

BAB II

LANDASAN TEORI

2.1 Penelitian Terdahulu

Penelitian terdahulu merupakan penelitian yang telah dibuat sebelum skripsi ini tercipta yang digunakan sebagai referensi atau acuan bagi skripsi ini sehingga skripsi ini dapat memperbaiki kekurangan dari penelitian terdahulu tersebut. Berikut ini adalah tabel penelitian terdahulu di bawah ini

Tabel 2. 1 Tabel Penelitian Terdahulu

Nama Artikel	Nama Penulis	Teknik Penelitian	Hasil Penelitian
“Analisis Sentimen pada ulasan Aplikasi <i>Amazon Shopping</i> di <i>Google Play Store</i> Menggunakan <i>Naïve Bayes Classifier</i> ” JTS (Jurnal Teknik Dan Science), Vol 1 No. 3 Oktober 2022	<ul style="list-style-type: none">- Ernianti- Hasibuan- Elmo Allistair- Heriyanto	<i>Naïve Bayes Classifier</i>	Setelah melakukan analisis sentiment klasifikasi menggunakan algoritma <i>naïve Bayes</i> , terdapat 4 hasil yang didapat yaitu akurasi yaitu 82,15%, <i>presicion</i> yaitu 72,25%, <i>fi-score</i> yaitu 77,41% dan <i>recall</i> yaitu 83,49%. Akurasi terbaik yang dihasilkan adalah <i>Multinomial MB</i> sebesar 86,74%. Ulasan data terhadap analisis sentimen menghasilkan 53,8% ulasan positif dan 46,2% ulasan negatif [6]
“Implementasi Algoritma <i>Naïve Bayes</i> untuk Analisis Sentimen Ulasan <i>Shopee</i> pada <i>Google Play Store</i> ” Institut Riset dan Publikasi Indonesia (IRPI) MALCOM: <i>Indonesian Journal of Machine Learning and Computer Science</i> , Vol. 2 Iss. 1 April 2022	<ul style="list-style-type: none">- Nurhaliza- Agustina. C.A- Desy Herlina Citra- Wido Purnama- Amanda Rozi Kurnia	<i>Naïve Bayes</i>	Hasil akurasi algoritma pada pembagian data <i>Hold Out</i> sebesar 83% lebih unggul 1% dibandingkan dengan menggunakan Teknik pembagian data 10 <i>Cross Fold Validation</i> yang hanya sebesar 82 %. Dari hasil nilai <i>precision</i> sebesar 83%, <i>recall</i> 100 %, dan <i>f1-score</i> 91% menyatakan bahwa Sentimen pada

Nama Artikel	Nama Penulis	Teknik Penelitian	Hasil Penelitian
			pengguna Shopee cukup positif [7]
<p>“Analisis Sentimen Ulasan Aplikasi Samsat Digital Nasional Pada <i>Google Playstore</i> Menggunakan Algoritma <i>Naïve Bayes</i>”</p> <p>KLIK: Kajian Ilmiah Informatika dan Komputer, Universitas Muhammadiyah, Vol 4, No 4, Februari 2024</p>	<ul style="list-style-type: none"> - Deni Wijaya - Rizki Adi Saputra - Faldy Irwiensyah 	<i>Naïve Bayes</i>	Akurasi yang dihasilkan oleh algoritma <i>Naïve Bayes</i> ini mendapatkan hasil yang sebesar 63,61%, <i>precision</i> sebesar 92,19%, dan <i>recall</i> sebesar 61,52%. Hasil sentimen dari Aplikasi Samsat <i>Digital Nasional</i> cenderung positif [8]
<p>“<i>Sentiment Analysis Against the Dana E-wallet on Google Play Reviews Using the K-Nearest Neighbor Algorithm</i>”</p> <p>Jurnal Pilar Nusa Mandiri, Universitas Nusa Mandiri, Vol 17, No 1, Maret 2021</p>	<ul style="list-style-type: none"> - Siti Masturoh - Achmad Baroqah Pohan 	<i>K-Nearest Neighbors Algorithm</i>	Setelah melakukan sentiment analisis terhadap review aplikasi Dana pada <i>Google Playstore</i> , dapat diambil kesimpulan bahwa akurasi tertinggi didapat yaitu value k1 sebesar 86,64% pada kelas ke 2 (Bintang 2 dan Bintang 5) [9]
<p>“<i>Sentiment Analysis Online Shop On The Play Store Using Method Support Vector Machine (SVM)</i>”</p> <p>Seminar Nasional Informatika, Universitas Pembangunan Nasional Veteran Yogyakarta, Vol 1, No 1, 2020</p>	<ul style="list-style-type: none"> - Muhammad Iqbal Ahmadi - Fuji Apriani - Mia Kurniasari - Siti Handayani - Dudih Gustian 	<i>Support Vector Machine (SVM)</i>	Setelah 5 aplikasi <i>e-commerce</i> yang beredar di <i>Google Play Store</i> diteliti, didapatkan hasil kelima akurasi menggunakan metode <i>Support Vector Machine</i> . Aplikasi dengan akurasi tertinggi adalah Tokopedia sebesar 90,67%, JD.ID sebesar 75,33%, Blibli sebesar 74,00%, Shopee sebesar 70,00%, dan yang paling rendah yaitu Lazada sebesar 69,00% [10]
<p>“<i>App Review Sentiment Analysis Shopee Application In Google</i>”</p>	<ul style="list-style-type: none"> - Dany Prاتمanto - Rousyati Rousyati 	<i>Naïve Bayes</i>	Setelah menganalisis aplikasi Shopee pada <i>Google Play Store</i> , didapatkan bahwa

Nama Artikel	Nama Penulis	Teknik Penelitian	Hasil Penelitian
<p><i>Play Store Using Naive Bayes Algorithm</i> <i>Journal of Physics: Conference Series</i>, STMIK Nusa Mandiri Jakarta, Universitas Bina Sarana Informatika, Vol 1641, 6-7 Agustus 2020</p>	<ul style="list-style-type: none"> - Fanny Fatma Wati - Andrian Eko Widodo - Suleman Suleman - Ragil Wijianto 		<p>akurasi hampir mencapai hasil sempurna yaitu sebesar 96,667% [11]</p>
<p>“<i>Sentiment Analysis for Go-Jek on Google Play Store</i>” <i>Journal of Physics: Conference Series</i>, STMIK Amikom Purwokerto, Vol 1641, 2019</p>	<ul style="list-style-type: none"> - Sitaresmi Wahyu Handani - Dhanar Intan Surya Saputra - Hasirun Rizky Mega Arino - Gita Fiza Asyrofi Ramadhan 	<i>Naïve Bayes</i>	<p>Hasil sentimen yang didapatkan dari aplikasi Gojek di <i>Google Play Store</i> yaitu negatif dengan kemunculan kata yang paling besar sebesar 1.3125 [12]</p>
<p>“Analisis Sentimen Aplikasi <i>E-Government</i> Pada <i>Google Play</i> Menggunakan Algoritma <i>Naive Bayes</i>” JATISI (Jurnal Teknik Informatika dan Sistem Informasi), Universitas Kristen Satya Wacana, Vol. 9, No. 2, Juni 2022</p>	<ul style="list-style-type: none"> - Artanti Inez Tanggraeni - Melkior N. N. Sitokdana 	<i>Naïve Bayes</i>	<p>Hasil keseluruhan dari sentimen analisis ini mendapati bahwa Sebagian besar respon pengguna terhadap aplikasi ini adalah positif dengan jumlah positif sebesar 407 dan 235 yang negatif. Hasil algoritma <i>Naive Bayes</i> dengan menggunakan TF-IDF mendapati angka akurasi sebesar 89% dan <i>precision</i> sebesar 83%, dan <i>recall</i> sebesar 87% [13]</p>
<p>“Analisis Sentimen Review Aplikasi Peduli Lindungi Pada <i>Google Play Store</i> Menggunakan Nbc” JTIK (Jurnal Teknik Informatika Kaputama), Universitas Bunda Mulia, Jurnal Teknik Informatika Kaputama (JTIK) Vol.6, No. 2, Juli 2022</p>	<ul style="list-style-type: none"> - George Kenneth Locarso 	<i>Naïve Bayes</i>	<p>Setelah menggunakan <i>Naive Bayes</i>, dihasilkan akurasi sebesar 83,3% dengan perbandingan sebesar 70:30 [14]</p>

Nama Artikel	Nama Penulis	Teknik Penelitian	Hasil Penelitian
<p>“Analisis Sentimen Ulasan Aplikasi Digital Korlantas POLRI Menggunakan Naïve Bayes pada <i>Google Play Store</i>”</p> <p>KLIK: KAJIAN ILMIAH INFORMATIKA DAN KOMPUTER, Vol.4, No 4, Februari 2024</p>	<ul style="list-style-type: none"> - Syarif Hidayatulloh Wazir Putra - Dimas Febriawan 	Naïve Bayes	<p>Dengan dataset yang berjumlah 1500, terdapat 536 sentimen positif dan 964 sentimen negatif. Sebanyak 300 <i>data training</i> dipakai untuk pengujian ini yang menghasilkan <i>accuracy</i> sebesar 70.33%, nilai <i>recall</i> sebesar 74.77%, dan nilai <i>precision</i> sebesar 56.34% [15]</p>
<p>“Sentimental Analysis on Web Scraping Using Machine Learning Method”</p> <p><i>Journal of Information and Computational Science</i>, Vol 12 issue 8, 7 Agustus 2022 (<i>Scopus</i>)</p>	<ul style="list-style-type: none"> - Saurabh Sahu - Km Divya - Dr. Neeta Rastogi - Puneet Kumar Yadav - Dr. Yusuf Perwej 	CNN-LSTM	<p>Setelah menggunakan CNN-LSTM untuk menemukan akurasi terbaik untuk melihat review pada pasar e-commerce, hasil akurasi dapat tercapai pada angka tertinggi yaitu 97,8 % di mana akurasi ini nyaris mendekati hasil akurasi yang sempurna [16]</p>
<p>“Analisis Sentimen Terhadap Implementasi Program Merdeka Belajar Kampus Merdeka Menggunakan Naïve Bayes, K-Nearest Neighbors Dan Decision Tree”</p> <p>Jurnal Media Informasi Budidarma, Vol 6, No. 2, 25 April 2022</p>	<ul style="list-style-type: none"> - Abdul Rozaq - Yessi Yunitasari - Kelik Sussolaikah - Eka Resty Novieta Sari - Restyono Ilham Syahputra 	Naïve Bayes, <i>K-Nearest Neighbors Algorithm</i> , <i>Decision Tree</i>	<p>Didapatkan hasil dari penelitian menggunakan algoritma Naive Bayes mendapatkan akurasi sekitar 99,22%, KNN sebesar 96,90%, dan Decision Tree sebesar 37,21% [17]</p>

Pada jurnal penelitian yang berjudul “Analisis Sentimen pada Ulasan Aplikasi *Amazon Shopping* di *Google Play Store* Menggunakan *Naive Bayes Classifier*”, dihasilkan bahwa akurasi yang didapatkan cukup tinggi yaitu 82,15%, *presicion* yaitu 72,25%, *fi-score* yaitu 77,41% dan *recall* yaitu 83,49% [6]. Akurasi

terbaik yang dihasilkan adalah *Multinomial NB* sebesar 86,74%. Untuk mengisi *gap analysis* melalui penelitian skripsi ini, akurasi dapat diraih dengan jumlah yang lebih tinggi dengan menggunakan algoritma lain yang bertipe klasifikasi sehingga dapat memperoleh akurasi yang lebih akurat. Selain itu, algoritma *Naïve Bayes* terlalu umum dipakai sebagai algoritma sentimen analisis. Untuk itu, pada skripsi ini algoritma lain digunakan seperti *K-Nearest Neighbors* dan *Decision Tree*. Untuk tahap visualisasi, pada penelitian terbaru ini digunakan website sehingga pengguna awam dapat menggunakannya dengan mudah secara interaktif.

Untuk jurnal yang berjudul “Implementasi Algoritma *Naïve Bayes* untuk Analisis Sentimen Ulasan Shopee pada Google Play Store “ dihasilkan nilai *precision* sebesar 83%, *recall* 100 %, dan *f1-score* 91% di mana hasil tersebut sangat tinggi [7]. Untuk mengisi *gap analysis*, maka pada skripsi ini akan di tambahkan pembagian *data training* dan *data testing* yang berbeda yaitu 70:30 dan 90:10 selain yang digunakan oleh jurnal sebelumnya 80:20 untuk menguji akurasi dengan data split yang berbeda. Selain itu, algoritma *Naïve Bayes* juga terlalu umum untuk penelitian sejenisnya sehingga pada skripsi ini menggunakan algoritma lain selain *Naïve Bayes* yaitu *Decision Tree* dan *K-Nearest Neighbors*.

Untuk jurnal yang berjudul “Analisis Sentimen Ulasan Aplikasi Samsat *Digital Nasional* Pada *Google Playstore* Menggunakan Algoritma *Naïve Bayes*” akurasi yang dihasilkan oleh algoritma *Naïve Bayes* ini mendapatkan hasil yang cukup rendah sebesar 63,61%, *precision* sebesar 92,19%, dan *recall* sebesar 61,52% [8]. Untuk mengisi *gap analysis* ini, maka skripsi ini akan memakai algoritma selain *Naïve Bayes* yaitu *Decision Tree* dan *K-Nearest Neighbors*. Jurnal terdahulu ini juga tidak memiliki visualisasi yang cukup yaitu *heatmap* dan *website deployment* untuk mengetahui sentimen positif dan negatif melalui kata sehingga pada skripsi ini kedua visualisasi tersebut akan ditambahkan.

Untuk jurnal yang berjudul “*Sentiment Analysis Against the Dana E-wallet on Google Play Reviews Using the K-Nearest Neighbor Algorithm*” akurasi yang dihasilkan cukup tinggi yaitu sebesar 86,64% [9]. Akurasi tersebut cukup tinggi namun masih banyak *gap analysis* yang harus diperbaiki pada skripsi ini yaitu

kurangnya visualisasi seperti *plot*, *heatmap*, dan *website deployment* sehingga harus ditambahkan pada skripsi ini. Selain itu, jumlah *data* yang digunakan cenderung sedikit yaitu 500 *data* sehingga pada skripsi ini data yang digunakan lebih banyak yaitu 1500 *data*.

Untuk jurnal yang berjudul “*Sentiment Analysis Online Shop On The Play Store Using Method Support Vector Machine (SVM)*” hasil penelitian menunjukkan aplikasi dengan akurasi tertinggi adalah Tokopedia sebesar 90,67%, JD.ID sebesar 75,33%, Blibli sebesar 74,00%, Shopee sebesar 70,00%, dan yang paling rendah yaitu Lazada sebesar 69,00% [10]. Untuk mengisi *gap analysis* dan meningkatkan akurasi maka dibutuhkan TF-IDF sebagai metode yang mengubah kata atau kalimat menjadi sebuah bobot angka sehingga akurasi pada algoritma lebih meningkat. Selain itu, algoritma yang digunakan hanya 1 sehingga untuk mengisi *gap analysis* ini digunakan 2 algoritma sebagai perbandingan yaitu *K-Nearest Neighbors* dan *Decision Tree*.

Untuk jurnal yang berjudul “*App Review Sentiment Analysis Shopee Application In Google Play Store Using Naive Bayes Algorithm*” akurasi yang didapat sangatlah tinggi yaitu mencapai 96,667%, *precision* 100% , *recall* 93.33% sehingga algoritma *Naive Bayes* terbukti paling efektif dalam menangani sentimen analisis pada penelitian ini [11]. *Gap analysis* yang perlu dipenuhi adalah visualisasi yang lebih banyak seperti adanya *plot* atau *wordcloud*.

Untuk jurnal yang berjudul “*Analisis Sentimen Aplikasi E-Government Pada Google Play Menggunakan Algoritma Naive Bayes*”, Hasil algoritma *Naive Bayes* dengan menggunakan TF-IDF mendapati angka akurasi sebesar 89% dan *precision* sebesar 83%, dan *recall* sebesar 87% [13]. Akurasi yang didapatkan cukup tinggi. *Gap analysis* yang perlu dipenuhi adalah penggunaan data yang lebih banyak dari 674 *data* menjadi 1500 *data* dan perlunya pengujian menggunakan algoritma yang berbeda yaitu *Decision Tree* dan *K-Nearest Neighbors*.

Untuk jurnal yang berjudul “*Analisis Sentimen Review Aplikasi Peduli Lindungi Pada Google Play Store Menggunakan Nbc*”, dihasilkan akurasi sebesar 83,3% dengan perbandingan sebesar 70:30 [14]. *Gap analysis* yang perlu diperbaiki

pada penelitian ini adalah penggunaan data training dan data testing yang lebih banyak yaitu 80:20 dan 90:10 untuk menguji algoritma yang dijalankan untuk memperoleh hasil yang terbaik. Selain itu, algoritma yang dipakai yaitu *Naïve Bayes* terlalu umum digunakan untuk analisis sentimen sehingga diperlukan algoritma yang beragam seperti *Decision Tree* dan *K-Nearest Neighbors*. Pada tahap *data preprocessing*, kata-kata yang memiliki kesalahan dalam penulisan diubah menjadi kata yang lebih baku.

Untuk jurnal yang berjudul “Analisis Sentimen Ulasan Aplikasi Digital Korlantas POLRI Menggunakan *Naïve Bayes* pada *Google Play Store*” menghasilkan *accuracy* sebesar 70.33%, nilai *recall* sebesar 74.77%, dan nilai *precision* sebesar 56.34% [15]. Akurasi yang dihasilkan cukup rendah sehingga perlunya penggunaan algoritma lain selain *Naïve Bayes* untuk mencari akurasi yang lebih tinggi karena *Naïve Bayes* tidak terlalu efektif dalam menjalankan analisis sentimen pada jurnal ini. Selain itu, penggunaan *data training* dan *data testing* dengan rasio yang lebih banyak seperti 70:30 atau 90:10 dapat membantu mencari akurasi yang lebih tinggi.

Untuk jurnal yang berjudul Analisis Sentimen Terhadap Implementasi Program Merdeka Belajar Kampus Merdeka Menggunakan *Naïve Bayes*, *K-Nearest Neighbors* Dan *Decision Tree*, didapatkan hasil dari penelitian menggunakan algoritma *Naïve Bayes* mendapatkan akurasi sekitar 99,22%, *KNN* sebesar 96,90%, dan *Decision Tree* sebesar 37,21% [17]. *Research analysis gap* yang ingin dicapai oleh penelitian ini adalah peningkatan akurasi pada penggunaan algoritma *Decision Tree* serta penggunaan pembagian *data training* dan *data testing* yang lebih beragam dari sebelumnya 80:20 bertambah menjadi 2 yaitu 70:30 dan 90:10 sehingga pada penelitian ini menggunakan 3 *splitting* data yang berbeda. Selain itu, *splitting* data ini juga bertujuan untuk mengetahui tingkat konsistensi algoritma dalam skenario yang berbeda-beda serta seberapa efektif pembagian data ini terhadap jumlah data yang lebih besar.

2.2 Teori yang digunakan untuk Pengujian

Bagian teori skripsi ini berisi literatur dan tinjauan teori mengenai dasar-dasar pada pembuatan skripsi serta kerangka yang menjadi pondasi dari skripsi.

2.2.1 Analisis Sentimen

Analisis sentimen atau opinion mining adalah deteksi sikap-sikap terhadap objek atau orang. Sentimen analisis merupakan pendekatan menggunakan analisis yang bertujuan untuk menentukan subjektivitas suatu teks [18]. Kegunaan dari analisis sentimen adalah untuk mendapatkan gambaran sentimen positif dan negatif pada data yang sedang diteliti. Sentimen analisis ini memiliki tiga hasil yaitu sentimen positif dan sentimen negatif. Sentimen analisis ini dapat digunakan untuk evaluasi terhadap respon pengguna dalam menggunakan suatu produk atau aplikasi.

2.2.2 Teori Quora

Quora merupakan aplikasi social media yang berfokus pada kegiatan tanya jawab antar pengguna. Pengguna dapat mengajukan pertanyaan, memberikan jawaban, dan juga berdiskusi didalam forum yang telah disediakan oleh Quora. Pengguna dapat melihat pertanyaan dan jawaban yang telah dibuat oleh pengguna lain pada menu beranda yang telah disesuaikan dengan algoritma berdasarkan kebiasaan pengguna. Quora dikembangkan pada tahun 2009 dan sekarang mempunyai lebih dari 400.000 topik [19].

2.2.3 *Sentiment Labelling*

Sentiment labelling merupakan proses pelabelan sentimen. Berdasarkan jurnal yang berjudul Klasifikasi Opini Pengguna Aplikasi Tripadvisor Di Indonesia Dari Google Play Store Menggunakan Metode Svm, pelabelan positif dan negatif diatur berdasarkan rating atau bintang yang ada pada aplikasi Google Play Store itu sendiri. Untuk rating atau

score 4 dan 5 dikenakan label positif sementara untuk rating atau score 3 kebawah akan dikenakan label negatif [20].

2.2.4 *Text Mining*

Text mining merupakan proses pengolahan data dari data mentah menjadi data yang terstruktur. Tujuan utama dari *data mining* ini adalah untuk memperoleh informasi yang berguna sehingga informasi tersebut dapat digunakan untuk melakukan proses selanjutnya [21].

Berikut ini adalah langkah-langkah dalam melakukan *text mining*

1. *Case folding*

Case folding merupakan pembersihan kalimat dari sehingga mengurangi tambahan yang tidak perlu seperti spasi, huruf kapital, atau angka.

2. *Tokenization*

Tokenization merupakan pemecahan dari kalimat menjadi kata-kata sehingga menjadi lebih ringkas dan setiap kata-kata tersebut dapat dianalisis oleh algoritma.

3. *Normalization*

Normalization merupakan proses seleksi kalimat-kalimat dalam bahasa Indonesia sehingga kalimat tersebut sesuai dengan standar bahasa Indonesia yang benar. Normalization menggunakan *library* sastrawi pada Python.

4. *Filtering*

Filtering merupakan proses pengurangan bobot kata dengan membuang kata-kata yang tidak penting menggunakan fitur *stopwords*.

5. *Stemming*

Stemming merupakan proses penghilangan kata imbuhan pada bahasa Indonesia menjadi kata dasar.

2.3 Algoritma yang Digunakan untuk Pengujian

Pada bagian tinjauan teori ini berisi literatur dan teori yang berkaitan dengan algoritma yang akan digunakan untuk melakukan sentimen analisis. Berikut ini adalah teori mengenai *decision tree* dan teori *K-Nearest Neighbors*

2.3.1 Teori Decision Tree

Decision tree merupakan algoritma yang populer karena algoritma ini merupakan algoritma yang cara kerjanya mirip dengan otak manusia. *Decision tree* merupakan algoritma dengan tampilan grafis yang menampilkan berbagai cabang sesuai dengan pengambilan keputusan yang mengindikasikan peluang serta konsekuensi di setiap keputusannya [1]. *Decision tree* biasa digunakan untuk algoritma yang berbentuk regresi dan klasifikasi. *Decision tree* bekerja dengan cara seperti jalan pikiran yang setiap cabangnya membuat keputusannya masing-masing sehingga mempengaruhi hasil-hasil selanjutnya.

2.3.2 Teori K-Nearest Neighbors

K-Nearest Neighbors merupakan algoritma yang bertipe klasifikasi dan juga regresi. Pengertian dari algoritma *K-Nearest Neighbors* adalah algoritma *machine learning* yang serbaguna, yang memiliki cara kerja dengan menemukan nilai k terdekat dan menggunakan kelas dari grup terdekat untuk memprediksi label dari instansi baru. Prinsip kerja dari *K-Nearest Neighbors* ini dengan melakukan identifikasi persamaan antara grup yang sudah dipecah sebelumnya [5].

2.3.3 TF-IDF (*Term Frequency - Inverse Document Frequency*)

TF-IDF merupakan singkatan dari *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF). TF-IDF ini merupakan metode yang mengubah kata atau kalimat menjadi sebuah bobot angka yang dapat dibaca oleh mesin sehingga dapat dijalankan oleh algoritma klasifikasi maupun prediksi secara akurat [20].

Berikut ini adalah penjelasan beserta dengan rumus dari TF-IDF.

○ **Term Frequency (TF):**

- Merupakan frekuensi berapa sering kata tersebut keluar pada dokumen.
- Rumusnya yaitu :

$$TF(t,d) = \frac{\text{jumlah kemunculan term } t \text{ dalam dokumen } d}{\text{total jumlah term dalam dokumen } d}$$

Rumus 2. 1 Rumus TF (t,d)

○ **Inverse Document Frequency (IDF):**

- Jika kemunculannya banyak tersebar di dokumen maka bobot suatu *term* akan dikurangi
- Rumusnya yaitu :

$$IDF(t,D) = \log \frac{\text{Jumlah total dokumen dalam korpus}}{\text{jumlah dokumen yang mengandung term } t}$$

Rumus 2. 2 Rumus IDF(t,D)

○ **TF-IDF Score:**

- Rumusnya yaitu mengalikan nilai TF dengan nilai IDF:

$$TF-IDF(t,d,D) = TF(t,d) \times IDF(t,D)$$

Rumus 2. 3 Rumus TF-IDF(t,d,D)

2.4 Tools yang digunakan untuk Pengujian

Berikut ini adalah tinjauan teori dari alat-alat yang digunakan untuk melakukan *machine learning* sentimen analisis.

2.4.1 **Jupyter Notebook**

Jupyter Notebook adalah suatu aplikasi berbasis *web* digunakan untuk menulis *program* dengan pendekatan *console*, komputasi secara interaktif, bisa diakses dari komputer klien [22]. *Jupyter Notebook* ini

sangat berguna dalam eksplorasi *data*, analisis, dan pengolahan *data*. *Jupyter Notebook* ini dapat digunakan untuk perhitungan komputasi seperti kode dengan berbagai macam metode ke dalam satu dokumen yang sama, melakukan kolaborasi dengan pengguna lain, dan mampu melakukan visualisasi untuk memudahkan pengguna awam dalam membaca hasil komputasi. Bagi para *data scientist*, menguasai *Jupyter* adalah langkah penting untuk berkolaborasi dan menghasilkan wawasan dari data yang digunakan.

2.4.2 Python

Python adalah sebuah bahasa pemrograman yang serbaguna dan digunakan untuk berbagai tujuan, termasuk membangun situs web, mengembangkan perangkat lunak, melakukan analisis data, dan otomatisasi tugas. *Python* diciptakan oleh Guido Van Rossum pada tahun 1991. *Python* memiliki kelebihan seperti penulisan *syntax* yang mudah dan *python* memiliki *library* yang lengkap serta bersifat open source sehingga dukungan komunitas sangat kuat untuk proses update *library* [23]

