

**KOMPARASI K-MEANS CLUSTERING DAN LATENT DIRICHLET
ALLOCATION PADA KASUS SHORT TEXT TOPIC MODELING**

(Studi Kasus : Data Judul Berita Pemilu 2024)

SKRIPSI

Untuk Memenuhi Sebagian Persyaratan Mencapai Derajat Sarjana S-1
Program Studi Teknik Informatika



STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

Disusun Oleh:

Enny Fitriani Nasution

19106050022

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA
YOGYAKARTA**

2023



KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA
FAKULTAS SAINS DAN TEKNOLOGI

Jl. Marsda Adisucipto Telp. (0274) 540971 Fax. (0274) 519739 Yogyakarta 55281

PENGESAHAN TUGAS AKHIR

Nomor : B-1711/Un.02/DST/PP.00.9/07/2023

Tugas Akhir dengan judul : Komparasi K-Means Clustering dan Latent Dirichlet Allocation pada Kasus Short Text Topic Modeling (Studi Kasus : Data Judul Berita Pemilu 2024)

yang dipersiapkan dan disusun oleh:

Nama : ENNY FITRIANI NASUTION
Nomor Induk Mahasiswa : 19106050022
Telah diujikan pada : Senin, 26 Juni 2023
Nilai ujian Tugas Akhir : A

dinyatakan telah diterima oleh Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta

TIM UJIAN TUGAS AKHIR



Ketua Sidang

Muhammad Didik Rohmad Wahyudi, S.T., MT.
SIGNED

Valid ID: 6467a0b3a38



Penguji I

Muhammad Mustakim, S.T. M.T.
SIGNED

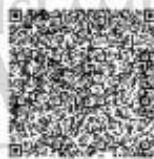
Valid ID: 6465a000f6d



Penguji II

Mandahadi Kusuma, M.Eng.
SIGNED

Valid ID: 6498f95600948



Yogyakarta, 26 Juni 2023
UIN Sunan Kalijaga
Dekan Fakultas Sains dan Teknologi

Dr. Dra. Hj. Khurul Wardati, M.Si.
SIGNED

Valid ID: 6465a000f6d

SURAT PERSETUJUAN SKRIPSI/TUGAS AKHIR

Hal : Persetujuan Skripsi

Lamp : -

Kepada:

Yth. Dekan Fakultas Sains dan Teknologi

UIN Sunan Kalijaga Yogyakarta

Di Yogyakarta

Assalamualaikum wr.wb.

Setelah membaca, meneliti, memberikan petunjuk dan mengoreksi serta mengadakan perbaikan seperlunya, maka saya selaku pembimbing berpendapat bahwa skripsi Saudara:

Nama : Enny Fitriani Nasution
NIM : 19106050022
Judul Skripsi : Komparasi K-Means Clustering dan Latent Dirichlet Allocation pada Kasus Short Text Topic Modeling

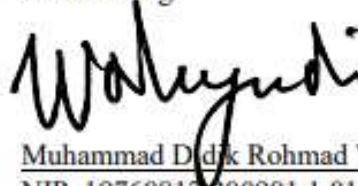
sudah dapat diajukan kembali kepada Program Studi Teknik Informatika Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta sebagai salah satu syarat untuk memperoleh gelar Sarjana Strata Satu dalam Program Studi Teknik Informatika.

Dengan ini kami mengharap agar skripsi tersebut di atas dapat segera dimunaqosyahkan. Atas perhatiannya saya ucapkan terimakasih.

Wassalamu'alaikum wr. wb.

Yogyakarta, 14 Juni 2023

Pembimbing:



Muhammad Didi Rohmad Wahyudi, S.T., MT.
NIP. 19760812 200901 1 015

SURAT PERNYATAAN KEASLIAN SKRIPSI

Yang bertanda tangan di bawah ini:

Nama : Eddy Fitriani Nasution
NIM : 19106050022
Program Studi : Teknik Informatika
Fakultas : Sains dan Teknologi

Menyatakan dengan sesungguhnya, bahwa skripsi saya yang berjudul: “: **Komparasi K-Means Clustering dan Latent Dirichlet Allocation pada Kasus Short Text Topic Modeling**” adalah hasil karya pribadi dan sepanjang pengetahuan penyusun tidak berisi materi yang dipublikasikan atau ditulis orang lain, kecuali bagian-bagian tertentu yang penyusun ambil sebagai acuan.

Apabila terbukti pernyataan ini tidak benar, maka sepenuhnya menjadi tanggungjawab penyusun.

Yogyakarta, 14 Juni 2023

menyatakan,



Eddy Fitriani Nasution
NIM. 19106050022

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

KATA PENGANTAR

Puji dan syukur saya panjatkan kepada Allah Swt. atas ridhonya saya dapat menyelesaikan penyusunan skripsi ini. Adapun judul skripsi yang saya ajukan adalah “Komparasi K-Means Clustering dan Latent Dirichlet Allocation pada Kasus Short Text Topic Modeling” dapat terselesaikan dengan baik. Tidak dapat disangkal bahwa butuh usaha yang keras dalam penyelesaian pengerjaan skripsi ini. Namun, skripsi ini tidak akan selesai tanpa orang-orang tercinta di sekeliling saya yang telah membantu dan mendukung. Terima kasih saya sampaikan kepada :

1. Bapak Prof. Dr.Phil. Al Makin, S.Ag., M.A., selaku Rektor UIN Sunan Kalijaga Yogyakarta.
2. Ibu Dr. Dra. Hj. Khurul Wardati, M.Si., selaku Dekan Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta.
3. Ibu Ir.Maria Ulfah Siregar, S.Kom., MIT., Ph.D., selaku Ketua Program Studi S1 Teknik Informatika.
4. Bapak Muhammad Didik Rohmad Wahyudi, S.T., MT., selaku Dosen Pembimbing yang telah sabar membimbing, mengarahkan, dan memberi nasihat selama penyusunan skripsi.
5. Kedua Orang tua saya Bapak Saripuddin Nasution, Ibu Lasmarolan dan keluarga Bang Risky, Kak Riska, Suci yang telah memberikan doa dan dukungan selama proses pembuatan skripsi
6. Bapak-Ibu Dosen Teknik Informatika UIN Sunan Kalijaga yang telah bersedia untuk membantu dalam memvalidasi data.
7. Seluruh staf dan karyawan Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta.
8. Staf dan karyawan Perpustakaan UIN Sunan Kalijaga Yogyakarta yang telah membantu dalam pencarian referensi penulisan skripsi.
9. Teman-teman seperjuangan Teknik Informatika 2019 UIN Sunan Kalijaga Yogyakarta
10. Teman-teman yang selalu peduli, saling mendukung selama proses penulisan skripsi Hana, Laras, Luthfia, Nazlah, Rahmi.
11. Teman diskusi dan seperjuangan skripsi Rima, Zulaicha, Nazlah

12. Teman-teman KKN 603 Temanggung Gunung Kidul

13. Semua pihak yang tidak dapat disebutkan satu persatu yang telah membantu memberi dukungan

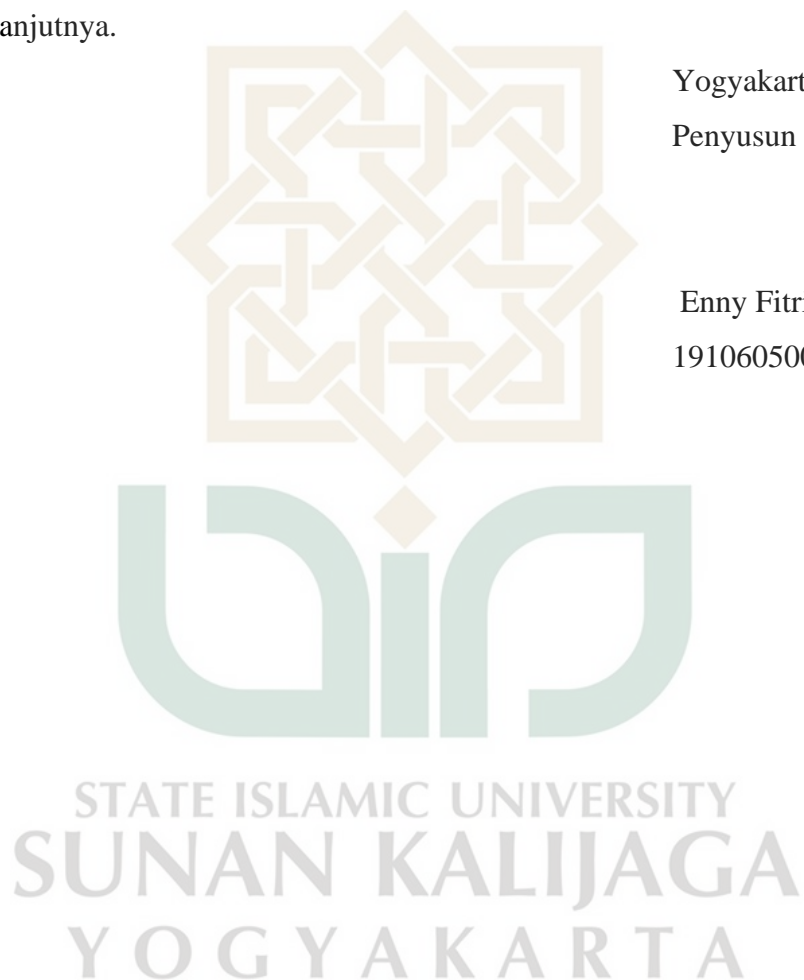
Penulis menyadari masih banyak kekurangan dalam skripsi ini, sehingga kritik dan saran yang membangun dari pembaca sangat penulis harapkan. Semoga skripsi ini memberikan manfaat untuk mendorong penelitian-penelitian selanjutnya.

Yogyakarta, Juni 2023

Penyusun

Enny Fitriani Nasution

19106050022



HALAMAN PERSEMBAHAN

Skripsi ini penulis persembahkan untuk almamater, kedua orang tua, keluarga, dan teman-teman seperjuangan yang selalu memberikan semangat support dan semangat kepada penulis serta semua pihak yang turut andil dalam proses penulisan ini yang tidak dapat disebutkan satu persatu.



HALAMAN MOTTO

“The important thing about a problem is not its solution, but the strength we gain
in finding the solution.”

-Seneca

Never say never because limits, like fears, are often just an illusion



DAFTAR ISI

PENGESAHAN TUGAS AKHIR.....	i
SURAT PERSETUJUAN TUGAS AKHIR.....	ii
SURAT PERNYATAAN KEASLIAN SKRIPSI.....	iii
KATA PENGANTAR.....	iv
HALAMAN PERSEMBAHAN.....	vi
HALAMAN MOTTO.....	vii
DAFTAR ISI.....	viii
DAFTAR TABEL.....	x
DAFTAR GAMBAR.....	xi
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah.....	5
1.3. Batasan Masalah.....	5
1.4. Tujuan Penelitian.....	6
1.5. Manfaat Penelitian.....	6
1.6. Keaslian Penelitian.....	6
1.7. Sistematika Penulisan.....	7
BAB II TINJAUAN PUSTAKA DAN LANDASAN TEORI.....	8
2.1. Tinjauan Pustaka.....	8
2.2. Landasan Teori.....	15
BAB III METODE PENELITIAN.....	26
3.1. Metode Penelitian.....	26
3.2. Obyek Penelitian.....	26

3.3. Perangkat Penelitian	26
3.4. Tahapan Penelitian	27
BAB IV HASIL DAN PEMBAHASAN	30
4.1. Pengumpulan Data	30
4.2. Preprocessing Data	35
4.3. Data Cleaning	42
4.4. Visualisasi Data	43
4.5. Pemodelan Topik Menggunakan K-Means Clustering	44
4.6. Pemodelan Topik Menggunakan LDA	53
4.7. Pelabelan Topik	60
4.8. Evaluasi Model	64
BAB V KESIMPULAN DAN SARAN	69
5.1 Kesimpulan	69
5.2 Saran	70

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

DAFTAR TABEL

Tabel 2.1 Tinjauan Pustaka	11
Tabel 2.2 Confusion Matrix	25
Tabel 4.1 Data Hasil Scraping	34
Tabel 4.2 Hasil Case Folding	36
Tabel 4.3 Punctitional Removal	37
Tabel 4.4 Remove Stopword	38
Tabel 4.5 Remove Number and Whitespace	40
Tabel 4.6 Stemming	41
Tabel 4.7 Data Cleaning	43
Tabel 4.8 Silhouette Score	45
Tabel 4.9 Cluster 0 K-Means Clustering	47
Tabel 4.10 Cluster 1 K-Means Clustering	48
Tabel 4.11 Cluster 2 K-Means Clustering	48
Tabel 4.12 Cluster 3 K-Means Clustering	49
Tabel 4.13 Cluster 4 K-Means Clustering	50
Tabel 4.14 Cluster 5 K-Means Clustering	51
Tabel 4.15 Cluster 6 K-Means Clustering	52
Tabel 4.16 Jumlah Berita Setiap Klaster	53
Tabel 4.17 Topik 0 LDA	54
Tabel 4.18 Topik 1 LDA	55
Tabel 4.19 Topik 2 LDA	56
Tabel 4.20 Topik 3 LDA	57
Tabel 4.21 Topik 4 LDA	58
Tabel 4.22 Coherence Score	60
Tabel 4.23 Nama Topik K-Means	61
Tabel 4.24 Pelabelan Topik K-Means.....	62
Tabel 4.25 Nama Topik LDA	63
Tabel 4.26 Pelabelan Topik LDA	63

DAFTAR GAMBAR

Gambar 2.1 Pendekatan TSTTM	19
Gambar 2.2 Latent Dirichlet Allocation	21
Gambar 3.1 Tahapan Penelitian	27
Gambar 4.1 Wordcloud Data	44
Gambar 4.2 Silhouette Score	45
Gambar 4.3 Coherence Score	59
Gambar 4.4 Confussion Matrix K-Means	66
Gambar 4.5 Confussion Matrix LDA	67

KOMPARASI K-MEANS CLUSTERING DAN LATENT DIRICHLET ALLOCATION PADA KASUS SHORT TEXT TOPIC MODELING

(Studi Kasus : Data Judul Berita Pemilu 2024)

Enny Fitriani Nasution

19106050022

INTISARI

Pemilu serentak 2024 adalah perhelatan demokrasi yang penting bagi bangsa Indonesia. Peran media dalam menyediakan informasi terkait pemilu menjadi sangat krusial, karena media menjadi sumber informasi tercepat dan terkini. Topic Modelling merupakan metode yang digunakan untuk mengidentifikasi dan menganalisis pola topik yang tersembunyi dalam koleksi data teks. Maka dari itu penelitian ini bertujuan untuk melakukan *topic modelling* terhadap judul berita terkait Pemilu 2024 menggunakan metode K-Means Clustering dan Latent Dirichlet Allocation pada portal berita online.

Penelitian ini bertujuan untuk membandingkan kedua model K-Means Clustering dan Latent Dirichlet Allocation. Data yang digunakan dalam penelitian ini adalah judul-judul berita yang berkaitan dengan Pemilu 2024, yang diperoleh melalui metode scraping. Analisa data dilakukan dengan menggunakan *confussion matrix* dan *Davies-Bouldin Index*.

Pada penelitian ini diperoleh pemodelan Latent Dirichlet Allocation (LDA) memiliki accuracy 83,62%, nilai percession 86,29% dan recall 83,62%. Sedangkan K-Means Clustering memiliki accuracy 80,03%, nilai percession 82,08% dan recall 80,00%. DBI dari masing-masing model yaitu 0.4979 untuk K-Means Clustering dan 0.4085 untuk Latent Dirichlet Allocation (LDA).

Kata Kunci : *Topic modelling, K-Means, LDA, pemilu 2024, berita*

COMPARISON OF K-MEANS CLUSTERING AND LATENT DIRICHLET ALLOCATION IN SHORT TEXT TOPIC MODELING

(Case Study: 2024 Election News Titles)

Eddy Fitriani Nasution

19106050022

ABSTRACT

The 2024 Simultaneous Elections are an important democratic event for the nation of Indonesia. The role of the media as a source of information related to the elections is crucial in providing the fastest and most up-to-date information. Topic modeling is a method used to identify and analyze hidden patterns, themes, or topics within a collection of textual data. Therefore, this research aims to conduct topic modeling on news titles related to the 2024 elections using the K-Means Clustering and Latent Dirichlet Allocation (LDA) methods on online news portals.

The objective of this research is to compare the two models and determine which of the two models performs better. The data used consists of news titles about the 2024 elections obtained through web scraping. Data analysis is performed using a confusion matrix and the Davies-Bouldin Index.

In this research, the Latent Dirichlet Allocation (LDA) modeling was found to have an accuracy of 83.62%, precision of 86.29%, and recall of 83.62%. Meanwhile, K-Means Clustering has an accuracy 80.03%, precision 82.08%, and recall 80.00%. The DBI (Davies-Bouldin Index) for each model is 0.4979 for K-Means Clustering and 0.4085 for Latent Dirichlet Allocation (LDA).

Keyword: *Topic modelling, K-Means, LDA, 2024 election, news*

BAB I

PENDAHULUAN

1.1. Latar Belakang

Berdasarkan jadwal penyelenggaraan pemilihan umum Tahun 2024 yang dikeluarkan Komisi Pemilihan Umum (KPU) pada Peraturan KPU Nomor 3 Tahun 2022 Indonesia akan menyelenggarakan pesta demokrasi akbar pada 14 Februari 2024 melalui pemilihan umum serentak. Pilpres 2024 menjadi bagian dari Pemilihan umum (Pemilu) serentak kedua di Indonesia setelah Pemilu 2019. Selain memilih Presiden dan Wakil Presiden, Pemilu 2024 juga menjadi momen bagi rakyat Indonesia untuk memilih anggota DPR, DPD, DPRD Provinsi, dan DPRD Kabupaten/Kota.

Menurut DP4 (Data Penduduk Potensial Pemilih Pemilu), proporsi pemilih yang berusia antara 17-39 tahun pada Pemilu 2024 mencapai sekitar 55-60 persen. Berdasarkan data tersebut, maka Pemilu 2024 akan didominasi oleh kelompok usia muda. Data ini kemudian bisa dijadikan acuan untuk partai politik memperoleh dukungan dari kelompok yang memiliki hak suara terbanyak.

Penelitian (Polii, Pati, dan Potabuga 2020) menyatakan bahwa partisipasi politik kaum milenial sebagian besar dipengaruhi media sosial yang sudah menjadi gaya hidup kaum milenial sekarang. Selain digunakan sebagai media berinteraksi antar teman saat ini, media sosial juga telah menjadi alat komunikasi politik yang digunakan oleh elite terutama partai politik karena biayanya yang rendah dan aksesnya yang tidak terbatas.

Pemanfaatan media sosial sebagai sarana dialog politik telah dilakukan dalam berbagai kesempatan.

Penelitian tersebut kemudian sejalan dengan penelitian (Abadi dan Putri 2016) yang menunjukkan bahwa faktor yang memiliki pengaruh signifikan terhadap perilaku pemilih adalah media yang menyediakan informasi mengenai kandidat. Informasi tersebut biasanya dikemas dalam berbagai media, salah satunya yaitu media massa yang menyajikan berita secara aktual.

Menurut (Hamad 2004) ada dua alasan yang berkaitan erat mengapa politik saat ini sangat dipengaruhi oleh media massa. Pertama, interaksi antara elit politik dan masyarakat saat ini banyak dilakukan melalui media massa sebagai perantaranya. Kedua, tindakan dan pernyataan para politisi selalu dianggap berita penting.

Interaksi tersebut kemudian semakin dipermudah dengan adanya media massa online yang menyajikan berita-berita aktual yang bisa diakses semua kalangan. Ketersediaan dan kemudahan dalam mengakses media massa secara online membuat media ini menjadi pilihan yang sangat diminati masyarakat (Nur 2021). Kondisi tersebut mengakibatkan media massa online memiliki jumlah data yang banyak dan melimpah, sehingga menjadi salah satu sumber data yang berguna bagi penelitian yang bertujuan untuk mengungkap topik atau tema yang sedang dibicarakan oleh media massa online terkait Pemilu 2024.

Dalam Islam, terdapat ayat Al-Qur'an yang menjelaskan bagaimana seharusnya seorang muslim bersikap jika menerima informasi yang belum jelas kebenarannya. Penjelasan tersebut terdapat pada Q.S. Al-Hujarat:6 yaitu sebagai berikut:

يَا أَيُّهَا الَّذِينَ آمَنُوا إِنْ جَاءَكُمْ فَاسِقٌ بِنَبَأٍ فَتَبَيَّنُوا أَنْ تُصِيبُوا قَوْمًا بِجَهَالَةٍ فَتُصْحَبُوا عَلَىٰ مَا
فَعَلْتُمْ نَادِمِينَ

Artinya: Wahai orang-orang yang beriman! Jika seseorang yang fasik datang kepadamu membawa suatu berita, maka telitilah kebenarannya, agar kamu tidak mencelakakan suatu kaum karena kebodohan (kecerobohan), yang akhirnya kamu menyesali perbuatanmu itu.

Dalam ayat ini, ditekankan pentingnya bagi umat muslim untuk berhati-hati dalam menerima berita terutama berita yang sumbernya tidak dapat dipercaya. Hal ini dilakukan untuk mencegah terjadinya penyesalan akibat tindakan yang diambil sebagai akibat dari berita yang belum diverifikasi kebenarannya.

Headline atau tajuk adalah kalimat pendek atau frasa yang ditempatkan secara mencolok pada sebuah berita, yang menggunakan huruf yang menonjol (Badan Pengembangan Bahasa dan Perbukuan, 2016). Judul berita biasanya berisi ringkasan informasi yang akan disampaikan dalam berita tersebut. Oleh karena itu untuk mendapatkan pesan pertama yang disampaikan dalam sebuah berita pembaca biasanya membaca judul berita.

Topic Modelling adalah salah satu teknik di dalam Natural Language Processing (NLP) untuk menganalisis teks (Aggarwal 2015). Algoritma *topic modelling* digunakan untuk mengidentifikasi pola tersembunyi dalam rangkaian kata yang terdistribusi di dalam sebuah kumpulan dokumen. Hasil dari teknik ini adalah sekumpulan topik yang terdiri dari beberapa kelompok kata yang muncul bersamaan dalam dokumen berdasarkan pola tertentu (Jacobi, Van Atteveldt, dan Welbers 2016).

Menurut (Murshed dkk. 2022) tujuan utama dari *Topic Modelling* adalah mengidentifikasi topik yang terkandung dalam dokumen teks yang panjang. Namun, beberapa studi menerapkan model tersebut kedalam *Short Text Topic Modeling* (STTM) yang berfokus pada teks yang lebih pendek. Dalam penelitian tersebut dijabarkan beberapa pendekatan dalam melakukan pemodelan topik diantaranya yang akan digunakan dalam penelitian ini adalah pendekatan probabilistik menggunakan model Latent Dirichlet Allocation (LDA) dan pendekatan dasar *machine learning* menggunakan K-Means Clustering.

LDA adalah algoritma probabilistik yang mengasumsikan bahwa setiap dokumen adalah campuran dari beberapa topik laten, dan setiap kata dalam dokumen berhubungan dengan salah satu topik laten tersebut. LDA dapat menangkap hubungan semantik antara kata-kata dan topik-topik, serta menghasilkan topik-topik yang lebih koheren dan mudah dimengerti (Jelodar dkk. 2019). Sedangkan K-Means Clustering merupakan metode pengelompokan data yang sering digunakan dalam analisis teks. Algoritma

ini memisahkan data menjadi k kluster dengan mengurangi jarak antara setiap titik data dengan pusat klasternya. Algoritma ini populer karena kemampuannya mengelompokkan data dengan efisien (Muningsih, Maryani, dan Handayani 2021).

Latent Dirichlet Allocation (LDA) dan *K-Means Clustering* merupakan algoritma populer yang banyak digunakan, oleh karena itu penulis merasa tertarik untuk membandingkan akurasi kedua algoritma tersebut dalam pemodelan topik judul berita tentang Pemilu 2024.

1.2. Rumusan Masalah

Berdasarkan latar belakang di atas, dapat ditentukan rumusan masalah yang dikaji adalah :

1. Bagaimana melakukan pemodelan topik menggunakan K-Means Clustering dan LDA pada studi kasus judul berita Pemilu 2024?
2. Bagaimana perbandingan kinerja metode *K-Means Clustering* dan *Latent Dirichlet Allocation (LDA)* dalam melakukan pemodelan topik berita Pemilu 2024 menggunakan *confussion matrix* dan *Davies-Bouldin Index*?

1.3. Batasan Masalah

1. Data yang akan digunakan merupakan data judul berita online dari portal berita antaranews, kompas, detik, viva, dan idntimes
2. Data yang dikumpulkan adalah data dalam Bahasa Indonesia
3. Data yang digunakan merupakan data judul berita dari tanggal 20 Oktober 2022 hingga 20 Februari 2023

4. Metode yang digunakan yaitu Latent Dirichlet Allocation (LDA) dan K-Means Clustering
5. Bahasa pemrograman yang digunakan dalam penelitian ini adalah Python

1.4. Tujuan Penelitian

Tujuan penelitian ini adalah menerapkan dan membandingkan kinerja model yang dihasilkan metode K-Means Clustering dan Latent Dirichlet Allocation (LDA) dalam melakukan pemodelan topik judul berita Pemilu 2024.

1.5. Manfaat Penelitian

1. Mengetahui Algoritma terbaik diantara LDA dan K-Means Clustering dalam memodelkan topik berita di berbagai portal berita online tentang Pemilu 2024
2. Penelitian ini dapat dijadikan studi literatur untuk penelitian selanjutnya.

1.6. Keaslian Penelitian

Berdasarkan studi pustaka yang dilakukan, Penelitian terkait pemodelan topik pada judul berita sudah pernah dilakukan. Namun penelitian mengenai studi komparasi Latent Dirichlet Allocation (LDA) dan K-Means Clustering pada teks berbahasa Indonesia belum pernah dilakukan sebelumnya.

1.7. Sistematika Penulisan

Sistematika penulisan yang digunakan dalam tugas akhir ini terbagi dalam beberapa pokok bahasan, yaitu:

1. BAB I PENDAHULUAN

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

2. BAB II TINJAUAN PUSTAKA DAN LANDASAN TEORI

Bab ini berisi tinjauan pustaka dari penelitian yang telah dilakukan sebelumnya serta landasan teori yang berhubungan dengan topik penelitian ini.

3. BAB III METODE PENELITIAN

Bab ini membahas tentang uraian penelitian. Pada bab ini dijelaskan dengan detail langkah-langkah yang harus dilalui untuk mencapai tujuan penelitian dan mendapatkan kesimpulan di akhir penelitian.

4. BAB IV HASIL DAN PEMBAHASAN

Bab ini menjelaskan proses yang dilakukan dalam penelitian mulai dari pengumpulan data hingga hasil yang didapatkan dari penelitian.

5. BAB V KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan yang didapatkan dari penelitian tugas akhir dan saran untuk penelitian selanjutnya.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan penelitian yang telah dilakukan dapat disimpulkan bahwa K-Means Clustering dan Latent Dirichlet Allocation (LDA) dapat melakukan pemodelan topik terhadap 9106 judul berita tentang Pemilu 2024 dengan baik.

Berdasarkan hasil percobaan menggunakan Silhouette Score untuk menentukan jumlah topik terbaik pada K-Means Clustering, ditemukan bahwa nilai Silhouette Score untuk jumlah topik $k=7$ adalah 0.03722, sedangkan untuk jumlah topik $k=5$ adalah 0.03575. Berdasarkan hasil tersebut diperoleh nilai k terbaik adalah 7. Hasil pemodelan topik menggunakan K-Means Clustering diantaranya yaitu, topik pertama secara garis besar membahas sistem proporsional pemilu, topik kedua membahas tentang capres, topik ketiga membahas koalisi partai Pemilu 2024, topik keempat tentang Partai, topik kelima membahas tentang KPU dan Bawaslu, topik keenam secara garis besar membahas tentang Golkar, dan topik ketujuh membahas Anies.

Dalam pemodelan LDA, jumlah topik yang optimal ditentukan dengan menggunakan perhitungan Coherence Score. Berdasarkan hasil perhitungan Coherence Score, diperoleh nilai 0.75958 untuk $t=5$ dan 0.74650 untuk $t=7$. Hasil tersebut menunjukkan jumlah topik terbaik adalah $t=5$. Kelima topik yang dihasilkan diantaranya yaitu topik pertama

membahas capres, topik kedua membahas sistem pemilu, topik ketiga membahas KPU, topik keempat membahas politik, topik kelima membahas tentang partai.

Berdasarkan hasil pengujian model didapatkan nilai accuracy K-Means Clustering sebesar 80,03%, nilai percession 82,08% dan recall 80,00%. Sedangkan akurasi dari model Latent Dirichlet Allocation (LDA) yaitu sebesar 83,62%, nilai percession 86,29% dan recall 83,62%. Nilai DBI dari keduanya masing masing 0.4979 untuk K-Means Clustering dan 0.4085 untuk Latent Dirichlet Allocation (LDA).

5.2 Saran

Penelitian ini masih terdapat kekurangan. Maka dari itu penulis menyarankan beberapa hal untuk penelitian selanjutnya, diantaranya:

1. Penelitian selanjutnya diharapkan melakukan pengembangan model khususnya pada model K-Means Clustering yang masih sederhana pada penelitian ini.
2. Penelitian selanjutnya diharapkan menggunakan data dengan jumlah yang lebih banyak sehingga dapat meningkatkan kualitas model yang digunakan.
3. Penelitian analisis *topic modeling* judul berita selanjutnya sebaiknya tidak hanya menggunakan bagian judul saja tetapi dapat menggunakan keseluruhan isi berita agar informasinya yang diperoleh lebih akurat.

DAFTAR PUSTAKA

- Abadi, Totok Wahyu, dan Ridlaty Ayu Oktaviana Putri. 2016. "Media dan Perilaku Pemilih Pemula pada Pilihan Presiden Tahun 2014 di Kabupaten Sidoarjo." *KANAL: Jurnal Ilmu Komunikasi* 5(1):87. doi: 10.21070/kanal.v5i1.1463.
- Aggarwal, Charu C., dan Charu C. Aggarwal. 2015. *Mining text data*. Springer.
- Altarawneh, Rasha. 2017. "Spelling Dtection Errors Techniques in NLP: A Survey."
- Berry, Michael W., dan Jacob Kogan. 2010. *Text mining: applications and theory*. John Wiley & Sons.
- Chalida, Muftia, dan Wahyudi R. Didik. 2019. *Analisis sentimen ujaran kebencian pemilihan presiden 2019 menggunakan algoritma Naïve Bayes*.
- Chen, Ying, Bo Xu, Hongwei Hao, Shiyu Zhou, dan Jie Cao. 2013. "User-defined hot topic detection in microblogging." dalam *International Conference on Internet Multimedia Computing and Service*.
- Deviacita, Dhita, Ayani #1, Helen Sasty, Pratiwi #2, Hafiz Muhardi, Jl Profesor, Dokter H. Hadari Nawawi, Bansir Laut, Pontianak Tenggara, Kota Pontianak, dan Kalimantan Barat. 2019. "Implementasi Web Scraping untuk Pengambilan Data pada Situs Marketplace." 7(4).
- Eka Rosyadi, Eka, Fikri Amrullah, Ronald David, dan Rizal Rahman. 2020. "Rancang Bangun Chatbot Informasi Lowongan Pekerjaan Berbasis Whatsapp dengan Metode NLP (Natural Language Processing)."
- Hamad, Ibnu. 2004. *KONSTRUKSI REALITAS POLITIK DALAM MEDIA MASSA (Studi Pesan Politik Dalam Media Cetak Pada Masa Pemilu 1999)*. Vol. 8.
- Hassani, Hossein, Christina Beneki, Stephan Unger, Maedeh Taj Mazinani, dan Mohammad Reza Yeganegi. 2020. "Text mining in big data analytics." *Big Data and Cognitive Computing* 4(1):1–34. doi: 10.3390/bdcc4010001.

- Jacobi, Carina, Wouter Van Atteveldt, dan Kasper Welbers. 2016. "Quantitative analysis of large amounts of journalistic texts using topic modelling." *Digital Journalism* 4(1):89–106. doi: 10.1080/21670811.2015.1093271.
- Jain, Aditya, Gandhar Kulkarni, dan Vraj Shah. 2018. "Natural Language Processing." *International Journal of Computer Sciences and Engineering* 6(1):161–67. doi: 10.26438/ijcse/v6i1.161167.
- Jelodar, Hamed, Yongli Wang, Chi Yuan, Xia Feng, Xiahui Jiang, Yanchao Li, dan Liang Zhao. 2019. "Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey." *Multimedia Tools and Applications* 78(11):15169–211. doi: 10.1007/s11042-018-6894-4.
- Kapil, Shruti, dan Meenu Chawla. t.t. *Performance Evaluation of K-means Clustering Algorithm with Various Distance Metrics*.
- Liu, Gongshen, Kui Meng, dan Jing Xie. 2013. "An Improved Topic Detection Method for Chinese Microblog Based On Incremental Clustering." *J. Softw.* 8(9):2313–20.
- Mandira, Soni, Herry Sujaini, dan Arif Bijaksana Putra Negara. 2016. "Perbaikan Probabilitas Lexical Model untuk Meningkatkan Akurasi Mesin Penerjemah Statistik."
- Mardalius. 2018. "PEMANFAATAN RAPID MINER STUDIO 8.2 UNTUK PENGELOMPOKAN DATA PENJUALAN AKSESORIS MENGGUNAKAN ALGORITMA K-MEANS Mardalius." *IV(2):123–32*.
- Muningsih, Elly, Ina Maryani, dan Vembria Rose Handayani. 2021. "Penerapan Metode K-Means dan Optimasi Jumlah Cluster dengan Index Davies Bouldin untuk Clustering Propinsi Berdasarkan Potensi Desa." *Jurnal Sains dan Manajemen* 9(1).
- Murshed, Belal Abdullah Hezam, Suresha Mallappa, Jemal Abawajy, Mufeed Ahmed Naji Saif, Hasib Daowd Esmail Al-ariki, dan Hudhaifa Mohammed Abdulwahab. 2022. "Short text topic modelling

- approaches in the context of big data: taxonomy, survey, and analysis.” *Artificial Intelligence Review*. doi: 10.1007/s10462-022-10254-w.
- Nikolenko, Sergey I., Sergei Koltcov, dan Olessia Koltsova. 2017. “Topic modelling for qualitative studies.” *Journal of Information Science* 43(1):88–102. doi: 10.1177/0165551515617393.
- Nur, Emilsyah. 2021. “PERAN MEDIA MASSA DALAM MENGHADAPI SERBUAN MEDIA ONLINE THE ROLE OF MASS MEDIA IN FACING ONLINE MEDIA ATTACKS.”
- Nur’aini, Khumaisa, Ibtisami Najahaty, Lina Hidayati, Hendri Murfi, dan Siti Nurrohma. 2015. *Combination of Singular Value Decomposition and K-means Clustering Methods for Topic Detection on Twitter*.
- O’Callaghan, Derek, Derek Greene, Joe Carthy, dan Pádraig Cunningham. 2015. “An analysis of the coherence of descriptors in topic modeling.” *Expert Systems with Applications* 42(13):5645–57. doi: 10.1016/j.eswa.2015.02.055.
- Polii, Ezra Zefanya Figo, Agustinus B. Pati, dan Jamin Potabuga. 2020. “Pengaruh Media Sosial Terhadap Partisipasi Politik Kaum Milenial Dalam Pemilihan Umum di Kecamatan Tareran Kabupaten Minahasa Selatan Tahun 2019.” *Politico: Jurnal Ilmu Politik* 9(3).
- Sahria, Yoga, dan Dhomas Hatta Fudholi. 2017. “Analisis Topik Penelitian Kesehatan di Indonesia Menggunakan Metode Topic Modeling LDA (Latent Dirichlet Allocation).” *masa berlaku mulai* 1(3):336–44.
- Suyanto, Machine Learning. 2018. “Tingkat Dasar dan Lanjut.” *Informatika Bandung*.
- Voytovich, Leah, dan Clayton Greenberg. 2022. “Natural language processing: practical applications in medicine and investigation of contextual autocomplete.” Hlm. 207–14 dalam *Machine Learning in Clinical Neuroscience: Foundations and Applications*. Springer.