ANALISIS SENTIMEN DAN KLASIFIKASI RESPON PENGGUNA MEDIA SOSIAL FACEBOOK TERHADAP PENDIDIKAN DARING PADA MASA PANDEMI COVID-19 DI INDONESIA

Heny Pratiwi[1)], Basrie[2)], dan **Artha Rizal[3)]**

[1]Sistem Informasi, STMIK Widya Cipta Dharma Samarinda
[2]Manajemen Informatika, STMIK Widya Cipta Dharma Samarinda
[3]Teknik Informatika, STMIK Widya Cipta Dharma Samarinda
[1,2,3]Jl. M. Yamin No.25, Samarinda, 75123
E-mail : prodisi@wicida.ac.id[1)],basrie@wicida.ac.id[2)],artharizal26@gmail.com[3)]

ABSTRAK

Pada saat ini media sosial telah menjadi alat komunikasi yang sangat popular dikalangan pengguna internet di Indonesia. Salah satu media sosial tersebut yakni Facebook dengan jumlah opini yang besar dan di dalamnya terdapat informasi yang sangat berharga sebagai alat penentu kebijakan dan ini bisa di lakukan dengan menggunakan text mining, Sistem klasifikasi komentar diharapkan dapat membantu untuk mengetahui respon positif dan negatif dari pengguna facebook yang memberikan komentarnya. Sentiment analysis digunakan untuk mengetahui sikap seseorang dalam konteks dokumen. Penelitian ini dilakukan berdasarkan komentar pengguna media sosial Facebook di Indonesia. Metode pengumpulan data yang digunakan yaitu dengan web crawling, selanjutnya data melalui tahapan preprocessing yang terdiri dari case folding, tokenizing, stopword removal, stemming. Kemudian menggunakan Naïve Bayes sebagai metode klasifikasinya dan supporting tools menggunakan RapidMiner. Adapun hasil akhir dari penelitian ini yakni berupa kesimpulan mengenai efektif atau tidaknya pendidikan daring pada masa pandemi Covid-19 di Indonesia berdasarkan respon pengguna media sosial facebook.

Kata Kunci: *Analisis Sentimen, Klasifikasi, Pendidikan Daring, COVID-19.*

1. PENDAHULUAN

Analisis sentimen adalah proses menentukan sentimen dan mengelompokkan polaritas teks dalam dokumen atau kalimat sehingga kategori dapat ditentukan sebagai sentimen positif, negatif, atau netral. Jejaring sosial, seperti Facebook, umumnya digunakan dalam analisis sentimen untuk menentukan persepsi publik. Dalam analisis sentimen, penambangan data dilakukan untuk menganalisis, memproses, dan mengekstrak data tekstual dalam suatu entitas.

World Health Organization (WHO) mendeklarasikan virus COVID-19 sebagai pandemi *global* pada 11 Maret 2020. WHO melaporkan lebih 52 juta orang terkonfirmasi positif Covid-19 dan 1,2 juta orang meninggal dunia pada minggu kedua bulan November 2020. Strategi melawan pandemi dengan pembatasan sosial memaksa semua institusi pendidikan menerapkan pembelajaran daring. Kementerian Pendidikan dan Kebudayaan Republik Indonesia Nadiem Makarim mempertegas kebijakan pembelajaran daring dalam masa pandemi dengan mengeluarkan kebijakan Belajar Dari Rumah (BDR). Kebijakan ini mewajibkan penggunakan jaringan internet melalui perantara *smartphone*, *gadget*, komputer, dan aplikasi sebagai pengganti tatap muka.

Penelitian ini diharapkan dapat menjelaskan bagaimana melakukan analisis sentimen dan klasifikasi respon pengguna media sosial Facebook terhadap pendidikan daring pada masa pandemi serta mengetahui seberapa efektif pendidikan daring pada masa pandemi berdasarkan komentar atau opini pengguna media sosial Facebook di Indonesia.

Facebook merupakan jejaring sosial atau sosial media yang memungkinkan para pengguna dapat menambahkan profil dengan foto, kontak, ataupun informasi. Penggunanya dapat bergabung dalam komunitas untuk melakukan koneksi dan berinteraksi dengan pengguna lainnya. Fitur yang ditawarkan facebook sebagai situs jejaring sosial atau media sosial membuat banyak orang menggunakannya (Nasrullah, 2016).

Dalam pengumpulan data pada penelitian ini menggunakan *Web Crawler*, *Web Crawler* adalah sebuah perangkat lunak yang digunakan untuk menjelajah serta mengumpulkan halaman-halaman *web* yang selanjutnya diindeks oleh mesin pencari. Sering juga disebut dengan *web spider* atau *web robot* yang menjelajahi halaman *website* dengan menggunakan *web browser* (Gatial E, 2013).

Kerangka kerja *Naïve Bayes* klasifikasi terdapat 2 (dua) proses yaitu model pelatihan dan model pengujian. Pada model pelatihan terdapat suatu kumpulan data yang memiliki kelas yang telah ditetapkan. Kumpulan data latih digunakan untuk membentuk model klasifikasi. Pada model pengujian, model klasifikasi yang sudah terbentuk diuji dengan data yang lainnya untuk mengetahui akurasi dari model

klasifikasi tersebut. Model klasifikasi yang dihasilkan dapat digunakan untuk menentukan kelas dari data yang baru (Aida Indriani, 2014).

Keuntungan penggunaan *Naïve Bayes* adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (*Training Data*) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. *Naïve Bayes* sering bekerja jauh lebih baik dalam kebanyakan situasi dunia nyata yang kompleks dari apa yang diharapkan. Persamaan dari teorema Bayes adalah sebagai berikut:

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)}$$

Dimana:

X = data dengan *class* yang belum diketahui

H = hipotesis data menggunakan suatu *class* spesifik

P(H|X) = probabilitas hipotesis H berdasarkan kondisi X (parteriori probabilitas)

P(H) = probabilitas hipotesis H (prior probabilitas)

P(X|H) = probabilitas X berdasarkan kondisi pada hipotesis H

P(X) = probabilitas H

Confusion matrix merupakan alat yang berguna untuk menganalisis seberapa baik *classifier* mengenali *tuple* dari kelas yang berbeda. TP dan TN memberikan informasi ketika *classifier* benar, sedangkan FP dan FN memberikan informasi ketika *classifier* salah. Terdapat 4 (empat) istilah sebagai representasi hasil proses klasifikasi pada *confusion matrix*. Keempat istilah tersebut adalah *True Positive* (TP), *True Negative* (TN), *False Positive* (FP) Dan *False Negative* (FN). Nilai *True Positive* (TP) dan *True Negative* (TN) adalah hasil klasifikasi yang benar. Nilai *False Positive* (FP) adalah nilai dimana hasilnya diprediksi sebagai *type I error* namun sebenarnya merupakan *type II error* sedangkan *False Negative* (FN) adalah nilai dimana prediksi mengklasifikasikan sebagai *type II error* namun faktanya termasuk dalam klasifikasi *type I error* (Elvianti, 2015).

Perhitungan yang dihasilkan pada Confusion Matrix meliputi:

Nilai *accuracy* dapat diperoleh dengan persamaan:

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

Nilai *precision* dapat diperoleh dengan persamaan:

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

Nilai *recall* dapat diperoleh dengan persamaan:

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

RapidMiner merupakan perangakat lunak yang bersifat terbuka (open source). RapidMiner adalah sebuah *tools* atau alat bantu untuk melakukan analisis terhadap *data mining, text mining* dan analisis prediksi. RapidMiner menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik. RapidMiner memiliki kurang lebih 500 operator *data mining*, termasuk operator untuk *input, output, data preprocessing* dan visualisasi. RapidMiner ditulis dengan bahasa *java* sehingga dapat bekerja di semua sistem operasi. RapidMiner adalah *platform* analisis modern yang meliputi *data mining*, *mechine learning*, analisis prediktif, *text mining* dan analisis bisnis. *Software* ini digunakan untuk mengukur kerja algoritma dan untuk menentukan algoritma terbaik yang akan berguna untuk klasifikasi, prediksi dan teknik lainnya di *data mining*. RapidMiner merupakan *software* yang *user friendly* dan memiliki GUI (*Graphic User Interface*) yang efektif digunakan untuk bekerja dengan mudah. RapidMiner memberikan *mechine learing* dan data prosedur termasuk *loading data* dan transformasi (*Extract, Transform, Load* (ELT), data *preprocessing* dan pemodelan statistik, visualisasi dan analisis prediktif, penyebaran dan evaluasi. (RapidMiner, 2001).

Wordcloud merupakan sebuah sistem yang memunculkan susunan kata sebagai citra visual terkait frekuensi kemunculan kata dalam suatu teks verbal. Visualisasi *Wordcloud* dari teks akan memudahkan pengamat dalam melihat gagasan dan pendirian sang penulis teks sehingga dapat menjadi alat bantu dalam melakukan analisis terhadap sebuah wacana tertulis (McNaught dan Lam, 2014).

Pada penelitian yang dilakukan oleh Abu Salam, dkk (2018), berjudul Analisis Sentimen Data Komentar Sosial Media Facebook Dengan K-Nearest Neighbor (Studi Kasus Pada Akun Jasa Ekspedisi Barang J&T Ekspress Indonesia) digunakan beberapa perbandingan terhadap rasio yang telah ditentukan, Pengujian menggunakan *Confusion matrix* untuk perbandingan data testing sebanyak 101 dengan data training sebanyak 234 yang diambil dari perusahaan yang sudah berlabel kelas positif dan kelas negatif. Menghasilkan nilai rata-rata akurasi sebesar 75.58%.

Sedangkan penelitian yang dilakukan Syahmia Gusrian, dkk (2016) dan Samsir, dkk (2020) metode Naïve Bayes digunakan ke dalam analisis sentiment untuk melakukan penelitian tersebut. Mengahasilkan masing-masing tingkat akurasi rata-rata sebesar 93.7% dan 97.15%.

Berdasarkan dari beberapa jurnal yang telah dijabarkan, maka pada penulisan penelitian ini digunakan metode Naïve Bayes dengan *tools* RapidMiner dalam melakukan analisis sentiment dan klasifikasi respon pengguna media sosial Facebook terhadap pendidikan daring di masa pandemic Covid-19 di Indoneisa.

2. RUANG LINGKUP

Agar penelitian ini tidak meluas lebih dari pembahasan yang dimaksud, maka dibuatlah batasan masalah pada ruang lingkup penelitian sebagai berikut:

1. Media yang digunakan adalah aplikasi *Facebook*.
2. Penelitian ini didasarkan pada pengguna *Facebook* di Indonesia.
3. Klasifikasi hanya mendeteksi respon positif dan negatif.
4. Menggunakan beberapa percobaan pengujian data latih dan data uji.
5. Metode yang digunakan adalah *Naïve Bayes.*
6. Menggunakan komentar atau opini pengguna *Facebook* sebagai sample.
7. Menggunakan beberapa perbandingan rasio, yaitu: 1:9, 3:7, 5:5, 7:3, dan 9:1.

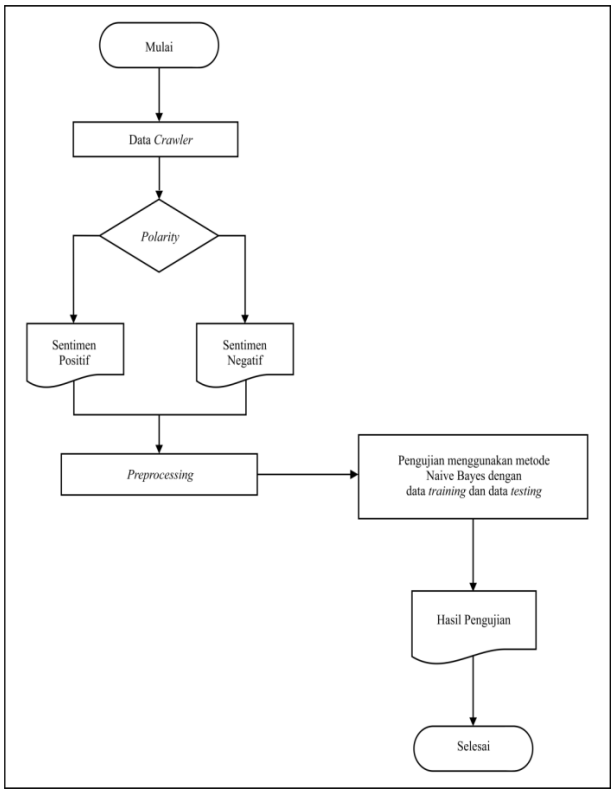
3. METODE PENELITIAN

3.1 Studi Pustaka

Pada metode ini penulis mengambil data melalui buku atau literatur-literatur lain yang ada kaitannya dengan materi penelitian sebagai refrensi yang menunjang dalam penelitian dan penulisan skripsi ini. Literatur-literatur yang digunakan untuk mengetahui dengan jelas gambaran permasalahan yang ada sehingga tidak menyimpang dari ketentuan dan teori yang telah ada dan diakui kebenarannya.

3.2 Alur Penelitian

Dari gambar 3.1 bisa dilihat proses keseluruhan pada implementasi metode penelitian ini.



Gambar 3.1 Alur Penelitian

Langkah pertama adalah mengumpulkan data, data dikumpulkan menggunakan data *crawler*, setelah data terkumpul kemudian dilakukan proses *polarity* dimana data akan terbagi menjadi sentimen positif dan negatif. Selanjutnya tahapan *preprocessing* adalah proses *tokenizing, case folding, filtering, stemming.* Selanjutnya akan memasukan rasio yang telah ditentukan untuk digunakan oleh *data testing* dan *training*. Terakhir, merupakan keluarannya yang berupa perhitungan *confusion matrix* meliputi nilai akurasi, *class recall* dan *class precision* hasil klasifikasi yang menentukan respon positif maupun negatif dari pengguna *facebook* terhadap pendidikan daring dimasa pandemi Covid-19 di Indonesia.

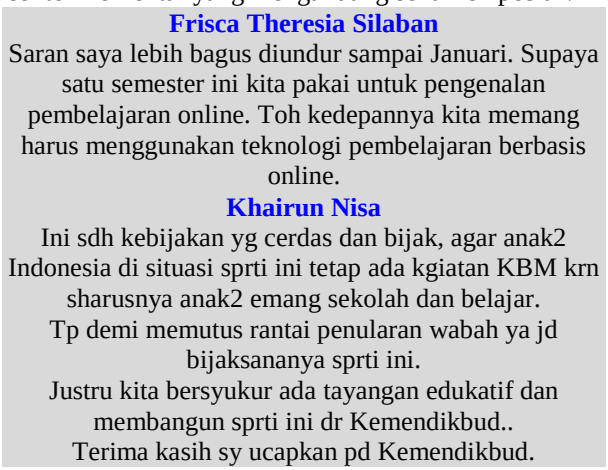
3.3 Data

Pada penelitian ini sumber data yang digunakan adalah komentar berbahasa Indonesia dan berbahasa Inggris yang terdapat di halaman akun *facebook* Kementerian Pendidikan Republik Indonesia. Komentar yang digunakan adalah komentar-komentar yang mengandung unsur respon positif dan negatif pengguna media sosial *facebook* terhadap pendidikan daring dimasa pandemi Covid-19 di Indonesia. Kemudian *data testing* akan diuji dengan *data training* menggunakan algoritma Naïve Bayes.

Pengumpulan data dilakukan dengan memanfaatkan fitur dari *google chrome* yaitu *simple scraper* pada *data crawler* untuk mendapatkan data komentar pada setiap postingan mengenai pendidikan daring di halaman akun Kementerian Pendidikan dan Kebudayaan Republik Indonesia. Pada penelitian ini didapatkan 10 postingan mengenai pendidikan daring yang di mulai pada tanggal 01 Maret 2020 sampai dengan tanggal 08 Juni 2020.

3.4 Analisis Sentimen

Komentar yang telah dipilih kemudian disimpan ke *file excel*, selanjutnya digunakan sebagai *input* pada sistem untuk menentukan sentimen positif dan negatif dari respon pengguna media sosial facebook sehingga akan didapat hasil klasifikasi. Gambar 3.2 adalah contoh komentar yang mengandung sentimen positif.



Gambar 3.2 Komentar Respon Positif

Gambar 3.3 adalah contoh komentar yang mengandung sentimen negatif.

Aditya Heriyanto

Ini gmn kami para mahasiswa yg mau praktek lewat daring mana ngerti, kalian mau buat mahasiswa lulus dalam keadaan buta gk punya keahlian

Handy

Ini mah Mentri mana tau susah nya rakyat yang sekolah online apalagi yg ga punya HP , duit , kuota sama sinyal susah dan sekolah online ga terlalu paham pelajaran nya , yang menting mah ini Mentri senang aja

Purdi Iyan

Saya pk sebagai orang tua boleh masuk sekolah .anak dirumah bukan belajar tp main terus..apalagi ortu yg kewalahan dalam keterbatasan sosial.apalgi ditempat saya beli beras aja susah apa lagi tuk beli kuota

Gambar 3.3 Komentar Respon Negatif

3.5 Tahap *Preprocessing*

Tahap *preprocessing* meliputi tahap *case folding, tokenizing, filtering* dan *stemming. Preprocessing* akan menghapus *link url*, *username*, *like* dan tanda *reply*. *Preprocessing* akan mengubah kata tidak baku atau kata yang disingkat menjadi kata yang baku menggunakan filter *stopword* (Bahasa Indonesia, Bahasa Inggris, Slang dan *Abbreviation*).

3.6 *Data Training* dan *Data Testing*

Pada tahap ini data akan dipilih secara acak, dalam melakukan proses pengecekan akurasi dari *data training* dan *data testing* serta akan melakukan pengujian terhadap rasio yang telah ditentukan pada metode yang digunakan sehingga dari data yang dihasilkan dapat diketahui hasil keakuratan data.

3.7 Hasil Pengujian

Hasil pengujian dari *data training* dan *data testing* akan berupa hasil akurasi dan hasil klasifikasi sentimen positif dan sentimen negatif. Hasil akurasi akan didapatkan dari hasil *Confusion Matrix* berupa akurasi data, *class recall* dan *class precision.* Hasil klasifikasi yaitu berupa jumlah sentimen positif dan sentimen negatif pengguna sosial media facebook terhadap pendidikan daring pada masa pandemi Covid-19 di Indonesia sehingga akan diperoleh kesimpulan mengenai efektif atau tidaknya pendidikan daring pada masa pandemi Covid-19 di Indonesia ini.

4. PEMBAHASAN

Tahap ini merupakan pembahasan proses analisis dan klasifikasi dengan menggunakan *tools* RapidMiner.

4.1 *Data Crawler*

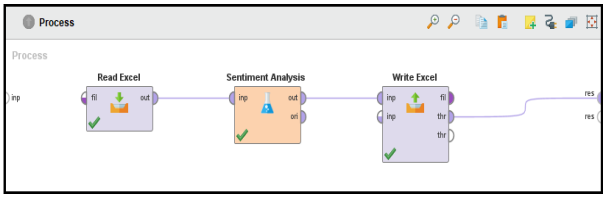
Proses *crawling* dilakukan menggunakan fitur yang terdapat pada *browser google chrome* yaitu *simplescraper*, dengan menggunakan *simplescraper* proses pengumpulan data dilakukan secara otomatis dan selanjutnya seleksi data secara manual untuk menghilangkan *spider trap* pada data komentar yang di ambil.

Gambar 4.1 Hasil *Data Crawling*

Pada Gambar 4.1 menunjukan banyaknya data komentar yang berhasil ditambang secara otomatis menggunakan *simplescraper*. Data yang telah berhasil dikumpulkan kemudian akan diolah kembali dengan merapikan data secara manual sehingga menghasilkan data sebanyak 621 data komentar untuk dianalisis.

4.2 *Polarity*

Gambar 4.2 adalah tahap *polarity* yaitu proses pemisahan antara sentimen positif dan negatif dari data yang telah didapatkan melalui *data crawler*, proses pemisahan sentimen dilakukan secara otomatis menggunakan fitur yang terdapat pada RapidMiner dan dikoreksi kembali secara manual.



Gambar 4.2 Proses *polarity*

Proses akan dimulai dari operator *Read Excel*, operator ini berfungsi untuk menginputkan *dataset*. Selanjutnya yaitu operator *Sentiment Analysis*, operator ini berfungsi untuk mengklasifikasi sentimen positif dan negatif secara otomatis. Tahap terakhir yaitu pada operator *Write Excel*, operator ini berfungsi untuk menghasilkan *output polarity* seperti gambar 4.3. *File* disimpan dalam bentuk *file excel* yang didalamnya akan menampilkan *text* dan *polarity*.

text	polarity
kami sebagai siswa sangat keberatan dalam pjj ini baik di segi kuota internet, pemahaman mat	negative
Pak Mentri....tolong perhatikan juga nasib para pedagang di kantin...sudah terlalu lama menung	negative
Tolong perhatikan juga orang orang yang nyari nafkan dari kegiatan sekolah..(angkutan umum,c	negative
Penyakit menular dari dulu sudah banyak...kenapa Corona spesial banget ya..ada apa?semua m	negative
Saya mau sedikit bertanya Orang tua murid sudah susah jangan lagi di bebani biaya kouta inter	negative
Ga ada bedanya dengan daring. Orang kita terlalu byk mengeluh dan mencari susah. Pdhal bel	positive
Km tdk mslah pak klu sekolah msh daring, tp fikirkan jg pak beli paket kuotanya...bagaimana klu	negative
Hai mentri kau suruh belajar daring,iuran tetap bayar,buku paket wajib beli,ditambah lagi paket	negative
Kalo saran saya sih lebih baik online kaya gini aja dulu karena apa, orgtua aja susah di atur unt	positive
Jangan kau per sulit , sekolah atau libur ?? Kalau belajar online signal internet diperbaiki dulu j	negative
Sekolah model apa iku pak. Anak kayak robot tdk bisa interaksi.	negative

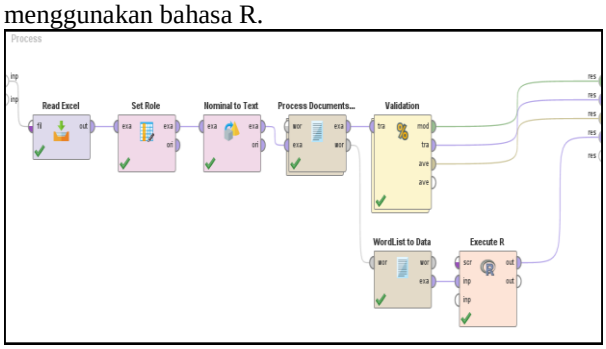
Gambar 4.3 *Output polarity*

Setelah mendapatkan hasil *polarity* dari fitur *Sentiment Analysis by MeaningCloud* maka pada penelitian ini akan dilakukan pengecekan kembali hasil

polarity tadi secara manual. Prediksi *polarity* yang dianggap kurang tepat akan di perbaiki secara manual.

4.3 Pemrosesan Data

Pada gambar 4.4 adalah tahap pemrosesan data, dimana data yang telah diperoleh dari hasil *polarity* akan di*input* kedalam *read excel*, selanjutnya pada proses *set role* yaitu memasukan atribut dan *target role* yang akan digunakan, kemudian *nominal to text* akan mengubah semua atribut *nominal* menjadi atribut *string*, didalam *process documents from data* akan melakukan proses *preprocessing* pada *text* yang hasilnya akan diuji menggunakan metode naïve bayes dengan *data training* dan *data latih* serta proses pembuatan *wordlist* pada *wordlist to data* yang hasilnya akan digunakan dalam pembuatan *wordcloud* menggunakan bahasa R.



Gambar 4.4 Pemrosesan Data

4.4 Hasil Analisis

Pada tahap ini akan dibahas mengenai hasil dari analisis sentimen dan hasil pengujian dengan menggunakan rasio tertentu agar diketahui seberapa akurat metode yang digunakan yaitu Naïve Bayes dalam melakukan analisis.

4.4.1 Hasil Analisis Sentimen

Berdasarkan data yang telah diuji pada penelitian ini yaitu berjumlah 621 data komentar, maka didapat jumlah sentimen positif sebanyak 124 data komentar dan sentimen negatif sebanyak 497 data komentar yang hasilnya diperoleh dari proses *polarity* secara otomatis menggunakan *Analysis Sentiment by MeaningCloud* dan dikoreksi kembali secara manual. Hasil *output* dari Rapidminer yaitu 20% sentiment positif dan 80% sentimen negatif (Gambar 4.5).

SimpleDistribution

```
Distribution model for label attribute polarity


Class negative (0.842)
2121 distributions

Class positive (0.158)
2121 distributions
```

Gambar 4.5 *Output Sentiment*

Pada gambar 4.5 dapat dilihat sentimen negatif mendapat jumlah data komentar terbanyak yaitu 80% atau sama dengan 497 data komentar dari total data yang diuji yaitu 621 data komentar, sehingga dapat ditarik kesimpulan bahwa respon masyarakat Indonesia terhadap pendidikan daring ini kurang efektif dengan banyaknya komentar negatif.

4.4.2 Hasil Pengujian

Dalam melakukan pengujian pada penelitian ini, data akan dibagi menjadi 2 yaitu *data training* dan *data testing*. Data yang digunakan sebagai *data training* dan *data testing* diambil secara acak dari 621 data komentar yang akan dianalisis. Dari 621 data komentar yang ada, maka data akan dibagi menjadi beberapa bagian untuk proses *training* dan *testing*. Dengan pembagian berdasarkan rasio, yaitu 90% *data training* dan 10% *data testing* yang ditulis dengan rasio 9:1, 70% *data training* dan 30% *data testing* yang ditulis dengan rasio 7:3, 50% *data training* dan 50% *data testing* yang ditulis dengan rasio 5:5, 30% *data training* dan 70% *data testing* yang ditulis dengan rasio 3:7 dan 10% *data training* dengan 90% *data testing* yang ditulis dengan rasio 1:9. Pada aplikasi RapidMiner rasio pembagian dimasukkan dengan nilai persentase yang diubah kedalam nilai desimal seperti 0.9, 0.7, 0.5, 0.3 dan 0.1.

Pada proses pembagian data apabila hasilnya desimal maka pembulatan akan disesuaikan dengan angka disebelah kiri koma, jika angka tersebut adalah 4, 3, 2 atau 1 maka akan dilakukan pembulatan kebawah atau mengikuti nilai disebelah kanan koma. Pembulatan keatas akan dilakukan apabalia angka disebelah kiri koma adalah 5, 6, 7, 8 atau 9. Hasil pengujian ini akan menunjukan jumlah nilai akurasi, *class precision* dan *class recall* yang dihasilkan dari perhitungan *confusion matrix* pada setiap rasio yang telah ditentukan dalam metode Naïve Bayes. Hasil perhitungan *confusion matrix* berdasarkan rasio adalah, sebagai berikut:

1. *Confusion Matrix* dengan Rasio 9:1

Pada gambar 4.6 dapat dilihat hasil dari pengujian *data training* dan *data testing* pada RapidMiner dengan rasio 9:1 menggunakan metode Naïve Bayes menghasilkan perhitungan *Confusion Matrix*. Nilai akurasi 83.87% adalah hasil dari gabungan nilai *class recall* dan *class precision*, yaitu *class recall* 30.00% dan *class precision* 50.00%. Karena nilai *class precision* lebih tinggi dari nilai *class recall* maka tingkat kekeliruan sistem dalam membaca *polarity* lebih tinggi.

accuracy: 83.87%

	true negative	true positive	class precision
pred. negative	49	7	87.50%
pred. positive	3	3	50.00%
class recall	94.23%	30.00%	

Gambar 4.6 *Confusion Matrix* dengan Rasio 9:1

Sedangkan jika dihitung secara manual menggunakan rumus *precision*, *recall* dan *accuracy* maka menghasilkan perhitungan sebagai berikut:

$$P = \frac{TP}{TP + FP} * 100\% = \frac{3}{3 + 3} * 100\% = 50.00\%$$

$$R = \frac{TP}{TP + FN} * 100\% = \frac{3}{3 + 7} * 100\% = 30.00\%$$

$$A = \frac{TP + TN}{TP + TN + FP + FN} * 100\% = \frac{3 + 49}{3 + 49 + 3 + 7} * 100\% = 83.35\%$$

2. *Confusion Matrix* dengan Rasio 7:3

Pada gambar 4.7 dapat dilihat hasil dari pengujian *data training* dan *data testing* dengan rasio 7:3 menggunakan metode Naïve Bayes menghasilkan perhitungan *Confusion Matrix*. Nilai akurasi 80.65% adalah hasil dari gabungan nilai *class recall* 44.83% dan *nilai class precision* 39.39%. Karena nilai *class recall* lebih tinggi dari nilai *class precision* maka tingkat ketepatan sistem dalam membaca *polarity* tinggi.

accuracy: 80.65%

	true negative	true positive	class precision
pred. negative	137	16	89.54%
pred. positive	20	13	39.39%
class recall	87.26%	44.83%	

Gambar 4.7 *Confusion Matrix* dengan Rasio 7:3

Sedangkan jika dihitung secara manual menggunakan rumus *precision*, *recall* dan *accuracy* maka menghasilkan perhitungan sebagai berikut:

$$P = \frac{TP}{TP + FP} * 100\% = \frac{13}{13 + 20} * 100\% = 39{,}40\%$$

$$R = \frac{TP}{TP + FN} * 100\% = \frac{13}{13 + 16} * 100\% = 44.82\%$$

$$A = \frac{TP + TN}{TP + TN + FP + FN} * 100\% = \frac{13 + 137}{13 + 137 + 10 + 16} * 100\% = 85.22\%$$

3. *Confusion Matrix* dengan Rasio 5:5

Pada gambar 4.8 dapat dilihat hasil dari pengujian *data training* dan *data testing* dengan rasio 5:5 menggunakan metode Naïve Bayes menghasilkan perhitungan *Confusion Matrix*. Nilai akurasi 87.42% adalah hasil dari gabungan nilai *class recall* dan nilai *class precision*, yaitu *class recall* 55.10% dan *class precision* 61.36%. Karena nilai *class precision* lebih tinggi dari nilai *class recall* maka tingkat kekeliruan sistem dalam membaca *polarity* lebih tinggi.

accuracy: 87.42%

	true negative	true positive	class precision
pred. negative	244	22	91.73%
pred. positive	17	27	61.36%
class recall	93.49%	55.10%	

Gambar 4.8 *Confusion Matrix* dengan Rasio 5:5

Sedangkan jika dihitung secara manual menggunakan rumus *precision*, *recall* dan *accuracy* maka menghasilkan perhitungan sebagai berikut:

$$P = \frac{TP}{TP + FP} * 100\% = \frac{27}{27 + 17} * 100\% = 61.36\%$$

$$R = \frac{TP}{TP + FN} * 100\% = \frac{27}{27 + 22} * 100\% = 55.10\%$$

$$A = \frac{TP + TN}{TP + TN + FP + FN} * 100\% = \frac{27 + 244}{27 + 244 + 17 + 22} * 100\% = 87.42\%$$

4. *Confusion Matrix* dengan Rasio 3:7

Pada gambar 4.9 dapat dilihat hasil dari pengujian *data training* dan *data testing* dengan rasio 3:7 menggunakan metode Naïve Bayes menghasilkan perhitungan *Confusion Matrix*. Nilai akurasi 87.42% adalah hasil dari gabungan nilai *class recall* 42.03% dan nilai *class precision* 46.77%. Karena nilai *class precision* lebih tinggi dari *class recall* maka tingkat kekeliruan sistem dalam membaca *polarity* lebih tinggi.

accuracy: 83.22%

	true negative	true positive	class precision
pred. negative	333	40	89.28%
pred. positive	33	29	46.77%
class recall	90.98%	42.03%	

Gambar 4.9 *Confusion Matrix* dengan Rasio 3:7

Sedangkan jika dihitung secara manual menggunakan rumus *precision*, *recall* dan *accuracy* maka menghasilkan perhitungan sebagai berikut:

$$P = \frac{TP}{TP + FP} * 100\% = \frac{29}{29 + 33} * 100\% = 46.50\%$$

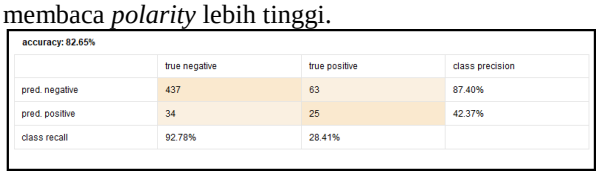
$$R = \frac{TP}{TP + FN} * 100\% = \frac{29}{29 + 40} * 100\% = 42.00\%$$

$$A = \frac{TP + TN}{TP + TN + FP + FN} * 100\% = \frac{29 + 333}{29 + 333 + 33 + 40} * 100\% = 83.50\%$$

5. *Confusion Matrix* dengan Rasio 1:9

Pada gambar 4.10 dapat dilihat hasil dari pengujian *data training* dan *data testing* dengan rasio 1:9 menggunakan metode Naïve Bayes menghasilkan perhitungan *Confusion Matrix*. Nilai akurasi 82.65%

adalah hasil dari gabungan nilai *class recall* dan *class precision*, yaitu *class recall* 28.41% dan *class precision* 42.37%. Karena nilai *class precision* lebih tinggi dari *class recall* maka tingkat kekeliruan sistem dalam membaca *polarity* lebih tinggi.



accuracy: 82.65%

	true negative	true positive	class precision
pred. negative	437	63	87.40%
pred. positive	34	25	42.37%
class recall	92.78%	28.41%	

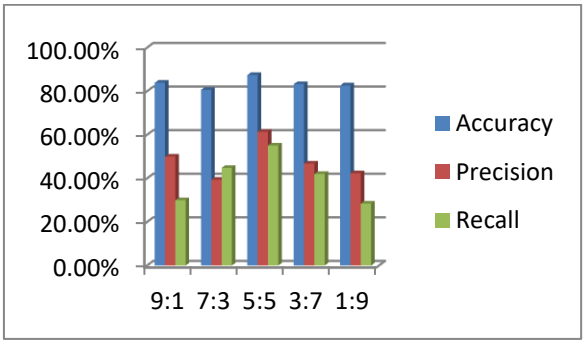
Gambar 4.10 *Confusion Matrix* dengan Rasio 1:9

Sedangkan jika dihitung secara manual menggunakan rumus *precision*, *recall* dan *accuracy* maka menghasilkan perhitungan sebagai berikut:

$$P = \frac{TP}{TP + FP} * 100\% = \frac{25}{25 + 34} * 100\% = 42.37\%$$

$$R = \frac{TP}{TP + FN} * 100\% = \frac{25}{25 + 63} * 100\% = 28.41\%$$

$$A = \frac{TP + TN}{TP + TN + FP + FN} * 100\% = \frac{25 + 437}{25 + 437 + 34 + 63} * 100\% = 82.65\%$$



Gambar 4.11 Diagram Batang *Confusion Matrix*

Pada gambar 4.11 dan tabel 4.1 menunjukan rekap dari hasil klasifikasi metode Naïve Bayes berupa hasil perhitungan *confusion matrix* dalam bentuk diagram batang dan tabel.

Tabel 4.1 Perbandingan rasio menggunakan metode Naïve Bayes

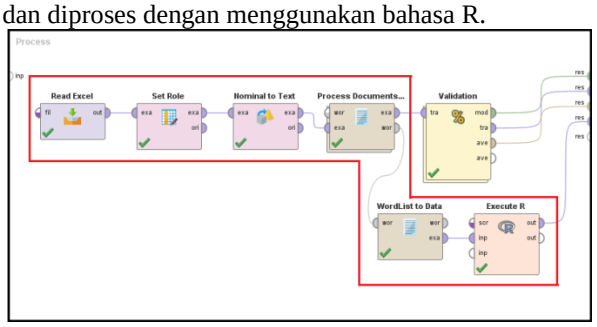
Rasio	*Data Training*	*Data Testing*	Akurasi
9:1	559 (90%)	62 (10%)	83.87%
7:3	435 (70%)	186 (30%)	80.65%
5:5	311 (50%)	310 (50%)	87.42%
3:7	186 (30%)	435 (70%)	83.22%
1:9	62 (10%)	559 (90%)	82.65%
Rata-rata			83.56%

Dapat dilihat pada gambar 4.11 dan tabel 4.1 hasil akurasi terbaik menggunakan metode Naïve Bayes berdasarkan data komentar yang telah diuji terdapat pada rasio 5:5 (50% *data training* dan 50% *data testing*) dengan hasil akurasi 87.42% dan hasil akurasi terendah terdapat pada rasio 7:3 (70% *data training* dan 30% *data testing*) dengan hasil akurasi 80.65%. Namun pada rasio 7:3 ini diperoleh nilai *class recall* lebih tinggi dari nilai *class precision*, maka dapat di artikan bahwa tingkat ketepatan sistem dalam membaca *polarity* tinggi. Sedangkan pada rasio 5:5 diperoleh nilai *class precision* paling tinggi, maka dapat diartikan bahwa tingkat kekeliruan sistem dalam membaca *polarity* tinggi.

Pada penelitian ini dapat dilihat pada gambar 4.11 nilai *class recall* sangat berpengaruh dalam menentukan nilai akurasi yang baik dari sebuah metode yang digunakan. Dapat dilihat juga pada tabel 4.1 diperoleh hasil rata-rata akurasi, yaitu 83.56%.

4.5 *WordCloud*

Dapat dilihat pada gambar 4.12 proses pembuatan *WordCloud* dilakukan pada operator *WordList to Data* dan diproses dengan menggunakan bahasa R.



Gambar 4.12 Proses *WordCloud*

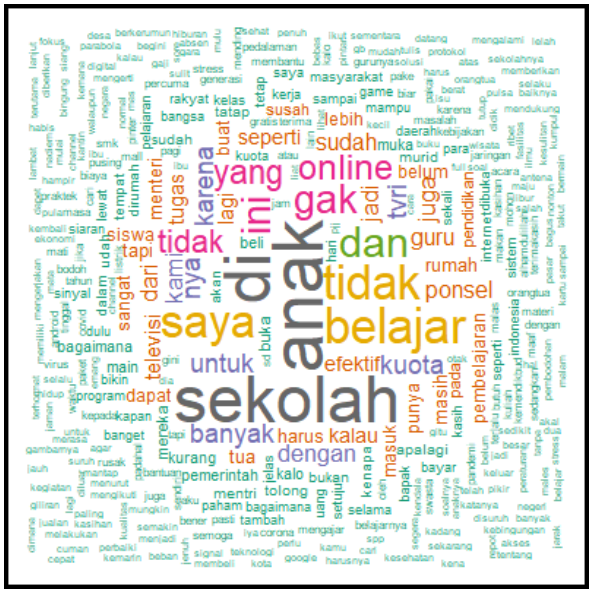
Pada operator *WordList to Data* akan diperoleh *list* kata terbanyak yang didapatkan dari dokumen yang digunakan. Selanjutnya dapat dilihat pada gambar 4.13 menunjukan data hasil *output* dari operator *WordList to Data*.

word	in.documents	total ↓
anak	190	297
sekolah	194	274
di	198	267
belajar	156	200
tidak	151	188
saya	113	187
dan	116	165
gak	111	148
ini	117	147
online	107	125

Gambar 4.13 *Output WordList to Data*

Pada *Output* akan tampil 3 kolom yang berisi *word* yaitu kata, *in documents* yang menampilkan jumlah kemunculan kata perbaris dan total yaitu jumlah keseluruhan penyebutan kata dalam setiap baris dokumen yang bisa lebih dari 1 kali penyebutan.

Tahap terakhir, yaitu kata akan diolah dengan semenarik mungkin yang dibantu dengan operator *Execute R*, pada operator ini kita dapat meng*input*kan *coding*/perintah untuk mengeksekusi *WordList* menjadi *WordCloud*.


anak
di
ini
sekolah
saya
belajar
tidak
dan
yang
online
gak
karena
tugas
menteri
tvri
juga
ponsel
rumah
guru
untuk
efektif
kuota
pembelajaran
banyak
harus
kalau
dengan
apalagi
pemerintah
kurang
tua
kalo
bukan
mentri
tolong
paham
bagaimana
selama
tambah
sudah
seperti
susah
lebih
masyarakat
saya
rakyat
kelas
tatap
kerja
sampai
game
buat
bangsa
muka
belum
murid
jadi
lagi
tapi
dari
kami
nya
televisi
bagaimana
dulu
buka
main
dapat
program
kapan
banget
bikin
virus
kepada
mereka
masuk
punya
masih
pada
kasih
indonesia
seperti
dibuka
internet
pendidikan
sistem
siswa
sinyal
sangat
udah
dalam
tempat
rumah
ibu
pagi
kuota
atau
daerah
kebijakan
mampu
karena
masalah
kecil
lebih
jaringan
wisata
para
sekali
acara
tugas
kuliah
dosen
materi
dengan
orangtua
pulsa
baik
nya
selaku
mendukung
kumpul
kesulitan
bermain
pesan
kartu
sampai
bapak
ibu
guru
dosen
teman
kuliah
tugas
ujian
nilai
kelas
sekolah
belajar
pandemi
covid
corona
virus
masker
cuci
tangan
jaga
jarak
rumah
keluarga
teman
main
game
hp
laptop
wifi
kuota
gratis
bayar
uang
biaya
mahal
murah
sulit
susah
mudah
bisa
tidak
gak
ga
nggak
enggak
ngga
kagak
kaga
gk
tdk
blm
udah
udh
sdh
sdah
sudah
sdh
lagi
lg
lgi
juga
jg
jga
aja
saja
sja
doang
doank
dong
deh
sih
kok
kan
nih
tuh
gitu
gini
begini
begitu
kayak
kaya
kayaknya
kayanya
kayaknya
kayanya
kayaknya
kayan

McFedries, Paul. 2013. *The Complete Idiot's Guide to Google Chrome and Chrome OS.* Alpha. A Member Of Penguin Group USA Inc.

Mcnaught, Carmel dan Lam, Paul. 2014. *Using Wordle as a Supplementary Research Tool. The Qualitative Report*. Volume 15, No.3. http://www.nova.edu/ssss/QR/QR15-3/mcnaught.pdf (diakses tanggal 18 Februari 2021).

Muktamar, Z. 2013. "*Sistem Informasi (Konsep Dasar, Analisi, Desain dan Implementasi)*". Graha Ilmu : Yogyakarta.

Mulyanto, Agus. 2011. *Sistem Informasi Konsep dan Aplikasi*. Yogyakarta : Cetakan I, Pustaka Pelajar.

Nasrullah, Rulli. 2016. *Media Sosial Perspektif Komunikasi, Budaya dan Sosioteknologi*. Bandung : Simbiosa Rekatama Media.

Nugroho, Kuncahyo Setyo. 2019. *Confusion Matrix untuk Evaluasi Model pada Supervised Learning*. https://medium.com/@ksnugroho/confusion-matrix-untuk-evaluasi-model-pada-unsupervised-machine-learning-bc4b1ae9ae3f. (diakses tanggal 17 Februari 2021).

Palucoumputer. 2015. Pengertian Sejarah Dan Pendiri Facebook. http://www.palucomputer.com/2015/06/pengertian-sejarah-danpendiri - facebook-terlengkap.html (diakses tanggal 23 Mei 2021).

Prasetyo, E. 2012. *Data Mining Konsep dan Aplikasi Menggunakan Matlab*. Yogyakarta : Andi Offset.

R. A. Paul. 2016. *Perennial Forages as Second Generation Bioenergy Crops.* International Journal of Molecular Sciences.

Ratniasih, N. L., Sudarma, M., dan Gunantara, N. 2017. *Penerapan Text Mining Dalam Spam Filtering Untuk Aplikasi Chat*. Teknologi Elektro.

Saleh, Ahmad. 2015. *Klasifikasi Gejala Depresi Pada Manusia Dengan Metode Naïve Bayes Menggunakan Java*. Yogyakarta : Andi Offset.

Schroeder, Jonathan E. 2012. *Visual Consumption.* London : Routledge.

Sugiyono. 2017. *Metode Penelitian Kuantitatif, Kualitatif dan R&D*. Bandung : Alfabeta, CV.

Van Dijck, J. 2013. *The Culture Of Connectivity: A Critical History Of Social Media*. UK : Oxford University Press.

Wibawanto, W. 2017. *Desain dan Pemrograman Multimedia Pembelajaran Interaktif*. Jember : Cerdas Ulet Kreatif.

Yatini, Indra. 2014. *Flowchart, Algoritma dan Pemrograman Menggunakan Bahasa C++ Builder*. Yogyakarta: Graha Ilmu.