ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP COVID-19 DI INDONESIA MENGGUNAKAN METODE *NAÏVE BAYES CLASSIFIER* (NBC)

Maria Mega Mala Olhang, Sentot Achmadi, F.X Ariwibisono
Program Studi Teknik Informatika S1, Fakultas Teknologi Industri
Institut Teknologi Nasional Malang, Jalan Raya Karanglo km 2 Malang, Indonesia
1618134@scholar.itn.ac.id

ABSTRAK

Media sosial khususnya Twitter pada saat ini banyak membahas mengenai penyebaran virus corona atau lebih dikenal dengan COVID-19. Diawali dengan ditemukan kasus pertama di Wuhan, China, pemberitaan mengenai virus corona terus berlanjut hingga penyebarannya sampai ke Indonesia. Pemberitaan melalui artikel di Twitter mengenai dampak dari adanya COVID-19 ini antra lain persediaan bahan pokok yang mulai meningkat harganya termasuk harga masker dan *hand sanitizer* juga penyampaian setuju dan tidak setujunya masyarakat terhadap kebijakan pemerintah yang dianggap kurang tanggap dalam menangani kasus ini sangat banyak diminati dan dikritik oleh masyarakat.

Pada penelitian ini, dilakukan proses menganalisis sentimen masyarakat terhadap aspirasi yang disampaikan melalui Twitter yaitu mengembangkan sistem dengan mengacu pada berbagai sistem yang sudah ada sebelumnya dengan menggunakan metode *Naïve Bayes Classifier* untuk mengklasifikasikan sentimen. Masukan pada sistem ini berupa *tweet* yang diperoleh dari Twitter menggunakan *keyword* seperti #coronavirusindonesia atau #covid-19 dengan jumlah data tidak melebihi 500 data *tweet*. Sedangkan outputnya berupa pengelompokkan sentimen positif dan negatif dari setiap *tweet* yang sudah melewati tahap *pre proceessing.*

Dari hasil pengujian, dokumen dengan jumlah sebanyak 75 *tweet* diperoleh hasil pengukuran akurasi *recall* 32%, *precission* 80%, *F-Measure* 45% serta rata-rata akurasi 36%.

Kata Kunci : Klasifikasi Sentimen, *Crawling Tweet, Naïve Bayes Classifier*

1. PENDAHULUAN

Pada akhir tahun 2019, dunia dikejutkan dengan munculnya COVID-19 di Wuhan, Ibu Kota Provinsi Hubei Tiongkok, China. Di Indonesia sendiri, kasus COVID-19 tercatat pertama kali pada 2 Maret 2020 di Depok, Jawa Barat yaitu dua WNI yang merupakan seorang ibu (64 tahun) dan putrinya (31 tahun). Keduanya diduga tertular virus corona karena melakukan kontak langsung dengan warga negara Jepang yang datang ke Indonesia. Hingga kini penyebaran virus corona di Indonesia meningkat sangat cepat. Pada 20 Maret 2020, tercatat pasien yang diperiksa sebanyak 1898 orang, pasien positif COVID-19 sebanyak 369 orang, pasien sembuh sebanyak 17 orang, pasien meninggal sebanyak 32 orang, pasien negatif sebanyak 1570 orang dan sedang dalam proses pemeriksaan sebanyak 21 orang [1].

Dilihat dari cepatnya peningkatan kasus pasien positif COVID-19 di Indonesia, tentunya meningkatkan keresahan masyarakat sehingga terjadinya peningkatan harga masker dan *hand sanitizer* yang tidak normal, adanya sentimen negatif yang terkesan mempolitisasi dalam pemberitaan, dikeluarkannya kebijakan – kebijakan pemerintah yang dinilai masyarakat terlalu santai dan masih kurang tanggap dalam penanganan penyebaran virus, tidak transparannya pemberitaan saat awal kasus terjadi, banyaknya pernyataan dari tokoh publik yang dinilai tidak sesuai dalam kondisi ini, juga tidak tertibnya masyarakat terhadap kebijakan yang sudah ditetapkan permerintah terkait *social distancing* sehingga menimbulkan semakin banyaknya kasus terjadi dan kepanikan masyarakat yang semakin menjadi – jadi. (Perwitasari, 2020)

Dalam melaksanakan penelitian ini diperlukannya sebuah sistem penganalisa opini baik itu opini positif atau negatif atau biasa disebut analisis sentimen dari banyaknya pengguna media sosial khususnya Twitter yang memberikan pengaruh besar bagi masyarakat dalam penilaian berupa pendapat atau komentar terhadap penyebaran Virus Corona di Indonesia terkait apa yang sedang dilakukan, dirasakan, percakapan, penyebaran informasi, pelaporan berita yang juga bisa menjadi tolak ukur pemerintah terhadap kebijakan yang akan diambil berdasarkan aspirasi masyarakat yang bisa dilakukan dengan *Text Mining*.

Oleh karena itu, penelitian ini mencoba menganalisis *tweet* berbahasa Indonesia yang membicarakan tentang virus corona di Indonesia. Analisis dilakukan dengan melakukan klasifikasi *tweet* yang berisi sentimen masyarakat tentang pemberitaan mengenai penyebaran virus corona di Indonesia dan kebijakan pemerintah. Metode klasifikasi yang digunakan dalam penelitian ini adalah *Naive Bayes Classifier*. Pada *Lexicon Based*, seleksi fitur tidak selalu meningkatkan akurasi pada

metode klasifikasi. Terdapat banyak kata tidak baku yang tidak terdeteksi sehingga menurunkan nilai evaluasi. Metode *Support Vector Machine* menghasilkan akurasi performa yang sama baiknya dengan metode *Naive Bayes* baik dalam klasifikasi sentimen maupun dalam klasifikasi kategori. Namun demikian, secara keseluruhan penggunaan metode *Support Vector Machine* dan *Naive Baye* sama-sama memiliki performansi yang cukup baik untuk melakukan klasifikasi tweet. Tetapi karena *Naïve Bayes* memiliki waktu pemrosesan yang cepat serta mudah diimplementasikan dengan tingkat efektifitas yang tinggi. Dengan bahasa yang lebih sederhana, *Naïve Bayes Classifier* mengasumsikan bahwa keberadaan maupun ketidakberadaan sebuah fitur dalam sebuah kelas tidak memiliki keterkaitan dengan keberadaan maupun ketidakberadaan fitur lainnya. Sehingga pada penelitian ini, penulis menggunakan metode *Naïve Bayes Classifier* untuk melakukan pmbobotan pada setiap kemunculan kata dan mengklasfikasikan sentimen dari opini dokumen *tweets* yang tersedia.

2. TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu

Penelitian mengenai klasifikasi masyarakat terhadap *e-commerce* pada media sosial sebagai bahan pertimbangan dalam Analisa pasar secara *online* dengan menggunakan metode *Naïve Bayes Classifier* dengan seleksi fitur *Information Gain* dalam pengkalsifikasian opini negatif dan positif. [2].

Pada penelitian yang dilakukan oleh Yasid Moh dan Junaedi, L, menggunakan metode *Naïve Bayes* untuk melihat penilaian masyarakat terhadap layanan maskapai penerbangan Citilink Indonesia melalui media sosial twitter dengan dataset sebanyak 2000 tweet menghasilkan akurasi yang mencapai 0,778, waktu proses dua menit dua puluh tiga detik[3].

Penelitian tentang klasifikasi juga dilakukan oleh Pakpahan, M, R, S., Indriati., & Marji (2019) dengan membangun sistem analisis sentimen terhadap opini performa klub sepak bola ada twitter menggunakan metode *Support Vector Machine* dan perbaikan kata *Levenshtein Distance*. Tahapan yang dilakukan dimulai dengan melakukan *preprocessing* pada data, kemudian melakukan perbaikan kata dengan *Levenshtein Distance*, pembobotan *Term Frequency-Inverse Document Frequency*, dilanjutkan dengan klasifikasi menggunakan *Support Vector Machine*[4].

Penelitian mengenai klasifikasi sentimen lainnya dilakukan oleh Nurhada, F., Sihwi, W, S., & Doewes, A (2014) diawali dengan tahap pengumpulan data, praprocesing data, POS Tagging, ekstraksi opini menggunakan *rule based* dan klasifikasi opini menggunakan metode *Naïve Bayes Classifier*. Kemudian penelitian yang dilaksanakan Antinasari, et al. (2017) mengenai analisis sentimen pada opini terhadap film di Twitter sudah dibuktikan bahwa perbaikan kata tidak baku memberikan pengaruh akurasi yang lebih baik terhadap hasil klasifikasi. Pada penelitian ini digunakan kamus kata tidak baku dan normalisasi *Levenshtein Distance* untuk memperbaiki kata yang tidak baku menjadi kata baku dengan pengklasifikasian *Naive Bayes*[5].

Kemudian penelitian yang dilaksanakan Antinasari, et al. (2017) mengenai analisis sentimen pada opini terhadap film di Twitter sudah dibuktikan bahwa perbaikan kata tidak baku memberikan pengaruh akurasi yang lebih baik terhadap hasil klasifikasi. Pada penelitian ini digunakan kamus kata tidak baku dan normalisasi *Levenshtein Distance* untuk memperbaiki kata yang tidak baku menjadi kata baku dengan pengklasifikasian *Naive Bayes*[6].

Pada penelitian yang dilakukan oleh Firdaus, Y.M. (Maret 2019) dengan metode *Naïve Bayes Classifier* untuk mengklasifikasi tingkat prestasi akademik santri pondok pesantren mahasiswa, kriteria yang digunakan pada sistem ini adalah nilai materi pokok, nilai pemahaman konsep dan praktikum, nilai sikap dan perilaku, nilai lima sukses santri, nilai kegian ekstrakurikuler, absensi dan keaktifan dan nilai prestasi kuliah. Dari hasil pengujian aplikasi yang telah dibuat dengan metode *Naive Bayes Classifier* untuk memprediksi predikat prestasi akademik Santri, mampu menghasilkan akurasi dengan nilai 76%. Pengujian algoritma ini dilakukan menggunakan perhitungan probabilitas yaitu membandingkan predikat pada data lama dengan predikat hasil prediksi[7].

Penelitian yang dilakukan oleh I Komang Somawirata, A.F. (Februari 2018) mengenai sistem pakar diagnosa diabetes mellitus dengan metode *Naïve Bayes* berbasis online. Penelitian yang dilakukan dengan cara mengumpulkan semua gejala dari diabetes tipe A dan tipe B dan dimasukkan kedalam database bernama data training, selanjutnya sesuai dengan keilmuan pakar data tersebut disusun menjadi rule/aturan pengetahuan yang nanti akan digunakan untuk menentukan diagnose penyakit diabetes tipe A atau tipe B[8].

2.2 Twitter

Twitter diluncurkan sebagai situs *micro-blogging* pada Maret 2006 yang memungkinkan pengguna untuk mengirim *update status* hingga 140 karakter, yang dikenal sebagai *tweets*. Sejak diluncurkan, *twitter* telah mengumpulkan basis pengguna yang besar dan sekarang memiliki lebih dari 300 juta pengguna per Juni 2011. Banyaknya pengguna Twitter memberikan pengaruh besar bagi masyarakat dalam memberikan penilaian berupa pendapat atau komentar terhadap suatu peristiwa yang sudah atau sedang terjadi.

2.3 Pre Processing Text

Prerocessing merupakan proses untuk mempersiapkan data mentah sebelum dilakukan proses lain. Pada umumnya, *preprocessing* data

dilakukan dengan cara mengeliminasi data yang tidak sesuai atau mengubah data menjadi bentuk yang lebih mudah diproses dalam sistem. *Preprocessing* sangat penting dalam pembuatan analisis sentimen terutama untuk media sosial yang sebagian besar berisi kata – kata atau kalimat yang tidak formal dan tidak terstruktur serta memiliki *noise* yang besar.

2.4 Text Mining

Text mining adalah proses mengeksplorasi dan menganalisis sejumlah besar data teks tidak terstruktur yang dibantu oleh perangkat lunak yang dapat mengidentifikasi konsep, pola, topik, kata kunci, dan atribut lainnya dalam data. Ini juga dikenal sebagai analisis teks, meskipun beberapa orang menarik perbedaan antara dua istilah. Dalam pandangan itu, analitik teks adalah aplikasi yang diaktifkan oleh penggunaan teknik *text mining* untuk memilah - milah set data[9].

2.5 Naïve Bayes Classifier

Naive Bayes Classifier merupakan salah satu metode yang populer untuk keperluan *data mining* karena penggunaannya yang mudah dan dalam pemroresan memiliki waktu yang cepat, mudah diimplementasikan dengan strukturnya yang cukup sederhana dan memiliki tingkat efektifitas yang tinggi. Teorema Bayes merupakan teorema yang mengacu pada konsep probabilitas bersyarat. Secara umum teorema bayes dapat dinotasikan pada persamaan 1 berikut :

$$P(H|X) = \frac{P(X|H)}{P(X)} . P(H) \ldots\ldots(1)$$

Terdapat dua kejadian terpisah misalkan X dan H. X adalah data dengan *class* yang belum dikathui, H adalah hipotesis yang merupakan suatu *class* spesifik, P(H|X) merupakan Probabilitas hipotesis H berdasar kondisi X (*posteriori probabilitas*), P(H) merupakan Probabilitas hipotesis H (*prior probabilitas*), P(X|H) merupakan Probabilitas X berdasarkan kondisi pada hipotesis H, dan P(X) adalah Probabilitas X. Teorema Bayes sering pula dikembangakn mengingat berlakunya hukum probabilitas total menjadi seperti berikut ini :

$$P(H|X) = \frac{P(X|H)}{\sum_{i}^{n}=1\, P(H_{i}|X)} . P(H) \ldots(2)$$

i : 1,2,3, n merupakan jumlah data hipotesis (*prior probabilitas*). Untuk menjelaskan Teorema Naïve Bayes, perlu diketahui bahwa proses klasifikasi memerlukan sejumlah petunjuk untuk menentukan kelas apa yang cocok bagi *sampel* yang dianalisis tersebut. Karena itu, Teorema Bayes di atas disesuaikan sebagai berikut:

$$P(C|F_{1,\ldots,}F_n) = \frac{P(F_{1,\ldots,}F_n|C)}{P(F_{1,\ldots,}F_n)} . P(C) \ldots(3)$$

Di mana Variabel C merepresentasikan kelas, sementara variabel F_1 ... F_n merepresentasikan karakteristik petunjuk yang dibutuhkan untuk melakukan klasifikasi. Maka rumus tersebut menjelaskan bahwa peluang masuknya sampel karakteristik tertentu dalam kelas C (*Posterior*) adalah peluang munculnya kelas C (sebelum masuknya *sampel* tersebut, seringkali disebut *prior*), dikali dengan peluang kemunculan karakteristik-karakteristik *sampel* pada kelas C (disebut juga *likelihood*), dibagi dengan peluang kemunculan karakteristik-karakteristik *sampel* secara global (disebut juga *evidence*). Karena itu, rumus di atas dapat pula ditulis secara sederhana sebagai berikut:

$$Posterior = \frac{prior\ x\ likehood}{evidence} \ldots(4)$$

Nilai *Evidence* selalu tetap untuk setiap kelas pada satu *sampel*. Nilai dari *posterior* tersebut nantinya akan dibandingkan dengan nilai-nilai posterior kelas lainnya untuk menentukan ke kelas apa suatu *sampel* akan diklasifikasikan. Penjabaran lebih lanjut rumus Bayes tersebut dilakukan dengan menjabarkan (C,F_1, ... , F_n) menggunakan aturan perkalian sebagai berikut:

$$P(C|F_{1,\ldots,}F_n) = P(C).P(F_{1,\ldots,}F_n|C) \ldots(5)$$

Dapat dilihat bahwa hasil penjabaran tersebut menyebabkan semakin banyak dan semakin kompleksnya faktor - faktor syarat yang mempengaruhi nilai probabilitas, yang hampir mustahil untuk dianalisa satu persatu. Akibatnya, perhitungan tersebut menjadi sulit untuk dilakukan. Disinilah digunakan asumsi independensi yang sangat tinggi (*naif*), bahwa masing-masing petunjuk (F_1,F_2 , ..., F_n) saling bebas (*independen*) satu sama lain. Dengan asumsi tersebut, maka berlaku suatu kesamaan sebagai berikut:

$$P(c|F) = P(f_1|C).P(f_2|C).P(f_3|C) \ldots P(f_n|C).P(C) \ldots(6)$$

Persamaan di atas merupakan model dari teorema Naïve Bayes yang selanjutnya akan digunakan dalam proses klasifikasi.

2.6 Evaluasi

Permormance measure merupakan tahapan terakhir dalam klasifikasi teks. Pada tahap ini akan mengevaluasi hasil percobaan, membandingkan dan menganalisis terhadap kinerja klasifikasi teks. Pada penelitian ini metode evaluasi yang digunakan adalah Recall, Precision, dan F-measure.

Tabel 1 Confusion Matrix

Predicted Class	**True Class**	
	Positif	Negatif
Positif	True Positif(TP)	False Positif (FP)
Negatif	False Negatif(FN)	True Negatif (TN)

Keempat parameter diatas digunakan untuk menghitung 3 metode evaluasi yakni :

1. *Recall*, yaitu perbandingan jumlah dokumen yang relevan terkenali dengan jumlah seluruh dokumen relevan. *Recall* memiliki rumusan sebagai berikut :

$$Recall = \frac{TP}{TP+FP} X100\% \quad \text{.....(7)}$$

2. *Precision*, yaitu perbandingan jumlah dokumen yang relevan terkenali dengan jumlah dokumen yang terkenali. Precision memiliki persamaan sebagai berikut :

$$Precision = \frac{TP}{TP+FN} x100\% \text{.....(8)}$$

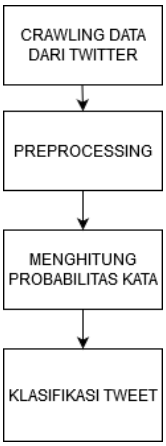
3. *F-Measure*, merupakan nilai yang mewakili sluruh kinerja sistem yang merupakan penggabungan nilai *Recall* dan *Presicion*. *F-measure* memiliki persamaan sebagai berikut :

$$F - Measure = \frac{2PR}{P+R} x100\% \text{.....(9)}$$

Nilai *recall*, *precision* dan *F-measure* dinyatakan dalam persen. Semakin tinggi persentase ketiga nilai tersebut menunjukan semakin baik kinerja sistem klasifikasi teks otomatis.

3. METODE PENELITIAN

3.1 Gambaran Umum Sistem

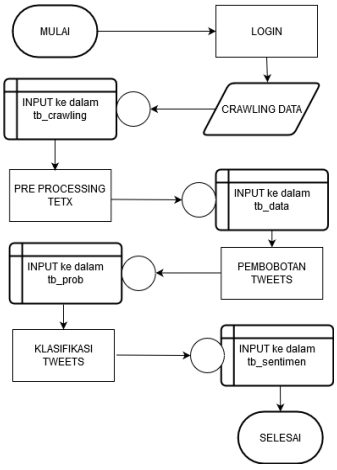


Gambar 1 Gambaran Umum Sistem

Sistem ini dimulai dengan melakukan *crawling* data secara otomatis menggunakan API twitter untuk mendapatkan *tweets*. Lalu melewati tahap *pre processing*. Output dari sistem ini merupakan data dengan nilai ouput positif atau negatif yang diklasifikasikan oleh sistem berdasarkan nilai probabilitas yang ada.

3.2 Flowchart Sistem

Flowchart sistem ini menjelaskan proses berjalananya aplikasi seperti ditunjukkan pada pada Gambar 2



Gambar 2 Flowchart Sistem

Berdasarkan flowchart pada Gambar 2, sistem dimulai dengan login oleh admin selaku pengelola sistem tersebut. Kemudian dilakukan *crawling* data secara realtime yang langsung diinputkan ke dalam tb_crawling. Selanjutnya dilakuakn *pre processing text* dimana data yang sudah di *crawling* dibersihkan terlebih dahulu untuk diinputkan ke dalam tb_data sebagai tempat penampungan data yang sudah bersih untuk dihitung nilai probabilitasnya. Setelah dilakukan perhtiungan probabilitas pada dokumen *tweets* yang sudah bersih, proses berikutnya adalah melakukan klasifikasi sentimen ke kelas manakah dokumen *tweets* itu akan diklasifikasikan. Ke dalam sentimen positif atau negatif yang kemudian disimpan ke dalam tb_sentmen dan proses pun selesai.

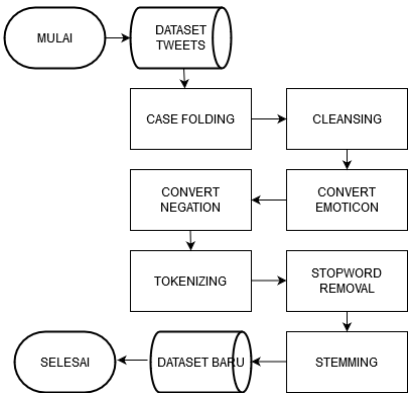
3.3 Pengumpulan Dataset

Data didapatkan dari hasil proses Crawling data *tweet*. Crawling merupakan proses pengumpulan data dari media sosial twitter dengan melakukan pencarian melalui suatu kata kunci tertentu. Pada proses crawling ini menggunakan hashtags #coronavirusindonesia dan #covid19. Data akan diklasifikasi secara manual dengan memberikan label positif atau negatif. Klasifikasi label positif dan negatif pada data training dilakukan secara manual Berikut contoh dokumen hasil *crawling* terdapat pada tabel 2.

Tabel 2 Contoh Data Hasil Crawling

Tweet	Kategori
RT@KRjogjadotcom: Pertama Kali DIY 'Zero' Kasus Positif	?
Kebersihan Lampaui DKI, Pasien Positif Corona ☺@anisbaswedan!!	?

3.4 PreProcessing Text



Gambar 3 Tahapan Preprocessing

3.5 Case Folding

Tahap *case folding* merupakan proses penyamaan huruf pada dokumen *tweet* dengan mengubah huruf kapital jadi huruf kecil, supaya merenggangkan kata berlebih - lebihan dikarenakan perbedadaan huruf, dan akan menghasilkan dokumen *tweet* baru. Berikut contoh *tweet* yang dihasilkan dari tahap *case folding* bisa dilihat pada tabel 3.

Tabel 3 Tahap Case Folding

Input Proses	Output Proses
Kebersihan Lampaui DKI, Pasien Positif Corona☺@anisbaswedan!!	kebersihan lampaui dki, pasien positif corona☺ @aniesbaswedan!!

3.6 Cleansing

Tahap *cleansing* adalah proses yang bertujuan untuk menghilangkan tweet dari kata yang kurang dibutukan seperti simbol HTML, kata kunci, hastags(#), username, email dan simbol (!@#%<>[]=;:",'&^$*+_-/?), yang akan menghasilkan dokumen *tweet* baru. Berikut contoh dokumen *tweet* yang menjadi keluaran dari tahap *cleansing* bisa dilihat pada tabel 4

Tabel 4 Tahap Cleansing

Input Proses	Output Proses
kebersihan lampaui dki, pasien positif corona ☺@aniesbaswedan!!	kebersihan lampaui dki pasien positif corona☺

3.7 Convert Emoticon

Pada tahap ini, mengubah bentuk *emoticon* seperti '☺' atau '☹' menjadi kata yaitu 'emot senang' dan 'emot sedih'. Sebagai gambaran dari proses *covert emoticon* berikut contoh *tweet* yang dihasilkan terdapat pada tabel 5

Tabel 5 Tahap Convert Emoticon

Input Proses	Output Proses
kebersihan lampaui dki, pasien positif corona ☺@aniesbaswedan!!	kebersihan lampaui dki pasien positif corona

3.8 Convert Negation

Proses *convert negation* bertujuan untuk transformasi kata negasi yang tergabung dalam *tweet*. Kata tersebut akan mengganti arti suatu sentimen pada tweet tersebut, kata negasi akan tergabung dengan kata setelahnya. Contoh kata negasi 'tidak', banyak lagi kata negasi yang akan menghasilkan dokumen *tweet* baru. Berikut contoh *tweet* yang dihasilkan dari tahap *convert negation* terdapat pada tabel 6.

Tabel 6 Tahap Convert Negation

Input Proses	Output Proses
kebersihan lampaui dki, pasien positif corona ☺@aniesbaswedan!!	kebersihan lampaui dki pasien positif corona emot senang

3.9 Tokenizing

Pada tahap tokenzing yaitu pemotongan setiap kata dalam *tweet* menjadi kata tunggal. Sebagai gambaran dari proses *tokenizing* berikut contoh *tweet* yang dihasilkan terdapat pada tabel 7

Tabel 7 Tahap Tokenizing

Input Proses	Output Proses
kebersihan lampaui dki, pasien positif corona ☺@aniesbaswedan!!	'kebersihan' 'lampaui' 'dki' 'pasien' 'positif' 'corona' 'emot' 'senang'

3.10 Stopword Removal

Stopword Removal, yakni teknik yang bertujuan untuk menghapus kata di dalam dokumen dataset, kata tersebut seperti "di", "ke", "dari", "yang", "sedang", "ini", karena dianggap tidak memberikan deskripsi sesuatu

dalam bahasa Indonesi. Tapi pada teks klasifikasi kata semacam "tidak", "bukan", "tanpa" kurang penting sehingga kata tersebut tidak ikut dibuang. Berikut contoh hasil tahap *stopword removal* terdapat pada tabel 8

Tabel 8 Tahap Stopword Removal

Input Proses	Output Proses
kebersihan lampaui dki, pasien positif corona ☺@aniesbaswedan!!	kebersihan lampaui dki pasien positif corona senang

3.11 Stemming

Pada tahap *stemming* yaitu mengubah kata sesuai aturan yang ada dengan menghilangkan prefix, sufix, konfix pada dokumen atau mengubah kata - kata tersebut menjadi kata dasar. Sebagai gambaran dari proses *stemming* berikut contoh *tweet* yang dihasilkan terdapat pada tabel 9

Tabel 9 Tahap Stemming

Input Proses	Output Proses
kebersihan lampaui dki, pasien positif corona☺@aniesbaswedan!!	bersih lampaui dki pasien positif corona senang

4. HASIL DAN PEMBAHASAN

4.1 Pengujian Sistem

Pada pengujian sistem klasifikasi ini, digunakan 100 dokumen *tweet*. Untuk mengetahui pengaruh jumlah data latih terhadap efektifitas *naïve bayes classifier* maka diambil beberapa kombinasi jumlah dokumen latih dan dokumen uji. Hasil perhitungan pada data testing dapat dilihat pada tabel 10.

Tabel 10 Nilai Probabilitas

Data	Probabilitas	
	Positif	**Negatif**
Data1	1.03182216	1.9177117
Data2	2.5486026	4.9668784
Data3	2.5486026	4.96687848
Data4	2.5486026	4.96687848
Data5	2.5486026	4.96687848

Berdasarkan hasil yang diperoleh pada tabel 10, kedua nilai probabilitas kemudian dibandingkan nilai terbesar untuk menentukan klasifikasi sentimen dari data tersebut. Dapat dilihat pada tabel 11 berikut

Tabel 11 Klasifikasi Data Testing

Data	Sentimen
Data1	Negatif
Data2	Negatif
Data3	Negatif
Data4	Negatif
Data5	Negatif

4.2 Halaman Preprocessing

Pada halaman ini disajikan proses *preprocessing* dengan cara menekan tombol proses yang akan diolah sistem dari database pada tabel crawling yang hasilnya akan ditampung ke dalam tabel hasil *processing* dan hasilnya akan ditampilkan pada kolom dibawah tombol proses tersebut. Tampilan halaman *pre processing* dapat dilihat pada gambar 4

Gambar 4 Tampilan Halaman Preprocesing

4.3 Halaman Pembobotan

Halaman ini terdapat tombol proses yang ketika ditekan akan memproses perhitungan probabilitas positif dan negatif dokumen *testing*. Hasil dari probabilitas ini akan dibandingkan mana nilai yang lebih besar diantara kedua probabilitas tersebut, sehingga jika probabilitas positif lebih besar, maka nilai sentimennya bisa diketahui positif dan jika sebaliknya probabilitas negatif bernilai lebih tinggi, maka termasuk sentimen negatif. Tampilan pembobotan dapat dilihat pada gambar 5

Gambar 5 Tampilan Halaman Pembobotan

4.4 Halaman Klasifikasi

Halaman ini merupakan halaman hasil klasifikasi *naïve* bayes, dimana halaman ini ditampilkan dokumen *tweets* serta sentimennya yaitu positif atau negatif. Tampilan halaman klasifikasi dapat dilihat pada gambar 6

Gambar 6 Tampilan Halaman Klasifikassi

4.5 Evaluasi

Evaluasi bertujuan untuk mengecek kebenaran hasil klasifikasi dengan menghitung berapa angka yang diperoleh. Acuan yang dipakai untuk tahap

evaluasi yakni perhitungan tabel matriks klasifikasi (*confussion matrix*). Dari dokumen *tweet* sebanyak sebanyak 100 dokumen, menghasilkan perhitugan *confussion matrix* seperti pada tabel 12.

Tabel 12 Confussion Matrix

Predicted Class	**True Class**		Total
	Positif	**Negatif**	
Positif	21	43	64
Negatif	5	6	9
Total	26	49	75

Tabel 13 Pengukuran Akurasi

Measure	**Result**	**Percent**
Recall	$\frac{21}{21+43} X100\%$	32
Precission	$\frac{21}{21+5} X100\%$	80
F-Measure	$\frac{2 \text{ x } 32 \text{ x } 80}{32+80} X100\%$	45
Rata-rata akurasi		**77%**

$$Accuracy\ (A) = \frac{TP + TN}{(TP + TN + FP + FN)} \text{x } 100\ \% = \frac{21 + 6}{21 + 6 + 43 + 5} = 36\%$$

Berdsarkan tabel 12, diperoleh dokumen dengan prediksi positif dan faktanya positif sebanyak 21 *tweet*, dokumen dengan prediksi positif dan faktanya negatif sebanyak 43 *tweet*, dokumen dengan prediksi negatif dan faktanya negatif sebanyak 6 *tweet*, serta dokumen dengan prediksi negatif dan faktanya positif sebanyak 5 *tweet*. Berdasarkan hasil rata-rata akurasi yang ditunjukkan pada tabel 4.8, dengan dokumen sebanyak 75 *tweet* diperoleh hasil pengukuran akurasi *recall* 32%, *precission* 80%, *F-Measure* 45% serta rata-rata akurasi 36%.

5. KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan penelitian yang telah dilaksanakan, dapat disimpulkan bahwa:

1. Sistem yang dibangun telah mampu mentransformasi sentimen yang berupa teks opini dari Twitter sehingga dapat menampilkan informasi sentimen masyarakat terhadap COVID-19 di Indonesia yang bersifat positif dan negatif.
2. Penggunaan metode *Naïve Bayes Classifier* pada penelitian ini tidak menjamin ketepatan dalam proses klasifikasi.
3. Dari dokumen sebanyak 75 dokumen, hasil pengukuran akurasi *recall* 32%, *precission* 80%, *F-Measure* 45% serta rata-rata akurasi 36%.
4. Jumlah data yang diproses dan ketepatan tahap pemberishan teks atau *processing text* dalam sistem mempengaruhi tingkat keakuratan kinerja klasifikasi teks menggunakan metode *Naïve Bayes Classfier*.
5. Tingkat kemiripan antar dokumen *tweet* dan dokumen berulang yang disebabkan oleh fitur *retweet* pada Twitter menyebabkan terjadinya pengulangan dokumen yang sama sehingga dalam satu daftar data yang diperoleh terdapat beberapa tumpukan data yang sama.

5.2 Saran

Berdasarkan penelitian yang telah dilakukan, maka penulis dapat memberikan saran-saran untuk pengembangan selanjutnya antar lain :

1. Meningkatkan tahap *processing text* menjadi lebih maksimal.
2. Menambah *stopword* dan *stemming* dengan menyesuaikan terhadap sentimen yang dikerjakan sehingga dalam *processing text* bisa ditingkatkan dalam pembersihan dokumen *tweet*.
3. Menambah jumlah data yang akan diproses agar tingkat akurasi semakin baik, juga sistem bisa dikembangkan untuk mengklasifikasikan topik lainnya.

DAFTAR PUSTAKA

[1] Indonesia.com, C. (2020, Maret Jumat). Update Corona 20 Maret: 369 Kasus, 32 Meninggal, 17 Sembuh. DKI Jakarta, Indonesia.

[2] Syakuro, A. (2017). Analisis Sentimen Masyarakat Terhadap E-Commerce Pada Media Sosial Menggunakan Metode Naive Bayes Classifier (NBC) Dengan Seleksi FItur Information Gain (IG). 1-72.

[3] Moh Yasid, L. J. (2019). Analisis Sentimen Maskapai Citilink Pada Twitter Dengan Metode Naive Bayes. *Jurnal Ilmiah Informatika (JIF)*, 83 - 85.

[4] Swandy Raja Manaek Pakpahan, I. M. (2019). Analisis Sentimen Tentang Opini Performa Klub Sepak Bola Pada Dokumen Twitter Menggunakan Support Vector Machine Dengan Perbaikan Kata Tidak Baku. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 7259-7267.

[5] Faishol Nurhuda, S. W. (2014). Analisis Sentimen Masyarakat terhadap Calon Presiden Indonesia 2014 berdasarkan Opini dari Twitter Menggunakan Metode Naive Bayes Classifier. *Jurnal Itsmart*, 35-42.

[6] Prananda Antinasari, R. S. (2017). Analisis Sentimen Tentang Opini Film Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Naive Bayes Dengan Perbaikan Kata Tidak Baku. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 1733-1741

[7] Firdaus, Y. M. (Maret 2019). Penerapan Metode Naive Bayes Classifier Untuk Mengklasifikasi TIngkat Perstasi Akademik Santri Pondok Pesantren Mahasiswa (PPM) Baitul Jannah Malang. *Jurnal Mahasiswa Teknik Informatika*, 327-336.

[8] I Komang Somawirata, A. F. (Februari 2018). Sistem Pakar Diagnosa Diabetes Mellitus Dengan Metode Naive Bayes Berbasis Online. *Jurnal MNEMONIC*, 1-3.

[9] Arni, U. D. (2018, 10 25). *Apa itu Text Mining?* Retrieved from Garuda Cyber Indonesia: https://garudacyber.co.id/artikel/1254-apa-itu-text-mining