

## **Bab II**

### **Tinjauan Pustaka**

#### **2.1 Penelitian Terdahulu**

Penelitian terdahulu digunakan sebagai bahan referensi dan perbandingan dalam melakukan penelitian. Dengan adanya ini, penulis dapat menambah wawasan yang nantinya akan digunakan dalam mengkaji penelitian. Selain itu, juga akan memudahkan penulis dalam menentukan langkah-langkah yang sistematis untuk penyusunan penelitian dari segi teori dan konsep. Biasanya penelitian terdahulu yang digunakan adalah penelitian yang terkait langsung dengan penelitian yang sedang dilakukan. Berikut ini merupakan penjelasan dari beberapa penelitian terdahulu yang terkait dengan analisis sentimen.

1. Yusra, Dhita Olivita, dan Yelfi Vitrian (2016)

Penelitian yang dibuat berjudul “Perbandingan Klasifikasi Tugas Akhir Mahasiswa Jurusan Teknik Informatika Menggunakan Metode Naïve Bayes Classifier Dan K-Nearest 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. Untuk kelas sentimen yang digunakan yaitu positif dan negatif. 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 metode klasifikasi Naïve Bayes, KNN, dan *10-fold-crossvalidation*. Dari analisis tersebut, didapatkan hasil bahwa akurasi metode Naïve Bayes sebesar 87%, sedikit lebih baik dibandingkan dengan metode KNN sebesar 84%

2. Prananda Antinasari, Rizal Setya Perdana, dan M. Ali Fauzi (2017)

Penelitian yang dibuat berjudul “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. Untuk

kelas sentimen yang digunakan yaitu positif dan negatif. 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. Namun ketiga tahap tersebut menggunakan metode klasifikasi Naïve Bayes. Berdasarkan hasil dari pengujian yang telah dilakukan, diperoleh hasil akurasi, *precision*, *recall*, dan *f-measure* sebesar 98.33%, 96.77%, 100%, dan 98.36%.

Tabel 2.1 Penelitian Terdahulu

| Peneliti  | Sumber Data   | Kelas               | Metode              | Hasil  |
|---|---|---------------------|---------------------|--|
| Yusra, Dhita Olivita, dan Yelfi Vitrian (2016)                    | Kumpulan abstrak Bahasa Indonesia dan daftar isi laporan tugas akhir jurusan Teknik Informatika | Positif dan Negatif | Naïve Bayes dan KNN | Akurasi Naïve Bayes 87%, akurasi KNN 84%   |
| Prananda Antinasari, Rizal Setya Perdana, dan M. Ali Fauzi (2017) | Kumpulan tweet mengenai opini film berbahasa Indonesia  | Positif dan Negatif | Naïve Bayes         | Akurasi 98.33%, <i>precision</i> 96.77%, <i>recall</i> 100%, <i>f-measure</i> 98.36% |

Tabel 2.1 merupakan perbedaan dari beberapa penelitian sebelumnya. Dari hasil penelitian terdahulu, didapatkan nilai akurasi yang tinggi (lebih dari 85%). Artinya metode Naïve Bayes dapat melakukan klasifikasi sentimen dengan baik, sehingga pada penelitian ini penulis akan menggunakan metode klasifikasi Naïve Bayes. Untuk kelas yang akan digunakan sedikit berbeda dengan penelitian

sebelumnya, yaitu positif, netral, dan negatif. Sedangkan untuk sumber data yang akan digunakan yaitu data ulasan aplikasi Zoom, Google Meet, dan Microsoft Teams yang ada di Play Store.

## 2.2 Data Mining

Data mining atau biasa disebut sebagai *Knowledge Discovery in Database* (KDD) adalah suatu proses penemuan informasi penting berdasarkan kumpulan data yang jumlahnya sangat besar (*big data*). 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, Apalagi dalam data yang besar, terdapat banyak *missing value*, *duplicate data*, *typo* dan sebagainya. Untuk menyelesaikan permasalahan tersebut, diperlukan proses data mining secara bertahap, mulai dari *selection*, *preprocessing*, *transformation*, hingga *knowledge*. Secara umum, tahapan pada data mining dapat dilihat pada gambar 2.1 berikut.



Gambar 2.1 *Knowledge Discovery in Database* (KDD)

(Sumber: [www.course-net.com/peranan-data-mining-dalam-perusahaan](http://www.course-net.com/peranan-data-mining-dalam-perusahaan))

Berikut ini merupakan penjelasan dari setiap tahapan KDD:

### 1. *Selection*

Pemilihan data perlu dilakukan sebelum tahap penggalian informasi. Hal itu perlu dilakukan karena tidak semua data yang ada akan diproses. Misalnya,

pada ulasan aplikasi terdapat nama pengguna, tanggal dan waktu, serta isi ulasan. Jika ingin melakukan analisis sentimen, maka data yang akan dipilih hanya isi ulasan saja.

## 2. *Preprocessing*

Sebelum melakukan data mining, perlu dilakukan *preprocessing* terlebih dahulu. Tahapan ini bertujuan untuk menghilangkan data-data yang tidak diperlukan. Diantaranya seperti *missing value*, *inconsistent data*, *typo*, dan *duplicate data*. Hasil dari tahapan *preprocessing* akan mempengaruhi *performance* dalam menggunakan metode data mining karena kompleksitas dan jumlah data menjadi berkurang.

## 3. *Transformation*

Sebelum dilakukan proses data mining, maka data yang sudah ada perlu diubah atau digabung terlebih dahulu ke dalam format yang sesuai. Hal tersebut dikarenakan beberapa metode pada data mining memerlukan format khusus sebelum diproses lebih lanjut.

## 4. Data mining

Tahapan ini merupakan tahapan utama untuk mendapatkan pengetahuan baru berdasarkan data yang diproses. Metode data mining yang tepat akan mempengaruhi pada tujuan, proses, dan hasil keseluruhan. Secara umum, pengelompokan pada proses data mining meliputi *prediction*, *clustering*, *classification*, *association*, dan *description*.

## 5. *Evaluation/Interpretation*

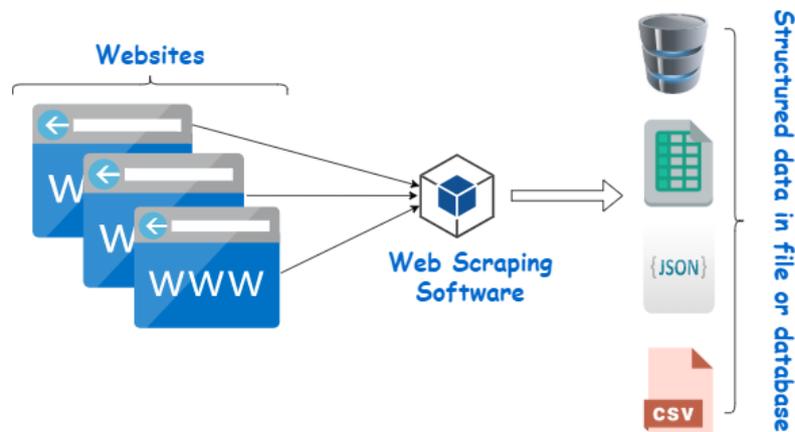
Evaluasi digunakan untuk mengidentifikasi pola-pola yang menarik ke dalam *knowledge based*. Hasil dari data mining perlu dievaluasi untuk mengetahui apakah sudah sesuai dengan hipotesis yang ada atau tidak.

## 6. *Knowledge*

Setelah dilakukan proses evaluasi, maka pola-pola yang dihasilkan akan dipresentasikan kepada pengguna. Visualisasi juga bisa membantu dalam menyampaikan hasil data mining, sehingga lebih menarik dan lebih mudah dimengerti. Pengetahuan baru yang dihasilkan bisa dipahami semua orang, serta dapat dijadikan acuan dalam pengambilan keputusan.

### 2.3 Web Scraping

*Web scraping* adalah 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).



Gambar 2.2 Cara Kerja *Web Scraping*

Terdapat 3 teknik yang umum digunakan dalam melakukan *web scraping*, diantaranya sebagai berikut:

1. HTML Parsing

HTML parsing ini dilakukan dengan cara menggunakan Javascript dan menargetkan halaman HTML. Teknik ini sering digunakan dalam proses parsing (penguraian data) karena dapat mengekstraksi data dengan cepat dan mudah.

2. DOM Parsing

DOM (*Document Object Model*) menentukan gaya, struktur serta konten yang terkandung dalam file XML (*Extensible Markup Language*). Biasanya teknik ini digunakan untuk mendapatkan informasi yang lebih mendalam mengenai struktur halaman website dan mengekstrak data menggunakan sebuah software tertentu, seperti XPath.

### 3. Google Sheets

Google Sheets dapat digunakan sebagai alat *web scraping*. Pengguna dapat memakai formula khusus untuk mengekstrak data dari halaman website, kemudian mengimpornya langsung ke dalam Google Sheets. Beberapa formula sheet yang biasa digunakan untuk *web scraping* di Google Sheet adalah ImportXML, ImportHTML, ImportFEED, dan ImportDATA.

## 2.4 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). Pendapat tersebut bisa mengenai produk, aplikasi, pelayanan publik, dan lain-lain. Analisis sentimen dibutuhkan untuk mendapatkan informasi yang bermanfaat bagi pengguna, serta evaluasi bagi perusahaan/developer.

Ada beragam tipe analisis sentimen yang digunakan untuk mengidentifikasi ulasan, hingga mengidentifikasi niat dan emosi pengguna. Berikut ini merupakan beberapa tipe analisis sentimen itu, diantaranya yaitu:

#### 1. *Fine-Grained Sentiment Analysis*

Tipe ini merupakan salah satu tipe yang paling umum digunakan, yaitu dengan cara mengelompokkan respon atau pendapat ke dalam beberapa kategori seperti positif, netral, dan negatif.

#### 2. *Intent Sentiment Analysis*

Tipe ini biasanya digunakan untuk mengidentifikasi dan menggali motivasi dari suatu pesan pengguna, untuk melihat apakah itu termasuk kategori keluhan, saran, pendapat, pertanyaan ataupun penghargaan.

#### 3. *Aspect - Based Sentiment Analysis*

Tipe ini berfokus pada elemen yang lebih spesifik dari suatu produk atau layanan. Analisis sentimen berbasis aspek ini juga dapat menghubungkan sentimen spesifik dengan berbagai aspek produk atau layanan.

Secara umum, terdapat 5 langkah utama yang perlu diperhatikan dalam melakukan analisis sentimen, Berikut ini merupakan paparan 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. Setelah dilabeli dan dicek kembali, maka data tersebut siap dilakukan *preprocessing*.

## 2. *Preprocessing*

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

a. *Case Folding*: Mengubah semua kalimat menjadi *lower case*.

b. *Tokenization*

Tahapan ini bertujuan untuk memecah dari suatu kalimat menjadi kumpulan kata dengan cara menghilangkan semua tanda baca dan memisahkannya berdasarkan spasi.

c. *Stopword Removal*

Pada tahapan ini, kata-kata yang dirasa tidak penting dan tidak ada kaitannya akan dihapus, dengan cara mengecek apakah kata-kata tersebut termasuk dalam daftar kata yang tidak penting (*stopword list*) atau tidak. Jika termasuk, maka kata tersebut akan dihapus.

d. *Stemming*

Tahapan ini bertujuan untuk mengubah kata menjadi kata dasar dengan cara menghilangkan semua imbuhan-imbuhan yang ada. Penggunaan algoritma untuk proses *stemming* bergantung pada bahasa yang digunakan.

Perlu diperhatikan bahwa tahapan ini dilakukan dengan tidak mengurangi makna yang terkandung di dalamnya. Jangan sampai kata-kata yang memiliki makna penting hilang dalam proses ini sehingga data tersebut tidak memiliki makna sama sekali.

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

Komputer tidak dapat mengolah data selain data numerik, sehingga dibutuhkan ekstraksi fitur untuk mengubah data yang berupa kata menjadi data numerik. Tahapan ini 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:

- a. *Bag of Word* (TF, TF-IDF)
- b. *Word Embedding* (Glove, Word2vec, FastText)
- c. *Character Embedding*

*Bag of Word* akan merubah kata menjadi besaran skalar, *Word Embedding* akan merubah kata menjadi vektor dengan dimensi tertentu, sedangkan *Character Embedding* akan merubah huruf menjadi vektor dengan dimensi tertentu. *Character Embedding* juga mampu dalam menangani masalah *Out of Vocabulary* (OOV).

#### 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 2x2).

Tabel 2.2 Model *Confusion Matrix*

| Actual Class | Prediction Class     |                      |
|--------------|----------------------|----------------------|
|              | Positive             | Negative             |
| Positive     | True Positives (TP)  | False Negatives (FN) |
| Negative     | False Positives (FP) | True Negatives (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 dan juga tingkat kedekatan antara hasil klasifikasi dengan hasil aktual. Presisi menggambarkan seberapa besar tingkat “keseragaman” dari semua hasil klasifikasi. *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*.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.1)$$

$$precision = \frac{TP}{TP + FP} \quad (2.2)$$

$$recall = \frac{TP}{TP + FN} \quad (2.3)$$

$$F1\ score = \frac{2 \times precision \times recall}{precision + recall} \quad (2.4)$$

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 yang memiliki kelas aktual positif

Data negatif = data uji yang memiliki kelas aktual negatif

Hasil dari *confusion matrix* dapat divisualisasikan dengan ROC (*Receiver Operating Characteristic*). ROC merupakan grafik dua dimensi dengan FP sebagai garis horizontal dan TP sebagai garis vertikal. ROC dapat menunjukkan akurasi dan membandingkan klasifikasi secara visual. ROC memiliki rentang nilai akurasi, yaitu:

- 0,90 – 1,00 = *Excellent classification*
- 0,80 – 0,90 = *Good classification*
- 0,70 – 0,80 = *Fair classification*
- 0,60 – 0,70 = *Poor classification*
- 0,50 – 0,60 = *Failure*

## 2.5 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 untuk mengetahui seberapa sering suatu kata muncul di dalam dokumen. Semakin banyak frekuensi kemunculan dari kata tersebut, maka semakin besar nilai TF-nya. Terdapat beberapa jenis TF yang dapat digunakan, yaitu:

1. TF biner, hanya memperhatikan apakah suatu kata atau term ada atau tidak dalam dokumen (1 atau 0).
2. TF murni (*raw TF*), dihitung berdasarkan jumlah kemunculan kata dalam suatu dokumen. Jika muncul tiga kali maka kata tersebut akan bernilai 3.
3. TF normalisasi, merupakan perbandingan antara frekuensi pada suatu *term* dengan nilai maksimum dari keseluruhan frekuensi *term* yang ada di dalam suatu dokumen.

$$TF = 0,5 + 0,5 \times \frac{f_{t,d}}{\max(f_{t,d})} \quad (2.5)$$

Keterangan:

$f_{t,d}$  = frekuensi *term* (t) dalam suatu dokumen (d).

TF = *Term Frequency*

4. TF logaritmik, digunakan untuk menghindari dominansi dokumen yang mengandung sedikit *term* dalam query, namun mempunyai frekuensi tinggi.

$$TF = 1 + \log (f_{t,d}) \quad ( 2.6 )$$

Keterangan:

$f_{t,d}$  = frekuensi *term* (t) dalam suatu dokumen (d).

TF = *Term Frequency*

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

$$IDF_j = \log (D/df_j) \quad ( 2.7 )$$

Keterangan:

D = jumlah semua dokumen dalam *database*

$df_j$  = frekuensi *term* j.

$IDF_j$  = *Inverse Document Frequency* untuk *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

$$w_{ij} = TF_{ij} \times IDF_j \quad ( 2.8 )$$

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

## 2.6 Naïve Bayes

Naïve Bayes merupakan salah satu algoritma data mining yang digunakan untuk mengklasifikasi suatu objek dengan menggunakan probabilitas dan statistik, serta dapat memprediksi peluang di masa depan 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 kategori *supervised learning* karena membutuhkan *data training* sebelum melakukan proses klasifikasi. Secara umum rumus dasar persamaan teorema Bayes adalah:

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)} \quad \text{Posterior} = \frac{\text{prior x likelihood}}{\text{evidence}} \quad (2.9)$$

Keterangan:

X = Data dengan kelas (label) yang belum diketahui

H = Hipotesis bahwa data X merupakan suatu kelas spesifik

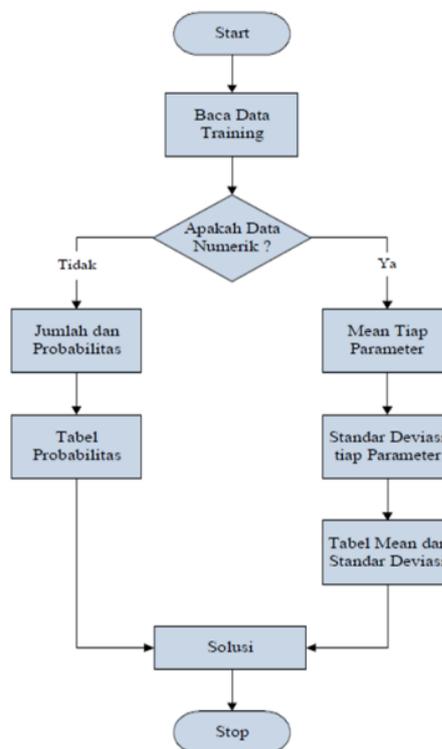
$P(H|X)$  = Probabilitas hipotesis H berdasarkan pada kondisi X

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

$P(X|H)$  = Probabilitas pada kondisi 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). Sedangkan mengenai alur dari metode Naïve Bayes secara umum dapat dilihat pada gambar 2.3 berikut.



Gambar 2.3 Alur Metode Naïve Bayes

(Sumber: Muliadi, Syarif & Salim, 2019)

Penjelasan mengenai gambar 2.3, yaitu sebagai berikut:

1. Membaca data training terlebih dahulu
2. Menghitung jumlah dan probabilitas. Apabila termasuk data numerik, maka
  - a. Mencari nilai *mean* dan standar deviasi dari setiap parameter yang merupakan data numerik.
  - b. Menghitung jumlah data yang sesuai dari kategori yang sama, dibagi dengan jumlah data pada kategori tersebut untuk mendapatkan nilai probabilitas.
3. Mendapatkan nilai tertentu dari tabel *mean*, standar deviasi dan probabilitas

Dalam penerapannya terdapat 3 varian Naive Bayes yang umum digunakan, yaitu BernoulliNB, GaussianNB, dan MultinomialNB. Berikut ini merupakan penjelasan dari setiap varian Naive Bayes.

- Gaussian Naive Bayes

Gaussian Naïve 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.

$$p(x_i | y_j) = \frac{1}{\sqrt{2\pi\sigma_j^2}} e^{-\frac{(x_i - \mu_j)^2}{2\sigma_j^2}} \quad (2.10)$$

Keterangan:

$P(x|y)$  : Probabilitas term x berada di dokumen pada kelas y

$\pi$  : Konstanta dengan nilai 3,14159...

$e$  : Konstanta eksponensial dengan nilai 2,7183...

$\mu_j$  : rata-rata pada kelas - j

$\sigma_j$  : standar deviasi pada kelas - j

- Bernoulli Naïve Bayes

Bernoulli Naïve Bayes menggunakan angka biner (hanya 0 dan 1) dalam melakukan klasifikasi. Naïve 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. Hal ini berbeda dengan varian lain, yang melakukan perhitungan term frekuensi pada setiap term.

- Multinomial Naïve Bayes

Multinomial Naïve 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.

$$\hat{P}(c) = \frac{N_c}{N} \quad (2.11)$$

$$\hat{P}(t|c) = \frac{T_{ct}}{\sum_{t' \in V} T_{ct'}} \quad (2.12)$$

$$\hat{P}(t|c) = \frac{T_{ct}+1}{(\sum_{t' \in V} T_{ct'})+B} \quad (2.13)$$

Keterangan:

$P(c)$  : Probabilitas suatu dokumen berada di kelas c

$N_c$  : Jumlah dokumen dengan kelas c

$N$  : Jumlah seluruh dokumen

$P(t|c)$  : Probabilitas term t berada di dokumen pada kelas c

$T_{ct}$  : Frekuensi term t pada dokumen dengan kategori c

$\sum_{t' \in V} T_{ct'}$  : frekuensi seluruh term pada kelas c

$B$  : Jumlah seluruh term pada *vocabulary*

Persamaan di atas digunakan untuk membangun pemodelan klasifikasi dari data latih. Namun persamaan tersebut memiliki satu kekurangan, yaitu hasil perhitungannya akan menjadi 0 jika suatu kata tidak ada dalam kelas tertentu. Hasil pemodelan akan berpengaruh pada proses klasifikasi data uji. Untuk menghindari kasus tersebut diperlukan *laplace smoothing (add one)*, sehingga menjadi persamaan 2.13. Sedangkan untuk Multinomial Naïve Bayes dengan pembobotan TF-IDF, sama seperti persamaan 2.13. Hanya saja jumlah kemunculan term diganti dengan bobot TF-IDF. Untuk lebih jelasnya, dapat dilihat pada persamaan berikut.

$$\hat{P}(t|c) = \frac{W_{ct}+1}{(\sum_{w' \in V} W_{ct'})+B} \quad (2.14)$$

Keterangan:

- $P(t|c)$  : Probabilitas term t berada di dokumen pada kelas c  
 $W_{ct}$  : Bobot term t pada dokumen dengan kategori c  
 $\sum_{w' \in V} W_{ct'}$  : Jumlah bobot seluruh term pada kelas c  
 $B$  : Jumlah W kata unik (nilai IDF tidak dikali dengan TF) pada seluruh dokumen.

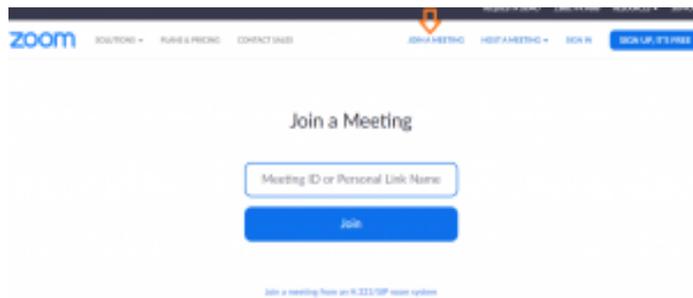
## 2.7 Zoom

Zoom (*Zoom Cloud Meeting*) merupakan sebuah aplikasi *video conference* berbasis *cloud computing*. Aplikasi tersebut dibuat oleh Eric Yuan, pendiri Zoom Video Communications, Inc. Zoom (berada di San Jose, California) dan dirilis pada tahun 2013 (Haqien & Rahman, 2020). Hingga saat ini, Zoom masih dipakai oleh banyak orang karena dikenal sebagai aplikasi yang andal dan jarang *down*. Selain itu, dengan *latency* yang cukup rendah juga membuat aplikasi ini tidak terganggu dengan jeda pembicaraan, serta bisa mempertahankan kualitas video dan audio meski koneksi internet labil.

Zoom sendiri juga menawarkan berbagai fitur yang bisa membantu perorangan, komunitas, hingga perusahaan-perusahaan besar dalam berinteraksi. Berikut ini merupakan fitur- fitur yang terdapat pada Zoom :

1. Memiliki kualitas *video conference* dan *screen sharing* yang bagus
2. Dapat mengirim berbagai file teks, suara, gambar, video dari Android ke laptop atau komputer
3. Bisa *join meeting* atau hanya melihat peserta webinar
4. Bekerja melalui jaringan WiFi, 4G LTE, dan 3G
5. Aman saat mode mengemudi di jalan
6. Dapat mengganti background saat *meeting*
7. Bagi peserta yang ingin bertanya, dapat menggunakan fitur *raise hand*
8. Fitur untuk membuat polling dan memfilter wajah
9. Memberikan komentar gambar
10. Pengguna dapat merekam dan menyimpan video saat *meeting*

Aplikasi ini tersedia gratis dan versi berbayar. Tentunya yang versi gratis mempunyai keterbatasan tersendiri seperti jumlah kapasitas yang dibatasi hanya 100 orang dalam satu *meeting*, waktu *meeting* hanya sekitar 40 menit, serta pembatasan fitur-fitur lainnya. Namun bagi pengguna yang ingin menggunakan aplikasi zoom untuk skala besar, maka harus berlangganan terlebih dulu.



Gambar 2.4 Tampilan Zoom Pada Website

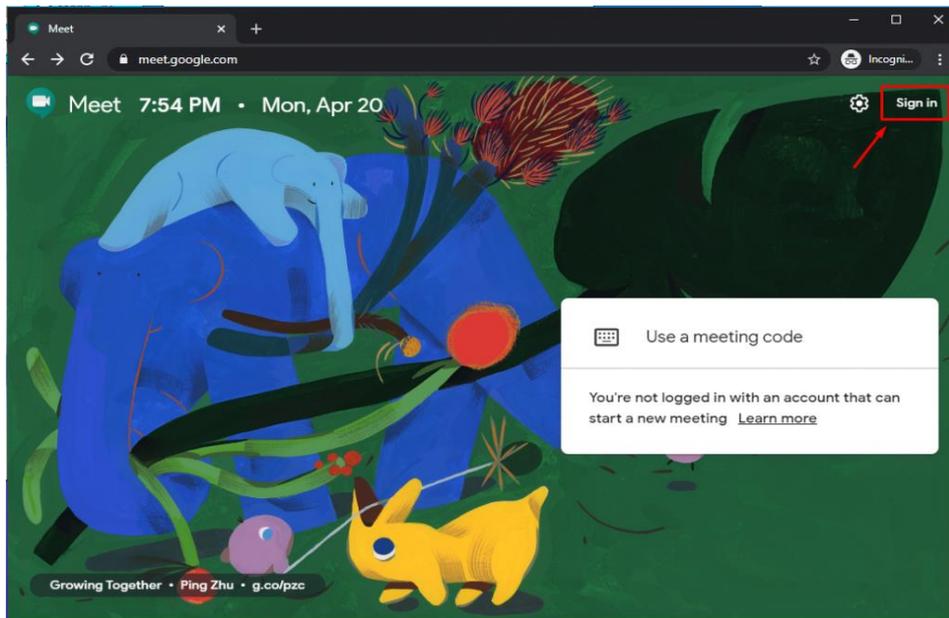


Gambar 2.5 Tampilan Zoom Saat *Video Conference*

## 2.8 Google Meet

Google Meet merupakan salah satu produk dari Google yang digunakan untuk *video conference*. Aplikasi ini merupakan kombinasi dari Google Hangouts dan Google Chat (versi sebelumnya). Pada mulanya, Google Meet dirilis hanya untuk pengguna iOS yang dilakukan secara diam-diam. Namun, pada bulan Maret 2017, pihak Google secara resmi meluncurkan aplikasi Google Meet. Aplikasi ini dapat memuat kapasitas lebih dari 30 peserta dalam suatu *meeting*, dan dapat

digunakan secara gratis. Selain itu, Google Meet juga memiliki UI yang minimalis dan menarik, serta proses operasional yang tidak rumit sehingga pemula dapat menggunakan aplikasi Google Meet tanpa merasa kesulitan (*user friendly*). Sejak saat itu, Google Meet diklaim menjadi versi yang lebih baik dibandingkan dengan Google Hangout (Sawitri, 2020).

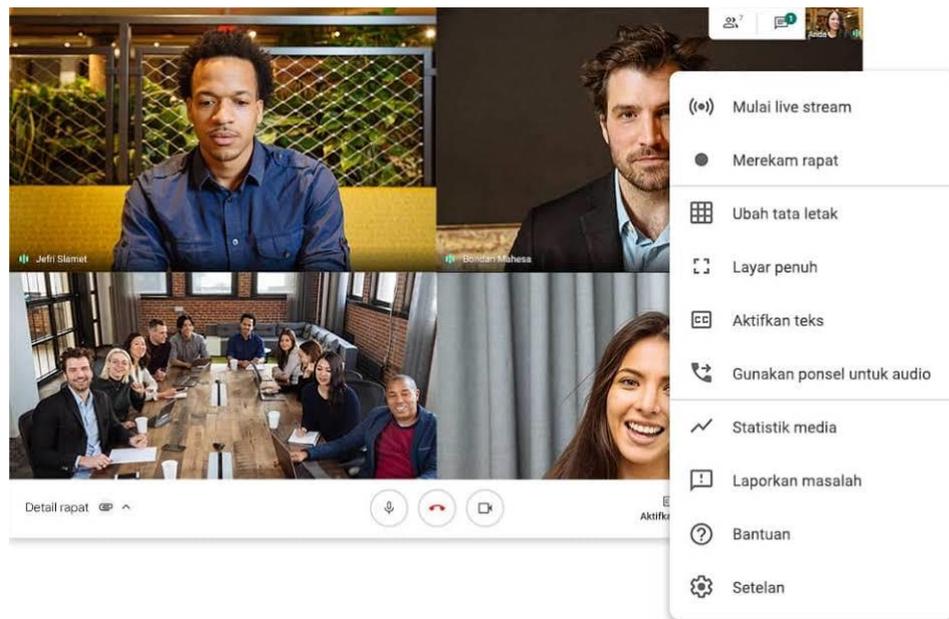


Gambar 2.6 Tampilan Google Meet Pada Website

Google Meet sendiri juga menawarkan berbagai fitur yang bisa membantu perorangan, komunitas, hingga perusahaan-perusahaan besar dalam berinteraksi. Berikut ini merupakan fitur- fitur terdapat pada Google Meet :

1. Dapat mengundang hingga 100 peserta (G Basic Suite), 150 orang (G Suite Business) dan 260 orang (G Suite Enterprise) per *meeting*.
2. Bisa *join meeting* melalui website atau aplikasi yang di instal oleh pengguna Android dan iOS.
3. Dapat melakukan rapat online dengan nomor dial-in.
4. Jika pengguna layanan G Suite Enterprise memakai nomor dial-in, maka *password* pengguna akan terlindungi.
5. Terintegrasi dengan Google Calendar untuk melakukan panggilan rapat dalam sekali klik.
6. Dapat menyajikan gambar, dokumen, *spreadsheet* atau presentasi.

7. Panggilan terenkripsi di antara semua penggunanya.
8. Filter *background noise*, bagi pengguna G Suite Enterprise dan G Suite for Education
9. Mode *low-light* (menerangi pencahayaan yang kurang ideal di lingkungan pengguna berada)
10. Berbagi video dan audio berkualitas tinggi

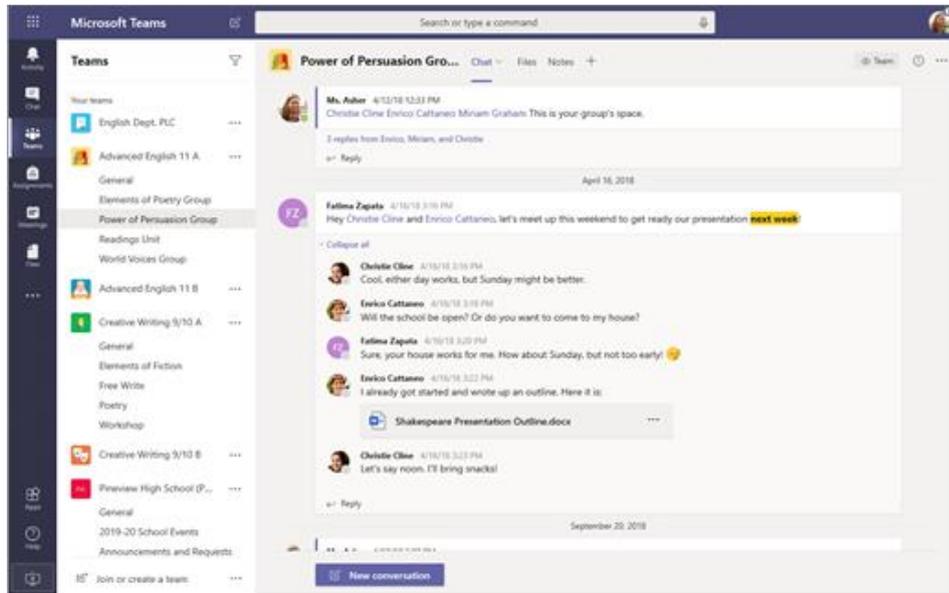


Gambar 2.7 Tampilan Google Meet Saat *Video Conference*

## 2.9 Microsoft Teams

Microsoft Teams merupakan salah satu aplikasi dari Microsoft yang digunakan untuk berkomunikasi, bekerja sama, dan melakukan semua aktivitas dari satu tempat. Aplikasi ini juga terintegrasi langsung dengan Office di dalam Microsoft 365. Pengguna dapat berbagi file langsung dan situs website, melakukan *meeting class*, berkomunikasi dengan pengguna yang lain, membuat catatan kelas OneNote, mendistribusikan dan menilai tugas, menyimpan rekaman *meeting*, dan sebagainya (Situmorang, 2020). Biasanya, perusahaan-perusahaan besar sudah berlangganan sehingga karyawan memiliki akses menggunakan Word, PowerPoint, Excel, OneDrive, dan juga Teams. Selain itu, pengguna juga dapat membuat *channel* sendiri di dalam aplikasi Microsoft Teams (contohnya *channel* untuk

komunitas sepeda, kelas matematika, grup mahasiswa, dan seterusnya) agar komunikasi di dalam tim menjadi lebih mudah, terfokus, dan tertata dengan baik. Tak hanya itu saja, pengguna juga bisa mengintegrasikan Teams dengan Office Outlook, sehingga semua jadwal dan kegiatan dapat diakses melalui satu aplikasi.



Gambar 2.8 Tampilan Diskusi Pada Microsoft Teams

Microsoft Teams dibuat dengan mengusung empat pilar. Pertama, memberikan interaksi yang modern dengan *thread* yang memungkinkan semua orang terlibat untuk dapat membaca pesan. Kedua, Microsoft Teams terintegrasi langsung dengan Word, Excel, PowerPoint, SharePoint, OneNote, dan PowerBI, sehingga memungkinkan semua orang memiliki informasi dan perangkat yang mereka butuhkan. Ketiga, aplikasi ini juga menawarkan kemampuan untuk menyesuaikan ruang kerja dari mitra pihak ketiga, beserta dengan layanan Microsoft lainnya. Keempat, standar keamanan yang tinggi membuat aplikasi yang ditawarkan ini sesuai dengan harapan pengguna.

Berikut ini merupakan fitur- fitur terdapat pada Microsoft Teams :

1. Chat dan pencarian yang *unlimited*
2. Komunikasi dengan tim kecil maupun besar (*channel*)
3. *File storage* yang besar

4. Kolaborasi langsung dengan aplikasi Office (Word, Excel, PowerPoint, dan OneNote)
5. Dapat menyajikan gambar, dokumen, spreadsheet atau presentasi
6. Akses untuk pihak luar kantor
7. Menggunakan lebih dari 450 aplikasi yang sudah terintegrasi
8. Keamanan data (*security*)
9. Fitur *raise hand* bagi peserta yang ingin bertanya atau berpendapat
10. Mengganti *video background* dengan gambar yang sudah disediakan



Gambar 2.9 Tampilan Microsoft Teams Saat *Video Conference*