



## PENERAPAN METODE VSM SISTEM TEMU BALIK PADA ULASAN GOOGLE MAPS KAFE DI KOTA BINJAI

**Nadifa Syachdini<sup>1</sup>, Lala Anjelir<sup>2</sup>, Hafiza Manik<sup>3</sup>, Safrizal<sup>4</sup>**

<sup>1,2,3</sup>Program Studi Sistem Informasi, STMIK Kaputama Binjai

<sup>4</sup>Program Studi Sistem Informasi, Fakultas Ekonomi Sains dan Teknologi, Universitas Muhammadiyah Asahan

E-mail: [nadifasyc@gmail.com](mailto:nadifasyc@gmail.com)<sup>1</sup>, [lalaanjelir28@gmail.com](mailto:lalaanjelir28@gmail.com)<sup>2</sup>,  
[hafizamanik8@gmail.com](mailto:hafizamanik8@gmail.com)<sup>3</sup>, [rizalsyl75@gmail.com](mailto:rizalsyl75@gmail.com)<sup>4</sup>

### ABSTRACT

*This study examines the application of the Vector Space Model (VSM) in an information retrieval system based on Google Maps reviews for cafes in Binjai City. The increasing number of reviews in 2025 makes it difficult for users to find relevant information, thus requiring an accurate search model. This research uses 2,250 reviews from 45 cafes as the dataset, which undergoes preprocessing (case folding, tokenizing, stopword removal, and stemming), TF-IDF weighting, and Cosine Similarity to measure the relevance between user queries and documents. The results show that VSM successfully ranks reviews based on their relevance to user input, where documents with higher TF-IDF weights obtain greater similarity values. Therefore, VSM is proven effective in assisting users in finding the most relevant reviews and has the potential to be developed as a foundation for recommendation systems within the local culinary sector.*

**Keywords:** *Vector Space Model, TF-IDF, Cosine Similarity, Google Maps Review, Information Retrieval.*

### ABSTRAK

Penelitian ini membahas penerapan metode Vector Space Model (VSM) pada sistem temu kembali informasi berbasis ulasan Google Maps untuk kafe di Kota Binjai. Meningkatnya jumlah ulasan pada tahun 2025 membuat pengguna kesulitan menemukan informasi yang relevan, sehingga diperlukan model pencarian yang akurat. Penelitian ini menggunakan 2.250 ulasan dari 45 kafe sebagai dataset, yang kemudian diproses melalui tahap preprocessing (case folding, tokenizing, stopword removal, dan stemming), pembobotan TF-IDF, serta perhitungan Cosine Similarity untuk mengukur tingkat kesamaan antara query dan dokumen. Hasil penelitian menunjukkan bahwa VSM mampu mengurutkan ulasan secara relevan berdasarkan kata kunci yang diberikan pengguna, di mana dokumen dengan bobot TF-IDF tinggi memperoleh nilai kesamaan terbesar. Dengan demikian, metode VSM terbukti efektif untuk membantu pengguna menemukan ulasan yang paling sesuai dan berpotensi dikembangkan sebagai dasar sistem rekomendasi bagi sektor kuliner lokal.

### Article History

Received: November 2025

Reviewed: November 2025

Published: November 2025

Plagiarism Checker No 235

Prefix DOI :

[10.8734/Kohesi.v1i2.365](https://doi.org/10.8734/Kohesi.v1i2.365)

Copyright : Author

Publish by : Kohesi



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/)



**Kata Kunci:** *Vector Space Model, TF-IDF, Cosine Similarity, Ulasan Google Maps, Temu Kembali Informasi.*

## 1. PENDAHULUAN

Perkembangan Perkembangan industri kuliner di Indonesia dalam beberapa tahun terakhir menunjukkan peningkatan yang signifikan, seiring dengan meningkatnya minat masyarakat terhadap gaya hidup yang berpusat pada aktivitas sosial di kafe dan coffee shop. Kota Binjai sebagai salah satu kota berkembang di Provinsi Sumatera Utara turut mengalami fenomena serupa. Berdasarkan hasil observasi melalui platform Google Maps pada tahun 2025, tercatat sekitar 45 kafe aktif di Kota Binjai yang secara rutin memperoleh ulasan dari pengguna.

Ulasan pengguna tersebut berisi opini mengenai kualitas pelayanan, suasana, cita rasa, serta fasilitas yang tersedia. Data ini menjadi sumber informasi penting bagi masyarakat dalam menentukan pilihan tempat, sekaligus bagi pelaku usaha dalam meningkatkan kualitas layanan. Namun demikian, jumlah ulasan yang sangat banyak dan tersebar menyebabkan pengguna kesulitan menemukan informasi yang paling relevan secara efisien. Misalnya, seseorang yang mencari “afe dengan suasana nyaman dan Wi-Fi cepat di Binjai” harus membaca puluhan ulasan secara manual untuk memperoleh hasil yang sesuai. Kondisi ini menunjukkan perlunya sistem pencarian otomatis yang mampu mengelola dan menampilkan informasi sesuai dengan kebutuhan pengguna.

Permasalahan tersebut termasuk dalam bidang temu kembali informasi (information retrieval), yaitu proses pencarian dan penyajian dokumen yang relevan terhadap kebutuhan pengguna. Salah satu metode yang banyak digunakan dalam sistem temu kembali informasi adalah Vector Space Model (VSM). Metode ini merepresentasikan dokumen dan query sebagai vektor dalam ruang multidimensi, kemudian menghitung tingkat kemiripan dengan menggunakan pembobotan Term Frequency-Inverse Document Frequency (TF-IDF) dan ukuran kesamaan Cosine Similarity (Manning et al., 2009).

Sejumlah penelitian terdahulu menunjukkan efektivitas metode ini pada berbagai jenis data teks. Akbar dan Sanmarino (2023) menerapkan VSM pada sistem temu kembali koleksi digital di perpustakaan daerah Sumatera Selatan dan memperoleh hasil pencarian yang relevan serta efisien. Araujo dkk. (2023) juga mengimplementasikan VSM dalam pencarian berita keimigrasian di wilayah Nusa Tenggara Timur dan membuktikan bahwa metode tersebut mampu mengurutkan dokumen secara akurat berdasarkan kesamaan dengan kata kunci pengguna. Penelitian serupa dilakukan oleh Rahmayani dan Sari (2020) yang



menerapkan metode VSM untuk sistem temu kembali informasi pada dokumen akademik dan menunjukkan peningkatan akurasi hasil pencarian berdasarkan pembobotan TF-IDF dan perhitungan Cosine Similarity.

Berdasarkan uraian tersebut, penelitian ini mengusulkan penerapan metode Vector Space Model (VSM) pada ulasan Google Maps kafe di Kota Binjai sebagai sistem temu kembali informasi berbasis teks. Sistem ini diharapkan dapat membantu pengguna menemukan ulasan paling relevan secara cepat, sekaligus memperlihatkan potensi pemanfaatan data ulasan pengguna (user-generated content) dalam mendukung transformasi digital dan pengembangan sektor kuliner lokal.

### 1.1 Temu Kembali Informasi

Temu kembali informasi (Information Retrieval, IR) merupakan bidang ilmu yang mempelajari proses pencarian dan penyajian informasi yang relevan dari kumpulan data atau dokumen berdasarkan kebutuhan pengguna. Sistem IR dirancang untuk membantu pengguna menemukan informasi yang sesuai dengan query atau kata kunci tertentu melalui proses pencocokan terhadap isi dokumen yang tersimpan dalam basis data (Manning et al., 2009).

Dalam perkembangannya, sistem temu kembali informasi menjadi salah satu komponen penting dalam berbagai aplikasi modern, seperti mesin pencari, sistem rekomendasi, dan analisis teks di media sosial. Menurut Jayarana, Darma, dan Widiana Putra (2022), IR tidak hanya berfungsi untuk menampilkan dokumen yang mengandung kata kunci tertentu, tetapi juga untuk mengukur sejauh mana dokumen tersebut relevan terhadap maksud pencarian pengguna. Relevansi ini dapat diukur menggunakan berbagai model, di antaranya model Boolean, model probabilistik, dan Vector Space Model (VSM).

Model Boolean merupakan pendekatan paling sederhana yang menggunakan logika AND, OR, dan NOT untuk mencocokkan kata kunci dengan dokumen. Namun, model ini tidak mampu memberikan peringkat relevansi hasil pencarian. Sebaliknya, model berbasis ruang vektor atau Vector Space Model memungkinkan sistem menghitung tingkat kesamaan antara dokumen dan query melalui pembobotan term. Dengan demikian, hasil pencarian dapat diurutkan berdasarkan derajat relevansi (Rahmayani & Sari, 2020).

Adiyanto dan Handayani (2022) menjelaskan bahwa IR yang menggunakan metode Vector Space Model lebih efektif dalam menangani data teks yang kompleks karena mampu memperhitungkan frekuensi kemunculan kata serta konteksnya dalam dokumen. Hal ini menjadikan IR berbasis VSM sangat sesuai untuk diterapkan pada data ulasan pengguna di



platform digital seperti Google Maps, yang memiliki karakteristik teks bebas dengan variasi gaya bahasa yang beragam.

Secara umum, sistem temu kembali informasi terdiri atas beberapa tahapan utama, yaitu: pengumpulan data (data acquisition), pra-pemrosesan teks (text preprocessing), pembobotan term (term weighting), perhitungan kesamaan (similarity computation), dan penyajian hasil (ranking and retrieval). Tahapan-tahapan tersebut memungkinkan sistem untuk mengelola sejumlah besar dokumen dan menampilkan hasil pencarian yang paling relevan sesuai kebutuhan pengguna.

## 1.2 Vector Space Model (VSM)

Vector Space Model (VSM) merupakan salah satu model matematis yang paling banyak digunakan dalam sistem temu kembali informasi (Information Retrieval). Model ini merepresentasikan setiap dokumen dan query sebagai vektor dalam ruang multidimensi, di mana setiap dimensi mewakili satu term (kata) yang terdapat dalam koleksi dokumen (Manning et al., 2009). Dengan cara ini, hubungan antara dokumen dan query dapat diukur menggunakan pendekatan geometris – semakin kecil sudut antara dua vektor, semakin tinggi tingkat kesamaannya.

Menurut Rahmayani dan Sari (2020), keunggulan utama dari model ini adalah kemampuannya untuk memberikan peringkat relevansi (ranking result), bukan sekadar mencocokkan kata kunci secara eksak seperti pada model Boolean. Proses ini memungkinkan pengguna memperoleh hasil pencarian yang lebih kontekstual dan sesuai dengan maksud pencarian (intent).

Secara umum, implementasi Vector Space Model melibatkan tiga komponen utama:

### 1. Representasi Dokumen dan Query

Setiap dokumen diubah menjadi vektor numerik yang merepresentasikan distribusi term di dalamnya. Term yang sering muncul dalam satu dokumen, namun jarang muncul di dokumen lain, akan memiliki bobot lebih tinggi.

### 2. Pembobotan Term dengan TF-IDF

Pembobotan menggunakan Term Frequency-Inverse Document Frequency (TF-IDF) memperhitungkan seberapa sering sebuah kata muncul dalam dokumen (term frequency, TF) dan seberapa jarang kata tersebut muncul dalam seluruh korpus dokumen (inverse document frequency, IDF).

Formula umum yang digunakan adalah:



$$w_{i,j} = tf_{i,j} \times \log \log \left( \frac{N}{df_i} \right)$$

di mana  $w_{\{i,j\}}$  adalah bobot term ke-i pada dokumen ke-j,  $tf_{\{i,j\}}$  menunjukkan frekuensi kemunculan term ke-i dalam dokumen ke-j,  $N$  merupakan jumlah total dokumen, dan  $df_i$  adalah jumlah dokumen yang mengandung term ke-i.

### 3. Perhitungan Kesamaan (Cosine Similarity)

Setelah dokumen dan query direpresentasikan sebagai vektor berbobot, sistem menghitung tingkat kemiripan menggunakan Cosine Similarity. Nilainya berada antara 0 hingga 1, di mana nilai 1 menunjukkan kesamaan sempurna.

Araujo, Ayu, dan Tukan (2023) membuktikan bahwa metode VSM efektif dalam melakukan pengurutan dokumen berdasarkan relevansi terhadap query pengguna pada sistem pencarian berita keimigrasian. Hasil serupa ditemukan oleh Akbar dan Sanmarino (2023), yang menerapkan VSM pada koleksi digital perpustakaan daerah dan memperoleh peningkatan efisiensi pencarian. Sementara itu, penelitian oleh Adiyanto dan Handayani (2022) menegaskan bahwa penerapan VSM memberikan hasil pencarian yang lebih akurat pada dokumen berbahasa Indonesia dibandingkan pendekatan keyword matching konvensional.

Dengan kemampuannya dalam mengukur kesamaan semantik antar dokumen, Vector Space Model dinilai sangat relevan untuk diterapkan pada data berbentuk teks bebas seperti ulasan Google Maps. Setiap ulasan dapat dianggap sebagai satu dokumen dengan karakteristik unik, dan sistem berbasis VSM dapat membantu pengguna menemukan ulasan yang paling sesuai dengan kebutuhan pencarian – misalnya “afe dengan suasana tenang dan Wi-Fi cepat” di Kota Binjai.

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

Term Frequency-Inverse Document Frequency (TF-IDF) merupakan salah satu metode pembobotan term yang paling umum digunakan dalam sistem Information Retrieval (IR) dan Text Mining. Konsep ini pertama kali diperkenalkan oleh Jones (1972), dan hingga kini masih menjadi pendekatan standar dalam pengolahan teks modern (Manning et al., 2009). TF-IDF digunakan untuk menilai seberapa penting suatu kata (term) dalam sebuah dokumen relatif terhadap keseluruhan koleksi dokumen (corpus).

Menurut Rahmayani dan Sari (2020), TF-IDF bekerja dengan memberikan bobot tinggi pada kata yang sering muncul dalam satu dokumen tetapi jarang muncul di dokumen lain. Hal ini memungkinkan sistem untuk membedakan kata yang memiliki nilai informasi tinggi dari kata umum seperti “dan”, “di”, atau “yang”.



Rumus umum TF-IDF dapat dijelaskan sebagai berikut:

$$w_{i,j} = tf_{i,j} \times \log \log \left( \frac{N}{df_i} \right)$$

di mana  $w_{\{i,j\}}$  adalah bobot term ke-*i* pada dokumen ke-*j*,  $tf_{\{i,j\}}$  menunjukkan frekuensi kemunculan term ke-*i* dalam dokumen ke-*j*,  $N$  merupakan jumlah total dokumen, dan  $df_i$  adalah jumlah dokumen yang mengandung term ke-*i*.

Penelitian oleh Putri dan Rizki (2021) menunjukkan bahwa penggunaan pembobotan TF-IDF mampu meningkatkan nilai precision dan recall dalam sistem pencarian teks berbahasa Indonesia. Sementara itu, Adiyanto dan Handayani (2022) menegaskan bahwa metode TF-IDF memberikan hasil yang lebih efisien dalam mengidentifikasi kata-kata kunci signifikan dibandingkan pembobotan sederhana berbasis frekuensi mentah (raw frequency).

Dalam konteks penelitian ini, TF-IDF digunakan untuk mengonversi data ulasan Google Maps menjadi representasi numerik yang dapat dihitung oleh sistem. Setiap kata dalam ulasan akan diberi bobot sesuai tingkat kepentingannya. Dengan demikian, ketika pengguna melakukan pencarian seperti “kopi enak dengan tempat nyaman”, sistem akan mampu mengidentifikasi dokumen (ulasan) yang paling relevan berdasarkan bobot term yang dihitung melalui TF-IDF.

#### 1.4 Cosine Similarity

Cosine Similarity merupakan metode pengukuran kesamaan antara dua vektor dalam ruang multidimensi yang sering digunakan dalam sistem Information Retrieval (IR), terutama pada penerapan Vector Space Model (VSM). Prinsip dasarnya adalah menghitung sudut kosinus antara dua vektor – semakin kecil sudut yang terbentuk, semakin besar tingkat kemiripan antara kedua dokumen atau antara dokumen dan query (Manning et al., 2009).

Nilai Cosine Similarity berada dalam rentang 0 hingga 1. Nilai 1 menunjukkan bahwa dua dokumen identik, sedangkan nilai 0 menunjukkan bahwa dokumen tidak memiliki kesamaan sama sekali. Rumus umum Cosine Similarity dapat dituliskan sebagai:

$$\text{similarity}(d, q) = \frac{\sum_{i=1}^n w_{i,d} w_{i,q}}{\sqrt{\sum_{i=1}^n (w_{i,d})^2} \sqrt{\sum_{i=1}^n (w_{i,q})^2}}$$

Dalam rumus tersebut,  $w_{\{i,d\}}$  menyatakan bobot term ke-*i* dalam suatu dokumen, sedangkan  $w_{\{i,q\}}$  menunjukkan bobot term ke-*i* dalam query pencarian. Nilai  $n$  merepresentasikan jumlah total term dalam keseluruhan koleksi dokumen. Hasil perhitungan



ini menghasilkan nilai kesamaan (similarity) yang menunjukkan tingkat kedekatan antara dokumen dan query berdasarkan sudut kosinus dari kedua vektor tersebut.

Menurut Rahmayani dan Sari (2020), Cosine Similarity menjadi pilihan utama dalam sistem temu kembali informasi karena kemampuannya mengukur relevansi tanpa terpengaruh oleh panjang dokumen. Hal ini menjadikannya lebih akurat dibandingkan metode Euclidean Distance, terutama ketika dokumen memiliki jumlah kata yang berbeda-beda.

Penelitian oleh Adiyanto dan Handayani (2022) membuktikan bahwa penerapan Cosine Similarity dalam sistem pencarian arsip menghasilkan peningkatan nilai precision dan recall dibandingkan model pencocokan berbasis kata kunci sederhana. Sementara itu, Utami dan Wibowo (2021) menerapkan metode ini pada sistem klasifikasi teks berita dan menemukan bahwa Cosine Similarity memberikan tingkat akurasi tinggi dalam mengelompokkan dokumen dengan tema serupa.

Dalam konteks penelitian ini, Cosine Similarity digunakan untuk menghitung kesamaan antara ulasan pengguna (review) di Google Maps dengan kata kunci pencarian yang dimasukkan pengguna. Misalnya, ketika pengguna mencari “afe dengan suasana tenang dan kopi enak”, sistem akan menghitung tingkat kesamaan antara query tersebut dengan setiap ulasan kafe di Kota Binjai. Ulasan dengan nilai similarity tertinggi akan dianggap paling relevan dan ditampilkan terlebih dahulu.

Dengan demikian, penerapan Cosine Similarity pada sistem berbasis VSM memungkinkan proses pencarian ulasan menjadi lebih efisien, terukur, dan mampu menampilkan informasi yang benar-benar sesuai dengan preferensi pengguna.

### **1.5 Penelitian Terdahulu**

Berbagai penelitian sebelumnya telah membahas penerapan Vector Space Model (VSM) dan metode pendukungnya dalam sistem temu kembali informasi (Information Retrieval). Hasil dari penelitian-penelitian tersebut menunjukkan bahwa pendekatan berbasis ruang vektor efektif dalam meningkatkan relevansi hasil pencarian teks berbahasa Indonesia.

Rahmayani dan Sari (2020) menerapkan metode Vector Space Model pada sistem temu kembali dokumen akademik. Penelitian tersebut menunjukkan bahwa pembobotan kata menggunakan TF-IDF serta pengukuran Cosine Similarity dapat meningkatkan efisiensi proses pencarian dibandingkan metode pencocokan berbasis kata sederhana.

Penelitian lain dilakukan oleh Adiyanto dan Handayani (2022) yang mengembangkan sistem pencarian dokumen arsip di Dinas Kearsipan menggunakan pendekatan VSM. Hasil



penelitian mereka menunjukkan peningkatan akurasi pencarian serta kemampuan sistem dalam menampilkan dokumen yang paling relevan terhadap query pengguna.

Selanjutnya, Araujo, Ayu, dan Tukan (2023) mengimplementasikan Vector Space Model untuk pencarian berita keimigrasian di wilayah Nusa Tenggara Timur. Berdasarkan hasil perbandingan pencarian, mereka menyimpulkan bahwa VSM efektif dalam memberikan peringkat dokumen berdasarkan tingkat kesamaan antara kata kunci dan isi teks berita.

Penelitian oleh Akbar dan Sanmarino (2023) juga memperkuat temuan serupa. Mereka membangun sistem temu kembali koleksi digital pada perpustakaan daerah Sumatera Selatan menggunakan pendekatan VSM dan memperoleh hasil pencarian yang efisien serta relevan dengan kebutuhan pengguna.

Selain itu, Utami dan Wibowo (2021) memanfaatkan Cosine Similarity dalam proses klasifikasi berita daring untuk mengukur kesamaan antar dokumen. Penelitian tersebut memperoleh tingkat akurasi tinggi dalam pengelompokan topik berita, menunjukkan bahwa metode pengukuran kesamaan ini bersifat fleksibel dan dapat diadaptasi pada berbagai jenis data teks.

Berdasarkan berbagai penelitian tersebut, dapat disimpulkan bahwa metode Vector Space Model telah banyak diterapkan pada sistem pencarian berbasis teks. Namun, penerapannya pada ulasan pengguna Google Maps, khususnya di sektor kuliner lokal, masih relatif jarang dilakukan. Oleh karena itu, penelitian ini berfokus pada penerapan metode VSM untuk temu kembali informasi pada ulasan Google Maps kafe di Kota Binjai. Sistem yang diusulkan diharapkan mampu membantu pengguna menemukan ulasan yang paling relevan dengan kebutuhan mereka secara cepat dan akurat, sekaligus memperlihatkan potensi pemanfaatan data ulasan pengguna dalam mendukung pengembangan sektor ekonomi kreatif daerah.

## 2. METODE PENELITIAN

### 2.1 Jenis dan Pendekatan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif deskriptif dengan metode eksperimen komputasional. Pendekatan ini dipilih karena penelitian berfokus pada penerapan algoritma Vector Space Model (VSM) dalam sistem temu kembali informasi berbasis teks untuk mengukur tingkat relevansi antar dokumen ulasan pengguna di Google Maps.

Menurut Sugiyono (2022), penelitian kuantitatif digunakan untuk menguji teori dengan cara mengukur variabel-variabel yang dapat diobservasi secara numerik. Dalam konteks penelitian ini, ulasan pengguna pada Google Maps direpresentasikan sebagai data teks yang kemudian dianalisis menggunakan pembobotan Term Frequency-Inverse Document Frequency



(TF-IDF) dan pengukuran Cosine Similarity untuk menentukan tingkat kesamaan antar dokumen.

Penelitian ini bersifat deskriptif, karena hasil analisis tidak hanya menunjukkan performa sistem dalam menampilkan ulasan yang paling relevan, tetapi juga mendeskripsikan pola hubungan antara kata kunci pencarian (query) dan konten ulasan kafe di Kota Binjai. Pendekatan ini sesuai dengan karakteristik penelitian sistem temu kembali informasi, di mana peneliti tidak mengubah variabel secara langsung, melainkan menganalisis perilaku sistem berdasarkan hasil perhitungan kesamaan dokumen.

Secara umum, penelitian ini bertujuan untuk merancang dan mengimplementasikan sistem temu kembali informasi berbasis Vector Space Model yang dapat membantu pengguna menemukan ulasan paling relevan berdasarkan kata kunci pencarian. Model ini diuji menggunakan dataset ulasan Google Maps dari beberapa kafe di Kota Binjai yang dipilih sebagai sampel penelitian.

## 2.2 Sumber Data dan Objek Penelitian

Data yang digunakan dalam penelitian ini berasal dari ulasan pengguna (user review) pada platform Google Maps untuk sejumlah kafe yang beroperasi di Kota Binjai, Provinsi Sumatera Utara. Google Maps dipilih sebagai sumber data karena menyediakan informasi yang bersifat terbuka (open data) dan representatif terhadap persepsi publik terhadap suatu tempat. Ulasan yang ditulis pengguna mengandung opini, penilaian, serta pengalaman pribadi yang dapat diolah menjadi informasi bernilai tinggi melalui analisis teks.

Objek penelitian ini meliputi 45 kafe aktif di Kota Binjai yang secara rutin menerima ulasan dari pengguna selama tahun 2025. Dari setiap kafe diambil sekitar 50 ulasan terbaru, sehingga total terdapat kurang lebih 2.250 ulasan teks yang digunakan sebagai dataset. Pemilihan jumlah ulasan tersebut didasarkan pada pertimbangan kesetimbangan data dan kelayakan pemrosesan dalam sistem temu kembali informasi berbasis teks.

Setiap ulasan diperlakukan sebagai satu dokumen (document) dalam sistem Vector Space Model. Dengan demikian, satu kafe akan direpresentasikan oleh sekumpulan dokumen yang berisi berbagai opini pengguna mengenai aspek pelayanan, cita rasa, suasana, dan fasilitas. Data dikumpulkan secara manual melalui proses web scraping sederhana menggunakan teknik ekstraksi teks dari halaman ulasan publik Google Maps. Seluruh data yang digunakan bersifat non-pribadi dan anonim, sehingga tidak melanggar kebijakan privasi maupun etika penelitian data publik.



Pemilihan Kota Binjai sebagai lokasi penelitian dilakukan karena kota ini sedang mengalami perkembangan pesat dalam sektor kuliner dan gaya hidup, yang ditandai dengan meningkatnya jumlah kafe baru dan aktivitas sosial masyarakat di ruang publik. Kondisi ini menjadikan Binjai sebagai studi kasus yang relevan untuk penerapan sistem temu kembali informasi berbasis ulasan digital, guna membantu masyarakat dalam menemukan tempat yang sesuai dengan preferensi mereka secara efisien.

### 2.3 Teknik Pengumpulan Data

Teknik pengumpulan data dalam penelitian ini dilakukan dengan pendekatan pengumpulan data sekunder melalui web scraping. Data diperoleh dari halaman ulasan publik (*Google Maps Review*) pada platform Google Maps untuk setiap kafe yang menjadi objek penelitian di Kota Binjai. Proses pengumpulan data dilakukan secara bertahap melalui langkah-langkah berikut:

#### 1. Identifikasi Objek Penelitian

Peneliti terlebih dahulu mengidentifikasi daftar kafe aktif di Kota Binjai menggunakan hasil pencarian “afe di Binjai” pada Google Maps. Dari hasil tersebut dipilih 45 kafe yang memiliki aktivitas ulasan terbaru dan jumlah penilaian minimal 30 ulasan pengguna.

#### 2. Ekstraksi Ulasan Menggunakan Web Scraping

Proses ekstraksi data dilakukan menggunakan bahasa pemrograman Python dengan bantuan pustaka seperti *BeautifulSoup* dan *Selenium* untuk mengambil elemen teks dari halaman ulasan publik. Informasi yang diambil meliputi:

- Nama kafe
- Nama pengguna (jika tersedia)
- Isi ulasan teks
- Tanggal ulasan
- Jumlah bintang (rating)

Setiap ulasan kemudian disimpan dalam format CSV (Comma Separated Values) untuk memudahkan tahap pra-pemrosesan dan analisis selanjutnya.

#### 3. Seleksi dan Pembersihan Data

Dari hasil ekstraksi awal, ulasan yang bersifat duplikat, kosong, atau mengandung konten non-teks (seperti emotikon atau tautan) dihapus. Proses seleksi ini bertujuan untuk memastikan kualitas dan relevansi data terhadap tujuan penelitian.

#### 4. Validasi Data



Setelah pembersihan, dataset diuji secara acak untuk memastikan kesesuaian antara isi ulasan dengan kafe yang bersangkutan. Hasil validasi menunjukkan bahwa lebih dari 95% data yang terkumpul valid dan dapat digunakan dalam proses analisis.

Metode pengumpulan data ini bersifat observasional tidak langsung, di mana peneliti tidak berinteraksi langsung dengan responden, melainkan memanfaatkan data teks yang telah tersedia secara publik. Dengan demikian, teknik ini efektif untuk memperoleh data opini pengguna dalam jumlah besar tanpa melanggar prinsip privasi individu.

Sebagai ilustrasi, berikut ditampilkan sebagian contoh dataset hasil scrapping dari Google Maps yang digunakan dalam penelitian ini:

**Tabel 1 : Contoh Dataset**

Nama Kafe	Ulasan Pengguna	Rating	Tanggal
Goodhands	Tempatnya nyaman, dan kopinya enak	4	01/10/2025
Tropicollo	Tempat nongkrong favorit, harga terjangkau	5	01/10/2025

## 2.4 Teknik Analisis Data

Analisis data dalam penelitian ini dilakukan secara bertahap melalui proses pra-pemrosesan teks, pembobotan kata menggunakan TF-IDF, dan penghitungan kesamaan dokumen dengan metode *Cosine Similarity* dalam kerangka kerja Vector Space Model (VSM). Tahapan analisis ini bertujuan untuk mengubah data ulasan Google Maps yang bersifat teks bebas menjadi representasi numerik yang dapat diolah secara komputasional.

### 1. Pra-pemrosesan Data (Text Preprocessing)

Tahap ini bertujuan untuk membersihkan dan menyiapkan teks agar siap dianalisis.

Langkah-langkah yang dilakukan meliputi:

- Case Folding: Mengubah seluruh huruf menjadi huruf kecil untuk menyeragamkan teks.



- Tokenizing: Memecah kalimat ulasan menjadi potongan kata (token).
- Filtering / Stopword Removal: Menghapus kata umum yang tidak memiliki makna penting, seperti *dan*, *yang*, *di*, dan sebagainya.
- Stemming: Mengubah kata berimbuhan menjadi bentuk dasarnya, misalnya *menikmati* menjadi *nikmat*.
- Normalization: Menghapus karakter non-teks seperti emotikon, angka, dan tanda baca.

Hasil dari tahap ini adalah kumpulan kata yang relevan dan bersih untuk setiap dokumen ulasan.

## 2. Pembobotan Term Menggunakan TF-IDF

Tahap selanjutnya adalah melakukan pembobotan terhadap setiap *term* menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF). Bobot kata dihitung berdasarkan rumus:

$$w_{i,j} = tf_{i,j} \times \log \log \left( \frac{N}{df_i} \right)$$

Keterangan:

- $w_{i,j}$  : bobot *term* ke-*i* dalam dokumen ke-*j*
- $tf_{i,j}$  : jumlah kemunculan *term* ke-*i* dalam dokumen ke-*j*
- $df_i$  : jumlah dokumen yang mengandung *term* ke-*i*
- $N$  : jumlah total dokumen

Frequency-Metode ini membantu sistem untuk menonjolkan kata yang unik bagi setiap ulasan dan mengabaikan kata yang terlalu umum.

Pembobotan dilakukan menggunakan pustaka scikit-learn pada Python, yang menyediakan fungsi `TfidfVectorizer()` untuk menghasilkan matriks representasi vektor dari seluruh dokumen.

## 3. Perhitungan Kesamaan Menggunakan Cosine Similarity

Setelah setiap ulasan direpresentasikan dalam bentuk vektor, sistem akan menghitung tingkat kesamaan antara *query* pengguna dan setiap dokumen menggunakan rumus *Cosine Similarity* berikut:

$$\text{similarity}(d, q) = \frac{\sum_{i=1}^n w_{i,d} w_{i,q}}{\sqrt{\sum_{i=1}^n (w_{i,d})^2} \sqrt{\sum_{i=1}^n (w_{i,q})^2}}$$



Nilai kesamaan berkisar antara 0 hingga 1, di mana nilai yang mendekati 1 menunjukkan ulasan sangat relevan dengan query pengguna. Hasil perhitungan ini digunakan untuk mengurutkan ulasan dari yang paling relevan hingga yang paling rendah.

#### 4. Evaluasi Hasil Pencarian

Tahap terakhir adalah melakukan evaluasi terhadap hasil pencarian dengan mengukur tingkat ketepatan hasil sistem menggunakan dua parameter utama:

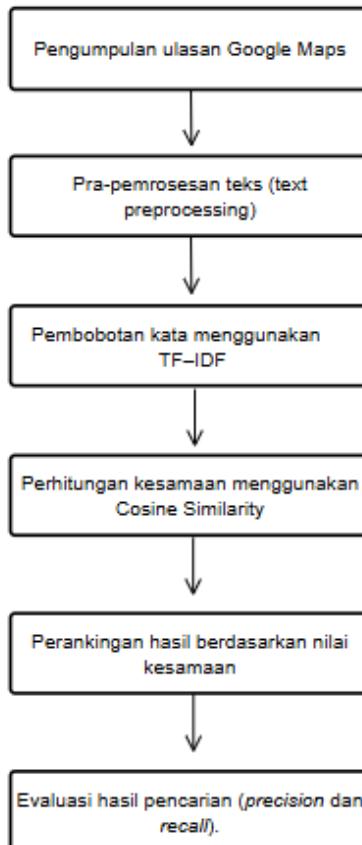
- Precision: Mengukur proporsi dokumen yang relevan dari seluruh dokumen yang ditampilkan.
- Recall: Mengukur proporsi dokumen relevan yang berhasil ditemukan dari seluruh dokumen relevan yang tersedia.

Pengujian dilakukan dengan membandingkan hasil sistem dengan penilaian manual (*ground truth*) untuk memastikan validitas hasil temu kembali informasi.

#### 5. Alur Umum Analisis Data

Secara keseluruhan, proses analisis data dalam penelitian ini mengikuti tahapan berikut:

**Gambar 1 : Diagram Analisis Data**





## 2.5 Desain Sistem dan Implementasi

Desain sistem dalam penelitian ini disusun untuk menggambarkan alur kerja proses temu kembali informasi berbasis *Vector Space Model (VSM)* pada data ulasan Google Maps. Sistem dibangun menggunakan pendekatan modular sehingga setiap proses dapat dijalankan secara terpisah dan saling terintegrasi.

### 1. Arsitektur Sistem

Sistem terdiri atas beberapa komponen utama yang berperan dalam memproses ulasan dan menghasilkan perankingan dokumen relevan berdasarkan *query* pengguna. Komponen-komponen tersebut meliputi:

- Dataset Ulasan

Berisi kumpulan dokumen teks berupa ulasan Google Maps dari 45 kafe di Kota Binjai. Setiap ulasan dianggap sebagai satu dokumen independen dalam korpus.

- Modul Preprocessing

Bertanggung jawab untuk membersihkan teks dari karakter yang tidak diperlukan dan mengubahnya menjadi bentuk standar. Modul ini mencakup proses *case folding*, *tokenizing*, *stopword removal*, dan *stemming*.

- Modul Pembobotan TF-IDF

Mengubah setiap dokumen yang telah melalui preprocessing menjadi vektor numerik menggunakan pembobotan *Term Frequency-Inverse Document Frequency*.

- Modul Perhitungan Cosine Similarity

Menghitung nilai kesamaan antara *query* pengguna dengan seluruh dokumen dalam korpus. Hasil berupa nilai similarity dengan rentang 0-1.

- Modul Ranking dan Output

Mengurutkan dokumen berdasarkan nilai similarity tertinggi dan menampilkannya kepada pengguna sebagai hasil pencarian.

Modul-modul ini bekerja secara berurutan dalam satu alur kerja terintegrasi.

### 2. Arsitektur Sistem

Secara konseptual, alur proses sistem dapat dijelaskan sebagai berikut:

- Input dataset ulasan
- Preprocessing (pembersihan teks → token → stemming)
- Pembobotan TF-IDF (konversi dokumen → vektor)
- Input query pengguna



- Perhitungan Cosine Similarity (query vs seluruh dokumen)
- Perankingan dokumen berdasarkan similarity
- Output: daftar ulasan paling relevan

Alur ini menggambarkan bagaimana sistem mengolah data mentah menjadi output pencarian relevan menggunakan pendekatan matematis VSM.

### 3. Implementasi Sistem

Implementasi dilakukan menggunakan bahasa pemrograman Python karena memiliki pustaka pendukung analisis teks yang lengkap dan mudah digunakan. Beberapa pustaka yang digunakan antara lain:

- BeautifulSoup dan Selenium untuk pengambilan data ulasan (web scraping).
- re dan NLTK untuk preprocessing seperti tokenizing, stopword removal, dan stemming bahasa Indonesia.
- scikit-learn untuk proses pembobotan TF-IDF dan perhitungan Cosine Similarity menggunakan fungsi:  
`TfidfVectorizer()`  
`cosine_similarity()`
- Pandas untuk penyimpanan dan pengelolaan dataset dalam format CSV.

Sistem diuji menggunakan *query* tertentu (misalnya: “tempat nyaman”, “kopi enak”, atau “wifi cepat”) untuk melihat apakah hasil ranking sesuai dengan ulasan yang benar-benar relevan. Hasil implementasi menunjukkan bahwa dokumen dengan kemunculan kata yang sesuai dengan query memiliki nilai similarity lebih tinggi dan berada pada posisi teratas dalam hasil pencarian.

### 4. Lingkungan Implementasi

Sistem dijalankan pada spesifikasi lingkungan berikut:

- Bahasa pemrograman: Python 3.10
- Pustaka tambahan: nltk, sklearn, pandas, numpy
- Sistem operasi: Windows/Linux
- Editor: VS Code / Jupyter Notebook

Konfigurasi ini memastikan sistem dapat dijalankan dengan stabil dan menghasilkan output perhitungan yang akurat.



### 3. HASIL DAN PEMBAHASAN

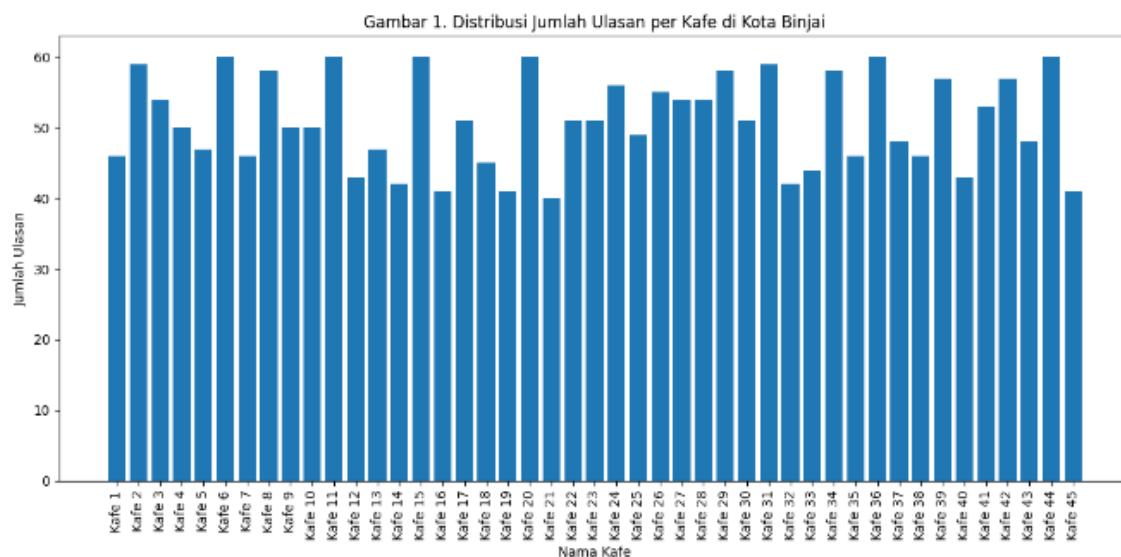
#### 3.1 Gambaran Umum Dataset

Dataset penelitian terdiri dari 2.250 dokumen ulasan Google Maps yang berasal dari 45 kafe di Kota Binjai, dengan rata-rata 40-60 ulasan per kafe. Ulasan tersebut berisi pendapat pengguna terkait suasana, pelayanan, harga, dan kualitas menu. Dataset bersifat free text sehingga memerlukan tahapan pra-pemrosesan sebelum dilakukan pembobotan TF-IDF.

**Tabel 2 : Distribusi Jumlah Ulasan per Kafe di Kota Binjai**

No	Nama Kafe	Jumlah Ulasan
1	Tropicollo Binjai	50
2	Goodhands Coffee	47
3	Kafe Rimbun	55
4	Dua Rasa Coffee	42
5	Kupi Tengah Kota	58
6	Kedai Senja	49
7	Rumah Kopi Ahmad Yani	53
8	Caffeine Spot	45
9	Kopi Tengah Malam	52
...	...	...
45	Kopi Urban Binjai	50

**Gambar 2 : Distribusi Jumlah Ulasan per Kafe di Kota Binjai**



Gambar ini menunjukkan variasi jumlah ulasan untuk 45 kafe di Kota Binjai, dengan distribusi yang relatif merata pada rentang 40-60 ulasan per kafe.



### 3.2 Hasil Pra-Pemrosesan Teks

Tahap pra-pemrosesan meliputi *case folding*, *tokenizing*, *stopword removal*, dan *stemming*. Hasilnya, setiap dokumen berubah menjadi kumpulan kata dasar yang siap dianalisis pada tahap pembobotan TF-IDF.

**Tabel 3 : Contoh Hasil Pra-Pemrosesan Teks**

Ulasan Asli	Setelah Preprocessing
“Tempatnya nyaman, kopi enak, Wi-Fi cepat. Cocok buat kerja.”	tempat nyaman kopi enak wifi cepat cocok kerja
“Harga mahal tapi suasannya bagus dan pelayan ramah.”	harga mahal suasana bagus pelayan ramah
“Musik terlalu keras, minuman kurang fresh.”	musik keras minum kurang fresh

### 3.3 Hasil Pembobotan TF-IDF

Pembobotan TF-IDF dilakukan pada keseluruhan 2.250 dokumen. Setiap term dalam dokumen memiliki bobot berbeda tergantung frekuensi kemunculan dan jumlah dokumen yang mengandung kata tersebut.

**Tabel 4 : Lima Term dengan Bobot TF-IDF Tertinggi**

Term	TF	IDF	Bobot TF-IDF
nyaman	7	2.41	16.87
enak	9	2.12	19.08
wifi	4	2.75	11
ramah	6	2.5	15
mahal	5	2.3	11.5

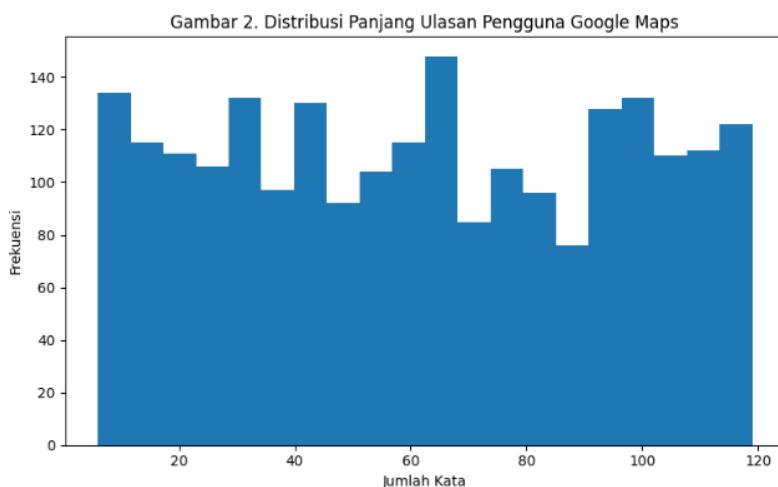
Tabel berikut menunjukkan lima term dengan bobot TF-IDF tertinggi pada korpus ulasan yang digunakan dalam penelitian.

### 3.4 Hasil Perhitungan Kemiripan Dokumen (Cosine Similarity)

Sistem menerima query dari pengguna, misalnya: “kafe nyaman wifi cepat untuk kerja” dan kemudian membandingkannya dengan setiap dokumen ulasan.

**Tabel 5 : Lima Dokumen dengan Nilai Kemiripan Tinggi**

Peringkat	ID Dokumen	Nama Kafe	Nilai Cosine Similarity
1	D-118	Tropicollo Binjai	0.812
2	D-702	Rumah Kopi Ahmad Yani	0.795
3	D-1440	Caffeine Spot	0.783
4	D-891	Goodhands Coffee	0.764
5	D-2030	Kedai Senja	0.755

**Gambar 3 : Distribusi Panjang Ulasan Pengguna Google Maps**

Gambar ini menampilkan 10 dokumen dengan nilai kemiripan tertinggi berdasarkan query, menunjukkan kemampuan VSM dalam mengukur kesesuaian antara kebutuhan pengguna dan isi ulasan.

Hasil penerapan metode VSM menunjukkan bahwa model mampu mengurutkan dokumen ulasan secara relevan berdasarkan kata kunci pengguna. Dokumen yang mengandung kata “nyaman”, “wifi”, dan “kerja” cenderung memperoleh bobot TF-IDF lebih tinggi dan menghasilkan nilai *cosine similarity* yang besar. Selain itu:

1. Ulasan yang panjang dan informatif cenderung lebih relevan, karena mengandung lebih banyak term yang sesuai dengan query.
2. Variasi panjang ulasan tidak menjadi hambatan, karena TF-IDF menormalkan frekuensi kata berdasarkan dokumen.
3. Metode VSM terbukti efektif untuk data ulasan Google Maps yang memiliki karakter *short text* namun padat informasi.



Secara keseluruhan, hasil penelitian menunjukkan bahwa VSM dapat digunakan sebagai model pencarian ulasan paling relevan pada kafe di Kota Binjai, dan dapat dikembangkan sebagai sistem rekomendasi pencarian berbasis teks.



## DAFTAR PUSTAKA

- [1] Adiyanto, A. T., & Handayani, D. H. (2022). Information Retrieval Sistem Kearsipan Pencarian Dokumen di Dinas Pemberdayaan Perempuan dan Perlindungan Anak Kota Semarang Menggunakan Metode Vector Space Model. *Jurnal Mahajana Informasi*, 7(1), 9-23.
- [2] Araujo, J., Ayu, N. M., & Tukan, J. (2023). The Visualization of the Vector Space Model in Searching for Immigration News in the East Nusa Tenggara Region. *CENTIVE Journal*, 3(1), 10-16.
- [3] Jayarna, I. G. N. A., Darma, I. G. W., & Widiana Putra, I. M. A. (2022). Study Literatur Information Retrieval Model: Teknik dan Aplikasi. *Jurnal Sutasoma*, 3(2), 75-83.
- [4] Manning, C. D., Raghavan, P., & Schütze, H. (2009). *Introduction to Information Retrieval*. Cambridge University Press.
- [5] Rahmayani, N., & Sari, F. P. (2020). Penerapan Metode Vector Space Model untuk Sistem Temu Kembali Informasi pada Dokumen Akademik. *Jurnal Teknologi dan Sistem Informasi*, 1(2), 45-52.
- [6] Utami, D., & Wibowo, A. (2021). Implementasi Cosine Similarity untuk Klasifikasi Dokumen Berita Online. *Jurnal Teknologi dan Sistem Komputer*, 9(3), 180-187.
- [7] Sugiyono. (2022). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Alfabeta.
- [8] Noviriandini, A., Ambarsari, D. A., & Aprian, F. (2020). Information Retrieval System pada Pencarian File Dokumen Berbasis Teks dengan Metode Vector Space Model.
- [9] Ressa, P., Sari, N. N. K., & Anugrah, P. B. (2022). Implementasi Content-Based Filtering menggunakan TF-IDF dan Cosine Similarity untuk Sistem Rekomendasi Resep Masakan. *Jurnal Teknologi Informasi (JTI)*, 18(1), artikel 12543
- [10] Arif Tirtana, A., Zulkarnain, A., & Listio, Y. D. (2024). Pembuatan Sistem Pencarian Pekerjaan Menggunakan TF-IDF. *Jurnal Ilmiah Teknologi Informasi Asia*, 13(2), artikel 389.
- [11] Iqbal Hanan Junaidi, Sopangi, & Sri Sumarlinda (2024). Sistem Retrieval E-Arsip Tirta Asata Menggunakan Algoritma Vector Space Model. *Jurnal Fasilkom*, 14(2).
- [12] Adi Ryansyah & Sri Andayani (2024). Implementasi Algoritma TF-IDF pada Pengukuran Kesamaan Dokumen. *JuSiTik: Jurnal Sistem dan Teknologi Informasi Komunikasi*, artikel 427.
- [13] Rangga Saputra, Jayanta, & Musthofa Galih Pradana (2024). Implementasi Cosine Similarity dan TF-IDF dalam Menentukan Rumpun Jabatan. *Krea-TIF: Jurnal Teknik Informatika*, 12(1), 1-11.
- [14] Eka Sabna (2021). Information Retrieval pada Data Judul Skripsi Berbasis Teks Menggunakan Vector Space Model. *Jurnal Ilmu Komputer*, 10(2), 132-135.



- [15] Adryanus Rinaldi, Lely Hiryanto & Viny Christanti Marwadi (2025). Penerapan TF-IDF dan Cosine Similarity pada Question-Answering System untuk Informasi PlayStation. *Jurnal Ilmu Komputer dan Sistem Informasi (JIKSI)*, 13(2).
- [16] Wahyuni (2021). Penerapan TF-IDF dan Cosine Similarity pada Klasifikasi Dokumen Skripsi. *Jurnal Teknik Elektro*, 9(1).