# BAB I

# PENDAHULUAN

## Latar Belakang

Perkembangan teknologi informatika, khususnya dalam pemrosesan bahasa alami (natural language processing/NLP), telah mengubah cara pengguna berinteraksi dengan platform digital, termasuk toko buku online. Tren adopsi buku digital meningkat pesat, dengan pasar global diproyeksikan mencapai $20,4 miliar pada 2027 (Grand View Research, 2022). Di Indonesia, Gramedians Book Store sebagai salah satu platform terkemuka menawarkan ribuan judul digital, namun volume data yang besar menyebabkan kesulitan bagi pengguna dalam menemukan buku yang relevan, sebuah masalah yang dikenal sebagai information overload. Sistem rekomendasi berbasis konten, yang memanfaatkan analisis teks untuk menemukan kesamaan antar item, menjadi solusi efektif untuk meningkatkan pengalaman pengguna.

Penelitian oleh Zhang et al. (2020) menunjukkan bahwa semantic search, didukung model embedding teks seperti transformer, dapat menangkap makna semantik secara mendalam, menghasilkan rekomendasi yang lebih akurat dibandingkan pencarian berbasis kata kunci. Namun, banyak platform buku online, termasuk di Indonesia, masih mengandalkan pencarian sederhana yang tidak memahami konteks atau preferensi pengguna, menyebabkan rekomendasi yang kurang relevan. Masalah ini diperparah oleh inefisiensi pengolahan data teks, seperti judul dan deskripsi buku, yang sering kali tidak terstruktur atau bervariasi panjangnya, menyulitkan representasi teks yang bermakna.

Urgensi pengembangan sistem rekomendasi berbasis semantic search terletak pada potensinya untuk meningkatkan kepuasan pengguna dan keterlibatan platform. Dengan memanfaatkan model text-embedding-004 dari Google GenAI untuk menghasilkan embedding teks dan cosine similarity untuk mengukur kesamaan, penelitian ini mengimplementasikan solusi informatika yang relevan untuk Gramedians Book Store. Sistem ini memungkinkan pencarian dan rekomendasi buku yang lebih intuitif melalui aplikasi web Streamlit, mendukung efisiensi operasional toko dan pengalaman pengguna yang lebih personal.

Proyek ini sejalan dengan tren teknologi informatika terkini, yakni penerapan NLP dalam sistem rekomendasi, dan berkontribusi pada kebutuhan praktis platform buku digital di Indonesia. Dengan fokus pada pengembangan aplikasi yang terintegrasi, penelitian ini menawarkan solusi teknologi yang tidak hanya mengatasi masalah information overload tetapi juga memperkuat daya saing Gramedia di pasar digital.

## Identifikasi Masalah

1. Kurangnya representasi teks bermakna dari data buku yang tidak terstruktur (judul dan deskripsi).
2. Keterbatasan menghasilkan embedding teks berkualitas tinggi untuk menangkap makna semantik.
3. Inefisiensi dalam menghitung kesamaan antar buku untuk rekomendasi yang relevan.
4. Kurangnya integrasi sistem rekomendasi ke dalam antarmuka web yang ramah pengguna.

## Batasan Masalah

1. Pendekatan terbatas pada sistem rekomendasi berbasis konten menggunakan teks judul dan deskripsi.
2. Dataset hanya mencakup data Gramedia (1.877 entri setelah preprocessing).
3. Teknologi yang digunakan adalah Python dengan Streamlit, Pandas, NumPy, dan Scikit-learn.
4. Tidak mempertimbangkan faktor seperti kategori buku, penulis, atau ulasan pengguna.
5. Pengujian fokus pada fungsionalitas sistem, tanpa evaluasi kuantitatif seperti precision atau recall.

## Rumusan Masalah

1. Bagaimana merancang pipeline preprocessing untuk menghasilkan representasi teks bermakna dari data buku?
2. Bagaimana mengembangkan metode menghasilkan embedding teks menggunakan text-embedding-004?
3. Bagaimana menganalisis kesamaan antar buku dengan cosine similarity untuk rekomendasi?
4. Bagaimana merancang aplikasi web Streamlit untuk mengintegrasikan semantic search dan rekomendasi?

## Tujuan Penelitian

1. Merancang pipeline preprocessing untuk menyiapkan data buku.
2. Mengembangkan metode generasi embedding teks dengan text-embedding-004.
3. Menganalisis kesamaan buku menggunakan cosine similarity untuk rekomendasi.
4. Merancang aplikasi web Streamlit untuk semantic search dan rekomendasi.

## Manfaat Penelitian

### Manfaat Praktis

1. Meningkatkan efisiensi pencarian dan rekomendasi buku di Gramedians Book Store.
2. Meningkatkan kepuasan pengguna melalui personalisasi berbasis semantic search.
3. Mendukung potensi peningkatan penjualan melalui rekomendasi yang relevan.

### Manfaat Teoritis

1. Memberikan kontribusi pada pengembangan sistem rekomendasi berbasis NLP.
2. Menyediakan studi kasus penerapan semantic search dalam konteks toko buku online.
3. Memperkaya literatur informatika tentang integrasi embedding teks dan antarmuka web.

Grand View Research. (2022). *E-book Market Size, Share & Trends Analysis Report*. https://www.grandviewresearch.com/industry-analysis/e-book-market

Zhang, S., Yao, L., Sun, A., & Tay, Y. (2020). Deep learning based recommender system: A survey and new perspectives. *ACM Computing Surveys*, 52(5), 1-38. DOI: 10.1145/3285029