

# ANALISIS SENTIMEN PADA ULASAN APLIKASI VIDEO CONFERENCE MENGGUNAKAN NAÏVE BAYES

Michael Suhendra<sup>1</sup>, Windra Swastika<sup>2</sup>, Mochamad Subianto<sup>3</sup>

Program Studi Teknik Informatika, Universitas Ma Chung

Email: [311610012@student.machung.ac.id](mailto:311610012@student.machung.ac.id) , [windra.swastika@machung.ac.id](mailto:windra.swastika@machung.ac.id) ,  
[mochamad.subianto@machung.ac.id](mailto:mochamad.subianto@machung.ac.id)

## Abstrak

Video conference merupakan aplikasi yang memungkinkan pengguna untuk berkomunikasi, melakukan rapat, belajar dan berbagi informasi. Hingga saat ini, sudah banyak aplikasi video conference yang digunakan, seperti Zoom, Google Meet, dan Microsoft Teams. Tentunya disetiap aplikasi terdapat ulasan mengenai aplikasi yang digunakan. Dengan adanya ulasan, maka pengguna lain dapat mempertimbangkan dan mengetahui lebih jauh terkait aplikasi yang digunakan. Namun karena jumlah ulasan di internet sangat banyak, maka diperlukan analisis sentimen untuk dapat mengklasifikasikan ke dalam sentimen positif, netral, atau negatif. Pada penelitian ini, sebanyak 400 data ulasan setiap aplikasi diambil dari Play Store. Data ulasan tersebut menggunakan Bahasa Indonesia dan diambil berdasarkan versi aplikasi terbaru. Sebelum digunakan, data perlu dilabeli dan dilakukan preprocessing terlebih dahulu untuk dapat melakukan klasifikasi sentimen. Proses pengerjaannya menggunakan pembobotan kata TF-IDF, klasifikasi Naïve Bayes, dan confusion matrix sebagai bahan evaluasi. Nantinya data ulasan dari ketiga aplikasi tersebut akan diproses secara terpisah. Dari hasil pengujian analisis sentimen menggunakan 100 data uji, didapatkan perbandingan jumlah sentimen positif antara kelas aktual (55 ulasan Zoom, 52 ulasan Google Meet, dan 47 ulasan Microsoft Teams) dengan kelas prediksi (90 ulasan, 76 ulasan, dan 71 ulasan). Selain itu, juga didapatkan nilai akurasi rata-rata dari ketiga aplikasi sebesar 69%.

**Kata kunci:** video conference, analisis sentimen, TF-IDF, naïve bayes, confusion matrix

## Abstract

Video conferencing is an application that allows users to communicate, conduct meetings, learn and share information. Until now, many video conferencing applications have been used, such as Zoom, Google Meet, and Microsoft Teams. Of course, in every application there is a review of the application used. With reviews, other users can consider and find out more about the applications used. However, because the number of reviews on the internet is very large, a sentiment analysis is needed to be able to classify sentiment into positive, neutral, or negative sentiments. In this study, 400 data reviews for each application were taken from the Play Store. The review data uses Indonesian and is taken based on the latest application version. Before being used, the data needs to be labeled and preprocessed in order to be able to classify sentiments. The process uses TF-IDF word weighting, Naïve Bayes classification, and confusion matrix as evaluation materials. Later, the review data from the three applications will

be processed separately. From the results of the sentiment analysis test using 100 test data, a comparison of the number of positive sentiments was obtained between the actual class (55 reviews for Zoom, 52 reviews for Google Meet, and 47 reviews for Microsoft Teams) and the prediction class (90 reviews, 76 reviews, and 71 reviews). In addition, the average accuracy value of the three applications is 69%.

**Keywords:** video conference, sentiment analysis, TF-IDF, naïve bayes, confusion matrix

## 1. PENDAHULUAN

Dengan kehadiran pandemi Covid-19 telah memberikan dampak yang sangat signifikan dalam segala bidang. Bahkan dalam waktu singkat, pandemi ini telah menyebar ke seluruh dunia dan menimbulkan banyak korban jiwa (Lombok Post, 2020). Berdasarkan data yang diperoleh per 16 Februari 2021, jumlah kasus positif di dunia sudah mencapai 109.491.385 dan korban jiwanya sudah mencapai 2,418,543. Sementara di Indonesia, jumlah kasus positif sudah mencapai 1.233.959 dan korban jiwanya sudah mencapai 33.596, dengan tambahan kasus positif sebanyak 10.029 per harinya. Parahnya, negara Indonesia sudah menjadi peringkat ke-19 dari negaranegara yang terpapar Covid-19 (Satuan Tugas Penanganan Covid-19, 2021). Oleh karena itu, banyak tempat-tempat umum dan penginapan yang ditutup, ekonomi dunia yang merosot drastis, kehidupan sosial masyarakat dunia yang berubah, tenaga medis dan rumah sakit yang terbatas. Apalagi ditambah dengan kebijakan PSBB, work from home, serta protokol kesehatan yang ketat (Warta Ekonomi, 2021). Hal tersebut membuat sebagian besar orang tidak dapat bertemu atau berkegiatan secara fisik sehingga semua kegiatan dilakukan secara online di rumah, termasuk sekolah.

Kegiatan sekolah yang dilakukan secara daring (online) dilaksanakan guna memberikan pengalaman belajar yang bermakna bagi peserta didik (Kementerian Riset Teknologi dan Pendidikan Tinggi, 2016). Namun dalam penerapannya, muncul berbagai masukan dan keluhan mengenai belajar dari rumah. Berdasarkan survei cepat yang dilakukan oleh kemendikbud dan UNICEF di Indonesia, sekitar 45 juta siswa telah didukung melalui pembelajaran jarak jauh. Namun keluhan yang didapatkan

yaitu sekitar 35% siswa melaporkan koneksi internet yang buruk, sedangkan sisanya mengalami kesulitan saat belajar. Kesulitan yang dirasakan yaitu kurangnya konsentrasi, belajar menjadi tidak maksimal, mudah jenuh dan lelah, serta lingkungan belajar yang tidak mendukung (UNICEF, 2020). Menanggapi hal tersebut, tentunya diperlukan aplikasi *video conference* yang dapat diandalkan, mengingat dampak dari pandemi yang begitu besar.

Hingga saat ini, sudah tersedia banyak aplikasi *video conference* yang beredar dan dapat digunakan secara gratis. Namun karena setiap aplikasi memiliki kelebihan dan kekurangan tersendiri, membuat pengguna merasa kesulitan untuk menentukan aplikasi yang terbaik. Misalnya, aplikasi Zoom dinilai kurang aman sehingga data pengguna sangat rawan disalahgunakan oleh pihak-pihak yang tidak berwenang, dan sebagainya.

Disetiap aplikasi, juga terdapat rating dan ulasan pengguna mengenai pelayanan dan fitur-fitur yang diberikan. Ulasan tersebut bisa berupa saran, kritik, maupun keluhan. Hal tersebut menjadi sangat bermanfaat dan membantu bagi pengguna lain yang akan menggunakan aplikasi tersebut. Tapi dalam penerapannya, tidak sedikit dijumpai adanya ketidaksesuaian antara rating dengan ulasan pengguna (Farki, Baihaqi, & Wibawa, 2016). Dari aplikasi yang ada saat ini, belum ditemukan adanya fitur analisis sentimen untuk mengelompokkan atau menyaring antara ulasan positif dan negatif, mengingat jumlah ulasan aplikasi yang sangat banyak dan akan terus bertambah setiap saat.

Penelitian terdahulu yang terkait dengan masalah analisis sentimen, yaitu. "Perbandingan Klasifikasi Tugas Akhir Mahasiswa Jurusan Teknik Informatika Menggunakan Metode Naïve Bayes Classifier Dan KNearest Neighbor". Sumber data berasal dari kumpulan abstrak Bahasa Indonesia dan daftar isi laporan tugas akhir jurusan Teknik Informatika. Pengambilan data tersebut diambil secara acak dalam rentang tahun 2012-2015. Sedangkan jumlah data yang akan diklasifikasikan sebanyak 100 data. Persentasi pembagian datasetnya yaitu 90% data latih dan 10% data uji. Mengenai tahap pengujiannya, dilakukan dengan menggunakan *10-fold cross validation*. Dari analisis tersebut, didapatkan hasil bahwa akurasi metode Naïve Bayes sebesar 87%, sedikit lebih baik dibandingkan dengan metode KNN sebesar 84% (Yusra, Olivita, & Vitriani, 2016).

Penelitian yang lainnya, yaitu "Analisis Sentimen Tentang Opini Film pada Dokumen Twitter Berbahasa Indonesia Menggunakan Naive Bayes dengan Perbaikan Kata Tidak Baku". Sumber data yang digunakan yaitu kumpulan tweet mengenai opini film berbahasa Indonesia, berjumlah 200 data opini. Kemudian dibagi menjadi 140 data latih dan 60 data uji, dimana jumlah pembagian kelasnya antara positif dengan negatif sama besar. Mengenai tahap pengujiannya, terdapat 3 macam pengujian yang dilakukan. Pengujian pertama dilakukan dengan tujuan untuk mengetahui pengaruh dari preprocessing.

Pengujian kedua dilakukan untuk mengetahui pengaruh dari proses perbaikan kata tidak baku. Pengujian ketiga dilakukan untuk mengetahui pengaruh dari preprocessing dengan perbaikan kata tidak baku dan normalisasi Levenshtein Distance. Berdasarkan hasil dari pengujian yang telah dilakukan, diperoleh hasil akurasi terbaik dari analisis sentimen tentang opini film pada dokumen twitter berbahasa Indonesia menggunakan Naive Bayes dengan perbaikan kata tidak baku sebesar 98.33%. Dan untuk nilai precision, recall, dan f-measure secara berurutan adalah 96.77%, 100%, dan 98.36% (Antinasari, Perdana, & Fauzi, 2017).

Berdasarkan hasil dari penelitian terdahulu, maka didapatkan nilai akurasi yang lebih dari 85%. Akurasi menunjukkan bahwa hasil prediksi kelas dari data uji sesuai dengan kelas yang sesungguhnya (aktual). Artinya metode Naïve Bayes dapat melakukan klasifikasi sentimen dengan sangat baik. Oleh karena itu pada penelitian ini, penulis akan menggunakan metode Naïve Bayes

Pada penelitian ini, penulis melakukan analisis sentimen pada aplikasi *video conference* yang paling populer, yaitu Microsoft Teams, Google Meet, dan Zoom. Data ulasan yang akan diambil merupakan data ulasan terbaru pada waktu tertentu, yang didapat dari aplikasi Play Store. Nantinya data ulasan tersebut akan dilabeli kedalam 3 kelas (positif, netral, dan negatif) dan diproses sendirisendiri sehingga dapat mengetahui hasil sentimen dan confusion matrix dari setiap aplikasi. Pengguna juga dapat mengetahui manakah aplikasi *video conference* yang terbaik berdasarkan hasil sentimen yang sudah ada.

## 2. TINJAUAN PUSTAKA

### 2.1 Data Mining

Data mining atau biasa disebut sebagai *Knowledge Discovery in Database* (KDD) adalah proses penemuan informasi penting berdasarkan kumpulan data yang jumlahnya sangat besar. Seringkali dalam proses data mining menerapkan berbagai metode probabilitas dan statistik, *database systems*, hingga memanfaatkan *artificial intelligence*, dan *machine learning* (Suyanto, 2017).

Banyak sekali permasalahan data mining yang dapat ditemukan dalam kehidupan sehari-hari, seperti transaksi bank, permainan saham, penjualan barang, dan sebagainya. Hal tersebut disebabkan karena data-data yang terus bertambah banyak dan semakin kompleks setiap harinya, sehingga data-data tersebut tidak dapat ditangani dengan baik. Untuk menyelesaikan permasalahan tersebut, diperlukan proses data mining secara bertahap, mulai dari *selection*, *preprocessing*, *transformation*, hingga *knowledge*.

### 2.2 Web Scraping

*Web scraping* merupakan suatu teknik untuk mengumpulkan data yang bersumber dari internet,



umumnya berupa halaman website dalam bahasa markup seperti HTML atau XHTML (Mitchell, 2018). *Web scraping* ini tidak dapat dimasukkan dalam bidang data mining karena tidak menyiratkan upaya untuk memahami pola tertentu dari sejumlah besar data yang telah diperoleh. Cara kerja dari *web scraping* yaitu dengan mengakses halaman website, memilih informasi tertentu yang ingin didapat, kemudian dilakukan proses ekstraksi data, dan terakhir menyimpan data tersebut menjadi dataset yang terstruktur (Boeing & Waddell, 2017).

### 2.3 Analisis Sentimen

Analisis sentimen (*opinion mining*) merupakan salah satu cabang dari data mining yang bertujuan untuk melakukan analisis terhadap pendapat, emosi, kritik, saran, maupun evaluasi yang disampaikan secara tekstual (Falahah & Nur, 2015). Analisis sentimen dibutuhkan untuk mendapatkan informasi yang bermanfaat bagi pengguna, serta evaluasi bagi perusahaan/developer. Secara umum, terdapat 5 langkah yang perlu diperhatikan dalam melakukan analisis sentimen, Berikut merupakan penjelasannya:

#### 1. Data

Data merupakan langkah awal yang harus dibutuhkan. Tanpa adanya data, maka analisis sentimen tidak dapat dilakukan. Pada kasus ini, data yang dimaksud adalah data tekstual yang sudah dilabeli (positif, negatif, dan netral) pada setiap data.

#### 2. Preprocessing

*Preprocessing* digunakan untuk membentuk “golden dataset”. Pada kasus ini, terdapat 4 langkah *preprocessing* yang dilakukan, yaitu:

- Case Folding
- Tokenization
- Stopword Removal
- Stemming

#### 3. Feature Extraction (Ekstraksi Fitur)

Merupakan tahapan awal komputasi yang berguna untuk menggali informasi penting dan merepresentasikan data secara menyeluruh. *Feature extraction* inilah yang nantinya digunakan sebagai input untuk *machine learning* pada tahap selanjutnya.

Secara umum terdapat 3 teknik ekstraksi fitur yang umum dilakukan, yaitu:

- Bag of Word
- Word Embedding
- Character Embedding

#### 4. Modelling

Setelah ekstraksi fitur, maka dilanjutkan dengan tahapan pemodelan (*modelling*). Analisis sentimen ini

menggunakan metode *supervised learning* untuk menyelesaikan masalah.

Berikut merupakan beberapa metode *supervised learning* yang umum digunakan:

#### a. Machine learning algorithm:

- Support Vector Classification (SVM)
- Naive Bayes
- K- Nearest Neighbor (KNN)
- Maximum Entropy
- Ensemble Classifier

#### b. Deep learning:

- Long Short Term Memory — RNN
- Bidirectional LSTM — RNN

### 5. Evaluation

Evaluasi (pengujian) digunakan untuk mengetahui seberapa efektif model tersebut bekerja dalam menyelesaikan permasalahan. Terdapat beberapa kriteria evaluasi model dalam masalah *supervised learning* yaitu akurasi, presisi, *recall*, dan *F1 score*. Untuk mengetahui evaluasi dari metode yang digunakan, diperlukan tabel confusion matrix. Berikut ini merupakan model dasar dari tabel confusion matrix (berukuran 2 x 2).

Tabel 1. Model Confusion Matrix

Actual Class	Prediction Class	
	Positive	Negative
Positive	True Positive (TP)	False Negative (FN)
Negative	False Positive (FP)	True Negative (TN)

Dari tabel tersebut, maka nilai akurasi, presisi, recall, dan F1 score bisa didapat. Akurasi menggambarkan seberapa besar akurat suatu model dapat mengklasifikasikan dengan benar. Presisi menggambarkan seberapa besar tingkat “keseragaman” dari semua hasil prediksi. *Recall* menggambarkan seberapa besar sistem dapat mengelompokkan data uji dari suatu kelas tertentu ke dalam kelas yang benar. Sedangkan *F1 score* merupakan kombinasi antara presisi dengan *recall*.

$$\text{accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \quad (1)$$

$$\text{precision} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{recall} = \frac{TP}{TP + FN} \quad (3) \text{ F1 Score}$$

$$= \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (4)$$

precision + recall

Keterangan:

TP = Banyaknya data positif yang sudah dikelompokkan dengan benar

TN = Banyaknya data negatif yang sudah dikelompokkan dengan benar

FN = Banyaknya data positif namun salah dikelompokkan oleh sistem

FP = Banyaknya data negatif namun salah dikelompokkan oleh sistem

Data positif = data uji kelas aktual positif

Data negatif = data uji kelas aktual negatif

## 2.4 Pembobotan Kata TF-IDF

TF - IDF (*Term Frequency-Inverse Document Frequency*) merupakan suatu metode yang berguna untuk menghitung bobot dari setiap kata yang digunakan (Riyani, Naf'an, & Burhanuddin, 2019). Metode ini menghitung nilai TF dan IDF pada setiap kata. Semakin banyak frekuensi kemunculan dari kata tersebut, maka semakin besar nilai TF-nya.

IDF menunjukkan bagaimana suatu term didistribusikan secara luas dalam dokumen, Semakin sedikit frekuensi kemunculan dari kata tersebut, maka semakin besar nilai IDF-nya. Untuk menghitung nilai IDF, digunakan rumus

$$IDF_j = \log (D/df_j) \quad (5)$$

Keterangan: D = jumlah semua dokumen

df<sub>j</sub> = frekuensi term j

IDF<sub>j</sub> = Inverse Document Frequency term j

Sedangkan pembobotan TF-IDF merupakan penggabungan dari TF dan IDF. Dalam penerapannya, jenis TF yang seringkali digunakan yaitu raw TF sehingga untuk perhitungannya dilakukan dengan cara mengalikan kedua nilai tersebut. Untuk perhitungan rumusnya, dapat ditulis sebagai berikut

$$w_{ij} = TF_{ij} \times IDF_j \quad (6)$$

dengan  $w_{ij}$  merupakan bobot dari term (j) terhadap dokumen (i),  $TF_{ij}$  merupakan frekuensi term (j) dalam dokumen.

## 2.5 Naïve Bayes

Naïve Bayes merupakan salah satu algoritma data mining untuk mengklasifikasi suatu objek menggunakan probabilitas dan statistik, serta dapat memprediksi peluang berdasarkan dari pengalaman sebelumnya. Ciri dari algoritma Naïve Bayes ini adalah independensi yang sangat kuat (*naïve*) dari masing-masing kondisi atau kejadian

(Nafalski & Wibawa, 2016). Algoritma ini termasuk dalam *supervised learning* karena membutuhkan data training sebelum proses klasifikasi. Secara umum rumus dasar persamaan teorema Bayes adalah:

$$P(H|X) = \frac{p(X|H) \times P(H)}{p(X)} \quad (7)$$

Keterangan:

X = Data dengan kelas yang belum diketahui

H = Hipotesis X merupakan suatu kelas spesifik

P(H|X) = Probabilitas H berdasarkan kondisi X

P(H) = Probabilitas pada hipotesis H (prior)

P(X|H) = Probabilitas X berdasarkan hipotesis H

P(X) = Probabilitas X (data sampel yang diamati)

Berdasarkan rumus tersebut, maka peluang masuknya sampel tertentu ke dalam kelas H adalah peluang munculnya kelas H (sebelum masuknya sampel tersebut), dikalikan dengan peluang kemunculan karakteristik sampel pada kelas H, dibagi dengan peluang kemunculan karakteristik sampel keseluruhan (Bustami, 2013).

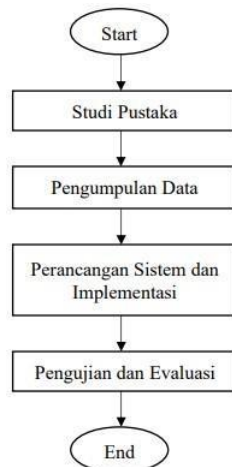
Dalam penerapannya terdapat 3 varian Naive Bayes yang umum digunakan, yaitu BernoulliNB, GaussianNB, dan MultinomialNB. Gaussian Naive Bayes digunakan untuk menghitung probabilitas dari suatu data kontinyu terhadap kelas tertentu. Perhitungannya menggunakan rumus dari Densitas Gauss dan ditandai dengan dua buah parameter, yaitu rata-rata dan standar deviasi

Bernoulli Naive Bayes menggunakan angka biner dalam melakukan klasifikasi. Naive Bayes jenis ini menggunakan data diskrit dan menerima fitur hanya sebagai nilai biner, seperti benar atau salah, ya atau tidak, berhasil atau gagal, 0 atau 1 dan seterusnya.

Multinomial Naive Bayes mengasumsikan tentang independensi diantara kemunculan kata-kata dalam dokumen, tanpa memperhitungkan urutan kata dan konteks informasi yang ada. Selain itu, juga memperhitungkan jumlah kemunculan kata dalam dokumen.

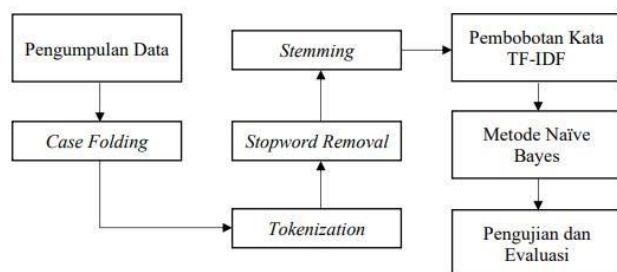
## 3. METODOLOGI

Metodologi penelitian menjelaskan bagaimana cara kerja suatu sistem secara keseluruhan, mulai dari tahap studi pustaka, pengumpulan data, perancangan sistem dan implementasi, hingga tahap pengujian dan evaluasi. Alur penelitian yang akan digunakan yaitu seperti pada gambar 1 berikut. Selain itu, akan dijelaskan secara lengkap mengenai masing-masing tahapan dalam alur penelitian



**Gambar 1. Alur Penelitian**

Pada penelitian ini, data yang akan digunakan yaitu data ulasan pengguna pada aplikasi Zoom, Google Meet, dan Microsoft. Dikarenakan jumlah ulasan yang sangat banyak dari suatu aplikasi, maka hanya diambil sebagian data saja (sampel data), yang berjumlah 400 ulasan dari setiap aplikasi. Mengenai perbandingan jumlah data latih dan data uji yang dilakukan, yaitu sebesar 75% dan 25%. Dataset tersebut akan dilabeli secara manual ke dalam sentimen positif, netral, dan negatif. Kemudian dilanjutkan proses analisis sentimen, seperti yang terlihat pada gambar 2 berikut.



**Gambar 2. Proses Analisis Sentimen**

(Sumber: Antinasari, Perdana, & Fauzi, 2017)

*Preprocessing* dilakukan untuk mengurangi tingkat kompleksitas data, tanpa mengurangi makna kalimat yang ada. Berikut ini merupakan penjelasan tahapan *preprocessing* yang dilakukan.

#### a. Case Folding

Merupakan tahapan *preprocessing* yang paling sederhana. Tidak semua data ulasan konsisten dalam hal penggunaan huruf kapital sehingga diperlukan tahapan ini. Tujuannya yaitu mengubah semua huruf kapital menjadi *lower case*. Hanya huruf kecil 'a' sampai 'z' saja yang diterima, sedangkan karakter selain huruf dianggap sebagai *delimiter*.

Pada tahapan ini, data ulasan yang ada akan dibersihkan terlebih dahulu. Jika tidak, maka

karakterkarakter selain huruf juga ikut diproses. Contoh-contoh pembersihan data yang umum digunakan, diantaranya yaitu tanda baca, angka, emoji, dan whitespace. Setelah dilakukan pembersihan data, maka data ulasan akan dipecah menjadi kumpulan kata-kata (token). Biasanya, yang menjadi acuan dalam proses *tokenizing* adalah tanda spasi.

#### c. Stopword Removal

Tahapan ini bertujuan untuk mengambil kata-kata penting dari hasil token yang sudah dilakukan sebelumnya. Dalam penerapannya, masih ditemukan banyak sekali kata-kata yang tidak mempunyai makna penting dalam suatu kalimat sehingga perlu dihapus. Selain itu, juga dapat mengurangi level noise. Secara umum, kata-kata tersebut sangat sering digunakan dan hampir ditemukan disemua ulasan yang ada.

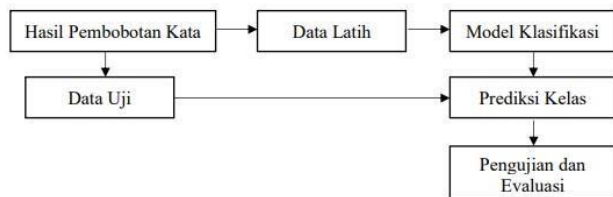
#### d. Stemming

*Stemming* digunakan untuk mengubah semua kata menjadi kata dasar. *Stemming* antara satu bahasa dengan bahasa yang lain tentunya berbeda. Untuk isi ulasan dengan berbahasa Indonesia tergolong lebih rumit/kompleks karena terdapat berbagai variasi imbuhan yang harus dibuang untuk mendapatkan kata dasar. Pada umumnya penggunaan kata bahasa Indonesia terdiri dari kombinasi prefiks (awalan), kata dasar, dan sufiks (akhiran).

Hasil dari tahap *stemming* tidak bisa langsung di proses karena masih berupa kumpulan array of string, sehingga semua array-nya perlu digabungkan terlebih dahulu. Dari sinilah, tahap pembobotan kata TF-IDF baru bisa dilakukan. Mengenai proses pengerjaannya dilakukan dengan menggunakan fungsi *TfidfVectorizer* Fungsi ini dipilih karena dapat menghitung hasil TFIDF dengan efisien, mudah, dan akurat. Nantinya luaran dari fungsi ini akan berupa vektor yang berisi indeks dan nilai TF-IDF nya.

Setelah melakukan tahap *preprocessing* dan pembobotan kata, maka dilanjutkan dengan proses klasifikasi Naïve Bayes. Namun sebelum melakukan klasifikasi, dataset akan dipecah menjadi data latih dan data uji terlebih dahulu. Pada tahapan ini, data latih diproses terlebih dahulu untuk membangun model klasifikasi. Dari model klasifikasi tersebut nantinya digunakan untuk memprediksi kelas pada data uji. Output dari proses ini yaitu pengelompokan ke dalam sentimen positif, netral, maupun negatif. Mengenai proses klasifikasinya, dapat dilihat pada gambar 3 berikut.





**Gambar 3. Proses Klasifikasi Naïve Bayes**  
(Sumber: Ratnawati, 2018)

Tahap pengujian diperlukan untuk mengetahui kinerja dari suatu sistem. Pengujian sistem pada analisis sentimen ini dilakukan dengan model *confusion matrix*, yaitu sebuah matriks yang dapat membandingkan hasil dari keluaran sistem dengan hasil yang diharapkan.

**4. HASIL DAN PEMBAHASAN**

Pada ulasan suatu aplikasi terdapat beragam informasi yang bisa didapat. Informasi tersebut nantinya sangat membantu baik bagi pengguna maupun bagi developer. Namun pada penelitian ini, isi dataset yang akan diambil hanya data ulasan dan tanggal saja. Sedangkan jumlah sampelnya sebanyak 400 untuk setiap aplikasi. Hal tersebut dikarenakan aplikasi yang dipilih merupakan aplikasi yang populer dan banyak digunakan hingga saat ini, sehingga jumlah ulasan yang ada sangat banyak (ratusan ribu) dan pengolahannya memakan waktu yang sangat lama. Gambar 4 berikut merupakan tampilan program untuk mendapatkan dataset.



**Gambar 4. Tampilan Aplikasi**

Setelah mendapatkan dataset, maka dilanjutkan dengan *preprocessing*. Tabel 2 berikut merupakan contoh ulasan dan penerapan *preprocessing*-nya.

**Tabel 2. Penerapan Preprocessing**

Tahapan	Contoh Ulasan
Case Folding	tolong nanti update selanjutnya untuk hp dikasih virtual background dan fitur lainnya dilengkapi

Tokenization	'tolong', 'nanti', 'update', 'selanjutnya', 'untuk', 'hp', 'dikasih', 'virtual', 'background', 'dan', 'fitur', 'lainnya', 'dilengkapi'
Stopword Removal	'update', 'selanjutnya', 'hp', 'dikasih', 'virtual', 'background', 'fitur', 'lainnya', 'dilengkapi'
Stemming	'update', 'lanjut', 'hp', 'kasih', 'virtual', 'background', 'fitur', 'lain', 'lengkap'

Berdasarkan contoh ulasan pada tabel 2, maka *feature names* yang didapat yaitu background, fitur, hp, kasih, lain, lanjut, lengkap, update, dan virtual. Sedangkan data ulasannya akan berubah menjadi "update lanjut hp kasih virtual background fitur lain lengkap". *Feature names* merupakan array yang berisi kumpulan kata-kata yang digunakan pada keseluruhan ulasan. Walaupun terdapat beberapa kata yang sama, namun pada array itu tetap ditulis sekali saja. Array tersebut sudah diurutkan sesuai abjad sehingga untuk indeks yang ke 0 akan berisi kata "background", indeks ke 1 berisi kata "fitur", dan seterusnya. Nantinya dalam proses TF-IDF akan menggunakan hasil dari *preprocessing* yang sudah dirubah dalam bentuk kalimat dan *feature names*. Output dari TFIDF akan digunakan pada tahap klasifikasi sentimen.

Klasifikasi sentimen merupakan bagian yang terpenting dalam proses analisis sentimen. Sebelum diproses lebih lanjut, dataset perlu dipecah menjadi data latih dan data uji terlebih dahulu. Mengenai pembagian dataset dapat dilihat pada tabel 3 berikut.

**Tabel 3. Pembagian Dataset**

Aplikasi	Data Latih			Data Uji		
	Positif	Netral	Negatif	Positif	Netral	Negatif
Zoom Cloud Meeting	161	66	73	55	15	30
Google Meet	124	76	100	52	15	33
Microsoft Teams	156	32	112	47	19	34

Namun pada penelitian ini, metode Naive Bayes dilakukan dengan menggunakan MultinomialNB(). Fungsi ini dipilih karena jumlah kelasnya bisa lebih dari 2 dan mudah untuk diimplementasikan. Model klasifikasi Naïve Bayes ini membutuhkan dua parameter, yaitu data latih dari hasil TF-IDF dan kelompok sentimennya.



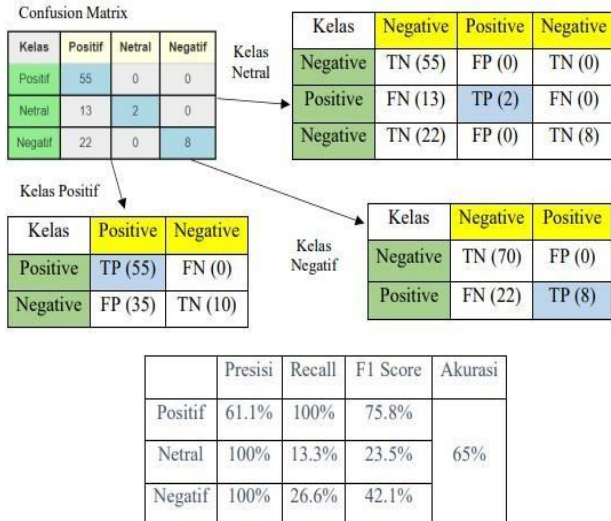
Actual	Prediction	Negatif	Netral	Positif
Netral	Positif	0.15319204721	0.110215782503	0.774988002775
Positif	Positif	0.152125862590	0.110441124819	0.734433850093
Positif	Positif	0.303338467132	0.071250862293	0.645488522084
Positif	Positif	0.22668621304	0.150618852536	0.622773700067
Negatif	Positif	0.230502981145	0.148834579550	0.620532438303
Positif	Positif	0.209669889899	0.11775	0.812860008800
Positif	Positif	0.163525785150	0.142633467307	0.693831751433
Positif	Positif	0.167878711900	0.148410519820	0.682758813204

Gambar 5. Tampilan Hasil Prediksi Naïve Bayes

Gambar 5 tersebut merupakan contoh hasil prediksi setiap kelas dalam bentuk probabilitas (sebelah kanan) dan kelas prediksinya (sebelah kiri). Terlihat bahwa setiap baris ulasan memiliki 3 nilai probabilitas.

Nantinya, nilai probabilitas yang paling besar akan dijadikan sebagai hasil kelas prediksi.

Dengan membandingkan hasil prediksi data uji dengan kelompok sentimen yang sebenarnya, maka perhitungan dari confusion matrix dapat dilakukan. Confusion matrix merupakan matriks yang berisi ketepatan prediksi data uji. Namun perlu diingat bahwa confusion matrix perlu disesuaikan dengan jumlah kelasnya sehingga untuk hasil confusion matrix-nya berupa matrix 3 x 3. Perhitungannya pun juga menjadi sedikit lebih kompleks karena perlu menghitung TP, FP, FN, dan TN dari setiap kelas yang ada.

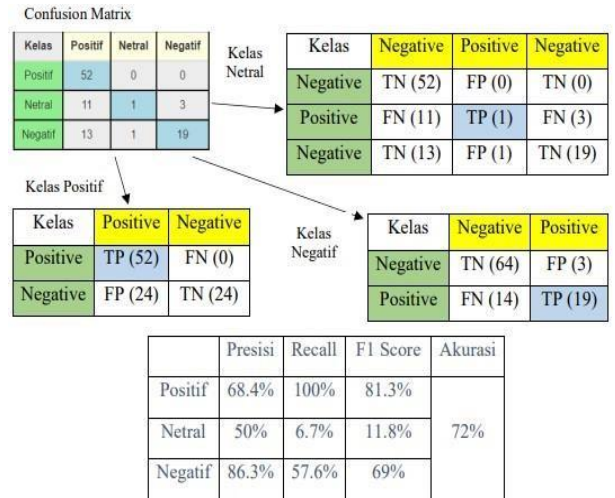


Gambar 6. Perhitungan Confusion Matrix Zoom

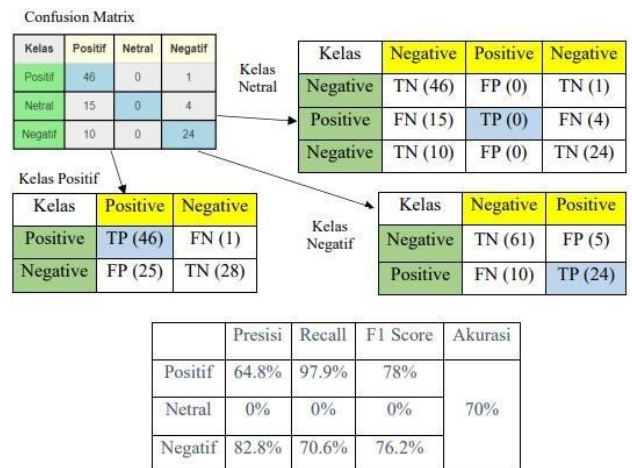
Gambar 6 tersebut menjelaskan bagaimana cara menghitung nilai TP, FP, FN, dan TN pada setiap kelas, beserta dengan hasilnya. Untuk warna hijau pada kolom merupakan kelas data yang sebenarnya (aktual), sedangkan warna kuning merupakan kelas data prediksi. Tabel kiri atas merupakan hasil confusion matrix dari proses klasifikasi sebelumnya, sedangkan tabel lainnya merupakan penjabaran setiap kelasnya. Nantinya dalam perhitungan ke empat variabel tersebut, kelas yang dipilih akan dianggap

sebagai "Positive" sedangkan kelas lainnya akan dianggap sebagai "Negative".

Berdasarkan data-data dari hasil perhitungan di atas, maka setidaknya bisa mengetahui seberapa baik atau buruk kinerja dari metode klasifikasi yang telah digunakan. Hal yang sama juga berlaku untuk aplikasi Google Meet dan Microsoft Teams. Dengan menggunakan contoh perhitungan pada aplikasi Zoom sebelumnya, maka hasil confusion matrix dari kedua aplikasi tersebut juga dapat dilihat pada gambar 7 dan gambar 8 berikut.

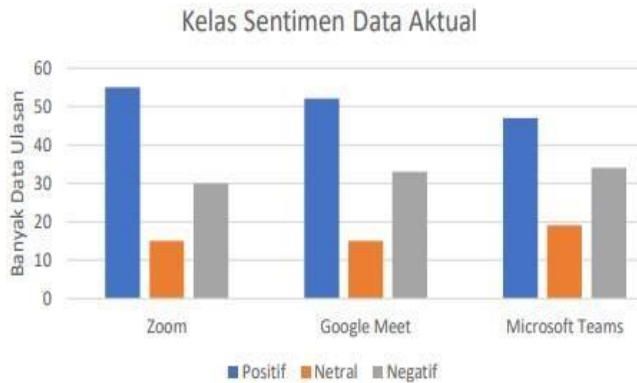


Gambar 7. Perhitungan Confusion Matrix Google Meet

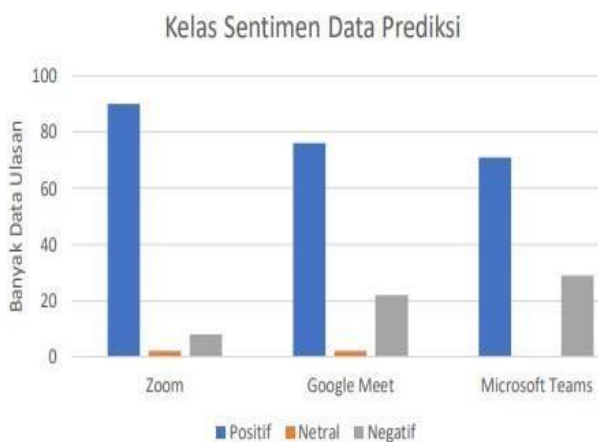


Gambar 8. Perhitungan Confusion Matrix Microsoft Teams

Sebagai penutup, pada gambar 9 dan gambar 10 berikut merupakan hasil grafik perbandingan jumlah sentimen dari setiap aplikasi, baik data aktual maupun data hasil prediksi.



Gambar 9. Hasil Grafik Data Aktual



Gambar 10. Hasil Grafik Data Prediksi

Dari ke dua gambar tersebut, terlihat adanya perbedaan jumlah sentimen yang cukup signifikan. Hal itu menunjukkan bahwa tingkat akurasi dari metode yang digunakan tidak terlalu tinggi, hanya sekitar 69%. Walaupun tidak terlalu tinggi, namun sebenarnya cukup mampu untuk mengklasifikasikan sentimen dengan baik. Hal tersebut dikarenakan terdapat *imbalance dataset*, beberapa ulasan yang sangat panjang dan sangat pendek, ulasan yang mengandung emoji, serta *labelling* dan hasil dari *preprocessing* yang masih “kurang sempurna”.

Bisa dilihat pula aplikasi mana yang memiliki jumlah sentimen positif terbanyak. Semakin banyak sentimen positif, berarti pengguna semakin “senang dan puas” dengan aplikasi tersebut. Pada kasus ini, Zoom memiliki jumlah sentimen positif terbanyak. Terlihat bahwa Zoom memiliki jumlah sentimen positif lebih banyak daripada aplikasi lain, baik dari data aktual maupun data dari hasil prediksi

## 5. KESIMPULAN

Berdasarkan hasil penelitian dari analisis sentimen yang telah dilakukan dengan menggunakan Naïve Bayes, maka dapat disimpulkan bahwa:

1. Nilai akurasi rata-rata yang didapat dari ke tiga aplikasi tersebut yaitu sekitar 69% (*poor classification*). Hal tersebut karena masih terdapat imbalance dataset, serta proses pelabelan dan preprocessing yang “kurang sempurna”.
2. Berdasarkan ulasan dari ke tiga aplikasi tersebut, didapatkan jumlah sentimen positif dari kelas aktual sebesar 55 ulasan Zoom, 52 ulasan Google Meet, dan 47 ulasan Microsoft Teams. Sedangkan jumlah sentimen positif dari kelas prediksi sebesar 90, 76, dan 71 ulasan. Dari hasil tersebut, maka Zoom memiliki perbedaan jarak yang paling besar antara data aktual dengan data prediksi. Namun jika berdasarkan jumlah kelas aktual, maka Zoom memiliki lebih banyak ulasan positif dibandingkan dengan aplikasi yang lain.

## DAFTAR PUSTAKA

- Antinasari, P., Perdana, R. Fauzi, M. (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 (JPTIIK)*, 1: 1733–1741.
- Boeing, G., Waddell, P. (2016). New Insights into Rental Housing Markets across the United States: Web Scraping and Analyzing Craigslist Rental Listings. *Journal of Planning Education and Research*, 37(4), 457–476.
- Bustami. (2013). Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Data Nasabah Asuransi, *TECHSI: Jurnal Penelitian Teknik Informatika*, Vol. 3, No. 2, 127-146.
- Falahah, Nur, D. D. A. (2015). Pengembangan Aplikasi Sentiment Analysis Menggunakan Metode Naïve Bayes. *Seminar Nasional Sistem Informasi Indonesia*, 335–340.
- Farki, A., Baihaqi, I., Wibawa, M. (2016). Pengaruh online customer review rating terhadap kepercayaan place di indonesia. *Jurnal Teknik ITS*, 5(2), A614–A619.
- Kementrian Riset Teknologi dan Pendidikan Tinggi (2016). Kebijakan Pendidikan Jarak Jauh dan E-Learning di Indonesia Pembelajaran berbasis Teknologi Group growth Individual growth. Ministry of Research, Technology, & Higher Education, 1–21.
- Lombok Post. (2020). Perubahan Sosial di Era Pandemi. (17 Februari 2021). *Citing Internet sources URL* <https://lombokpost.jawapos.com/opini/15/07/2020/perubahan-sosialdi-era-pandemi/>



- Mitchell, R. (2018). Web scraping with Python: collecting data from the modern web, Beijing: O'Reilly.
- Muliadi, Syarif, S., Salim, A. (2019). Penerapan Algoritma Naive Bayes pada Penilaian Kinerja Pemerintah Desa dalam Pengelolaan Dana Desa. Jurnal Riset Informatika. Vol. 1, No. 2, Maret 2019, 71-80.
- Nafalski, A., Wibawa, A. P. (2016). Machine Translation With Javanese Speech Levels Classification. Informatics, Control Measurement in Economy and Environment Protection, 6(1), 21–25.
- Ratnawati, F. (2018). Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter. INOVTEK Polbeng - Seri Informatika, 3(1)
- Riyani, A., Nafan, M. Z., Burhanuddin, A. (2019). “Penerapan Cosine Similarity dan Pembobotan TF-IDF untuk Mendeteksi Kemiripan Dokumen”. Jurnal Linguistik Komputasional. Vol. 2, No. 1, Maret 2019, 23-27.
- Satuan Tugas Penanganan Covid-19 (2021). Peta Sebaran COVID19 16 Februari 2021. (17 Februari 2021).
- Suyanto (2017), Data Mining Untuk Klasifikasi dan Klasterisasi Data. Bandung: Informatika.
- UNICEF (2020), COVID-19: Laporan baru UNICEF mengungkap setidaknya sepertiga anak sekolah di seluruh dunia tidak dapat mengakses pembelajaran jarak jauh selama sekolah ditutup. (21 Januari 2021). *Citing Internet sources URL*  
<https://www.unicef.org/indonesia/id/pressreleases/covid-19-laporan-baruunicef-mengungkapsetidaknya-sepertiga-anak-sekolah-di-seluruh>
- Warta Ekonomi (2021). Pemerintah Terapkan Kebijakan Ketat PSBB, Apa Saja Kegiatan yang Dibatasi?. (17 Februari 2021). *Citing Internet sources URL*  
<https://www.wartaekonomi.co.id/read321594/pemerintah-terapkankebijakan-ketat-psbb-apa-saja-kegiatan-yangdibatasi>
- Yusra., Olivita, D., Vitriani, Y. (2016). Perbandingan Klasifikasi Tugas Akhir Mahasiswa Jurusan Teknik Informatika Menggunakan Metode Naïve Bayes Classifier dan KNearest Neighbor. Jurnal Sains, Teknologi dan Industri, Vol. 14, No. 1, Desember 2016, 79 –85.

