakan *Count Vectorizer* yaitu untuk dapat mengubah fitur teks menjadi representasi matriks.

D. Ekstraksi *Dataset*

Term Frequency-Inverse Document Frequency atau yang biasa disebut dengan TF-IDF merupakan teknik reduksi variabel pada tahap ekstraksi *dataset* yang digunakan dalam penelitian ini. Dilakukannya TF-IDF untuk melakukan pembobotan kata dalam dokumen yang dihitung dari frekuensi kemunculan term dalam dokumen. Hal ini ditekankan pada seringnya kata dalam dokumen yang muncul di berita. Terjadi peningkatan nilai TF-IDF secara seimbang sesuai dengan waktu kemunculan sebuah kata dalam dokumen dengan diimbangi dengan frekuensi kata dalam berita. Tujuannya untuk melakukan penyesuaian fakta pada beberapa kata yang sering muncul dalam berita secara keseluruhan. Kata-kata yang bersifat umum akan mendapatkan bobot yang lebih rendah dan memberi makna pada kata-kata tertentu yang jarang ditemukan dalam sebuah dokumen.

E. Implementasi Algoritma

Pada tahap implementasi, tahapan ini menggunakan algoritma *naive bayes classifier* dan menggunakan teknik *multinomial naive bayes*. Tahapan ini mempunyai tiga proses, yaitu proses *training*, *validation*, dan *testing*. Proses *training* dilakukan untuk pelatihan data, kemudian, pada saat itu, proses *validation* akan dilakukan dengan mengacu probabilitas dari *dataset training*. Setelah dilakukan proses *validation*, maka akan dilakukan proses *testing* dengan menggunakan data yang dikumpulkan dari kuisloner *googleform* sebagai percobaan dari data yang telah di *validation*. Klasifikasi dilakukan dengan teori probabilitas dan teorema *Bayesian*, yang beranggapan semua variabel penentu keputusan adalah bebas dengan begitu keberadaan tiap variabel tidak berhubungan dengan adanya atribut dari variabel lain.

F. Evaluasi Sistem

Setelah proses implementasi dilakukan akan didapatkan sebuah hasil, maka hasil tersebut akan dihitung nilai *f-measure* (*f1-score*) yang merupakan salah satu perhitungan evaluasi dengan mengkombinasikan *recall* dan *precision*.

$$f - measure = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Recall mengacu pada jumlah data yang secara akurat dikelompokkan dalam suatu kelas, dibagi dengan total keseluruhan data yang ada didalam kelas tersebut.

$$Recall (R) = \frac{TP}{(TP + FN)}$$

Keterangan:

TP = *True Positive*, adalah jumlah data positif yang dikelompokkan benar oleh sistem

FN = *False Negative*, adalah jumlah data negatif yang dikelompokkan salah oleh sistem

Precision dihitung dari jumlah data benar, dibagi total keseluruhan data yang telah dikelompokkan dalam sebuah kelas.

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

Keterangan:

TP = *True Positive*, adalah jumlah data positif yang dikelompokkan benar oleh system.

FP = *False Positive*, adalah jumlah data positif yang dikelompokkan salah oleh system.

Pada penelitian ini, alat dan bahan dibutuhkan untuk menunjang proses penelitian ini yang dibagi menjadi beberapa bagian, seperti perangkat keras (*hardware*), perangkat lunak (*software*), *library* dan *googleform*. Pada penelitian ini, peneliti menggunakan perangkat lunak pada Tabel.1

TABEL I
DAFTAR SOFTWARE

No.	Nama Software	Versi
1	Python	3.7 (32 bit)
2	Anaconda	2019.7
3	OS	Windows 10

III. HASIL DAN PEMBAHASAN

A. Proses Data

Data berita yang dikumpulkan pada penelitian ini berjumlah 200 dari beberapa sumber berita seperti Kompas, CNN Indonesia dan yang lainnya. Berita yang diambil dibagi menjadi 2 bagian yaitu 100 berita positif dan 100 berita negatif, seperti yang ditunjukkan dapat dilihat pada Tabel 2.

TABEL II
JUMLAH BERITA DIKUMPULKAN

Berita	Jumlah
Positif	100
Negatif	100
Total	200

Sedangkan untuk data yang dikumpulkan melalui *googleform*, peneliti mengumpulkan 40 data yang akan digunakan untuk pengujian, contoh data *googleform* dapat dilihat dibawah ini Gambar.2.



Gambar 2. Contoh Sebagian Data dari *Googleform*

sedangkan Tabel.3 dan untuk keseluruhan data dapat dilihat pada bit.ly/PendapatOrangtua.

TABEL III
DATA INPUT DARI KUISIONER

No	Kolom Data Kuisioner	
	Dampak Daring	Pendapat dari Orangtua
1	Negatif	Tetap semangat dalam belajar
2	Positif	Kendala dr metode daring yaitu tdk adanya kepemilikan hp bagi sebagian siswa , gangguan jaringan/ sinyal,kuota dll
3	Positif	Kendalanya di jaringan /sinyal dan kuota
4	Positif	Lama kelamaan pun anak mnjadi jenuh belajar dari rumah... Karna tempat terbaik untuk belajar memang lebih baik di sekolah.
5	Negatif	Belajar d rmh kurang memadai karena ank banyak kurang paham lebih baik ketemu langsung dengan guru
6	Negatif	Menurut saya anak" sudah mulai jenuh dengan adanya PJJ yg berkepanjangan. Mereka harus mengerjakan tugas" tanpa adanya oenjelasan secara langsung tentang materi oembelajaran, sehingga anak" harus mengembangkan sendiri materi yg d ajarkan saat ini

No	Kolom Data Kuisisioner	
	Dampak Daring	Pendapat dari Orangtua
7	Negatif	Anak merasa guru kurang dalam menyampaikan materi sehingga membuat anak menjadi tidak paham dengan materi dibahas, interaksi guru dan murid juga kurang efektif
8	Positif	Dengan adanya metode daring saya merasa kurang efektif dalam mengajarkan materi anak dirumah dan anak jadi kurang memahami materi yang diberikan.
9	Negatif	Anak jd kurang disiplin dan lebih bnyak waktu main hp
10	Negatif	Tidak adanya penguasaan atau pemahaman karena tidak interkasi antara guru dan murid, serta tidak ssuainya pokok bahasan pembelajaran karena membingungkan , serta tidak adanya pantuan psikologi antara murid, guru dan orang tua murid.
11	Positif	Belajar daring msh kurang efektif
12	Positif	Belajar masih kurang efektif klo dari rmh
13	Positif	Sangat bagus Klo suasana msh covid
14	Negatif	sebenarnya bagus tetapi metode yg diajarkan oleh pengajar kurang efektif
15	Negatif	pengajar memberikan materi tidak efektif, selalu memberikan tugas
16	Negatif	pengajar sulit untuk menyampaikan materi
17	Negatif	Menurut saya, belajar dari rumah membuat anak menjadi tidak interaktif dikarenakan banyak materi yang tidak dipahami oleh orangtua sehingga anak tidak dapat bertanya kepada orangtua, hanya mengandalkan google saja
18	Negatif	Dari sisi pengajaran orang tua kepada anak tidak efisien dikarenakan orang tua yang sibuk bekerja.
19	Negatif	Kesehatan anak dari virus memang terjaga namun kesehatan mental anak menjadi terganggu karena mereka sudah mulai bosan dan jenuh dengan belajar di rumah. Mereka butuh berinteraksi dengan lingkungan sekolah dan teman-teman lainnya
20	Positif	Karena anak belajar dari rumah sehingga memudahkan saya untuk memantau kesehatan anak
21	Negatif	Pendapat saya kurang bagus karena belajar daring ini terkadang membuat anak jadi malas , karna merasa tidak diawasi dan kurang nya efektifitas dalam mencerna materi yg di berikan guru karena suasana dan lingkungan dirumah biasanya dilakukan sambil tiduran
22	Positif	Beberapa dosen kurang jelas dalam menyampaikan materinya, sehingga anak kadang kuliah daring hanya dapat tugas saja tanpa mendapat materi, bila dosen nya bisa dengan bagus maka akan positif karna waktu yg fleksibel dan tidak buang waktu dan tenaga harus ke kampus
23	Negatif	Tingkatan lagi daring
24	Negatif	Pembelajaran menjadi kurang efektif karena beberapa masalah, seperti koneksi, kejelasan menyampaikan materi dari guru kepada peserta didikmenjadi kurang jelas karena delay.
25	Negatif	Perlu metode yang lebih efektif selain hanya melalui video call.
26	Negatif	Menyebabkan anak saya tidak mau keluar rumah hanya main ps 5 dan gak lepas dari handphone nya

No	Kolom Data Kuisisioner	
	Dampak Daring	Pendapat dari Orangtua
27	Negatif	Anak mengalami kecanduan handphone sehingga membuat anak malah malas belajar, dia malah asik membuka aplikasi lain
28	Positif	Pembelajaran tatap muka memang lebih baik dibandingkan dengan pembelajaran daring dari sisi interaksi sehingga lebih mudah dipahami oleh murid terkait pelajaran yang diajarkan, Untuk jumlah pembelajaran daring pun di sekolah anak saya lebih menitik beratkan ke pemberian tugas bukan pembelajaran melalui zoom, mungkin karena pembelian kuota untuk guru dan murid yang memberatkan kedua pihak, jadi jarang dilakukan. Sehingga murid lebih banyak belajar dengan orang tua atau mencari di google daripada dengan gurunya
29	Negatif	Sukil untuk mengajarkan anak yang berusia dibawah 10 tahun karena kita sebagai orang tua belum memiliki pengalaman bagaimana menyampaikan suatu materi kepada anak sehingga dapat disimpulkan anak menjadi kurang dalam pemahaman materinya
30	Negatif	Saya pikir anak menjadi kurang dalam pemahaman materi dasar dikarenakan cara kita yang mungkin salah dalam menyampaikan materi
31	Negatif	Kesulitan dalam mengajarkan atau menyampaikan materi kepada anak, maka menurut saya sekolah tatap muka lebih baik dikarenakan ada interaksi yang cukup antara pengajar dengan murid. Bukan berarti saya tidak dapat mendidik anak namun sudah ada keahlian masing masing
32	Negatif	Menurut saya belajar dari rumah itu baik seperti belajar tatap muka pada umumnya apabila interaksi antara guru dan murid berjalan dengan baik, namun yang saya lihat interaksi tersebut tidak ada, yang ada ialah guru hanya memberikan tugas tanpa menjelaskan terkait materi tersebut
33	Positif	nilai menjadi semakin bagus, sisi negatifnya kegiatan belajar mengajar menjadi kurang maksimal
34	Positif	Pembelajaran daring dapat berdampak positif disaat situasi pandemi seperti ini,
35	Negatif	guru lebih mengutamakan tugas, dan interaksi ke anak kurang (negatif)
36	Negatif	Kurang efektif karna guru hanya memberikan tugas tanpa berinteraksi dengan murid
37	Negatif	Harusnya seimbang antara dosen dengan mahasiswa dari segi tugas nya jangan lebih membebankan dari yang seharusnya hanya karena belajar dari rumah
38	Positif	Belajar dari rumah sangat baik dikarenakan orangtua menjadi lebih mudah dalam mengawasi perkembangan anak namun saya rasa tidak seimbang antara materi yang disampaikan dengan tugas yang diberikan
39	Negatif	Cara belajar dari rumah kurang efektif untuk dijalani karena guru hanya memberikan tugas
40	Negatif	Menurut saya anak" sudah mulai jenuh dengan adanya PJJ yg berkepanjangan. Mereka harus mengerjakan tugas" tanpa adanya oenjelasan secara langsung tentang materi oembelajaran, sehingga anak" harus mengembangkan sendiri materi yg d ajarkan saat ini

B. Perancangan Preprocessing Data

Sub bab ini menjelaskan tahap pembuatan preprocessing data yang akan digunakan pada sistem yang berfungsi untuk melakukan pengolahan data berita. Tahapan ini mempunyai 3 proses yaitu *Case Folding*, *Cleansing* dan *Stemming*.

1. Data Masukan Berita

Tahap ini merupakan tahap awal dimana setiap data (berupa berita) yang dikumpulkan akan digunakan pada penelitian ini. Contoh dari beberapa *headline* data masukan yang digunakan akan dijelaskan pada tabel dibawah.

TABEL IV
DATA MASUKAN BERITA

No	Sumber	Waktu	Headline
1.	Kompas	26/10/2020	Anak Stres Belajar Daring? Ini Saran Psikolog untuk Mencegahnya
2.	Kompas	15/09/2020	Sulit Diajari Belajar Online, Bocah SD Dipukuli Sapu hingga Meninggal
3.	Kompas	15/12/2020	Cara Mengatasi Mata Lelah pada Anak Selama Belajar Daring
4.	Kompas	20/10/2020	Siswi SMA Bunuh Diri karena Beban Tugas Daring, Dinas Pendidikan Evaluasi Sistem Belajar Online
5.	CNNIndonesia	30/03/2020	7 Cara Membantu Anak Fokus saat Belajar di Rumah
6.	CNNIndonesia	23/07/2020	Proses Belajar Jarak Jauh Anak Masih Diselimuti Kendala
7.	CNNIndonesia	27/08/2020	Nadiem Akan Beri Kuota Internet Siswa 35 GB, Mahasiswa 50 GB
8.	CNNIndonesia	18/03/2020	Tak Semua Murid Punya Gadget, Belajar dari Rumah Terkendala

2. Case Folding

Tahap *case folding* bertujuan untuk menghilangkan seluruh huruf kapital di dalam berita menjadi huruf kecil. Berikut ini contoh dari tahap *case folding* yang dijabarkan pada table dibawah ini.

TABEL V
HASIL PROSES CASE FOLDING

No	Input	Output
1.	Anak Stres Belajar Daring? Ini Saran Psikolog untuk Mencegahnya	anak stres belajar daring? ini saran psikolog untuk mencegahnya
2.	Sulit Diajari Belajar Online, Bocah SD Dipukuli Sapu hingga Meninggal	sulit diajari belajar online, bocah sd dipukuli sapu hingga meninggal
3.	Cara Mengatasi Mata Lelah pada Anak Selama Belajar Daring	cara mengatasi mata lelah pada anak selama belajar daring
4.	Siswi SMA Bunuh Diri karena Beban Tugas Daring, Dinas Pendidikan Evaluasi Sistem Belajar Online	siswi sma bunuh diri karena beban tugas daring, dinas pendidikan evaluasi sistem belajar online
5.	7 Cara Membantu Anak Fokus saat Belajar diRumah	7 cara membantu anak fokus saat belajar dirumah
6.	Proses Belajar Jarak Jauh Anak Masih Diselimuti Kendala	proses belajar jarak jauh anak masih diselimuti kendala
7.	Nadiem Akan Beri Kuota Internet Siswa 35 GB, Mahasiswa 50 GB	nadiem akan beri kuota internet siswa 35 gb, mahasiswa 50 gb
8.	Tak Semua Murid Punya Gadget, Belajar dari Rumah Terkendala	tak semua murid punya gadget, belajar dari rumah terkendala

3. Cleansing

Pada tahap *cleansing* dilakukan penghapusan karakter atau elemen yang tidak diperlukan dalam penelitian analisis sentimen ini. Beberapa karakter dan elemen yang dihilangkan adalah tanda baca seperti (,), (?), angka (1) dan link ([//www.kompas.com/sains/read/](http://www.kompas.com/sains/read/)). Tahap ini dilakukan dengan tujuan menghilangkan karakter dan elemen yang tidak diperlukan ikut terproses dan menyebabkan kekeliruan. Berikut ini merupakan contoh dari tahap *cleansing*.

TABEL VI
HASIL PROSES CLEANSING

No	Input	Output
1.	anak stres belajar daring? ini saran psikolog untuk mencegahnya	anak stres belajar daring ini saran psikolog untuk mencegahnya
2.	sulit diajari belajar online, bocah sd dipukuli sapu hingga meninggal	sulit diajari belajar online bocah sd dipukuli sapu hingga meninggal
3.	cara mengatasi mata lelah pada anak selama belajar daring	cara mengatasi mata lelah pada anak selama belajar daring
4.	siswi sma bunuh diri karena beban tugas daring, dinas pendidikan evaluasi sistem belajar online	siswi sma bunuh diri karena beban tugas daring, dinas pendidikan evaluasi sistem belajar online
5.	7 cara membantu anak fokus saat belajar dirumah	cara membantu anak fokus saat belajar dirumah
6.	proses belajar jarak jauh anak masih diselimuti	proses belajar jarak jauh anak masih diselimuti

No	Input	Output
	kendala	kendala
7.	nadiem akan beri kuota internet siswa 35 gb, mahasiswa 50 gb	nadiem akan beri kuota internet siswa gb mahasiswa gb
8.	tak semua murid punya gadget, belajar dari rumah terkendala	tak semua murid punya gadget belajar dari rumah terkendala

4. *Stemming*

Tahap *stemming* merupakan tahapan untuk merubah kata yang mempunyai imbuhan menjadi kata dasar dengan menggunakan *library* Python PySastrawi. Berikut ini contoh dari tahapan *stemming*.

TABEL VII

CONTOH PROSES STEMMING

No	Input	Output
1.	anak stres belajar daring ini saran psikolog untuk mencegahnya	anak stres ajar daring ini saran psikolog untuk cegah
2.	sulit diajari belajar online bocah sd dipukuli sapu hingga meninggal	sulit ajar ajar online bocah sd pukul sapu hingga tinggal
3.	cara mengatasi mata lelah pada anak selama belajar daring	cara atas mata lelah pada anak lama ajar daring
4.	siswi sma bunuh diri karena beban tugas daring, dinas pendidikan evaluasi sistem belajar online	siswi sma bunuh diri karena beban tugas daring dinas didik evaluasi sistem ajar online
5.	cara membantu anak fokus saat belajar dirumah	cara bantu anak fokus saat ajar rumah
6.	proses belajar jarak jauh anak masih diselimuti kendala	proses ajar jarak jauh anak masih limut kendala
7.	nadiem akan beri kuota internet siswa gb mahasiswa gb	nadiem akan beri kuota internet siswa gb mahasiswa gb
8.	tak semua murid punya gadget belajar dari rumah terkendala	tak semua murid punya gadget ajar dari rumah kendala

C. Pengambilan *Dataset* dan *Word Frequency*

Dataset yang telah diproses sebelumnya akan memasuki tahap dimana *dataset* tersebut akan dipanggil ke dalam program sehingga dapat diketahui jumlah *dataset* yang akan digunakan. selanjutnya melakukan perhitungan *word frequency* untuk mengetahui kata-kata yang muncul, setelah melakukan proses *word frequency*, kata-kata diubah menjadi bentuk matriks dengan menggunakan *count vectorizer*. Proses tersebut terdapat pada Gambar.3 dan Gambar.4

```
dataset = 'dataset/clean'

berita_train = load_files(dataset, shuffle=True)

len(berita_train.data)

200
```

Gambar 3. Proses Pengambilan *Dataset*

Setelah proses pemanggilan *dataset* dilakukan Gambar 3 dapat diketahui jumlah *dataset* sebanyak 200.

```
berita_vec = CountVectorizer(min_df=1, tokenizer=nlTK.word_tokenize)
berita_counts = berita_vec.fit_transform(berita_train.data)

print(berita_vec.vocabulary_.get('rumah'))

5210

berita_counts.shape

(200, 6969)
```

Gambar 4. Proses *Word Frequency* dan *CountVectorizer*

Sebagai contoh dari proses *word frequency* peneliti melakukan perhitungan berapa banyak kata “rumah” yang ada pada 200 *dataset* dan diketahui hasil sebanyak 5210 kata “rumah” yang terdapat dari 200 *dataset* tersebut. Untuk *countvectorizer* dari 200 terdapat 6969 kata yang tersedia.

D. Ekstraksi *Dataset*

Setelah dilakukannya tahap *word frequency*, untuk pencarian kata, selanjutnya dilakukan penambahan bobot kata dengan teknik *Term Frequency-Inverse Document Frequency*. Dilakukannya *TF-IDF* untuk memberi bobot pada kata di dalam dokumen dengan menghitung frekuensi kemunculan term dalam dokumen. Hal ini dititik beratkan dengan semakin seringnya pengulangan kata dalam dokumen muncul di dalam berita. Fungsi dari *TF-IDF* terlihat pada Gambar.5

```
tfidf_transformer = TfidfTransformer()
berita_tfidf = tfidf_transformer.fit_transform(berita_counts)
berita_tfidf.toarray()
```

Gambar.5 Fungsi *TF-IDF*

Setelah fungsi *TF-IDF* dilakukan maka pada Gambar 6 menunjukkan hasil keluaran dari proses fungsi *TF-IDF* berupa array. Array merupakan bobot dari kata-kata yang muncul dalam *dataset* semakin banyak kata yang ditampilkan, semakin kecil bobot kata dari kata tersebut. Kemudian array akan digunakan untuk dijadikan acuan bobot pada data baru yang akan dimasukkan kedalam *machine*.

```
array([[0.0, 0.0, 0.0, ..., 0.0, 0.0, ...],
       [0.0, 0.0, 0.0, ..., 0.0, 0.0, ...],
       [0.0, 0.0, 0.0, ..., 0.0, 0.0, ...],
       ...,
       [0.0, 0.0, 0.0, ..., 0.0, 0.0, ...],
       [0.0026272, 0.0, ..., 0.0, 0.0, ...],
       [0.0, 0.0, 0.0, ..., 0.0, 0.0, ...],
       [0.0, ...]])
```

Gambar 6. Hasil Perhitungan *TF-IDF*

E. Implementasi Algoritma

Setelah melakukan proses pembobotan, tahap selanjutnya data terlebih dahulu di bagi 80:20 dengan 80% *training* dan 20% data *validation*, selanjutnya dilakukan

fitting dengan algoritma *Naïve Bayes Classifier*. Prediksi data *validation* yang dilakukan untuk mendapatkan hasil yang akurat. Implementasi algoritma ditunjukkan pada Gambar 7.

```
docs_train, docs_test, y_train, y_test = train_test_split(
    berita_tfidf, berita_train.targat, test_size = 0.2, random_state = 12)

clf = MultinomialNB().fit(docs_train, y_train)

y_pred = clf.predict(docs_test)
accuracy_score(y_test, y_pred)

0.65
```

Gambar 7. Implementasi Algoritma Naïve Bayes Classifier

Setelah algoritma *Naïve Bayes Classifier* diimplementasikan kedalam model maka dapat dilihat pada Gambar 4.6 bahwa nilai akurasi yang didapatkan sebesar 65% yang didapat dari 200 *dataset* dimana 100 data positif dan 100 data negatif dengan perbandingan *training* dan *testing* 80:20.

F. Evaluasi

Setelah dilakukannya proses implementasi algoritma tahap selanjutnya ialah mengevaluasi model yang telah dibuat. Dilakukan evaluasi terhadap model guna memastikan model tersebut sudah memenuhi harapan dalam menjalankan tugas. evaluasi mempunyai beberapa cara dalam melakukan tugas klasifikasi seperti melakukan pencarian nilai *precision*, *recall* dan *f1-score*. Tahap evaluasi ditunjukkan pada Gambar 8.

```
print(classification_report(y_test, y_pred))
```

Gambar 8. Tahap Evaluasi

G. Hasil Percobaan

Setelah dilakukan proses *training* dan *validation* data pada model algoritma *Naïve Bayes Classifier* seperti yang ditunjukkan pada Gambar 7 sebesar 65% kemudian tahap evaluasi seperti Tabel 8.

TABEL VIII
ARRAY PREDICTION DAN TESTING

n=40 (Data Testing)	Testing Positif (1)	Testing Negatif (0)
Prediction Positif (1)	TP=[1.1] 15	FP=[1.0] 10
Prediction Negatif (0)	FN=[0.1] 4	TN=[0.0] 11

Hasil evaluasi pada Gambar.8, skor *precision* negatif 73%, *recall* sebesar 52% dan *f1-score* sebesar 61% sedangkan untuk skor *precision* positif 60%, *recall* sebesar 79% dan *f1-score* sebesar 68%. Untuk lebih jelasnya dapat dilihat pada Gambar.9.

	precision	recall	f1-score	support
0	0.73	0.52	0.61	21
1	0.60	0.79	0.68	19
accuracy			0.65	40
macro avg	0.67	0.66	0.65	40
weighted avg	0.67	0.65	0.64	40

Gambar 9. Hasil Evaluasi

Berdasarkan hasil evaluasi pada Gambar 9 nilai tersebut didapat dengan penghitungan menggunakan persamaan yang terdapat pada Sub Bab tentang Evaluasi Sistem dimana array yang terbentuk dapat terlihat pada Gambar.10.

```
y_pred
array([1, 1, 0, 1, 1, 1, 1, 1, 1, 0, 1, 0, 1, 0, 0, 1, 1, 1, 1, 0, 0, 0,
       1, 1, 1, 1, 1, 0, 0, 1, 1, 0, 1, 1, 0, 0, 0, 1, 0])

y_test
array([1, 1, 0, 1, 0, 1, 0, 1, 1, 1, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0,
       1, 1, 0, 1, 1, 1, 1, 1, 0, 1, 0, 0, 0, 0, 0, 0, 1, 1])
```

Gambar.10 Array Prediction dan Testing

sedangkan perhitungannya bisa dilihat pada Tabel 9.

TABEL IX

PERHITUNGAN EVALUASI

Kategori	Perhitungan		
	Precision $P = \frac{TP}{(TP + FP)}$	Recall $R = \frac{TP}{(TP + FN)}$	F1-Score $F1 = 2x \frac{(Precision \times Recall)}{(Precision + Recall)}$
Positif	$\frac{15}{(15 + 10)} = 0.60$	$\frac{15}{(15 + 4)} = 0.79$	$2x \frac{(0.60 \times 0.79)}{(0.60 + 0.79)} = 0.68$
Negatif	$\frac{11}{(11 + 4)} = 0.73$	$\frac{11}{(11 + 10)} = 0.53$	$2x \frac{(0.73 \times 0.53)}{(0.73 + 0.53)} = 0.61$
Accuracy	$A = \frac{(TP + TN)}{(TP + FP + TN + FN)} = \frac{(15 + 11)}{(15 + 10 + 11 + 4)} = 0.65$		

H. Hasil Pengujian

Hasil pengujian merupakan tahap dimana model yang sudah dievaluasi akan melakukan prediksi terhadap data teks yang dituliskan oleh orangtua murid pada googleform. Data teks dari *googleform* terlebih dahulu diunduh dalam bentuk Microsoft Excel dengan format .csv dan data melalui tahap penamaan ulang pada kolom kuisioner dikarenakan nama yang terlalu panjang. Proses tersebut terlihat pada Gambar.11.

```
# Do = #.rename(columns=
    "kolom pengisian atau opsi jawaban terdapat metode pembelajaran daring (kita dan apa? apa?) dapat berupa ya yang positif
    "Ya/otk/ya/sari/ya metode pembelajaran daring sangat baik membantu saya yang positif atau negatif? kepada anak ku
    "terapan atau kegiatan lain?/ya terdapat metode pembelajaran di masa pandemi Covid-19? "terapan", inplace = False)

#_kolom = #_df[["ya_yg", "ya_pnngkat", "terapan"]]

#_kolom.to_csv('pendaf.csv')
```

Gambar.11 Proses Rename dan Pengunduhan Data Kuisioner

Setelah data dilakukan penamaan ulang dan disimpan dengan format .csv kemudian data dipanggil ke dalam model dengan menggunakan fungsi yang sama seperti ditunjukkan pada Gambar.12.

```
df = pd.read_csv('pendapat.csv')
df.columns
Index(['Unnamed: 0', 'pos_neg', 'pend_orangtua', 'harapan'], dtype='object')
df = df[['pos_neg', 'pend_orangtua', 'harapan']]
```

Gambar.12 Pemanggilan Data Kuisisioner

Pada Gambar.12 dapat dilihat proses pemanggilan data kemudian dilanjutkan dengan tahap melihat kolom dari data kuisisioner yang digunakan, kolom tersebut ialah 'pos_neg' atau positif negatif, 'pend_orangtua' atau pendapat orangtua, 'harapan' atau keinginan dari orangtua terkait pembelajaran online. Data kuisisioner ada pada Gambar.13.

```
df.head()
```

	pos_neg	pend_orangtua	harapan
0	Negatif	Tetap semangat dalam belajar	Kegiatan belajar mengajar kembali kepada tatap...
1	Positif	Kendala di metode pengajaran tidak adanya kepe...	Kegiatan belajar mengajar kembali kepada tatap...
2	Positif	Kendalanya di jaringan yang tidak lancar	Kegiatan belajar mengajar kembali kepada tatap...
3	Positif	Lama belajarnya pun anak menjadi jenuh belajar da...	Kegiatan belajar mengajar kembali kepada tatap...
4	Negatif	Belajar di rumah kurang memadai karena anak banyak...	Kegiatan belajar mengajar kembali kepada tatap...

Gambar.13 Kolom Data Kuisisioner

Data kuisisioner yang tersedia pada Gambar.13 dapat diketahui deskripsi data dengan menggunakan fungsi *describe* seperti Gambar 14.

```
df.describe()
```

	pos_neg	pend_orangtua	harapan
count	40	40	40
unique	2	39	2
top	Negatif	Menurut saya anak sudah mulai jenuh dengan ad...	Kegiatan belajar mengajar kembali kepada tatap...
freq	27	1	20

Gambar.14 Deskripsi Data Kuisisioner

Deskripsi dari data kuisisioner pada kolom tabel terdapat tiga kategori pertanyaan yaitu 'pos_neg', 'pend_orangtua' dan 'harapan'. Kemudian pada baris tabel terdapat 'count' yaitu data masuk, 'unique' pertanyaan unik, 'top' jawaban terbaik yang ada dan 'freq' frekuensi jawaban yang 'pos_neg', 'pend_orangtua' dan 'harapan' terdapat 40 jumlah jawaban yang masuk. 'unique' terdapat 2 jawaban tersedia, 'pend_orangtua' terdapat 39 jawaban yang masuk dan 'harapan' terdapat 2 jawaban tersedia. 'top' menampilkan jawaban terbaik dari jumlah 40 jawaban yang ada dengan 'pos_neg' yang menunjukkan jawaban negatif, 'pend_orangtua' Menurut saya anak sudah mulai jenuh dengan adanya PJJ yg berkepanjangan. Mereka harus mengerjakan tugas tanpa adanya penjelasan secara langsung tentang materi pembelajaran, sehingga anak harus mengembangkan sendiri materi yg diajarkan saat ini dan 'harapan' Kegiatan belajar mengajar kembali kepada

tatap muka namun tetap dengan protokol kesehatan dan skema pembelajaran yang baik (Misal pembagian hari masuk sekolah). Sedangkan untuk info dari data kuisisioner ditunjukkan pada Gambar 15.

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 40 entries, 0 to 39
Data columns (total 3 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   pos_neg          40 non-null     object
1   pend_orangtua    40 non-null     object
2   harapan          40 non-null     object
dtypes: object(3)
memory usage: 1.1+ KB
```

Gambar 15. Informasi Data Kuisisioner

Dari informasi yang ada pada data kuisisioner dapat diketahui jumlah data kuisisioner yang masuk terdapat 40 jawaban, memiliki 3 kolom dengan tipe data objek yang belum terdefinisi dan dari setiap butir pertanyaan tidak ada jawaban yang kosong atau 'non-null'. Data kuisisioner tersebut akan diproses oleh sistem untuk mengetahui banyaknya jawaban positif dan negatif dari pendapat orangtua. Pada tahap ini data kuisisioner akan dimasukkan kedalam machine learning untuk selanjutnya akan dikategorikan sebagai pendapat positif atau negatif menggunakan library pandas. berdasarkan data *training* yang telah ada dan dengan tingkat keakuratan berdasarkan data *testing* yang telah dievaluasi sebelumnya. Hasil dari proses ini akan disimpan kedalam excel dengan format .csv. Proses tersebut dapat dilihat pada Gambar.16.

```
final = pd.DataFrame({'category': y_pred})
final.to_csv('final.csv', index=False)
```

Gambar.16 Proses Data Kuisisioner

Hasil dari data pendapat orangtua yang telah disimpan dengan format .csv dengan machine learning akan divisualisasikan kedalam bentuk diagram seperti yang ditunjukkan dengan Gambar.17 dan keluaran diagram dapat dilihat pada Gambar.18.

```
data_1 = pd.read_csv("final.csv")
sizes = data_1['category'].value_counts(sort = True)
df['category'] = data_1['category']
colors = ["blue", "magenta"]
rcParams['figure.figsize'] = 5,5
labels = ["Positif", "Negatif"]

# Plot
plt.pie(sizes, labels=labels, colors=colors,
        autopct='%1.1f%%', shadow=True, startangle=270,)
plt.title('Persentase Perbandingan ')
plt.show()
```

Gambar.17. Proses Penampilan Hasil Data ke Bentuk Diagram

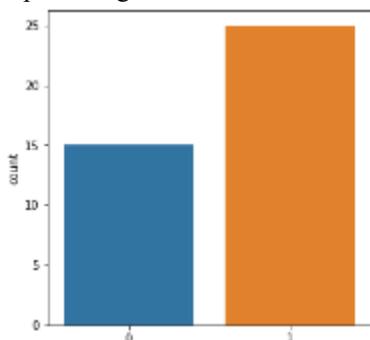
Berdasarkan Gambar.18 dapat dilihat warna biru tervisualisasi sebagai jawaban positif dan warna ungu

tervisualisasi sebagai jawaban negatif dengan persentase 62.5% untuk jawaban positif dan 37.5% untuk jawaban negatif dari jumlah data yang dimasukkan sebanyak 40 jawaban dengan jumlah jawaban positif sebanyak 25 jawaban dan jawaban negatif sebanyak 15 jawaban seperti yang terlihat pada Gambar.19.



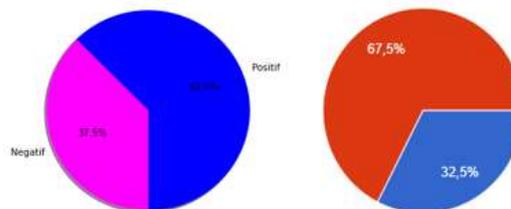
Gambar.18 Persentase Perbandingan Pendapat Orangtua

Data 40 masukan pada tahap pengujian ini merupakan tanggapan dari orangtua terkait pembelajaran daring seperti yang telah ditampilkan pada Gambar.19 dengan kategori “Pendapat Orangtua”.



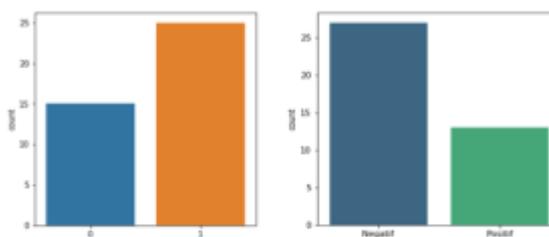
Gambar 19. Hasil Uji Pendapat Orangtua

Dari data kuisiner pendapat orangtua dapat dilakukan perbandingan untuk positif dan negatif. Berdasarkan jumlah jawaban pendapat orangtua pada googleform lebih banyak yang menyatakan bahwa pembelajaran daring dari rumah merupakan suatu yang negatif namun berbeda dengan hasil ketika data tersebut diproses dengan menggunakan model. Hasil pengujian menunjukkan bahwa pendapat orangtua lebih banyak yang bernilai positif. Perbandingan tersebut dapat dilihat pada Gambar.20 dimana sebelah kiri hasil pengujian dengan warna biru sebagai positif dan warna ungu sebagai negatif sedangkan sebelah kanan hasil survei kuisisioner dengan warna biru sebagai positif dan warna merah sebagai negatif.



Gambar.20 Perbandingan Hasil Uji dan Kuisisioner Pendapat Orangtua

Grafik perbandingan hasil pengujian dan sebelum pengujian dapat dilihat pada Gambar.21 dimana sebelah kiri hasil pengujian dengan warna kuning menunjukkan positif bernilai 1 dan warna biru menunjukkan negatif bernilai 0 sedangkan sebelah kanan hasil survey kuisisioner dengan warna hijau menyatakan positif dan warna biru menyatakan negatif.



Gambar.21 Grafik Perbandingan Pendapat Orangtua Sesudah dan Sebelum Diuji.

IV. KESIMPULAN

Hasil penelitian yang telah dilakukan dapat disimpulkan bahwa algoritma *Naïve bayes Classifier* (NBC) dapat memprediksi serta melakukan perbandingan analisis sentimen terhadap pendapat orangtua terhadap pembelajaran daring dari rumah dengan hasil yang cukup baik dikarenakan lebih banyak jawaban yang divisualisasikan sebagai jawaban positif dari pada negatif dimana perbandingannya 62.5% untuk jawaban positif dan 37.5% untuk jawaban negative dari 40 data kuisisioner yang dimasukkan kedalam model yang telah dibuat dan dilakukan tahap evaluasi. Nilai akurasi yang didapatkan pada penelitian ini sebesar 65%. Nilai akurasi ini didapatkan dari 200 *dataset* berita dimana 100 *dataset* positif dan 100 *dataset* negatif. Jadi nilai akurasi dari model untuk melakukan analisis sentimen bergantung pada kualitas penentuan kategori berita positif atau negatif dari *dataset* yang dikumpulkan. Data kuisisioner yang positif belum pasti terbaca positif pada model begitupun dengan sebaliknya. Hal ini dipengaruhi oleh kualitas dan ketepatan penentuan kategori positif dan negatif isi dari berita.

REFERENSI

- [1] A. Nafrin, "EDUKATIF : JURNAL ILMU PENDIDIKAN Perkembangan Pendidikan Indonesia di Masa Pandemi Covid-19 Abstrak," vol. 3, no. 2, pp. 456–462, 2021.
- [2] M. Siahaan *et al.*, "Halaman: 1-3 Terakreditasi Peringkat 5 (SINTA 5) sesuai SK RISTEKDIKTI Nomor," *Ed. Khusus*, vol. 1, no. 1, pp. 1410–9794, 2020.
- [3] F. N. Hasan and A. Febriandirza, "Perancangan Data Warehouse Untuk Data Penelitian di Perguruan Tinggi Menggunakan Pendekatan Nine Steps Methodology," *Pseudocode*, vol. VIII, no. 1, pp. 49–57, 2021.
- [4] W. Mahmudi and R. Fernandes, "Adaptasi Siswa Terhadap Pola Pembelajaran Daring pada Masa Pandemi Covid-19 di SMAN 1 Solok," *J. Perspekt.*, vol. 4, no. 3, pp. 395–407, 2021.
- [5] F. N. Hasan, "Pandemi Covid-19 Serta Pengaruhnya Terhadap Aktivitas Ibadah Dan Nilai-Nilai Al Islam Kemuhmadiyah," *Fikroh J. Pemikir. dan Pendidik. Islam*, vol. 14, no. 2, pp. 114–128, 2021.
- [6] Y. I. dan C. A. R. Lindawati, "Adaptasi Guru Dalam Implementasi Pembelajaran Daring Di Era Pandemi Covid-19," *Pros. Semin. Nas. Pendidik. FKIP*, vol. 3, no. 2252, pp. 58–66, 2019.
- [7] E. Susanto and A. P. Deapalupi, "Analisis Dampak Covid-19 terhadap Implementasi Study From Home (SFH) di Tingkat Sekolah Dasar," *J. Pendidik. Ris. dan Konseptual*, vol. 4, no. 4, p. 536, 2020.
- [8] S. Winarti, B. Taib, B. Alhadad, and F. Achmad, "Analisis Dampak Covid-19 Dalam Proses Pembelajaran Daring Pada Kelas B4 Di Paud Telkom Ternate," *J. Ilm. Cahaya Paud*, vol. 3, no. 1, pp. 150–159, 2021.
- [9] A. N. Puthree, D. W. Rahayu, M. Ibrahim, and M. S. Djazilan, "Analisis Faktor Penyebab Rendahnya Motivasi Belajar Siswa Sekolah Dasar selama Pembelajaran Daring," *J. Basicedu*, vol. 5, no. 5, pp. 3101–3108, 2021.
- [10] G. A. Buntoro, "Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter," *Integer J.*, vol. 2, no. 1, pp. 32–41, 2017.
- [11] I. Rozi, S. Pramono, and E. Dahlan, "Implementasi Opinion Mining (Analisis Sentimen) Untuk Ekstraksi Data Opini Publik Pada Perguruan Tinggi," *J. EECCIS*, vol. 6, no. 1, pp. 37–43, 2012.
- [12] M. A. Amrustian, W. Widayat, and A. M. Wirawan, "Analisis Sentimen Evaluasi Terhadap Pengajaran Dosen di Perguruan Tinggi Menggunakan Metode LSTM," vol. 6, pp. 535–541, 2022.
- [13] W. A. Prabowo and C. Wiguna, "Sistem Informasi UMKM Bengkel Berbasis Web Menggunakan Metode SCRUM," *J. Media Inform. Budidarma*, vol. 5, no. 1, p. 149, 2021.
- [14] F. N. Hasan and M. Dwijayanti, "Analisis Sentimen Ulasan Pelanggan Terhadap Layanan Grab Indonesia Menggunakan Multinomial Naïve Bayes Classifier," *J. Linguist. Komputasional*, vol. 4, no. 2, pp. 52–58, 2021.
- [15] N. Ramdhani and R. H. Al-Fadillah, "Analisis Sentimen Pengguna Twitter Terhadap Belajar Daring Selama Pandemi Covid-19 Dengan Deep Learning," vol. 7, no. 2, p. 2021, 2021.
- [16] A. N. Indrainsi, I. Ernawati, and A. Zaidah, "ANALISIS SENTIMEN TERHADAP PEMBELAJARAN DARING DI INDONESIA MENGGUNAKAN SUPPORT VECTOR MACHINE (SVM)."
- [17] S. Y. Nursyi'ah, A. Erfina, and C. Warman, "Analisis Sentimen Pembelajaran Daring Pada Masa Pandemi Covid-19 Di Twitter Menggunakan Algoritma Naïve Bayes," *J. Media Inform. Budidarma*, pp. 117–123, 2021.