

***TOPIC MODELLING DAN TOPIC GENERATION PADA
REVIEW GAME STEAM DENGAN FINE-TUNING BERTOPIC
DAN QWEN2.5***

Diajukan Sebagai Syarat untuk Menyelesaikan
Pendidikan Program Strata-1 pada
Jurusan Teknik Informatika Fakultas Ilmu Komputer UNSRI



Oleh :

Muhammad Rizalul Fiqri Syah Dani
NIM : 09021282126114

**Jurusan Teknik Informatika
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2025**

HALAMAN PENGESAHAN

SKRIPSI

Topic Modelling dan Topic Generation pada Review Game Steam dengan Fine-Tuning BERTopic dan Qwen2.5

Sebagai salah satu syarat untuk penyelesaian studi di
Program Studi S1 Teknik Informatika

Oleh:

MUHAMMAD RIZALUL FIQRI SYAH DANI
09021282126114

Pembimbing 1 : Hadipurnawan Satria, M.Sc., Ph.D.
NIP. 198004182020121001

Mengetahui

Ketua Jurusan Teknik Informatika



Hadipurnawan Satria, Ph.D
198004182020121001

TANDA LULUS UJIAN KOMPREHENSIF

Pada hari Jum'at tanggal 25 April 2025 telah dilaksanakan ujian komprehensif skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya

Nama : Muhammad Rizalul Fiqri Syah Dani
NIM : 09021282126114
Judul : Topic Modelling dan Topic Generation pada Review Game Steam dengan Fine-Tuning BERTopic dan Qwen2.5.

Dan dinyatakan LULUS

1. Ketua Penguji

Julian Supardi, M.T., Ph.D.

NIP. 197207102010121001



2. Penguji I

Samsuryadi, M.Kom., Ph.D.

NIP. 197102041997021003



3. Pembimbing

Hadipurnawan Satria, M.Sc., Ph.D.

NIP. 198004182020121001



Mengetahui,
Ketua Jurusan Teknik Informatika

Hadipurnawan Satria, M.Sc., Ph.D.

NIP. 198004182020121001

SURAT KETERANGAN PENGECEKAN SIMILARITY

Saya yang bertanda tangan di bawah ini

Nama : Muhammad Rizalul Fiqri Syah Dani
Nim : 09021282126114
Prodi : Teknik Informatika

Menyatakan bahwa benar hasil pengecekan similarity Skripsi/Tesis/Disertasi/Lap. Penelitian yang berjudul Topic Modelling dan Topic Generation pada Review Game Steam Dengan Fine-Tuning BERTopic dan Qwen2.5 adalah 1%.

Dicek oleh operator *: 1. Dosen Pembimbing

2. UPT Perpustakaan

Demikianlah surat keterangan ini saya buat dengan sebenarnya dan dapat saya pertanggung jawabkan.



Indralaya, 25 Maret 2025

Yang menyatakan,

Nama: Muhammad Rizalul Fiqri Syah Dani
NIM: 09021282126114

***Lingkari salah satu jawaban, tempat anda melakukan pengecekan Similarity**

MOTTO DAN PERSEMBAHAN

“Sesungguhnya Allah tidak akan mengubah keadaan diri suatu kaum sebelum mereka mengubah keadaan mereka sendiri.”

(Q.S Ar-Ra’d:11)

Karya tulis ini dipersembahkan kepada

- Allah S.W.T dan Nabi Muhammad S.A.W
- Kedua Orang tua
- Saudara kandung
- Fakultas Ilmu Komputer
- Universitas Sriwijaya

ABSTRACT

This study aims to develop and evaluate a Topic Modelling and Topic Generation system using user review data from the Steam platform. The research utilizes the BERTopic model to cluster reviews into topics in an unsupervised manner, and the Qwen2.5 Large Language Model to generate more informative topic representations. Various algorithm combinations covering embedding, dimensionality reduction, and clustering were tested to determine the optimal configuration using evaluation metrics such as C_V, U_Mass, and C_NPMI. Additionally, the quality of topics generated by Qwen2.5 was evaluated through Cosine Similarity scores. The results indicate that certain algorithm combinations produce more accurate and representative topic clusters. The developed system also features a user-friendly interface for result exploration. This research contributes to text review processing in the gaming domain and serves as a foundation for future studies in opinion analysis and natural language processing.

Keywords : *Topic Modelling, Large Language Model, BERTopic, Qwen2.5, Natural Language Processing, Topic Generation*

ABSTRAK

Penelitian ini bertujuan untuk mengembangkan dan mengevaluasi sistem Topic Modelling dan generasi topik pada data ulasan pengguna dari platform Steam. Penelitian memanfaatkan model BERTopic untuk mengelompokkan review berdasarkan topik tanpa supervisi manusia, serta model Large Language Model Qwen2.5 untuk menghasilkan representasi topik yang lebih informatif. Berbagai kombinasi algoritma seperti *embedding*, *dimensionality reduction*, dan *clustering* diuji untuk menemukan konfigurasi terbaik menggunakan metrik evaluasi C_V , U_{Mass} , dan C_{NPMI} . Selain itu, kualitas topik yang dihasilkan oleh Qwen2.5 dievaluasi menggunakan skor *Cosine Similarity*. Hasil penelitian menunjukkan bahwa kombinasi algoritma tertentu mampu memberikan hasil pengelompokan topik yang lebih akurat dan representatif. Sistem yang dikembangkan juga dilengkapi dengan antarmuka pengguna yang memudahkan eksplorasi hasil. Penelitian ini memberikan kontribusi dalam pemrosesan data teks ulasan gim serta dapat dijadikan dasar untuk penelitian lanjutan di bidang analisis opini dan Pemrosesan Bahasa Alami.

Kata Kunci : *Topic Modelling*, *Large Language Model*, *BERTopic*, *Qwen2.5*, Pemrosesan Bahasa Alami, Generasi Topik

KATA PENGANTAR

Puji syukur diaplikasikan kepada Allah S.W.T karena atas rahmat dan hidayah-Nya karya tulis yang berjudul “*Topic Modelling dan Topic Generation pada Review Game Steam dengan Fine-Tuning BERTopic dan Qwen2.5*” dapat diselesaikan sebagai tugas akhir penulis. Karya tulis ini disusun sebagai salah satu syarat menyelesaikan Program Studi Teknik Informatika Reguler di Fakultas Ilmu Komputer Universitas Sriwijaya.

Proses penulisan karya tulis ini tidak akan berjalan lancar tanpa bantuan dan dukungan dari berbagai pihak baik secara langsung ataupun tidak langsung. Oleh sebab itu, ucapan terima kasih disampaikan penulis kepada

1. Allah Subhanallahu Ta’ala atas rahmat dan hidayahnya
2. Kedua orang tua yang selalu memberi dukungan dan do’a
3. Bapak Prof. Erwin, S.Si., M.Si sebagai Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
4. Bapak Hadipurniawan Satria, M.Sc., Ph.D. sebagai Ketua Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya dan sebagai Dosen Pembimbing yang telah membimbing proses penulisan karya tulis ini dari awal sampai dengan selesai..
5. Seluruh dosen serta staff Fakultas Ilmu Komputer Universitas Sriwijaya
6. Teman terdekat yang telah memberi dukungan dan motivasi untuk menyelesaikan karya tulis ini, yang terdiri dari Deva Saumena, Kevin Putra Yudha Naserwan, serta Ahmad Naufal Muzakki.
7. Seluruh teman teman Program Studi Teknik Informatika angkatan 2021

Karya tulis ini telah diselesaikan penulis dengan kemampuan maksimal penulis, tetapi penulis sadar bahwa karya tulis ini tidak lepas dari ketidaksempurnaan, sehingga besar harapan penulis untuk menerima kritik, saran dan masukan dari pembaca demi peningkatan kualitas karya tulis kedepannya.

Penulis berharap karya tulis ini dapat memberikan manfaat terhadap bidang penelitian yang terkait, serta dapat menjadi basis untuk penelitian kedepannya.

Inderalaya, 4 Desember 2024

Penulis,

A handwritten signature in black ink, consisting of a stylized 'M' followed by a series of loops and a long horizontal stroke.

Muhammad Rizalul Fiqri Syah Dani

DAFTAR ISI

HALAMAN PENGESAHAN SKRIPSI.....	ii
TANDA LULUS UJIAN KOMPREHENSIF	iii
SURAT KETERANGAN PENGECEKAN SIMILARITY	iv
MOTTO DAN PERSEMBAHAN.....	v
ABSTRACT	vi
ABSTRAK	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	x
DAFTAR GAMBAR.....	xv
DAFTAR TABEL	xvi
BAB I PENDAHULUAN.....	I-1
1.1 Pendahuluan	I-1
1.2 Latar Belakang Masalah.....	I-1
1.3 Rumusan Masalah	I-3
1.4. Tujuan Penelitian	I-3
1.5 Manfaat Penelitian	I-4
1.6 Batasan Masalah.....	I-4
1.7 Sistematika Penulisan	I-5
1.8 Kesimpulan	I-6
BAB II KAJIAN LITERATUR	II-1

2.1 Pendahuluan	II-1
2.2 Landasan Teori.....	II-1
2.2.1 Steam.....	II-1
2.2.2 Topic Modelling.....	II-2
2.2.2.1 Word Embedding	II-3
2.2.2.2 Dimensionality Reduction.....	II-5
2.2.2.3 Clustering	II-9
2.2.2.3.1 HDBScan	II-9
2.2.2.3.2 KMeans	II-12
2.2.2.4 Tokenizing.....	II-15
2.2.2.5 Weighting Scheme	II-16
2.2.2.5.1 c-TF-IDF	II-17
2.2.2.5.2 c-TF-IDF + Normalization.....	II-17
2.2.3 Large Language Model	II-18
2.2.3.1 System Prompt	II-20
2.2.3.2 Main Prompt	II-20
2.2.4 Metrik Perhitungan Kualitas Topic Modelling	II-21
2.2.4.1 Coherence via Normalized PMI (C_NPMI)	II-21
2.2.4.2 Coherence Vector (C_V).....	II-23
2.2.4.3 Umass Coherence (U_Mass).....	II-24
2.3 Penelitian Lain yang Relevan	II-25

2.4 Kesimpulan	II-28
BAB III METODOLOGI PENELITIAN	III-1
3.1 Pendahuluan	III-1
3.2 Pengumpulan Data	III-1
3.2.1 Jenis Data	III-2
3.2.2 Sumber Data.....	III-2
3.2.3 Metode Pengumpulan Data	III-3
3.3 Tahapan penelitian	III-3
3.3.1 Kerangka kerja	III-5
3.3.2 Menentukan Kriteria Pengujian	III-9
3.3.3 Format Data Pengujian.....	III-9
3.3.4 Menentukan Alat Bantu Penelitian	III-11
3.3.5 Analisis Hasil Pengujian Penelitian	III-12
3.3.6 Melakukan Analisis dan Menarik Kesimpulan Penelitian	III-12
3.4 Manajemen Proyek Penelitian.....	III-13
3.5 Kesimpulan	III-15
BAB IV RANCANGAN PERANGKAT LUNAK.....	IV-1
4.1 Pendahuluan	IV -1
4.2 Fase Perencanaan	IV-1
4.2.1 Pemodelan Bisnis	IV-1
4.2.2 Kebutuhan	IV-2

4.2.3 Analisis Perangkat Lunak	IV-3
4.3 Fase Designing	IV-4
4.3.1 Use Case.....	IV-4
4.3.2 Sequence Diagram	IV-9
4.3.3 Activity Diagram.....	IV-9
4.3.4 Desain antarmuka pengguna	IV-11
4.4 Fase Implementasi.....	IV-13
4.4.1 Implementasi Antar Muka.....	IV-13
4.4.2 Implementasi Algoritma Topic Modelling	IV-15
4.4.3 Implementasi Large Language Model	IV-18
4.5 Fase Cutover	IV-19
4.5.1 Kebutuhan	IV-19
4.5.2 Kerangka Pengujian	IV-20
4.5.3 Hasil Pengujian	IV-21
4.6 Kesimpulan	IV-23
BAB V HASIL DAN ANALISIS PENELITIAN.....	V-1
5.1 Pendahuluan	V-1
5.2 Data Hasil Percobaan	V-1
5.2.1 Konfigurasi Percobaan.....	V-1
5.2.2 Data Hasil Konfigurasi I	V-2
5.2.3 Data Hasil Konfigurasi II.....	V-5

5.2.4 Data Hasil Konfigurasi III.....	V-7
5.2.5 Data Hasil Konfigurasi IV	V-9
5.2.6 Data Hasil Konfigurasi V.....	V-11
5.2.7 Data Hasil Konfigurasi VI	V-13
5.2.8 Data Hasil Konfigurasi VII.....	V-16
5.2.9 Data Hasil Konfigurasi VIII.....	V-18
5.3 Analisis Hasil Penelitian	V-20
5.4 Kesimpulan	V-23
BAB VI KESIMPULAN DAN SARAN.....	VI-1
6.1 Kesimpulan	VI-1
6.2 Saran.....	VI-2
DAFTAR PUSTAKA	xvii
LAMPIRAN.....	xix

DAFTAR GAMBAR

Gambar II-1. Contoh struktur word embedding dengan vektor acak.....	II-3
Gambar II-2. Contoh struktur word embedding dengan vektor yang sesuai	II-4
Gambar II-3. Graf perbandingan vektor.....	II-5
Gambar II-4. Contoh graf data vektor dua dimensi	II-8
Gambar II-5. Contoh graf data vektor satu dimensi.....	II-8
Gambar II-6. Contoh data vektor dua dimensi.....	II-10
Gambar II-7. Contoh graf data vektor yang sudah terhubung	II-11
Gambar II-8. Contoh graf data vektor yang telah dikelompokkan dengan HDBScan	II-12
Gambar II-9. pengambilan titik acak dengan nilai $k = 3$	II-13
Gambar II-10. Pengelompokkan setiap titik dan pengambilan lokasi titik tengah.....	II-13
Gambar II-11. Contoh dari pengulangan pengelompokkan sebanyak tiga kali.....	II-14
Gambar II-12. Contoh graf perbandingan nilai k dengan jumlah pengurangan varian.....	II-15
Gambar II-13. Contoh proses tokenizing pada suatu kalimat	II-16
Gambar II-14. Arsitektur Transformer.....	II-19
Gambar III-1. Tahapan penelitian	III-4
Gambar III-2. Diagram kerangka kerja penelitian	III-5
Gambar IV-1. Use Case diagram perangkat lunak.....	IV-5
Gambar IV-2. Sequence Diagram perangkat lunak	IV-9
Gambar IV-3. Activity Diagram perangkat lunak.....	IV-10
Gambar IV-4. Desain antarmuka halaman awal	IV-11
Gambar IV-5. Desain antarmuka saat menunjukkan hasil Topic Modelling dan Topic Generation	IV-11
Gambar IV-6. Desain antarmuka saat dibuka pada layar dengan ukuran kecil.....	IV-12
Gambar IV-7. Antarmuka halaman awal	IV-13
Gambar IV-8. Antarmuka saat menunjukkan hasil Topic Modelling dan Topic Generation	IV-14
Gambar IV-9. Antarmuka saat dibuka pada layar dengan ukuran kecil	IV-14
Gambar IV-10. Elbow Plot untuk embedding all-mpnet-base-v2	IV-16
Gambar IV-11. Elbow Plot untuk embedding all-MiniLM-L6-v2	IV-17

DAFTAR TABEL

Tabel III-1. Tabel Konfigurasi Kombinasi Algoritma Topic Modelling	III-10
Tabel III-2. Tabel Format Hasil Algoritma Topic Modelling.....	III-10
Tabel III-3. Tabel Format Hasil Evaluasi Algoritma Clustering	III-10
Tabel III-4. Tabel Alat Bantu yang digunakan	III-11
Tabel III-5. Tabel Manajemen Waktu Proyek Penelitian	III-13
Tabel IV-1. Kebutuhan Fungsional.....	IV-2
Tabel IV-2. Kebutuhan Non-Fungsional.....	IV-3
Tabel IV-3. Penjelasan aktor.....	IV-5
Tabel IV-4. Penjelasan use case.....	IV-5
Tabel IV-5. Skenario use case unggah file CSV	IV-6
Tabel IV-6. Skenario use case pengelompokkan review + pengambilan topik.....	IV-7
Tabel IV-7. Skenario use case melihat hasil data hasil pengelompokkan review	IV-8
Tabel IV-8. Kerangka pengujian perangkat lunak	IV-20
Tabel IV-9. Hasil pengujian perangkat lunak	IV-22
Tabel V-1. Konfigurasi Percobaan Penelitian Kombinasi Algoritma Topic Modelling	V-1
Tabel V-2. Hasil Pengujian Kualitas Topic Modelling Konfigurasi Percobaan I.....	V-3
Tabel V-3. Hasil Pengujian Kualitas Topic Generation Konfigurasi Percobaan I.....	V-3
Tabel V-4. Hasil Pengujian Kualitas Topic Modelling Konfigurasi Percobaan II	V-5
Tabel V-5. Hasil Pengujian Kualitas Topic Generation Konfigurasi Percobaan II	V-5
Tabel V-6. Hasil Pengujian Kualitas Topic Modelling Konfigurasi Percobaan III.....	V-7
Tabel V-7. Hasil Pengujian Kualitas Topic Generation Konfigurasi Percobaan III.....	V-7
Tabel V-8. Hasil Pengujian Kualitas Topic Modelling Konfigurasi Percobaan IV	V-9
Tabel V-9. Hasil Pengujian Kualitas Topic Generation Konfigurasi Percobaan IV.....	V-10
Tabel V-10. Hasil Pengujian Kualitas Topic Modelling Konfigurasi Percobaan V	V-11

Tabel V-11. Hasil Pengujian Kualitas Topic Generation	
Konfigurasi Percobaan V	V-12
Tabel V-12. Hasil Pengujian Kualitas Topic Modelling	
Konfigurasi Percobaan VI.....	V-14
Tabel V-13. Hasil Pengujian Kualitas Topic Generation	
Konfigurasi Percobaan VI.....	V-14
Tabel V-14. Hasil Pengujian Kualitas Topic Modelling	
Konfigurasi Percobaan VII	V-16
Tabel V-15. Hasil Pengujian Kualitas Topic Generation	
Konfigurasi Percobaan VII	V-16
Tabel V-16. Hasil Pengujian Kualitas Topic Modelling	
Konfigurasi Percobaan VIII	V-18
Tabel V-17. Hasil Pengujian Kualitas Topic Generation	
Konfigurasi Percobaan VIII	V-19

BAB I

PENDAHULUAN

1.1 Pendahuluan

Pada bab ini diuraikan mengenai latar belakang, perumusan masalah, tujuan dan manfaat penelitian, batasan masalah/ruang lingkup, metodologi penelitian, dan sistematika penulisan.

1.2 Latar Belakang Masalah

Di era digital ini, dimana sebuah komputer pribadi makin terjangkau, komputer tidak hanya digunakan untuk keperluan perhitungan yang kompleks, tetapi juga digunakan untuk keperluan hiburan, seperti menonton video, berbagi hal di sosial media, dan bermain permainan (gim). Karena perkembangan teknologi komputer pribadi yang sudah maju ini, banyak para pengembang perangkat lunak yang ingin memanfaatkan sektor hiburan pada komputer sebagai ladang bisnis, sehingga muncul beberapa platform untuk membantu para pengembang menyalurkan produk mereka ke pengguna.

Salah satu platform tersebut adalah *Steam*, pada platform ini pengembang dapat menyalurkan serta masukan dari pengguna tentang produk mereka, *Steam* juga menyediakan fitur pemisah review “positif” dan “negatif”. Tetapi fitur tersebut tidak dapat memberi wawasan yang cukup terhadap pengembang gim tentang apa

yang pemain rasakan saat memainkan gim tersebut, dan hal ini dapat menyebabkan proses pembaruan gim tidak sesuai dengan keluhan pemain gim tersebut. Solusi dari masalah ini adalah sebuah sistem untuk mengelompokkan review pemain sesuai topik yang dibicarakan pemain secara otomatis, hal ini dapat membantu pengembang gim untuk mengambil wawasan yang lebih banyak tentang apa yang diinginkan pemain dari gim tersebut.

Sistem pengelompokkan teks berdasarkan topik ini disebut *Topic Modelling* dan telah digunakan dalam berbagai hal. Tetapi kelemahan prosedur *Topic Modelling* pada umumnya adalah prosedur tersebut hanya melihat kata yang berkemungkinan besar menjadi topik di dalam suatu badan teks untuk mengelompokkan teks, dan tidak mempertimbangkan konteks kata yang diambil di dalam badan teks tersebut (Grootendorst, 2022).

Sehingga pada penelitian ini, akan dilakukan *Topic Modelling* dengan menggunakan pustaka *BERTopic*, dimana pustaka ini menggunakan teknik *Word Embedding*. Teknik ini merubah kata menjadi matriks yang dapat digunakan untuk proses komputasi oleh komputer, karena sifat matriks yang tidak komutatif, maka letak kata yang berbeda pada suatu kalimat akan menghasilkan hasil kalkulasi yang berbeda. Sifat matriks ini memungkinkan hasil komputasi untuk menyimpan konteks dari suatu kalimat.

Karena teknik *Word Embedding* dapat menyimpan konteks kalimat (Ghannay et al., 2016). Hal ini sesuai untuk mengatasi masalah *Topic Modelling* pada umumnya dengan mempertimbangkan konteks dari setiap kata dalam proses pengelompokkan. Sehingga kualitas dari hasil pengelompokkan dapat lebih

mempertimbangkan konteks kalimat tersebut, dan tidak hanya bergantung pada kata yang terdapat pada masing masing kalimat

1.3 Rumusan Masalah

Berdasarkan latar belakang yang telah disajikan, terdapat beberapa permasalahan yang muncul tentang sistem pengelompokan tersebut, beberapa diantaranya adalah sebagai berikut

1. Bagaimana melakukan *Topic Modelling* yang mempertimbangkan konteks kata dengan menggunakan *BERTopic*.
2. Bagaimana melakukan generasi topik dari proses pengelompokan *Topic Modelling* dengan menggunakan *Qwen2.5*.
3. Bagaimana mengevaluasi hasil pengelompokan *Topic Modelling* oleh *BERTopic*.
4. Bagaimana mengevaluasi hasil generasi topik oleh *Qwen2.5*

1.4. Tujuan Penelitian

Berdasarkan latar belakang dan tujuan masalah yang telah ditetapkan, dapat ditetapkan tujuan penelitian sebagai berikut

1. Mengetahui cara agar *Topic Modelling* mempertimbangkan konteks kata dengan menggunakan pustaka *BERTopic*.
2. Mengetahui cara melakukan generasi topik dari proses pengelompokan *Topic Modelling* dengan menggunakan *Qwen2.5*.

3. Mengetahui kinerja hasil pengelompokkan *Topic Modelling* dengan *BERTopic*.
4. Mengetahui kinerja hasil generasi topik dari hasil pengelompokkan *Topic Modelling* dengan *Qwen2.5*.

1.5 Manfaat Penelitian

1. Sebagai referensi penggunaan *BERTopic* sebagai alat untuk melakukan *Topic Modelling*.
2. Sebagai referensi penggunaan *Large Language Model Qwen2.5* sebagai alat untuk generasi topik dari suatu kelompok teks.
3. Menambah wawasan dalam kinerja algoritma yang digunakan pada proses *Topic Modelling* menggunakan *BERTopic*.
4. Menambah wawasan mengenai performa *Large Language Model Qwen2.5* dalam menghasilkan topik.

1.6 Batasan Masalah

Berdasarkan rumusan masalah yang telah disajikan, maka batasan masalah dari penelitian ini adalah sebagai berikut

1. Penelitian ini hanya dilakukan dengan dataset yang diambil dari platform *Steam*.
2. Dataset yang digunakan hanyalah review yang menggunakan bahasa inggris.

3. Perbandingan performa algoritma *Topic Modelling* dibatasi pada proses yang ditawarkan dalam pustaka *BERTopic*.
4. Implementasi *Topic Modelling* dan *Topic Generation* dilakukan menggunakan pustaka *BERTopic* serta *Large Language Model Qwen2*.

1.7 Sistematika Penulisan

Adapun sistematika penulisan laporan tugas akhir ini adalah sebagai berikut

BAB I. PENDAHULUAN

Bab ini berisi penjelasan latar belakang, rumusan masalah, batasan masalah, tujuan dan manfaat penelitian, serta sistematika penulisan penelitian yang dilakukan.

BAB II. KAJIAN LITERATUR

Bab ini berisikan penjelasan konsep dan teori yang dijadikan landasan selama proses penelitian, serta sumber terpercaya dari konsep dan teori tersebut.

BAB III. METODOLOGI PENELITIAN

Bab ini menjelaskan secara rinci setiap langkah yang akan dilakukan dari awal penelitian sampai dengan akhir penelitian.

BAB IV. PENGEMBANGAN PERANGKAT LUNAK

Bab ini menjelaskan secara detail proses yang dilakukan selama pembangunan perangkat lunak sampai dengan mendapatkan hasil dari penelitian tersebut.

BAB V. HASIL DAN ANALISIS PENELITIAN

Bab ini berisi penjelasan hasil dari penelitian yang telah dilakukan, serta kesimpulan yang dapat diambil dari hasil yang didapat tersebut.

1.8 Kesimpulan

Berdasarkan hal yang telah dilampirkan pada bab pendahuluan ini, dapat disimpulkan bahwa sistem *Topic Modelling* pada umpan balik gim merupakan sistem pengelompokkan yang efektif karena tidak perlunya campur tangan pengembang gim dalam proses pengelompokkan review tersebut, dan penelitian harus dilakukan untuk meningkatkan kualitas pengelompokkan review kedepannya.

DAFTAR PUSTAKA

- Al-Hawari, Feras, and Hala Barham. 2021. "A Survey on Text Pre-Processing & Feature Extraction Techniques in Natural Language Processing." *Journal of King Saud University - Computer and Information Sciences* 33 (6): 702–18. <https://doi.org/10.1016/j.jksuci.2019.04.001>.
- Dieng, Adji B., Francisco J. R. Ruiz, and David M. Blei. 2020. "Topic Modeling in Embedding Spaces." *Transactions of the Association for Computational Linguistics* 8 (December):439–53. https://doi.org/10.1162/tacl_a_00325.
- Ghannay, Sahar, Benoit Favre, Yannick Esteve, and Nathalie Camelin. n.d. "Word Embeddings Evaluation and Combination."
- Grootendorst, Maarten. 2022. "BERTopic: Neural Topic Modeling with a Class-Based TF-IDF Procedure." arXiv. <https://doi.org/10.48550/arXiv.2203.05794>.
- Jia, Weikuan, Meili Sun, Jian Lian, and Sujuan Hou. 2022. "Feature Dimensionality Reduction: A Review." *Complex & Intelligent Systems* 8 (3): 2663–93. <https://doi.org/10.1007/s40747-021-00637-x>.
- Kanungo, T., D.M. Mount, N.S. Netanyahu, C.D. Piatko, R. Silverman, and A.Y. Wu. 2002. "An Efficient K-Means Clustering Algorithm: Analysis and Implementation." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24 (7): 881–92. <https://doi.org/10.1109/TPAMI.2002.1017616>.

- McInnes, Leland, John Healy, and James Melville. 2020. "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction." arXiv. <https://doi.org/10.48550/arXiv.1802.03426>.
- Naveed, Humza, Asad Ullah Khan, Shi Qiu, Muhammad Saqib, Saeed Anwar, Muhammad Usman, Naveed Akhtar, Nick Barnes, and Ajmal Mian. 2024. "A Comprehensive Overview of Large Language Models." arXiv. <https://doi.org/10.48550/arXiv.2307.06435>.
- Sinaga, Kristina P., and Miin-Shen Yang. 2020. "Unsupervised K-Means Clustering Algorithm." *IEEE Access* 8:80716–27. <https://doi.org/10.1109/ACCESS.2020.2988796>.
- Stewart, Geoffrey, and Mahmood Al-Khassaweneh. 2022. "An Implementation of the HDBSCAN* Clustering Algorithm." *Applied Sciences* 12 (5): 2405. <https://doi.org/10.3390/app12052405>.
- Févotte, Cédric, and Jérôme Idier. "Algorithms for nonnegative matrix factorization with the β -divergence." *Neural computation* 23.9 (2011): 2421-2456.
- De Waal, Alta, and Etienne Barnard. "Evaluating topic models with stability." (2008).
- Röder, Michael, Andreas Both, and Alexander Hinneburg. 2015. "Exploring the Space of Topic Coherence Measures." In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, 399–408. Shanghai China: ACM. <https://doi.org/10.1145/2684822.2685324>.