

• • •

2025

PROJECT

RECOMMENDATION PRODUCT

Webtoon





Profile



<https://github.com/Dvagm>



www.linkedin.com/in/dava-gilang-maulana-b7a03a27b



[@dvaamlna](https://www.instagram.com/@dvaamlna)

...

PROBLEM



In the rapidly growing digital era, readers have access to thousands of manhwa titles across various platforms. However, the overwhelming number of options often leads to difficulties in finding stories that match individual preferences. A recommendation system can address this challenge, but several issues need to be overcome. Data limitations, including incomplete or inaccurate information on genres, ratings, and synopses, can hinder the accuracy of recommendations. Additionally, the cold start problem arises when new or lesser-known manhwas lack sufficient user interaction data, making it difficult to generate relevant suggestions. On the model side, inaccurate recommendations may occur due to overfitting or reliance on popularity bias, where mainstream titles dominate suggestions while niche or emerging works remain undiscovered. Furthermore, user preferences are often dynamic, changing over time, which traditional models may fail to capture. New users also face cold start issues, as the system has no historical data to base recommendations on.

...
Project



MODEL MACHINE LEARNING & DEEP LEARNING

MODEL



Models Used for Manhwa Recommendation System

For the recommendation system, a combination of traditional and advanced models was used to generate similarity scores based on manhwa synopses:

1. TF-IDF (Term Frequency-Inverse Document Frequency) — Converts text into numerical vectors by assessing word importance. Cosine similarity on the TF-IDF matrix provides content-based recommendations.
2. Sentence-BERT — Generates semantic embeddings using a pre-trained Sentence-BERT model (all-MiniLM-L6-v2). This captures contextual meaning for more accurate similarity comparisons.
3. Latent Semantic Analysis (LSA) — Reduces dimensionality of the TF-IDF matrix using Singular Value Decomposition (SVD) to uncover hidden semantic structures and compute similarity.
4. Doc2Vec — Learns vector representations of synopses using a distributed memory approach. This is effective for capturing contextual relationships between words and documents.
5. Transformer Fine-Tuning — A pre-trained BERT model (bert-base-uncased) was fine-tuned to generate high-quality embeddings. Cosine similarity on these embeddings enhances recommendation accuracy.

... →

model

EVALUATE

Evaluation Summary

- SBERT performed best with NDCG 0.4636 and Recall 0.2708, excelling in capturing semantic meaning.
- BERT showed balanced performance (NDCG 0.4187, Recall 0.2301) but lagged behind SBERT.
- TF-IDF and LSA had moderate results, with TF-IDF achieving NDCG 0.4120 and Recall 0.2059.
- Doc2Vec performed the worst (NDCG 0.3132, Recall 0.1354), struggling with contextual understanding.

Conclusion: SBERT is the most effective model for accurate manhwa recommendations.