

## BAB II

### LANDASAN TEORI

#### 2.1 Penelitian Terdahulu

Setelah mempelajari adapun penelitian terdahulu yang terkait dalam topik analisis sentimen, sehingga ditemukan beberapa jurnal yang terdapat berkaitan dalam penelitian ini, yaitu:

*Tabel 2. 1 Penelitian Terdahulu*

Nama Artikel	Nama Penulis	Objek Penelitian	Teknik Penelitian	Hasil Penelitian
“Analisis Sentimen terhadap Penggunaan Aplikasi MySAPK BKN di Google Play Store” Jurnal Nasional Teknik Elektro dan Teknologi Informasi (JNTETI), Vol 11, No 2, page 105 – 113 dan ISSN 2460 – 5719, 2022[16]	- Raksaka Indra - I Made Kurniawan - Yova Ruldeviyani	Analisis sentimen pada penggunaan aplikasi MySAPK BKN di Google Playstore	- google playstore - Naïve Bayes - Support Vector Machine (SVM)	- Pengguna dari aplikasi MySAPK BKN di google playstore memiliki respon negatif yang lebih besar dibandingkan respon positif. - Algoritma dari Naïve Bayes yang baik dalam penelitian ini,
“Perbandingan Metode Random Forest dan Naïve Bayes pada Analisis Sentimen Review Aplikasi BCA Mobile” Smart Comp: Jurnalnya Orang Pintar	- Akhmad Miftahusalam - Hasih Pratiwi - Isnandar Slamet	Analisis sentimen terhadap aplikasi MyBca <i>Mobile</i>	- Google playstore - MyBca Mobile - Random Forest - Naïve Bayes	- hasil analisis sentimen review aplikasi BCA Mobile didapatkan bahwa klasifikasi menggunakan metode Random Forest menghasilkan prediksi yang lebih baik

Nama Artikel	Nama Penulis	Objek Penelitian	Teknik Penelitian	Hasil Penelitian
Komputer, Vol. 12 No. 2, hal 767 – 772 ISSN: 2089676X. 2023 [15]				dibandingkan dengan metode Naïve Bayes.
“Analisa sentimen Pengguna Instagram di Indonesia Pada Review Smartphone Menggunakan Naive Bayes” Jurnal KLIK: Kajian Ilmiah Informatika dan Komputer, Vol 2, No 4, hal 148-155, ISSN 2723-3898, 2022 [17]	Kairil Anwar	Melakukan analisis sentimen terhadap pengguna aplikasi Instagram dengan metode Naïve Bayes terhadap review atau ulasan.	<ul style="list-style-type: none"> <li>- Instagram</li> <li>- Ulasan</li> <li>- Metode Naïve Bayes</li> </ul>	Pada penelitian ini memberikan hasil dari kurasi sebesar 73% sementara dengan menggunakan Multinomial Naïve Bayes menghasilkan akurasi sebesar 81%. Metode dari Naïve Bayes memiliki akurasi yang akurat dan baik.
“Komparasi Algoritma Data Mining untuk Analisis Sentimen Aplikasi Pedulilindungi”, Jurnal Keilmuan dan Aplikasi Bidang Teknik Informatika Vol 16, No 1, hal 1-11, ISSN 2656-0321, 2022 [18]	<ul style="list-style-type: none"> <li>- Hiras Parasian Doloksaribu</li> <li>- Yusran Timur Samuel</li> </ul>	Melakukan komparasi pada aplikasi Pedulilindungi dengan algoritma data mining dan analisis sentiment dengan metode Naive Bayes, Support Vector Machine	<ul style="list-style-type: none"> <li>- pedulilindungi</li> <li>- Naïve Bayes</li> <li>- Support Vector Machine</li> </ul>	Pada penelitian ini Algoritma dari Support Vector Machine yang dibantu TF IDF Vectorizer dalam kasus ini menjadi algoritma yang terbaik dalam tingkat akurasi dan perbandingan data.

Nama Artikel	Nama Penulis	Objek Penelitian	Teknik Penelitian	Hasil Penelitian
<p>“Analisis Sentimen Twitter Terhadap Program MBKM Menggunakan Decision Tree dan Support Vector Machine” Journal of Information System Research (JOSH), Vol. 4, hal 1145-1154, ISSN 2686-228X, 2023 [19]</p>	<ul style="list-style-type: none"> <li>- Lita Astri Pramesti</li> <li>- Nunik Pratiwi</li> </ul>	<p>Melakukan klasifikasi terhadap program MBKM dengan menggunakan metode algoritma DT dan SVM.</p>	<ul style="list-style-type: none"> <li>- Klasifikasi</li> <li>- Analisis sentimen</li> <li>- Decision Tree</li> <li>- SVM</li> </ul>	<p>Berdasarkan hasil tersebut dapat disimpulkan bahwa Algoritma Support Vector Machine lebih unggul dalam melakukan analisis sentimen terhadap program MBKM.</p>
<p>Analisis Sentimen Pada Ulasan Aplikasi WeTV di Google Play Store Menggunakan Algoritma NBC dan SVM Sentiment, “SISTEMASI: Jurnal Sistem Informasi” vol 12, tahun 2023. ISSN:2302-8149 [10]</p>	<ul style="list-style-type: none"> <li>- Petronilia Palinggik Allorerung</li> <li>- Rismayani</li> </ul>	<p>Melakukan analisis sentimen terhadap aplikasi WeTV dengan NBC dan SVM</p>	<ul style="list-style-type: none"> <li>- Analisis sentimen</li> <li>- Klasifikasi</li> <li>- Naïve Bayes</li> <li>- SVM</li> </ul>	<p>Hasil precision NBC dan SVM sebesar 70.83% dan 80.00%. Dan hasil recall NBC dan SVM sebesar 70.00% dan 80.00%.</p>
<p>Sentiment Analysis on Tokopedia Product Online Reviews Using Random Forest Method, E3S Web of Conferences, 2020, vol 202, ISSN 22671242 [20]</p>	<ul style="list-style-type: none"> <li>- Stephanie Budi Warsito</li> <li>- Alan Prahutama</li> </ul>	<p>Melakukan sentimen analisis terhadap aplikasi Tokopedia dengan Random Forest</p>	<ul style="list-style-type: none"> <li>- Analisis sentimen</li> <li>- Random Forest</li> </ul>	<p>Akurasi klasifikasi sentimen menggunakan metode Random Forest dengan 70% data training dan 30% data testing adalah 97.38%. Parameter yang digunakan dalam analisis ini adalah mtry = 73 dan ntree = 50.</p>

Nama Artikel	Nama Penulis	Objek Penelitian	Teknik Penelitian	Hasil Penelitian
Support Vector Machine VS Information Gain: Analisis Sentimen Cyberbullying di Twitter Indonesia, Ultima InfoSys: Jurnal Ilmu Sistem Informasi, Vol 11, tahun 2020, ISSN 2085-4579 [13]	<ul style="list-style-type: none"> <li>- Christevan Destitus</li> <li>- Wella</li> <li>- Suryasari</li> </ul>	Melakukan sentimen analisis terhadap Cyberbullying dengan menggunakan data dari aplikasi Twitter atau X di Indonesia	<ul style="list-style-type: none"> <li>- Analisis sentimen</li> <li>- SVM</li> <li>- Information Gain</li> </ul>	hasil accuracy 80%, precision 75.1%, recall 96%, dan f-measure 85% pada Information Gain mendapatkan nilai yang cukup tinggi sebesar accuracy 86%, precision 81%, recall 95%, dan f-measure 87% sehingga hasil identifikasi tweet cyberbullying dengan kedua metode memiliki hasil yang cukup maksimal.
Analisis Sentimen Kepuasan Pengguna Terhadap Layanan Streaming Mola Menggunakan Algoritma Random Forest, Jurnal Aplikasi Teknologi Informasi dan Manajemen (JATIM), Vol 3, No 2, tahun 2022, ISSN 2722-435X [21]	<ul style="list-style-type: none"> <li>- Setya Nanda</li> <li>- Desti Mualfah</li> <li>- Diah Angraina</li> </ul>	Melakukan analisis sentimen pada layanan <i>streaming</i> dari aplikasi MolaTV	<ul style="list-style-type: none"> <li>- Analisis sentimen</li> <li>- Random Forest</li> </ul>	Mendapatkan hasil klasifikasi pada kelas positif dengan jumlah 126, negatif 196 dan netral 198. Pembagian data training 80% dan testing 20% menghasilkan nilai accuracy 98%, nilai precision 1,00%, nilai recall 95% dan nilai f1-score 98%.
Analisis Sentimen Aplikasi Youtube di Google Play Store Menggunakan Machine Learning, RESOLUSI : Rekayasa Teknik Informatika dan Informasi, Vol 4, No 4, Tahun 2024, ISSN 2745-7966	<ul style="list-style-type: none"> <li>- Jimmy Alga</li> <li>- Cindi Wulandari</li> <li>- Bunga Intan</li> </ul>	Melakukan analisis sentimen terhadap aplikasi Youtube	<ul style="list-style-type: none"> <li>- Analisis sentimen</li> <li>- SVM</li> </ul>	Proses klasifikasi dan penerapan algoritma support vector machine menggunakan perbandingan data training dan data testing yang telah penulis lakukan berdasarkan data training 90% dan data testing 10% menghasilkan nilai akurasi yaitu

Nama Artikel	Nama Penulis	Objek Penelitian	Teknik Penelitian	Hasil Penelitian
				75% dengan nilai presisi kelas negatif 76% beserta recall kelas negatif 97%. Pengujian K-Fold Cross Validation menggunakan nilai K=10 mendapatkan nilai rata-rata Akurasi sebesar 0.75 atau 75%.

UMMN

UNIVERSITAS

MULTIMEDIA

NUSANTARA

1. Pada penelitian pertama melakukan analisis sentimen terhadap penggunaan aplikasi MySAPK BKN dengan data dari ulasan atau *review* Google Play Store. Dengan menggunakan algoritma Naïve Bayes dan Support Vector Machine (SVM), dalam penelitian ini menggunakan data sebanyak 4778 ulasan. Dalam penelitian ini ulasan sentimen negatif sebesar 2660 (55,7%) dibandingkan dengan ulasan sentimen positif sebesar 2118 (44,3%), kemudian akurasi dari algoritma yang digunakan yaitu Naïve Bayes sebesar 92,47% dan SVM 94,14%. [16] Dalam penelitian ini memberikan kontribusi atau *references* dalam penggunaan algoritma dari Naïve Bayes. Dalam penelitian ini Naïve Bayes memiliki tingkat akurasi yang baik dibandingkan SVM.
2. Pada Perbandingan Metode Random Forest dan Naive Bayes pada Analisis Sentimen Review Aplikasi BCA Mobile. Dengan menggunakan algoritma Random Forest dan Naïve Bayes, dalam penelitian ini menggunakan data sebanyak 494 data ulasan atau *review*, dengan memperoleh akurasi dari Random Forest sebesar 93,93% dan akurasi dari Naïve Bayes sebesar 92,31% [22] Dalam penelitian ini memiliki kontribusi atau *references* dalam algoritma dan akurasi yang sudah didapatkan akan memiliki pengaruh dalam penelitian untuk menentukan analisis sentimen.
3. Pada penelitian ketiga melakukan analisis sentimen dalam penggunaan aplikasi Instagram pada ulasan atau *review smartphone* dengan menggunakan algoritma Naïve Bayes. Dalam penelitian ini *review* atau ulasan menggunakan *smartphone* dengan merk Vivo dan Oppo di Indonesia. Dalam penelitian ini melakukan pengujian analisis sentimen dari Gaussian Naïve Bayes dengan jumlah akurasi 73% sedangkan dengan Multinomial Naïve Bayes dengan jumlah akurasi 81%. [17] Dalam penelitian ini memberikan kontribusi atau *references* untuk

algoritma Naïve Bayes yang dapat memberikan dampak dalam penelitian ini.

4. Pada penelitian keempat melakukan sebuah komparasi dari algoritma Data Mining untuk analisis sentimen terhadap aplikasi PeduliLindungi. Dalam penelitian ini menggunakan dua algoritma yaitu Naïve Bayes dan SVM. Data yang digunakan dari penelitian ini sebanyak 1948 data. Data tersebut didapatkan dari ulasan atau *review* dari pengguna dari aplikasi PeduliLindungi yang didapatkan di Google Play Store. Dalam penelitian ini algoritma dari SVM memiliki tingkat keakurasian data lebih tinggi dibandingkan dengan algoritma dari Naïve Bayes. [18] Dalam penelitian ini memberikan kontribusi atau *references* dalam penentuan algoritma Decision Tree, bahwa Decision Tree memberikan dampak yang baik juga dalam penelitian terdahulu ini.
5. Pada penelitian kelima melakukan Analisis Sentimen Twitter Terhadap Program MBKM Menggunakan Decision Tree dan Support Vector Machine, dalam penelitian ini menggunakan algoritma Decision Tree dan Support Vector Machine. Dalam penelitian ini akurasi dari DT dan SVM sebesar 72.86% dan 84.76% . [4] Dalam penelitian ini memberikan kontribusi atau *references* dalam penentuan penggunaan algoritma Decision Tree.
6. Pada penelitian keenam melakukan Analisis Sentimen Pada Ulasan Aplikasi WeTV di Google Play Store Menggunakan Algoritma NBC dan SVM. Dalam penelitian ini menggunakan algoritma Naïve Bayes dan SVM. Pada penelitian ini Hasil precision NBC dan SVM sebesar 70.83% dan 80.00%. Dan hasil *recall* NBC dan SVM sebesar 70.00% dan 80.00%. Dari hasil tersebut, diketahui bahwa SVM mendapatkan performa tertinggi sehingga menjadi algoritma terbaik dalam penelitian ini, di mana tidak terdapat komentar yang mendominasi. [10] Oleh karena itu dalam penelitian ini memberikan kontribusi terhadap algoritma Naïve Bayes untuk digunakan pada *dataset* WeTV dengan menggunakan *data splitting* yang lebih baik.

7. Pada penelitian ketujuh Analisis Sentimen pada Ulasan Online Produk Tokopedia Menggunakan Metode Random Forest. Dalam penelitian ini menggunakan Algoritma Random Forest untuk akurasi. Dalam penelitian ini pelanggan Tokopedia memiliki ekspektasi yang baik terhadap produk mereka dan mayoritas pelanggan telah membeli dalam kategori elektronik. Akurasi klasifikasi sentimen menggunakan metode Random Forest dengan 70% data training dan 30% data testing adalah 97.38%. Parameter yang digunakan dalam analisis ini adalah  $mtry = 73$  dan  $ntree = 50$ . [20] Dalam penelitian ini memberikan kontribusi terhadap pengambilan algoritma yang digunakan yaitu Random Forest, karena memiliki hasil akurasi yang akurat dan baik.
8. Pada penelitian kedelapan Support Vector Machine VS Information Gain: Analisis Sentimen Cyberbullying di Twitter Indonesia. Dalam penelitian ini menggunakan teknik *scrapping* terhadap aplikasi Twitter, kemudian dalam tahap *data pre – processing* dilakukan seperti *case folding*, *stop words* dan *tokenizing*[13]. Sehingga penelitian ini memberikan kontribusi terhadap proses untuk *data cleaning* dan *data pre – processing* agar *dataset* yang digunakan lebih rapih dan mudah untuk diolah.
9. Pada penelitian kesembilan Analisis Sentimen Kepuasan Pengguna Terhadap Layanan Streaming Mola Menggunakan Algoritma Random Forest. Dalam penelitian ini menggunakan proses pengelolaan data CRISP – DM [21]. Sehingga penelitian ini memberikan kontribusi pada proses perhitungan dari algoritma Random Forest hingga *accuracy*, *precision*, *recall* dan *F1-Score*.
10. Pada penelitian kesepuluh Analisis Sentimen Aplikasi Youtube di Google Play Store Menggunakan Machine Learning. Dalam penelitian ini menggunakan teknik *scrapping* dengan mengambil data melalui Google Play Store [23]. Sehingga pada penelitian ini memberikan kontribusi pada proses pengambilan data dengan teknik Google *Scrapping*.

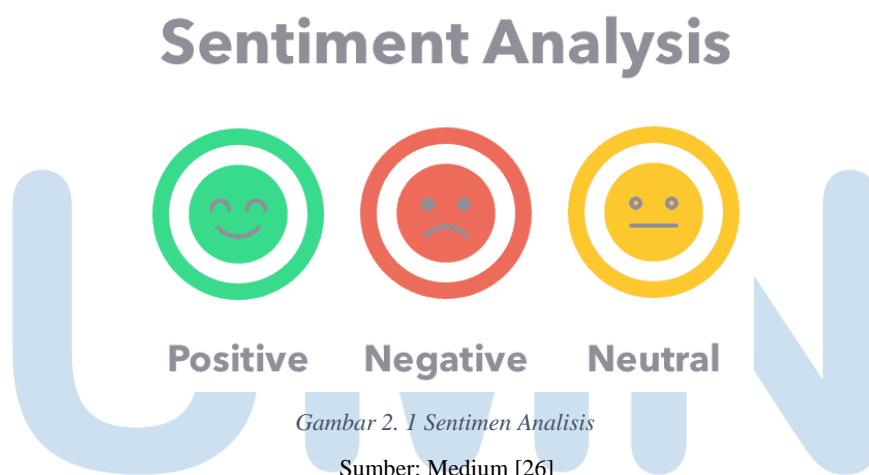


## 2.2 Teori Tentang Skripsi

### 2.2.1 Analisis Sentimen

Analisis Sentimen merupakan sebuah teknik dalam melakukan sebuah pengekstrakan data teks agar mendapatkan sebuah informasi tentang sentimen dengan nilai positif maupun negatif, selain itu adanya gambaran dari analisis sentimen yaitu sebuah opini dari penulis pada sebuah topik [24].

Analisis Sentimen ini dapat disebut juga dengan Opinion Mining adalah adanya komputerasi agar dapat mengetahui *feedback* dari *user* seperti ekspresi opini, evaluasi, emosi, hingga penilaian yang diberikan atau sudut pandang dari *user* dalam suatu teks atau gambar [25]. Dengan menggunakan Text Mining ini dapat menyelesaikan beberapa masalah dari Opinion Mining seperti menggunakan metode *Decision Tree*, *Naïve Bayes Classifier* dan *Random Forest*. Dengan menggunakan metode tersebut dapat melakukan klasifikasi beberapa opini ke berbagai kelompok agar dapat dilakukan sebuah klasifikasi opini positif atau opini negatif [15]. Berikut merupakan gambar dari Analisis sentimen.



Gambar 2. 1 Sentimen Analisis

Sumber: Medium [26]

Tujuan akhir dalam penelitian ini akan dilakukan analisis sentimen untuk memahami tingkat dari kepuasan dari pengguna dari objek yang dikonsumsi atau digunakan sehingga dapat dilakukan sebuah tindakan lebih lanjut seperti perbaikan – perbaikan dalam beberapa bagian dimana aspek – aspek yang dianggap masih kurang memuaskan, Hal ini dilakukan dengan tujuan meningkatkan kualitas objek tersebut dari sebelumnya.

Pada penelitian ini menggunakan analisis sentimen dengan basis *machine learning* sehingga membutuhkan penggunaan algoritma untuk mengkategorikan kata-kata ke dalam kategori positif dan negatif. Contohnya termasuk algoritma seperti Naïve Bayes (NB), Decision Tree (DT), dan Random Forest (RF).

### **2.2.2 Google Play Store**

Google Playstore merupakan sebuah aplikasi atau program yang mendistribusikan berbagai jenis konten dan aplikasi digital yang dikembangkan langsung oleh google. Google Play Store saat ini menjadi aplikasi install android terbesar di dunia. Google Play Store sendiri memiliki beberapa jenis layanan seperti buku, games, aplikasi hingga streaming film. Dengan memiliki fitur yang lengkap dan *user friendly* [27]. Selain itu terdapat fitur yang dapat melakukan *review* dan *ratings* agar para pengguna dapat melakukan penilaian terhadap aplikasi yang sudah di unduh dan digunakan sebelumnya. Fitur ini dapat membantu *developer* aplikasi tersebut mendengarkan *feedback* atau penilaian dari pengguna.

### **2.2.3 Text Mining**

Text Mining merupakan algoritma yang dapat digunakan untuk melakukan sebuah penggalian data, Text Mining sendiri bekerja dengan cara melakukan pengekstrakan dari informasi berbagai sumber yang *valid* dan secara otomatis dengan melakukan penghubungan dari berbagai informasi yang sudah berhasil dilakukan dalam pengesktrakan [28]. Pada Text Mining ini harus memiliki format yang terstruktur dikarenakan dalam *pre – processing* teks yang digunakan tidak memiliki format yang tidak beraturan [29].

Text Mining merupakan teknik yang diterapkan untuk menangani masalah klasifikasi, pengambilan informasi, ekstraksi informasi, dan pengelompokan. Secara umum, proses Text Mining banyak mengadopsi konsep dari Data Mining, namun perbedaannya terletak pada pola yang digunakan. Text Mining mengambil pola dari sejumlah bahasa alami yang bersifat tidak terstruktur, berbeda dengan data mining yang menggunakan pola dari basis data yang terstruktur [24] [29].

#### 2.2.4 Confusion Matrix

Confusion Matrix merupakan untuk alat pengukur performa pada masalah klasifikasi yang memberikan output berupa dua kelas atau lebih. Confusion matrix berfungsi untuk menganalisa sebuah classifier agar dapat diketahui classifier tersebut baik atau buruk [30]. Pada Confusion matrix ini memiliki tabel untuk klasifikasi biner yang ditunjukkan pada tabel dibawah ini

Tabel 2. 2 Tabel Confusion Matrix

	Actual		
	Positive	Negative	
Prediction	Positive	TP	FP
	Negative	FN	TN

Pada tabel 2.1 terdapat empat *point* sebagai hasil untuk representasi dari beberapa proses dalam sebuah klasifikasi pada *confusion matrix* sebagai berikut:

- TP (*true positive*): Memiliki nilai yang diprediksi dengan benar sebagai hasil yang positif.
- FP (*false positive*): Memiliki nilai yang diprediksi positif yang sebenarnya. Tetapi ada nilai negatif yang sudah diprediksi sebagai nilai positif.
- FN (*false negative*): Memiliki nilai positif tetapi diprediksi memiliki hasil yang negatif.
- TN (*true negative*): Memiliki nilai prediksi dengan benar tetapi diprediksi sebagai nilai negatif sebenarnya.

Dalam *confusion matrix* memiliki 4 rumus untuk pengukuran analisis sentimen yaitu:

- *Precision*

*Precision* merupakan proporsi dari prediksi positif yang tepat dibandingkan dengan seluruh prediksi positif yang dilakukan. Berikut merupakan rumus dari *precision*.

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

Rumus 2. 1 Perhitungan terhadap Precision

- *Recall*

*Recall* merupakan persentase dari jumlah prediksi positif yang berhasil dilakukan oleh model untuk kelas positif. Berikut merupakan rumus dari *Recall*.

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

Rumus 2. 2 Perhitungan terhadap Recall

- *F1-Score*

*F1-Score* merupakan sebuah rasio yang menggabungkan *precision* dan *recall* yang cenderung memiliki hubungan yang berlawanan. Berikut merupakan rumus dari *F1 – Score*.

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Rumus 2. 3 Perhitungan terhadap F1 - Score

- *Accuracy*

*Accuracy* akan memberikan informasi mengenai proporsi prediksi yang akurat dari keseluruhan prediksi yang telah dilakukan. Berikut merupakan rumus dari *accuracy*.

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

Rumus 2. 4 Perhitungan terhadap Accuracy

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

### 2.2.5 Term Frequency-Inverse Document Frequency (TF-IDF)

*TF-IDF* merupakan sebuah metode yang terdapat nilai frekuensi dari sebuah kata yang terdapat pada dokumen maupun dataset, dan memiliki perhitungan yang dapat melihat ketentuan seberapa relevan dari kata tersebut [15]. Dalam metode ini adanya fungsi untuk dapat menentukan nilai dari setiap bobot yang memiliki suatu kata sehingga dapat melakukan evaluasi seberapa penting dari kata tersebut yang terkandung dalam dokumen maupun dataset. Pada *TF-IDF* ini dibagi menjadi dua bagian yaitu TF adalah sebuah proporsi dimana hasil dari kepentingan kemunculan dari teks atau dokumen maupun dataset. Kemudian *TF-IDF* merupakan sebuah metode pembobotan token yang dapat berfungsi untuk melakukan sebuah monitoring kemunculan dari token ini dalam teks [31]. Sebagai contoh, TF-IDF mampu mengidentifikasi kata-kata yang paling umum muncul dalam sebuah ulasan, seperti "baik", "bagus", "buruk", "jelek", dan sebagainya. Umumnya, penerapan teknik *feature extraction* seperti TF-IDF ini dapat meningkatkan akurasi kinerja algoritma yang digunakan dalam menganalisis sentimen.

## 2.3 Teori Tentang Algoritma

### 2.3.1 Naïve Bayes Classifier

Naïve Bayes Classifier ini merupakan sebuah metode untuk probabilitas yang mampu mengerjakan dalam jumlah frekuensi dan melakukan perhitungan dari probabilitas hingga kombinasi yang acak dari nilai maupun dataset [25]. Pada metode Naïve Bayes ini memiliki beberapa tahapan yaitu *training* dan *testing*. Ditahapan *training* dataset yang sudah ada dilakukan pelatihan pada dokumen yang sudah terdapat beberapa kategori, kemudian pada tahap *testing* dataset akan dilakukan klasifikasi terhadap dokumen yang belum tersedia kategori [24].

Berikut merupakan rumus dari algoritma Naïve Bayes

$$P(H|E) = \frac{P(E|H) \times P(H)}{P(E)}$$

Sumber: [28]

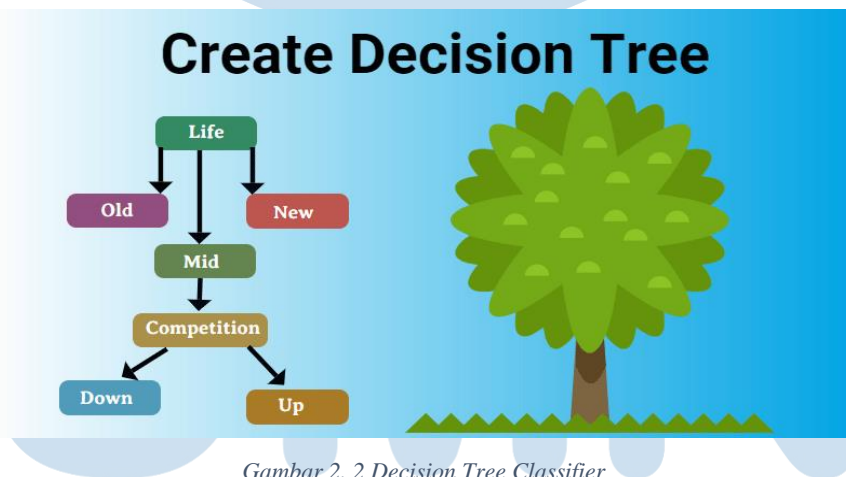
Rumus 2.5 Perhitungan terhadap Algoritma Naive Bayes

Berikut merupakan keterangan dari rumus Naïve Bayes:

- $P(H|E)$  merupakan peristiwa H terjadi pada saat peristiwa E terjadi
- $P(E|H)$  menyatakan seberapa sering terjadi kejadian dari E ketika kejadian H terjadi terlebih dahulu
- $P(H)$  merupakan peluang terjadinya kejadian H
- $P(E)$  merupakan peluang terjadinya kejadian E

### 2.3.2 Decision Tree

Decision Tree merupakan sebuah metode klasifikasi dengan memiliki struktur yang mirip dengan pohon untuk memberikan hasil keputusan yang memiliki beberapa kemungkinan dan konsekuensi [12]. Pada metode *decision tree* ini dapat meningkatkan sebuah model prediktif dengan memiliki kemudahan dalam melakukan akurasi hingga interpretasi serta stabilitas. *Decision tree* merupakan metode dengan pembelajaran algoritma yang baik dan efektif dalam menyesuaikan dari hubungan yang *non – linear* karena dapat menyesuaikan data seperti klasifikasi dan regresi.



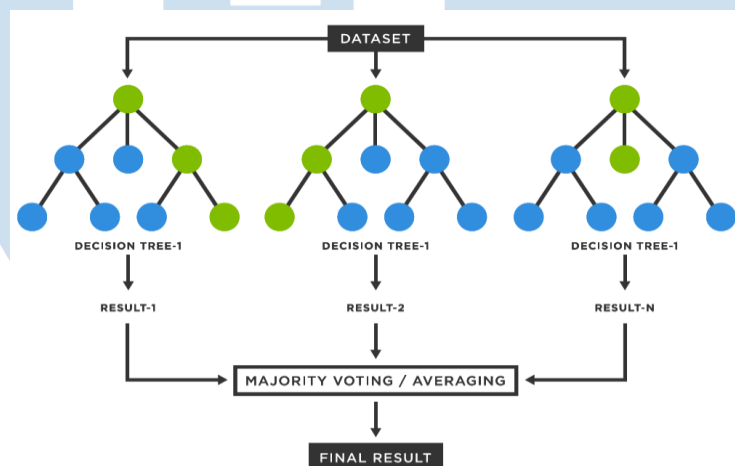
Gambar 2. 2 Decision Tree Classifier

Sumber: EduCba [32]

*Life* adalah kepala dari seluruh percabangan sehingga *life* akan selalu hanya ada di paling atas dari pohon keputusan. Kemudian *Old*, *Mid* dan *New* merupakan percabangan yang baru dari *life* sehingga adanya pengujian dari setiap atribut. Garis – garis yang menghubungkan dari *life* hingga *old*, *mid* dan *new* merupakan *branches* atau cabang yang menghubungkan antara satu atribut dengan atribut yang lainnya.

### 2.3.3 Random Forest

Algoritma Random Forest adalah sebuah klasifikasi peningkatan dari metode *Classification and Regression Tree* (CART) dengan menggunakan teknik *bootstrap aggregating* (bagging) dan adanya pemilihan fitur secara acak. Dengan menggunakan sampel *bootstrap*, memiliki alur dari setiap simpul, pemilihan fitur dilakukan secara acak untuk menentukan pemisahan, di mana jumlah variabel yang diambil secara acak sebagai kandidat pada setiap pemisahan harus lebih kecil dari total variabel [20].



Gambar 2. 3 Random Forest Classifier

Sumber: Medium [33]

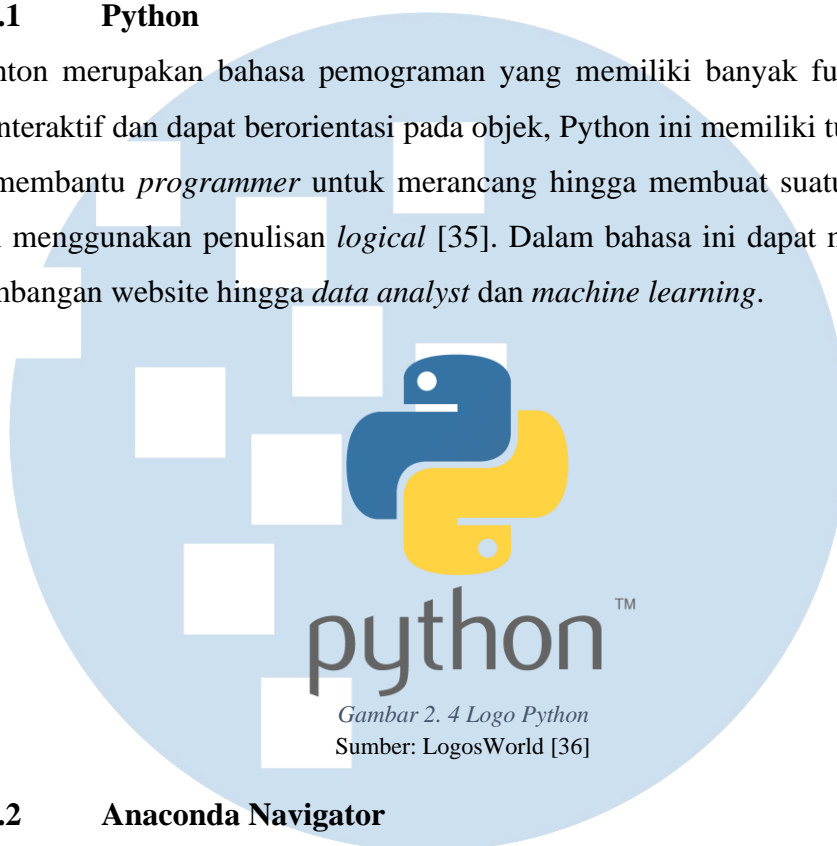
Dukungan dari beberapa Decision Trees (DT) dalam Random Forest (RF) menyebabkan RF menjadi model dengan tingkat akurasi klasifikasi yang tinggi. RF juga memiliki toleransi yang baik terhadap *outlier* dan *noise*, serta mampu menangani sebagian besar kasus *overfitting*, yang mana terjadi ketika akurasi pada data pelatihan tinggi tetapi akurasi prediksi pada data lainnya rendah [34].

U N I V E R S I T A S  
M U L T I M E D I A  
N U S A N T A R A

## 2.4 Teori Tentang Tools

### 2.4.1 Python

Python merupakan bahasa pemrograman yang memiliki banyak fungsi yang dapat interaktif dan dapat berorientasi pada objek, Python ini memiliki tujuan agar dapat membantu *programmer* untuk merancang hingga membuat suatu program dengan menggunakan penulisan *logical* [35]. Dalam bahasa ini dapat melakukan pengembangan website hingga *data analyst* dan *machine learning*.



Gambar 2. 4 Logo Python  
Sumber: LogosWorld [36]

### 2.4.2 Anaconda Navigator

Anaconda Navigator merupakan sebuah *package* untuk melakukan distribusi dari bahasa pemrograman dari Python, pada *package* ini sudah termasuk keperluan dalam pemrograman *data science* dan matematika. Anaconda ini *software* ini dapat membantu *programmer* dalam merancang program sampai teknik *code* dalam melakukan sebuah distribusi pada *platform*. Dalam *anaconda navigator* terdapat beberapa tambahan *package* seperti R-studio, CMD dan lain sebagainya [37].



Sumber: Medium [38]



### 2.4.3 Jupyter Notebook

Jupyter Notebook merupakan *software* yang menjadi penghubung dari berbagai bahasa pemrograman. *Jupyter notebook* sebuah web yang bersifat *open source* sehingga dapat melakukan sebuah perancangan dan berbagai dokumen yang interaktif yang terdiri dari seperti *code*, *live* dan visualisasi [39]. *Jupyter notebook* ini dapat melakukan penggabungan antar *code – code* dari dataset atau dokumen yang terpisah atau berbeda dan ditampilkan pada *library* atau aplikasi seperti Visual Studio dan Eclipse.



Gambar 2. 6 Jupyter Notebook Logo  
Sumber: SeekLogo [40]

# UMMN

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA