

**PENGELOMPOKAN ARTIKEL ILMIAH MENGGUNAKAN  
ROBERTA DAN K-MEANS**

Diajukan Sebagai Syarat untuk Menyelesaikan  
Pendidikan Program Sastra-1 Pada  
Jurusan Teknik Informatika



Oleh:

**JIHAAN JASMINE JAHROO**

**NIM: 09021282126058**

**JURUSAN TEKNIK INFORMATIKA  
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA**

**2025**

# HALAMAN PENGESAHAN

## SKRIPSI

### PENGELOMPOKAN ARTIKEL ILMIAH MENGGUNAKAN ROBERTA DAN K-MEANS

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

Oleh:

**JIHAAN JASMINE JAHROO**  
**09021282126058**

**Pembimbing 1** : **Novi Yusliani S.Kom., M.T.**  
**NIP. 198211082012122001**

**Pembimbing 2** : **Desty Rodiah, S.Kom., M.T.**  
**NIP. 198912212020122011**

Mengetahui

**Ketua Jurusan Teknik Informatika**



**Hadipurnawan Satria, Ph.D**  
**198004182020121001**

## TANDA LULUS UJIAN KOMPREHENSIF SKRIPSI

Pada hari Jumat tanggal 28 Februari 2025 telah dilaksanakan ujian komprehensif skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya

Nama : Jihaan Jasmine Jahroo

NIM : 09021282126058

Judul : Pengelompokan Artikel Ilmiah Menggunakan RoBERTa dan K-means

dan dinyatakan LULUS.

1. Ketua Penguji



Kanda Januar Miraswan, S.Kom., M.T.

NIP. 199001092019031012

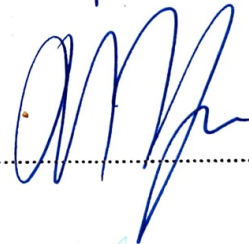
2. Penguji



Muhammad Qurhanul Rizqie, S.Kom., M.T.

NIP. 198712032022031006

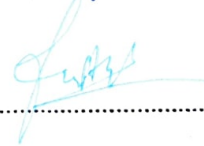
3. Pembimbing



Novi Yusliani S.Kom., M.T.

NIM. 198211082012122001

4. Pembimbing



Desty Rodiah, S.Kom., M.T.

NIP. 198912212020122011



Mengetahui,  
Ketua Jurusan Teknik Informatika

Hadipurnawan Satria, Ph.D.

NIP. 198004182020121001

## HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Jihaan Jasmine Jahroo  
NIM : 09021282126058  
Program Studi : Teknik Informatika  
Judul Skripsi : Pengelompokan Artikel Ilmiah Menggunakan RoBERTa dan K-means

Hasil pengecekan Software Turnitin : 11%

Menyatakan bahwa laporan tugas akhir saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat/Apabila ditemukan unsur penjiplakan/plagiat dalam laporan proyek ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan dari pihak mana pun.



Palembang, 28 Februari 2025

isian

Jihaan Jasmine Jahroo  
NIM. 09021282126058

## **MOTTO DAN PERSEMBAHAN**

Motto:

*“Every adversity, every failure, every heartbreak, carries with it the seed of an equal or greater benefit”*

- Napoleon Hill

Kupersembahkan Karya Tulis ini Kepada:

- Allah SWT
- Orang Tua
- Keluarga Besar
- Fakultas Ilmu Komputer
- Universitas Sriwijaya

## **ABSTRACT**

*The significant increase in the number of scientific articles published each year presents challenges in managing and utilizing information efficiently. This research aims to develop an automatic scientific article clustering system using a combination of RoBERTa method and K-Means algorithm, as well as evaluate its performance based on the Silhouette Score. The research data consists of 8,675 scientific articles collected through scraping techniques from 12 Sinta-accredited journals (ranks 1-4). The results show variations in clustering performance, with the highest Silhouette Score of 0.8834 for the Obsession Journal: Early Childhood Education Journal and the lowest of 0.5523 for the Journal of Business and Accounting. Data injection testing was conducted to evaluate the relevance between journals, where articles were classified as in-scope or out-scope based on their distance to the cluster centroid.*

**Keywords:** *Scientific Journal, Scientific Article Clustering, RoBERTa, K-Means, Silhouette Score*

## ABSTRAK

Peningkatan signifikan dalam jumlah artikel ilmiah yang diterbitkan setiap tahun menimbulkan tantangan dalam pengelolaan dan pemanfaatan informasi secara efisien. Penelitian ini bertujuan untuk mengembangkan sistem pengelompokan artikel ilmiah otomatis menggunakan kombinasi metode RoBERTa dan algoritma K-Means, serta mengevaluasi kinerjanya berdasarkan *Silhouette Score*. Data penelitian terdiri dari 8.675 artikel ilmiah yang dikumpulkan melalui teknik scraping dari 12 jurnal terakreditasi Sinta (peringkat 1-4). Hasil penelitian menunjukkan variasi dalam kinerja pengelompokan, dengan *Silhouette Score* tertinggi 0,8834 pada Jurnal Obsesi: Jurnal Pendidikan Anak Usia Dini dan terendah 0,5523 pada Jurnal Bisnis dan Akuntansi. Pengujian *inject* data dilakukan untuk mengevaluasi relevansi antar jurnal, di mana artikel diklasifikasikan sebagai *in-scope* atau *out-scope* berdasarkan jarak ke *centroid cluster*.

**Kata Kunci:** Jurnal Ilmiah, Pengelompokan Artikel Ilmiah, RoBERTa, K-Means, *Silhouette Score*

## DAFTAR ISI

LEMBAR PENGESAHAN SKRIPSI .....	ii
TANDA LULUS UJIAN KOMPREHENSIF SKRIPSI.....	iii
HALAMAN PERNYATAAN .....	iv
MOTTO DAN PERSEMBAHAN .....	v
ABSTRACT.....	vi
ABSTRAK.....	vii
DAFTAR ISI.....	viii
DAFTAR TABEL.....	xi
DAFTAR GAMBAR .....	xiii
KATA PENGANTAR .....	xv
BAB I PENDAHULUAN.....	I-1
1.1    Pendahuluan .....	I-1
1.2    Latar Belakang .....	I-1
1.3    Rumusan Masalah .....	I-2
1.4    Tujuan Penelitian .....	I-3
1.5    Manfaat Penelitian .....	I-3
1.6    Batasan Masalah.....	I-3
1.7    Sistematika Penulisan .....	I-3
1.8    Kesimpulan .....	I-4
BAB II KAJIAN LITERATUR .....	II-1
2.1    Pendahuluan .....	II-1
2.2    Landasan Teori.....	II-1
2.2.1    Text Clustering.....	II-1
2.2.2    RoBERTa .....	II-3
2.2.3    K-means .....	II-4
2.2.4 <i>Silhouette Score</i> .....	II-5
2.2.5    Threshold .....	II-8
2.2.6    Rational Unified Process.....	II-9
2.3    Penelitian Lain yang Relevan.....	II-10
2.4    Kesimpulan .....	II-12
BAB III METODE PENELITIAN.....	III-1



3.1	Pendahuluan .....	III-1
3.2	Pengumpulan Data .....	III-1
3.2.1	Jenis dan Sumber Data .....	III-1
3.2.2	Metode Pengumpulan Data .....	III-1
3.3	Tahapan Penelitian .....	III-1
3.3.1	Mengumpulkan Data .....	III-2
3.3.2	Menentukan Kerangka Kerja Penelitian .....	III-4
3.3.3	Menentukan Kriteria Pengujian .....	III-6
3.3.4	Menentukan Format Data Pengujian.....	III-6
3.3.5	Menentukan Alat Bantu Penelitian .....	III-7
3.3.6	Melakukan Pengujian Penelitian.....	III-7
3.3.7	Melakukan Analisis dan Menarik Kesimpulan Penelitian .....	III-8
3.4	Kesimpulan .....	III-8
<b>BAB IV PENGEMBANGAN PERANGKAT LUNAK .....</b>		<b>IV-1</b>
4.1	Pendahuluan .....	IV-1
4.2	Fase Insepsi .....	IV-1
4.2.1	Pemodelan Bisnis .....	IV-1
4.2.2	Kebutuhan Sistem .....	IV-2
4.2.3	Analisis dan Desain.....	IV-2
4.3	Fase Elaborasi .....	IV-15
4.3.1	Pemodelan Bisnis .....	IV-16
4.3.2	Kebutuhan Sistem .....	IV-17
4.3.3	Analisis dan Perancangan .....	IV-18
4.4	Fase Konstruksi.....	IV-21
4.4.1	Kebutuhan Sistem .....	IV-21
4.4.2	Implementasi .....	IV-22
4.5	Fase Transisi .....	IV-26
4.5.1	Pemodelan Bisnis .....	IV-26
4.5.2	Rencana Pengujian .....	IV-27
4.5.3	Implementasi .....	IV-27
4.6	Kesimpulan .....	IV-29
<b>BAB V HASIL DAN ANALISIS .....</b>		<b>V-1</b>
5.1	Pendahuluan .....	V-1
5.2	Hasil Penelitian .....	V-1

5.2.1	Konfigurasi Pengujian.....	V-1
5.3	Analisis Hasil Penelitian .....	V-16
5.4	Kesimpulan .....	V-18
BAB VI KESIMPULAN DAN SARAN .....		VI-1
6.1	Pendahuluan .....	VI-1
6.2	Kesimpulan .....	VI-1
6.3	Saran.....	VI-2
DAFTAR PUSTAKA .....		xvii
Lampiran .....		xxi

## DAFTAR TABEL

<b>Tabel III-1.</b> Tabel Contoh Data .....	III-3
<b>Tabel III-2.</b> Hasil Pengujian Inject data.....	III-6
<b>Tabel III-3.</b> Alat Bantu Penelitian .....	III-7
<b>Tabel III-4.</b> Hasil Analisa Nilai k Setiap Jurnal .....	III-8
<b>Tabel IV-1.</b> Kebutuhan Fungsional.....	IV-2
<b>Tabel IV-2.</b> Kebutuhan Non-Fungsional.....	IV-2
<b>Tabel IV-3.</b> Contoh Data Artikel Ilmiah .....	IV-3
<b>Tabel IV-4.</b> Hasil Penggabungan Artikel Ilmiah .....	IV-4
<b>Tabel IV-5.</b> Hasil Teks menjadi Huruf Kecil.....	IV-5
<b>Tabel IV-6.</b> Hasil Penghapusan.....	IV-6
<b>Tabel IV-7.</b> Hasil Analisis Model RoBERTa.....	IV-7
<b>Tabel IV-8.</b> Contoh Evaluasi Hasil Pengelompokan .....	IV-10
<b>Tabel IV-9.</b> Definisi Aktor .....	IV-11
<b>Tabel IV-10.</b> Definisi Use Case .....	IV-12
<b>Tabel IV-11.</b> Skenario Use Case Clustering Artikel Ilmiah.....	IV-12
<b>Tabel IV-12.</b> Implementasi Class Perangkat lunak.....	IV-23
<b>Tabel IV-13.</b> Rencana Pengujian Use Case Clustering Artikel Ilmiah.....	IV-27
<b>Tabel IV-14.</b> Rencana Pengujian Use Case Inject Data.....	IV-27
<b>Tabel IV-15.</b> Pengujian Use Case Clustering Artikel Ilmiah.....	IV-28
<b>Tabel IV-16.</b> Pengujian Use Case Inject Data.....	IV-28
<b>Tabel V-1.</b> Data sampel jurnal 1 di injeksi ke jurnal 2.....	V-9
<b>Tabel V-2.</b> Data sampel pada jurnal 2 di injeksi ke jurnal 4.....	V-9

<b>Tabel V-3.</b> Data sampel pada jurnal 3 yang di injeksi ke jurnal 7. ....	V-10
<b>Tabel V-4.</b> Data sampel pada jurnal 4 yang di injeksi ke jurnal 10 .....	V-11
<b>Tabel V-5.</b> Data sampel pada Jurnal 5 yang di injeksi ke jurnal 6.....	V-11
<b>Tabel V-6.</b> Data sampel pada Jurnal 6 yang di injeksi ke Jurnal 12 .....	V-12
<b>Tabel V-7.</b> Data sampel pada jurnal 7 yang di injeksi ke jurnal 1 .....	V-12
<b>Tabel V-8.</b> Data sampel Jurnal 8 yang di injeksi ke jurnal 9 .....	V-13
<b>Tabel V-9.</b> Data sampel pada jurnal 9 yang di injeksi ke Jurnal 2.....	V-14
<b>Tabel V-10.</b> Data sampel pada Jurnal 10 yang di injeksi ke jurnal 3.....	V-14
<b>Tabel V-11.</b> Data sampel pada jurnal 11 yang di injeksi ke jurnal 4 .....	V-15
<b>Tabel V-12.</b> Data sampel pada jurnal 12 yang di injeksi ke Jurnal 8.....	V-15
<b>Tabel V-13.</b> Hasil Evaluasi Silhouette Score .....	V-16

## DAFTAR GAMBAR

<b>Gambar II-1.</b> Proses sederhana clustering.....	II-1
<b>Gambar II-2.</b> Arsitektur RUP.....	II-9
<b>Gambar III-1.</b> Tahapan penelitian.....	III-2
<b>Gambar III-2.</b> Kerangka Kerja Penelitian.....	III-4
<b>Gambar IV-1.</b> Hasil Analisis Jumlah Kluster.....	IV-8
<b>Gambar IV-2.</b> Visualisasi hasil clustering dengan PCA .....	IV-9
<b>Gambar IV-3.</b> Use Case Pengelompokan Artikel .....	IV-11
<b>Gambar IV-4.</b> Rancangan Antarmuka Pengelompokan Artikel Ilmiah .....	IV-16
<b>Gambar IV-5.</b> Rancangan Antarmuka Inject Data .....	IV-17
<b>Gambar IV-6.</b> <i>Activity Diagram Clustering</i> Artikel Ilmiah .....	IV-18
<b>Gambar IV-7.</b> <i>Activity Diagram Inject Data</i> .....	IV-19
<b>Gambar IV-8.</b> <i>Sequence Diagram Clustering</i> Artikel Ilmiah.....	IV-20
<b>Gambar IV-9.</b> <i>Sequence Diagram Inject data</i> .....	IV-21
<b>Gambar V-1.</b> Hasil nilai k dari Jurnal 1.....	V-2
<b>Gambar V-2.</b> Hasil nilai k Jurnal 2.....	V-3
<b>Gambar V-3.</b> Hasil nilai k untuk jurnal 3. ....	V-3
<b>Gambar V-4.</b> Hasil nilai k jurnal 4. ....	V-4
<b>Gambar V-5.</b> Hasil nilai k untuk Jurnal 5.....	V-4
<b>Gambar V-6.</b> Hasil nilai k Jurnal 6.....	V-5
<b>Gambar V-7.</b> Hasil nilai k untuk jurnal 7. ....	V-5
<b>Gambar V-8.</b> Hasil nilai k untuk Jurnal 8.....	V-6
<b>Gambar V-9.</b> Hasil nilai k untuk Jurnal 9.....	V-6

<b>Gambar V-10.</b> Hasil nilai k untuk Jurnal 10.....	V-7
<b>Gambar V-11.</b> Hasil nilai k untuk jurnal 11. ....	V-7
<b>Gambar V-12.</b> Hasil nilai k untuk jurnal 12. ....	V-8

## **KATA PENGANTAR**

Puji syukur kepada Allah SWT atas rahmat dan nikmat Nya yang telah diberikan kepada penulis sehingga dapat menyelesaikan skripsi ini dengan baik. Skripsi ini disusun sebagai salah satu syarat menyelesaikan Pendidikan program Strata-1 di Fakultas Ilmu Komputer Universitas Sriwijaya. Dalam menyelesaikan skripsi ini, penulis menerima bantuan, bimbingan dan dukungan dari banyak pihak, baik secara langsung maupun tidak langsung. Oleh karena itu, penulis ingin menyampaikan terima kasih kepada:

1. Allah SWT atas rahmat dan nikmat-Nya penulis dapat menyelesaikan skripsi ini dengan baik.
2. Kedua orang tua dan keluarga besar yang telah mendoakan, memberi semangat, memotivasi, dan nasihat untuk menyelesaikan skripsi ini.
3. Bapak Hadipurnawan Satria, Ph.D. selaku Ketua Jurusan Teknik Informatika Universitas Sriwijaya.
4. Ibu Yunita, S.Si., M.Cs. selaku Dosen Pembimbing Akademik yang telah memberikan banyak sekali bantuan dan arahan kepada penulis selama perkuliahan.
5. Ibu Novi Yusliani, M.T. selaku Dosen Pembimbing I dan Ibu Desty Rodiah, S.Kom., M.T. selaku Dosen Pembimbing II yang telah membimbing serta memberikan arahan kepada penulis selama proses pengerjaan skripsi.
6. Seluruh dosen program studi serta admin Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

7. Teman-teman Grup “Senja 2.0” yang telah menemani dan memotivasi penulis untuk semangat mengerjakan skripsi.
8. Seluruh Staf Administrasi dan Pegawai Fakultas Ilmu Komputer yang telah membantu dalam urusan administrasi tugas akhir penulis.
9. Seluruh teman-teman yang telah memberikan saran, motivasi, dan semangat kepada penulis.
10. Pihak-pihak lain yang tidak dapat penulis sebutkan satu per satu.

Penulis menyadari bahwa dalam penyusunan skripsi ini masih banyak sekali kekurangan dikarenakan kurangnya pengalaman dan pengetahuan penulis. Oleh karena itu penulis mengharapkan saran dan kritik yang membangun guna kemajuan penelitian selanjutnya. Semoga tugas akhir ini dapat bermanfaat. Terima kasih.

Palembang, 28 Februari 2025  
Penulis



Jihaan Jasmine Jahroo



# **BAB I**

## **PENDAHULUAN**

### **1.1 Pendahuluan**

Bab ini akan menguraikan latar belakang masalah, perumusan masalah, tujuan penelitian, manfaat penelitian, serta batasan masalah. Pembahasan dalam bab ini memberikan gambaran umum mengenai keseluruhan rangkaian penelitian.

### **1.2 Latar Belakang**

Dalam era perkembangan teknologi informasi dan internet, jumlah artikel ilmiah yang diterbitkan setiap tahun mengalami peningkatan yang signifikan. Fenomena ini menimbulkan tantangan bagi peneliti, pustakawan, dan profesional lain dalam mengelola, mengakses, serta memanfaatkan informasi secara efisien. Salah satu solusi yang dapat diterapkan untuk menghadapi tantangan tersebut adalah dengan melakukan pengelompokan artikel ilmiah secara otomatis menggunakan teknik pemrosesan bahasa alami dan pembelajaran mesin (Wenando et al., 2023).

Pengelompokan artikel ilmiah bertujuan untuk mengelompokkan artikel berdasarkan kesamaan konten, sehingga artikel yang membahas topik serupa dapat dikelompokkan bersama (Nurhayati & Mayesti, 2023). Proses ini memiliki banyak manfaat, seperti pembuatan basis data pengetahuan, sistem rekomendasi, dan analisis tren penelitian. Namun, metode pengelompokan tradisional yang mengandalkan teknik statistik sederhana sering kali tidak mampu menangkap kompleksitas dan nuansa dalam teks ilmiah. Oleh karena itu, diperlukan pendekatan yang lebih canggih untuk memahami konten secara mendalam.

Salah satu algoritma yang sering digunakan dalam klusterisasi data adalah K-Means karena kesederhanaan dan efisiensinya (Nugroho & Hendrawan, 2022). Algoritma ini membagi data ke dalam beberapa kelompok berdasarkan jarak vektor, di mana data dalam satu kelompok memiliki kesamaan yang tinggi (Lase et al., 2019). Namun, penerapan K-Means pada data teks, seperti artikel ilmiah, memerlukan representasi teks dalam bentuk angka menggunakan teknik word representation. Salah satu metode representasi yang dapat digunakan adalah *Robustly Optimized BERT Pretraining Approach* (RoBERTa).

RoBERTa merupakan pengembangan dari model *Bidirectional Encoder Representations from Transformers* (BERT) yang dioptimalkan melalui pelatihan ulang menggunakan data yang lebih besar dan teknik yang lebih baik. Hal ini membuat RoBERTa mampu memahami konteks dan nuansa bahasa dengan lebih baik (Faqih et al., 2023). Penelitian Acheampong (2020) menunjukkan keunggulan RoBERTa dibandingkan model lainnya, seperti BERT, DistilBERT, dan XLNet, dalam mengidentifikasi emosi dari teks menggunakan dataset ISEAR. RoBERTa memperoleh tingkat akurasi tertinggi sebesar 0,7431 di antara model yang diuji (Acheampong et al., 2020). Dalam pengelompokan artikel ilmiah, RoBERTa dapat menghasilkan representasi vektor teks yang lebih akurat (Saragih et al., 2022). Penelitian ini bertujuan untuk mengembangkan dan mengevaluasi metode pengelompokan artikel ilmiah menggunakan RoBERTa dan K-Means.

### **1.3 Rumusan Masalah**

Rumusan masalah pada penelitian ini adalah sebagai berikut:

1. Bagaimana mengembangkan sistem pengelompokan artikel ilmiah berdasarkan kemiripan topik menggunakan metode RoBERTa dan K-Means?
2. Bagaimana kinerja metode RoBERTa dan K-means dalam melakukan pengelompokan artikel ilmiah berdasarkan *Silhouette Score*?

#### **1.4 Tujuan Penelitian**

Tujuan pada penelitian ini adalah sebagai berikut:

1. Menghasilkan sistem pengelompokan artikel ilmiah berdasarkan kemiripan topik dengan menggunakan metode RoBERTa dan K-Means.
2. Mengetahui kinerja metode RoBERTa dan K-means dalam melakukan pengelompokan artikel ilmiah berdasarkan *Silhouette Score*.

#### **1.5 Manfaat Penelitian**

Manfaat yang diperoleh dari penelitian ini sebagai berikut:

1. Sistem dapat menghasilkan distribusi topik artikel ilmiah melalui proses klusterisasi berdasarkan kesesuaian konten.
2. Hasil dari penelitian ini dapat dijadikan sebagai referensi untuk penelitian selanjutnya.

#### **1.6 Batasan Masalah**

Batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Teks yang digunakan merupakan teks berbahasa Inggris.

#### **1.7 Sistematika Penulisan**

Sistematika penulisan skripsi ini adalah sebagai berikut:

### **BAB I. PENDAHULUAN**

Bab ini membahas latar belakang, perumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan.

## **BAB II. KAJIAN LITERATUR**

Bab ini menguraikan landasan teori yang mendukung penelitian, termasuk penjelasan mengenai RoBERTa dan K-means, serta literatur terkait yang relevan dengan penelitian ini.

## **BAB III. METODOLOGI PENELITIAN**

Bab ini membahas proses pengumpulan data, metode yang digunakan dalam pengumpulan data, serta kerangka kerja penelitian yang dijelaskan secara mendalam.

## **BAB IV. METODOLOGI PENELITIAN**

Bab ini memaparkan metode pengembangan perangkat lunak yang digunakan, yaitu *Rational Unified Process* (RUP), untuk mendukung proses pengelompokan artikel ilmiah menggunakan RoBERTa dan K-means.

## **BAB V. HASIL DAN ANALISIS**

Bab ini menyajikan hasil serta analisis dari penerapan pengelompokan artikel ilmiah menggunakan RoBERTa dan K-means.

## **BAB VI. KESIMPULAN DAN SARAN**

Bab ini merangkum kesimpulan dan memberikan saran berdasarkan hasil analisis pengelompokan artikel ilmiah menggunakan RoBERTa dan K-means.

### **1.8 Kesimpulan**

Bab ini telah menguraikan latar belakang, perumusan masalah, tujuan dan manfaat penelitian, batasan masalah, serta sistematika penelitian.

## DAFTAR PUSTAKA

- Abualigah, L., Gandomi, A. H., Elaziz, M. A., Hamad, H. Al, Omari, M., Alshinwan, M., & Khasawneh, A. M. (2021). Advances in meta-heuristic optimization algorithms in big data text clustering. *Electronics (Switzerland)*, *10*(2), 1–29. <https://doi.org/10.3390/electronics10020101>
- Acheampong, F. A., Nunoo-Mensah, H., & Wenyu, C. (2020). *Comparative Analyses of BERT, RoBERTa, DistilBERT, and XLNet for Text-based Emotion Recognition*. <https://www.researchgate.net/publication/346443459>
- Ahmed, M. H., Tiun, S., Omar, N., & Sani, N. S. (2023). Short Text Clustering Algorithms, Application and Challenges: A Survey. In *Applied Sciences (Switzerland)* (Vol. 13, Issue 1). MDPI. <https://doi.org/10.3390/app13010342>
- Wenando, A. F., Fatma, Y., Ulfa, A., & Taurin, S. J. (2023). *Aplikasi dan Kerentanan Algoritma Probabilistic Neural Network (PNN): Systematic Literature Review*. *4*(2), 491–499. <https://doi.org/10.37859/coscitech.v4i2.5676>
- Ariyanto, D. (2022). Data Mining Menggunakan Algoritma K-Means untuk Klasifikasi Penyakit Infeksi Saluran Pernafasan Akut. *Jurnal Sistim Informasi Dan Teknologi*, 13–18. <https://doi.org/10.37034/jsisfotek.v4i1.117>
- Bezdan, T., Stoean, C., Naamany, A. Al, Bacanin, N., Rashid, T. A., Zivkovic, M., & Venkatachalam, K. (2021). Hybrid fruit-fly optimization algorithm with k-means for text document clustering. *Mathematics*, *9*(16), 1929.

- Delobelle, P., Winters, T., & Berendt, B. (2020). *RobBERT: a Dutch RoBERTa-based Language Model*. <http://arxiv.org/abs/2001.06286>
- Huang, Z., Low, C., Teng, M., Zhang, H., Ho, D. E., Krass, M. S., & Grabmair, M. (2021). Context-aware legal citation recommendation using deep learning. *Proceedings of the 18th International Conference on Artificial Intelligence and Law, ICAIL 2021*, 79–88. <https://doi.org/10.1145/3462757.3466066>
- Lase, Y., & Panggabean, E. (2019). *Implementasi Metode K-Means Clustering Dalam Sistem Pemilihan Jurusan Di SMK Swasta Harapan Baru*.
- Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). *RoBERTa: A Robustly Optimized BERT Pretraining Approach*. <http://arxiv.org/abs/1907.11692>
- Mo, Y., Li, S., Dong, Y., Zhu, Z., & Li, Z. (2024). Password Complexity Prediction Based on RoBERTa Algorithm. *Applied Science & Engineering Journal for Advanced Research Peer Reviewed and Refereed Journal ISSN*, 1–5. <https://doi.org/10.5281/zenodo.11180356>
- Faqih, M. I., Basir, F. A., Sallu, S., & Matematika dan Ilmu Pengetahuan Alam, F. (2023). Analisis Perbandingan Kinerja Model Arsitektur XLNet, ERNIE, dan RoBERTa Terhadap AI-Generated Text Detection. *Konferensi Nasional Ilmu Komputer*.
- Muhr, D., Affenzeller, M., & Küng, J. (2023). A Probabilistic Transformation of Distance-Based Outliers. *Machine Learning and Knowledge Extraction*, 5(3), 782–802. <https://doi.org/10.3390/make5030042>

- Mulyani, H., Setiawan, R., & Fathi, H. (2023). Optimization of K Value in Clustering Using Silhouette Score (Case Study: Mall Customers Data). *Journal of Information Technology and Its Utilization*, 6, 45–50.  
<https://doi.org/10.56873/jitu.6.2.5243>
- Nemkul, K. (2024). Use of Bidirectional Encoder Representations from Transformers (BERT) and Robustly Optimized Bert Pretraining Approach (RoBERTa) for Nepali News Classification. *Tribhuvan University Journal*, 39(1), 124–137. <https://doi.org/10.3126/tuj.v39i1.66679>
- Nurhayati, E. S., & Mayesti, N. (2023). Tinjauan Literatur Sistematis Terhadap Penerapan Sistem Klasifikasi Khusus Di Perpustakaan. *Jurnal Ilmu Informasi, Perpustakaan, Dan Kearsipan*, 25(1).  
<https://doi.org/10.7454/jipk.v25i2.1093>
- Ratiha, S., & Fiska, R. R. (2024). Sistem Perpustakaan Menggunakan Metode Rational Unified Process. *Pondasi: Journal of Applied Science Engineering*, 2(1), 21–28.
- Nugroho, R. M., & Hendrawan, I. E. (2022). Penerapan Algoritma K-Means Untuk Klasterisasi Data Obat Pada Rumah Sakit ASRI. 16(1).  
<https://journal.uniku.ac.id/index.php/ilkom>
- Saragih, L., Nababan, M., Simatupang, Y., Amalia, J., & Del, I. T. (2022). ANALISIS SELF-ATTENTION PADA BIDIRECTIONAL LSTM DENGAN FASTTEXT DALAM MENDETEKSI EMOSI BERDASARKAN TEXT (Vol. 4, Issue 2).

- Shahapure, K. R., & Nicholas, C. (2020). Cluster quality analysis using silhouette score. *2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA)*, 747–748.
- Sudarma, M., Ariyani, S., & Wicaksana, P. A. (2021). Implementation of the Rational Unified Process (RUP) Model in Design Planning of Sales Order Management System. *INTENSIF: Jurnal Ilmiah Penelitian Dan Penerapan Teknologi Sistem Informasi*, 5(2), 249–265.  
<https://doi.org/10.29407/intensif.v5i2.15543>
- Wang, Z., Cheng, J., Cui, C., & Yu, C. (2023). *Implementing BERT and fine-tuned RobertA to detect AI generated news by ChatGPT*.
- Widodo, E., & Hadikristanto, W. (2023). Pengelompokan Untuk Penjualan Obat Dengan Menggunakan Algoritma K-Means. *Bulletin of Information Technology (BIT)*, 4(2), 408–413. <https://doi.org/10.47065/bit.v3i1>
- Xu, Q., Gu, H., & Ji, S. W. (2023). Text clustering based on pre-trained models and autoencoders. *Frontiers in Computational Neuroscience*, 17.  
<https://doi.org/10.3389/fncom.2023.1334436>