ANALISIS SENTIMEN MENGENAI KULIAH DARING DI TWITTER MENGGUNAKAN METODE NAÏVE BAYES DAN PEMBENTUKAN STOPWORD DENGAN TERM BASED RANDOM SAMPLING

PROPOSAL SKRIPSI

Disusun oleh:

Raditya Rinandyaswara

NIM: 175150200111047



PROGRAM STUDI TEKNIK INFORMATIKA

JURUSAN TEKNIK INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS BRAWIJAYA

MALANG

2020

# PENDAHULUAN

Bab ini terdiri dari hal yang melatarbelakangi dari penelitian ini dilaksanakan, rumusan masalah yang diperoleh dari latar belakang hingga tujuan dan manfaat dari penelitian ini serta batasan yang dijabarkan sesuai dengan cakupan dan kemampuan penulis , maupun sistematika yang menuliskan secara rangkum isi dari tiap bab.

## Latar belakang

Pada saat ini kemajuan teknologi di dunia maupun di Indonesia mengalami perkembangan yang cukup pesat. Kemajuan teknologi ini salah satunya adalah ditandai dengan mudahnya proses pertukaran Informasi antara satu pengguna dan pengguna lainnya. Kemudahan proses pertukaran informasi ini ditandai dengan maraknya bermuncul sosial media di Internet. Twitter merupakan salah satu contoh dari sosial media yang banyak digunakan oleh masyarakat Indonesia yang digunakan sebagai sarana pertukaran informasi di dunia digital. Dalam pengunaannya, Twitter memberi istilah kepada pertukaran informasi tersebut dengan nama *Tweets,* yang mana *Tweets* adalah suatu teks atau kata yang dibatasi panjang nya hingga 280 karakter yang nanti akan diposting dalam platform Twitter tersebut.

Pada akhir tahun 2019 lalu, dunia dikejutkan dengan adanya wabah yang diakibatkan oleh virus corona yang berasal dari kota Wuhan, China. Penyakit yang disebut sebagai (COVID-19) ini adalah penyakit yang menyerang sistem pernapasan virus manusia. Menurut data pemerintah China, penduduk Hubei menjadi kasus pertama Covid-19 pada 17 November 2019 (Arnani, 2020). Setelah kasus pertama Covid-19 di dunia itu terjadi peningkatan pasien tiap bulannya. Hingga saat ini Indonesia sudah melewati angka 190 ribu kasus Covid-19 yang sudah terkonfirmasi yang terhitung dari sejak pasien pertama (Ramadhan, et al., 2020). Dengan adanya pandemi Covid-19 ini, pemerintah Indonesia mengadakan sistem New Normal dengan tujuan untuk mempercepat penanganan Covid-19 (Putsanra, 2020). Dalam penerapannya banyak perilaku yang biasa dilakukan secara luring berubah menjadi daring. Salah satu contohnya adalah perkuliahan. Tentunya kuliah daring ini menyebabkan banyak pro dan kontra dan salah satu sarana mahasiswa menuliskan opininya mengenai kuliah daring tersebut yaitu melalui Twitter. Namun twitter sendiri hanya berfungsi sebagai sosial media sehingga tidak menyediakan analisis sentimen dari tweet pengguna dan bagaimana masyarakat Indonesia menanggapi protokol baru tersebut.

Analisis Sentimen atau *Opinion Mining* adalah salah satu bidang studi yang menganalisis pendapat, sentimen, evaluasi, penilaian, sikap, dan emosi orang terhadap entitas seperti produk, layanan, organisasi, isu, peristiwa, topik, dan atributnya (Liu, 2012). Dengan proses analisis sentimen ini kita dapat mengetahui bagaimana pendapat orang apakah cenderung positif, negatif atau pun netral.

Pada penelitian ini, peneliti akan membangun sebuah sistem analisis sentimen terhadap kuliah daring yang dituliskan mahasiswa di Twitter menggunakan metode klasifikasi Naïve Bayes. Dalam penelitian ini klasifikasi akan dibuat menjadi 3 kelas yaitu opini netral, positif dan negatif sesuai dari saran penelitian analisis sentimen sebelumnya (Sa'rony, et al., 2019). Dalam proses analisis sentimen diperlukan preprocessing data agar data dapat dihitung dan diolah*.* Salah satu tahap yang penting dalam preprocessing adalah *Stopword Removal*, yaitu penghapusan kata yang tidak memiliki nilai keunikan dalam suatu dokumen. Biasanya dalam proses penghapusan stopword ini memanfaatkan *digital library* namun ada stopword yang kurang tepat sehingga hal itu dapat membuat kualitas proses stopword removal menjadi berkurang (Dila Purnama Sari, et al., 2020). Oleh karena itu perlu adanya suatu mekanisme sistem daftar stopword dinamik yang dapat menghasilkan daftar stopword yang sesuai dengan yang diperlukan sesuai saran dari penelitian sebelumnya (Rahutomo & Ririd, 2018). Oleh karena itu pada penelitian ini, peneliti akan melakukan pembuatan daftar stopword menggunakan metode *Term Based Random Sampling*.

## Rumusan masalah

Dari latar belakang yang telah diuraikan sebelumnya, berikut adalah rumusan masalah untuk penelitian ini:

1. Bagaimana mengimplementasikan metode *Naïve Bayes* *Classifiers* dengan pembentukan daftar stopword dengan Term Based Random Sampling pada analisis sentimen terhadap dengan Kuliah Daring?
2. Bagaimana hasil akurasi analisis sentimen terhadap Kuliah Daring dengan menggunakan metode *Naïve Bayes* *Classifiers* dengan pembentukan daftar stopword dengan Term Based Random Sampling?
3. Bagaimanakah pengaruh pembentukan daftar stopword dengan Term Based Random Sampling terhadap hasil klasifikasi analisis sentimen terhadap Kuliah Daring?

## Tujuan

Tujuan khusus penelitian ini adalah:

1. Mengetahui pengaruh pembuatan Stopword dengan Term Based Random Sampling terhadap klasifikasi menggunakan metode Naïve Bayes.
2. Mengetahui hasil pengujian dari penggunaan daftar stopword Term Based Random Sampling dan Naïve Bayes dalam klasifikasi tweet mengenai Kuliah Daring.

## Manfaat

Manfaat penelitian ini adalah:

1. Mengetahui akurasi dari klasifikasi Naïve Bayes yang menggunakan Term Based Random Sampling untuk pembuatan Stopword.
2. Dapat mengklasifikasikan positif, netral atau negatif dari tweet pengguna Twitter terhadap Kuliah Daring.

## Batasan masalah

Batasan masalah penelitian ini adalah:

1. Opini pengguna Twitter dan bagaimana pandangan pengguna Twitter mengenai Kuliah Daring ini.
2. Algoritma yang digunakan hanya Naïve Bayes Classifier tidak membandingkan dengan algoritma lain.
3. Hasil klasifikasi sentimen hanya dibagi menjadi tiga kelas yaitu positif, netral dan negatif.
4. Tweet yang merupakan datasets hanya tweet yang berbahasa Indonesia.

## Sistematika pembahasan

Berikut sistematika pembahasan dalam penelitian ini:

BAB I PENDAHULUAN

Bab ini menjelaskan mengenai latar belakang, rumusan masalah, tujuan, manfaat, batasan masalah dan sistematika pembahasan sesuai dengan aturan dalam peneliltian

BAB II LANDASAN KEPUSTAKAAN

Landasan Kepustakaan menjelaskan penelitian-penelitian sebelumnya yang serupa dengan penelitian dalam proposal ini, serta dasar-dasar teori yang akan diimplementasikan dalam penelitian ini seperti *preprocessing, term weighting, naïve bayes, term based random sampling,* serta tabel *confusion matrix* sehingga dapat mendukung penelitian.

BAB III METODOLOGI

Pada bab ini dijelaskan tentang bagaimana menerapkan penelitian seperti untuk mengimplementasikan Naïve Bayes dengan pembuatan daftar Stopword untuk analisis sentimen pengguna Twitter terhadap layanan Kuliah Daring.

BAB IV PERANCANGAN

Bab ini menjelaskan bagaimana proses perancangan dalam sistem yang akan dibangun.

BAB V IMPLEMENTASI

Pada bab ini menjelaskan bagimana implementasi sistem yang sudah dirancang di bab sebelumnya.

BAB VI PENGUJIAN DAN ANALISIS

Pada bab ini dilakukan pengujian terhadap sistem yang sudah dibangun dan menganalisa hasil yang didapatkan untuk menemukan kesimpulan dari hasil pengujian.

BAB VII PENUTUP

Pada bab terakhir ini menjelaskan tentang bagaimana kesimpulan dari penelitian ini dan saran untuk penelitian berikutnya.

# LANDASAN KEPUSTAKAAN

Pada bab ini akan dijelaskan bagaimana penelitian-penelitian sebelumnya yang serupa dengan penelitian yang sedang diajukan, serta dasar-dasar teori yang akan diimplementasikan dalam penelitian ini seperti *preprocessing, term weighting, Naïve Bayes, Term Based Random Sampling,* serta tabel *confusion matrix* yang dapat mendukung penelitian.

## Kajian Pustaka

Pada bagian ini akan dibahasnya mengeni penelitian atau kajian pustaka yang digunakan dalam penelitian ini berasal dari beberapa penelitian yang sudah dilakukan sebelumnya dan memiliki keterkaitan dengan judul skripsi Analisis Sentimen mengenai Kuliah Daring di Twitter menggunakan metode Naïve Bayes dan Pembentukan Stopword dengan Term Based Random Sampling.

Contoh salah satu penelitian terkait judul skripsi yang telah disebutkan adalah penelitian yang dilakukan oleh (Septian, et al., 2019) yaitu mengenai analisis sentimen pengguna Twitter terhadap polemik persepakbolaan Indonesia menggunakan pembobotan TF-IDF dan metode klasifikasi K-Nearest Neighbor. Penelitian ini menggunakan. Hasil pengujian yang didapatkan dari pengujian silang sebanyak 10 kali dan mendapatkan hasil akurasi optimal pada nilai k-23 sejumlah 79.99%.

Selain K-Nearest Neighbor, salah satu metode klasifikasi umum yang digunakan adalah Naïve Bayes. Pada Penelitian sebelumnya oleh (Rahman, et al., 2017) metode Multinomial Naïve Bayes digunakan untuk Klasifikasi Berita Online dengan menggunakan feature selection Document Frequency Thresholding dan menggunakan TF-IDF untuk pembobotan term dan menghasilkan akurasi tertinggi pada saat menggunakan TF-IDF 86,62%.

Lalu dilanjutkan penelitian oleh (Sa'rony, et al., 2019) analisis sentimen positif dan negatif yang dilakukan menggunakan Multinomial Naïve Bayes yang menggunakan Raw Term Frequency serta pembuatan stopword menggunakan Term Based Random Sampling dan berhasil mendapatkan *macroaverage* terbaik pada klasifikasi dengan stoplist 20 persen. Peneliti saat ini ingin melakukan analisis sentimen dengan 3 kelas yaitu positif, negatif, dan netral serta pembobotan term dilakukan dengan TF-IDF.

Lalu pada penelitian sebelumnya yang dilakukan oleh (Dila Purnama Sari, et al., 2020)

Lalu dilanjutkan penelitian oleh (Sa'rony, et al., 2019) analisis sentimen positif dan negatif yang dilakukan menggunakan Multinomial Naïve Bayes yang menggunakan Raw Term Frequency serta pembuatan stopword menggunakan Term Based Random Sampling dan berhasil mendapatkan *macroaverage* terbaik pada klasifikasi dengan stoplist 20 persen. Peneliti saat ini ingin melakukan analisis sentimen dengan 3 kelas yaitu positif, negatif, dan netral serta pembobotan term dilakukan dengan TF-IDF.

### Kuliah Daring

Kuliah daring merupakan salah satu dari efek kebijakan sistem New Normal yang terjadi karena pandemi Covid-19 ini. Kuliah daring merupakan metode pembelajaran kuliah yang dilakukan melalui daring seperti berupa video conference dan sebagainya.

### Twitter

Twitter merupakan sosial media besutan Amerika Serikat yang diluncurkan pada tahun 2006. Twitter merupakan salah satu contoh dari sosial media yang banyak digunakan oleh masyarakat Indonesia yang digunakan sebagai sarana pertukaran informasi di dunia digital. Dalam pengunaannya, Twitter memberi istilah kepada pertukaran informasi tersebut dengan nama *Tweets,* yang mana *Tweets* adalah suatu teks atau kata yang dibatasi panjang nya hingga 280 karakter yang nanti akan diposting dalam platform Twitter tersebut.

## Teks *Pre-processing*

Teks pre-processing merupakan langkah awal yang dilakukan dalam analisis sentimen untuk menyiapkan data yang berupa teks agar mudah untuk diproses nantinya (Gaddam, 2019). Teks pre-processing ini meliputi, case folding, cleaning, tokenisasi, stopword removal dan stemming.

### Case Folding

Case folding adalah suatu tahapan untuk menyeragamkan kalimat menjadi huruf kecil atau lowercase semua. Contohnya, jika ada kalimat “Saya suka bermain Komputer” menjadi “saya suka bermain komputer”.

### Cleaning

Cleaning adalah suatu tahapan pembersihan kalimat dari simbol-simbol, tanda baca, maupun angka. Contohnya, jika ada kalimat “Selamat pagi Adis, semoga harimu menyenangkan!” menjadi “Selamat pagi Adis semoga harimu menyenangkan”.

### Tokenisasi

Tokenisasi adalah suatu tahapan untuk memisahkan antar kata dari suatu kalimat sehingga kata-kata tersebut menjadi satu tidak tergabung dengan kata-kata lainnya (Gaddam, 2019). Contohnya, jika ada kalimat “saya sedang bermain gitar” menjadi [‘saya’, ‘sedang’, ‘bermain’, ‘gitar’].

### Stopword Removal

Stopword Removal adalah suatu tahapan untuk menghilangkan kata-kata yang kurang relevan berdasarkan kamus stopword yang digunakan (Gaddam, 2019). Kamus Stopword yang digunakan dalam penelitian ini adalah kamus stopword yang bersifat dinamis yang akan dibuat sesuai dengan kebutuhan sistem.4e

### Stemming

Stemming adalah suatu tahapan untuk mencari kata dasar dari suatu kata (Gaddam, 2019). Contohnya jika ada kata “bermain” menjadi main.

## Term Based Random Sampling

## Term Weighting

Pada tahap ini dilakukan pembobotan kata yang mempresentasikan kata-kata tersebut untuk dilakukan perhitungan nantinya. Salah satu metode dalam term weighting adalah *Term Frequency – Inverse Document Frequency* (𝑡𝑓. 𝑖𝑑𝑓) (Jones, 1972). TF-IDF merupakan hasil perkalian dari *term frequency* dengan *inverse document frequency.* Berikut rumus yang digunakan untuk perhitungan tf-idf.

Berikut perhitungan nilai log tf :

Berikut perhitungan nilai idf :

𝑖𝑑𝑓(𝑡) =

Keterangan :

: frekuensi term pada dokumen d

: nilai *Inverse Document Frequency* suatu term t

: total dokumen

: nilai *Document Frequency* suatu term t

## Algoritme Naïve Bayes

Algoritma Naïve Bayes yaitu algoritma klasifikasi supervised yang berbasis dengan teorema bayes dengan asumsi independensi tiap fitur (Sawla, 2018). Algoritma ini menggunakan metode probabilistik dan statistik.

Algoritma ini mencari probabilitas tertinggi untuk proses klasifikasi. Perhitungan Algoritma Naïve Bayes adalah sebagai berikut:

Keterangan :

P(c|d) : Posterior atau Probabilitas kelas c diberikan dokumen d

P(c) : Prior atau Probabilitas awal muncul kategori c

P(d|c) : Likelihood

## *Cosine Similarity*

*Cosine Similarity* digunakan untuk mengukur nilai sudut antara vektor term dari dua dokumen (Karbhari, 2020). Semakin tinggi nilai yang dihasilkan maka tingkat kemiripan antar kedua dokumen tersebut semakin mirip. Berikut rumus dari *cosine similarity*:

Keterangan :

* : term vektor dokumen latih
* : term vektor dokumen uji

## *Confusion Matrix*

Confusion Matrix adalah pengukur performa dari klasifikasi pembelajaran mesin (*Machine Learning*) (Narkhede, 2018). Confusion Matrix berisikan tabel untuk menampilkan hasil evaluasi yang didalamnya terdapat 2 kolom yaitu kelas hasil prediksi dan kelas sebenarnya.

**Tabel Confusion Matrix**

|  |  |  |  |
| --- | --- | --- | --- |
|  | | Predicted | |
| Negatif | Positif |
| Actual | Negatif | TN | FP |
| Positif | FN | TP |

Keterangan :

* True Negative (TN) : jumlah dokumen yang predicted negatif dan actual negatif
* False Positive (FP) : jumlah dokumen yang predicted positif namun actual negatif
* False Negative (FN) : jumlah dokumen yang predicted negatif namun actual positif
* True Postive (TP) : jumlah dokumen yang predicted positif dan actual positif.

Fungsi dari confusion matrix untuk mempermudah evaluasi hasil klasifikasi untuk mencari akurasi, precision, recall dan f-measure. Berikut rumus nya:

* *Accuracy* : kesesuaian nilai prediksi dengan nilai aktual
* Recall : jumlah banyak atau sedikitnya kesesuaian informasi yang didapatkan berdasarkan sudut pandang kelas atau label yang digunakan
* Precision : tingkat ketepatan antara informasi yang diminta
* *F-Measure* : bobot harmonic mean pada recall dan precision

# METODOLOGI

Pada bab ini akan dijelaskan metodologi yang digunakan pada penelitian ini. Metodologi yang digunakan berupa tipe penelitian, strategi penelitian, subjek penelitian, lokasi penelitian, teknik pengumpulan data, peralatan pendukung, implementasi algoritme, analisis, serta kesimpulan.



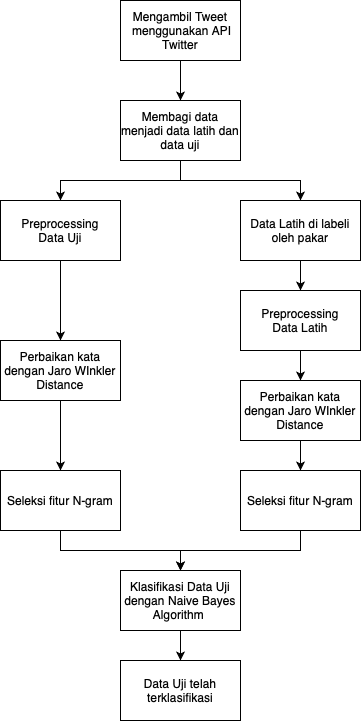
## Studi Literatur

Studi literatur yang dilakukan peneliti adalah mengumpulkan berbagai macam sumber teori atau penelitian yang sudah ada sebelumnya yang terkait dengan Analisis Sentimen, Preprocessing, Naïve Bayes, N-Gram, Jaro Winkler Distance, dan Confusion Matrix.

## Tipe Penelitian

Tipe penelitian yang dilakukan adalah bersifat non-implementatif dengan menggunakan pendekatan analitik. Penelitian bertipe non-implementatif adalah penelitian yang menguji hubungan terhadap suatu kejadian yang kemudian akan di analisis. Sedangkan pendekatan analitik memiliki fungsi untuk menjelaskan hubungan suatu kejadian dengan suatu objek penelitian yang sedang diteliti.

## Strategi Penelitian



Strategi penelitian ini diawali dengan pengambilan data dari twitter lalu dibagi menjadi data uji dan latih. Kedua tipe data dilakukan preprocessing, perbaikan kata, dan seleksi fitur, namun data latih dilabeli manual oleh pakar. Setelah itu data uji diklasifikasi menggunakan nilai yang didapat dari hasil training data latih.

## Subjek Penelitian

Subjek penelitian yang digunakan pada penelitian ini adalah pengguna media sosial khususnya Twitter yang membahas mengenai Layanan Provider By.U.

## Peralatan Pendukung

Kebutuhan yang diperlukan adalah data tweet dari pengguna Twitter dengan jumlah yang cukup untuk nantinya akan dibagi menjadi data training dan data testing.

Peralatan pendukung yang digunakan pada penelitian ini adalah:

|  |  |
| --- | --- |
| Spesifikasi | Keterangan |
| Laptop | Dell XPS 15 9575 |
| CPU | Core i7-8750G |
| GPU | NVIDIA GeForce GTX 1050 (4GB GDDR5) |
| RAM | 16 GB |
| Tipe Memori | DDR4 |
| SSD | 512GB SSD PCIe NVMe |

|  |  |
| --- | --- |
| Jenis | Keterangan |
| *Operating System* | Windows 10 |
| Bahasa Pemrograman | Python 3.7.5 |
| IDE | Visual Studio Code |
| Library | Sastrawi, Pandas, Numpy |

## Lokasi Penelitian

Lokasi penelitian ini bertempat di Laboratorium Komputasi Cerdas, Fakultas Ilmu Komputer Universitas Brawijaya.

## Teknik Pengumpulan Data

Teknik yang digunakan untuk pengumpulan data pada penelitian ini berasal dari Pengguna Twitter. Data diambil menggunakan API atau *Application Programming Interface* yang disediakan oleh Twitter melalui fitur Twitter Developer dengan mengambil tweets yang berisikan by.U.

## Data Penelitian

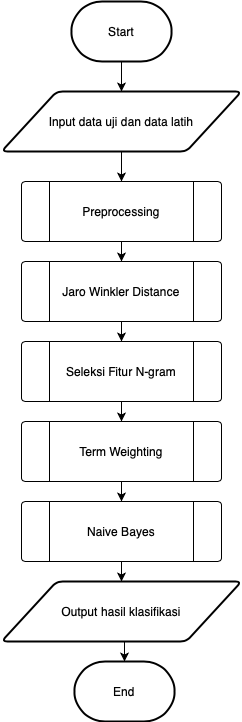
Pada penelitian kali ini, data yang digunakan adalah tweet berbahasa Indonesia. Total dokumen yang akan diambil dari Twitter adalah 1000 dokumen dimana dari 1000 dokumen akan dibagi menjadi 800 data latih, dan 200 data uji. Proses klasifikasi akan dibagi menjadi 2 yaitu positif dan negatif.

## Teknik Analisis Data

Teknik analisis data pada analisis sentimen mengenai layanan provider by.u ini merupakan analisa dari hasil uji tweet dengan metode *Naïve Bayes* dan penggunaan seleksi fitur *N-gram* serta perbaikan kata dengan *Jaro Winkler Distance*. Sehingga nantinya hasil yang diterima akan dimasukan ke dalam tabel Confusion Matrix dan dicari nilai *precision, recall, accuracy, dan f-measure.*

## Perancangan Algoritme

Perancangan Algoritme yang digunakan dalam penelitian ini adalah penerapan *preprocessing* pada data yang diambil dari Twitter hingga menjadi sebuah term. Lalu melakukan perbaikan kata menggunakan metode Jaro Winkler Distance. Setelah perbaikan akan dilanjutkan proses pembobotan kata dengan *tf.idf* untuk merubah term tersebut menjadi suatu numerik dan memilih fitur menggunakan N-gram yang setelah itu akan ditraining menggunakan metode Naïve Bayes dan menerapkan *Cosine* Similarity untuk klasifikasi dokumen uji untuk mencari kesamaan antara dokumen uji dengan dokumen latih.



## Analisis

Dari data yang telah peneliti analisa bahwa banyak pengguna sosial media khususnya Twitter mengungkapkan bagaimana pendapat mereka mengenai layanan provider by.U di sosial media. Sehingga analisis sentimen yang dalam penelitian ini memiliki data yang cukup untuk dilakukan sebuah penelitian karena tidak sedikit yang mengungkapan pendapat mereka di sosial media tersebut.

## Kesimpulan

Dengan dilihat dari penelitian sebelumnya, penggunaan metode Naïve Bayes dalam analisis sentimen memiliki nilai akurasi yang tinggi. Sehingga peneliti dapat mengambil hipotesa awal bahwa, akurasi dapat ditingkatkan jika seleksi fitur N-gram ditambahkan dan tweet sebagai data dilakukan perbaikan kata terlebih dahulu.

# Daftar Pustaka

Wulandari, D., 2015. *ANALISIS SENTIMEN PADA ULASAN “LAZADA” BERBAHASA INDONESIA MENGGUNAKAN BM25 DAN K-NEAREST NEIGHBOR (K-NN) DENGAN PERBAIKAN KATA MENGGUNAKAN JARO WINKLER DISTANCE.* Malang: FILKOM.

Gunawan, F., Fauzi, M. A. & Adikara, P. P., 2017. Analisis Sentimen Pada Ulasan Aplikasi Mobile Menggunakan Naive Bayes. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer,* 1(10), pp. 1082-1088.

Febrianti, Y. M., Indriati, I. & Widodo, A. W., 2018. Analisis Sentimen Pada Ulasan “Lazada” Berbahasa Indonesia Menggunakan K-Nearest Neighbor (K-NN) Dengan Perbaikan Kata Menggunakan Jaro Winkler Distance. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer,* 2(10), pp. 3689-3698.

Prasanti, A. A., Fauzi, M. A. & Tanzil, M., 2017. Klasifikasi Teks Pengaduan pada Sambat Online menggunakan N-gram dan Neighbor Weighted K-Nearest Neighbor (NW-KNN).. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer,* 2(2), pp. 594-601.

Nisa, A., Darwiyanto, E. & Asror, I., 2019. Analisis Sentimen Menggunakan Naive Bayes Classifier dengan Chi-Square Feature. *e-Proceeding of Engineering,* 6(2), p. 8650.

Putri, P. A., Ridok, A. & Indriati, I., 2013. Implementasi Metode Improved K-Nearest Neighbor pada Analisis Sentimen Twitter Berbahasa Indonesia. *Repositori Jurnal Mahasiswa PTIIK UB,* 2(2).

Gaddam, S. H. R., 2019. *Text Preprocessing in Natural Language Processing.* [Online]   
Available at: https://towardsdatascience.com/text-preprocessing-in-natural-language-processing-using-python-6113ff5decd8

Sawla, S., 2018. *Introduction to Naive Bayes for Classification.* [Online]   
Available at: https://medium.com/@srishtisawla/introduction-to-naive-bayes-for-classification-baefefb43a2d

Narkhede, S., 2018. *Understanding Confusion Matrix.* [Online]   
Available at: https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62

Arnani, M., 2020. *KOMPAS.* [Online]   
Available at: https://www.kompas.com/tren/read/2020/03/13/111245765/kasus-pertama-virus-corona-di-china-dilacak-hingga-17-november-2019

Ramadhan, A., Nugraheny, D. E. & Maharani, T., 2020. *KOMPAS.* [Online]   
Available at: https://nasional.kompas.com/read/2020/09/05/15204581/update-kembali-bertambah-di-atas-3000-kasus-covid-19-lewati-190000?page=all

Putsanra, D. V., 2020. *tirto.* [Online]   
Available at: https://tirto.id/apa-itu-new-normal-dan-bagaimana-penerapannya-saat-pandemi-corona-fCSg

Liu, B., 2012. *Sentiment Analysis and Opinion Mining.* Chicago: Morgan & Claypool.

Sa'rony, A., Adikara, P. P. & Wihandika, R. C., 2019. Analisis Sentimen Kebijakan Pemindahan Ibukota Republik Indonesia dengan Menggunakan Algoritme Term-Based Random Samplingdan Metode Klasifikasi Naïve Bayes. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer,* pp. 10086-10094.

Dila Purnama Sari, D. E., Sari, Y. A. & Furqon, M. T., 2020. Pembentukan Daftar Stopword menggunakan Zipf Law dan Pembobotan Augmented TF - Probability IDF pada Klasifikasi Dokumen Ulasan Produk. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer,* pp. 406-412.

Rahutomo, F. & Ririd, A. R. T. H., 2018. EVALUASI DAFTAR STOPWORD BAHASA INDONESIA. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK),* pp. 41-48.

Rahman, A., Wiranto & Doewes, A., 2017. Online News Classification Using Multinomial Naive Bayes. *ITSMART: Jurnal Ilmiah Teknologi dan Informasi.*

Organization, W. H. (2020, Januari 12). *Nove Coronavirus - China*. Diambil kembali dari World Health Organization: https://www.who.int/csr/don/12-january-2020-novel-coronavirus-china/en/