

TUGAS AKHIR

**ANALISIS ENGAGEMENT PENGUNJUNG SOSIAL MEDIA
UNIVERSITAS MULTI DATA PALEMBANG BERDASARKAN
TOPIK DENGAN MENGGUNAKAN LATENT DIRICHLET
ALGORITHM (LDA)**



Oleh :

Nathaniel Jerin Sisanto 1923250007

**PROGRAM STUDI INFORMATIKA (S1)
FAKULTAS ILMU KOMPUTER DAN REKAYASA
UNIVERSITAS MULTI DATA PALEMBANG
PALEMBANG
2023**

**Fakultas Ilmu Komputer dan Rekayasa
Universitas Multi Data Palembang**

Program Studi Informatika
Tugas Akhir Sarjana Komputer
Semester Genap Tahun 2022/2023

**ANALISIS ENGAGEMENT PENGUNJUNG SOSIAL MEDIA
UNIVERSITAS MULTI DATA PALEMBANG BERDASARKAN TOPIK
DENGAN MENGGUNAKAN LATENT DIRICHLET ALGORITHM (LDA)**

Nathaniel Jerin Sisanto

1923250007

Abstrak

Media sosial sangat memberikan dampak yang besar dalam kehidupan sehari-hari, salah satunya untuk berkomunikasi atau untuk mendapatkan informasi. Oleh karena itu semakin berkembangnya aplikasi media sosial membuat masyarakat menggunakan aplikasi media sosial untuk mencari informasi melalui internet. Aplikasi instagram merupakan salah satu media sosial yang paling populer karena memiliki topik berbeda berdasarkan dari postingan yang berupa gambar atau video oleh karena itu, sangat sulit untuk mengidentifikasi suatu topik tersebut secara manual. Salah satu cara untuk mendapatkan informasi tersirat di media sosial adalah melalui pemodelan topik. Penelitian ini dilakukan untuk menganalisis penerapan metode LDA untuk identifikasi topik apa saja yang ada di instagram Universitas Multi Data Palembang. Topik yang dipilih dalam penelitian ini diperoleh dari LDA berdasarkan nilai koherensi. Penelitian ini menggunakan 2 model yaitu *random forest* dan *decision tree* yang berfungsi untuk memprediksi topik. Setiap model yang diuji akan menghasilkan nilai *accuracy*, *precision*, *recall*, dan *f1-score* yang berbeda. Pengujian dilakukan pada dataset pelabelan LDA dan pelabelan manual, hasil pengujian pada dataset pelabelan LDA sangat baik dengan menggunakan model *random forest* dengan nilai *accuracy* 78%, *Precision* 80%, *Recall* 66.66%, dan *f1-score* 72.72%.

BAB 1

PENDAHULUAN

Bab 1 ini menyajikan dasar atau alasan yang mendasar dari penelitian ini. Bab ini membahas tentang latar belakang penelitian, permasalahan yang dikaji, batasan penelitian, tujuan dan manfaat penelitian, metode penelitian, dan struktur penulisan penelitian.

1.1 Latar Belakang

Seiring berjalannya waktu, pengguna teknologi saat ini sangat memberikan dampak yang besar dalam kehidupan sehari-hari, salah satunya adalah perkembangan teknologi informasi dan komunikasi. Semakin berkembangnya teknologi informasi dan komunikasi membuat masyarakat menggunakan media sosial untuk mencari informasi melalui internet dengan *smartphone*.

Media sosial adalah platform digital yang menawarkan setiap pengguna kesempatan untuk terlibat dalam kegiatan sosial. Beberapa kegiatan yang dapat dilakukan di media sosial, seperti berkomunikasi atau interaksi untuk memberikan informasi atau konten berupa teks, foto, dan video. Salah satu platform digital atau media sosial yang banyak digunakan untuk menerima informasi adalah *Instagram* (Umam, 2021).

Instagram juga dapat diartikan sebagai aplikasi yang berperan sebagai sarana berbagi foto dan video di jejaring sosial, memungkinkan pengguna untuk mengambil foto dan video serta menambahkan filter yang memberikan efek menarik pada foto. Pengguna juga dapat memberikan pendapat atau komentar negatif dan positif saat membahas suatu topik yang sedang tren saat ini (Rudi Dian Arifin, 2023). *Instagram* juga sering digunakan untuk media periklanan karena *instagram* memiliki fitur-fitur yang mendukung pemasaran produk dengan sangat baik. Besarnya peluang ini memicu maraknya akun elektronik atau toko online yang menjual berbagai produk di *instagram*. Toko online adalah toko virtual dimana pembeli dan penjual tidak dapat bertemu secara langsung, jadi dibutuhkan *customer engagement* (Rohadian & Amir, 2019). *Customer Engagement* adalah suatu interaksi emosional antara pelanggan dan perusahaan yang dibentuk oleh motivasi, rujukan dan pengalaman pelanggan melalui postingan media sosial yang kreatif dari merek yang memengaruhi pembeli (Ashari & Sitorus, 2023). Oleh karena itu, keterlibatan pengguna di media sosial sangat berpengaruh, sehingga postingan atau akun media sosial dianggap populer dan cenderung menarik lebih banyak perhatian, jika menghasilkan indikator engagement yang lebih tinggi untuk postingan atau akun tersebut. Sehingga peran digital di perguruan tinggi sangat penting untuk menjaring calon mahasiswa dan mempromosikan fasilitas kampus atau kegiatan kampus. Ada beberapa perguruan tinggi di Palembang yaitu Universitas Sriwijaya, Universitas PGRI Palembang, Universitas Multi Data Palembang, Universitas Katolik Musi Charitas, Universitas Tamansiswa, Universitas Kader Bangsa, Universitas

Tridinanti Palembang, Universitas IGM, Universitas Bina Darma, Universitas Muhammadiyah Palembang, dan Universitas Islam Negeri (Raihannisa Fitriah, 2021).

Banyaknya Universitas yang ada di SUMATERA SELATAN penulis memilih Universitas Multi Data Palembang untuk melakukan metode pemodelan topik pada *Instagram* Universitas Multi Data Palembang yang terletak di Jl. Rajawali No. 14,9 Ilir, Kec. Ilir Tim. II, Kota Palembang, SUMATERA SELATAN. Peneliti memilih Universitas Multi Data Palembang karena Universitas Multi Data Palembang menempati peringkat dua dalam pemeringkatan perguruan tinggi di SUMATERA SELATAN versi *webimerics* Juli 2021 dan *instagram* Universitas Multi Data Palembang memiliki 11,7 ribu followers, namun like dan komentar pada postingan *instagram* Universitas Multi Data Palembang sangat tidak sesuai dengan followers, sehingga peneliti memilih media sosial *instagram* Universitas Multi Data Palembang untuk melakukan pemodelan topik. Pemodelan topik merupakan salah satu teknik pembelajaran mesin yang digunakan untuk mengelompokkan dokumen ke dalam topik atau tema yang berkaitan. Pemodelan topik mencakup banyak metode yang dapat digunakan, seperti *Latent Semantic Analysis (LSA)*, *Probabilistic Latent Semantic Analysis (PLSA)*, *Latent Dirichlet Allocation (LDA)*. *Latent Dirichlet Allocation (LDA)* merupakan model campuran yang ditingkatkan untuk menangkap pertukaran kata dan dokumen sistem lama melalui PLSA dan LSA. Metode pemodelan topik yang paling populer hingga saat ini adalah *Latent Dirichlet Allocation (LDA)* (Raihannisa Fitriah, 2021). *Latent Dirichlet*

Allocation (LDA) adalah algoritma probabilistik yang memodelkan setiap dokumen sebagai kombinasi dari topik yang berbeda dan distribusi kata dari setiap topik dalam dokumen. Algoritma *Latent Dirichlet Allocation* (LDA) mencoba menemukan topik atau tema yang paling sering muncul dalam dokumen, lalu menentukan seberapa sering kata tertentu muncul dalam topik tersebut.

Pembuktian penerapan Algoritma *Latent Dirichlet Allocation* (LDA) dapat dibuktikan pada penelitian yang dilakukan oleh Akhsin Nurlyli dan Morch. Ari Nasichuddin (2019), mengenai Topic Modeling Penelitian Dosen JPTEI UNY Pada Google Scholar Menggunakan *Latent Dirichlet Allocation*. Data penelitian ini dilakukan dengan teknik *Scraping*. Hasil *Scraping* disimpan dalam bentuk format.csv dan data yang didapat sebanyak 909 judul publikasi. Proses klasifikasi dibagi menjadi 4 kluster atau 4 topik dan algoritma yang digunakan adalah *Latent Dirichlet Allocation*. Hasil penelitian paling banyak pada kluster ke-1 dan paling sedikit kluster ke-4 (Nurlyli & Nasichuddin, 2019).

Selanjutnya pada penelitian terdahulu yang dilakukan oleh Bagus Wicaksono Arianto dan Gangga Anuraga (2020), mengenai Pemodelan Topik Pengguna Twitter Mengenai Aplikasi “Ruangguru”. Data penelitian ini dilakukan dengan menggunakan tools *Wordcloud*. Algoritma yang digunakan adalah *Latent Dirichlet Allocation*. Metode pengklasteran *Latent Dirichlet Allocation* (LDA) pada aplikasi ruangguru mengelompokkan data twitter menjadi 28 topik dengan topik yang sering dibahas adalah diskon ruangguru (Arianto & Anuraga, 2020).

Berdasarkan uraian bahwa algoritma *Latent Dirichlet Allocation* (LDA) sangat tepat dan cocok digunakan dalam melakukan pemodelan topik. Dalam penelitian ini akan menggunakan *instagram* untuk melihat postingan-postingan dan membandingkan topik-topik yang dibahas di media sosial *instagram* Universitas Multi Data Palembang.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah dijelaskan, maka didapat sebuah rumusan masalah dalam penelitian ini adalah bagaimana *Latent Dirichlet Allocation* (LDA) dapat digunakan dalam penelitian topik yang sedang hangat diperbincangkan di *Instagram* Universitas Multi Data Palembang.

1.3 Ruang Lingkup

Ruang Lingkup yang digunakan dalam penelitian analisis sentiment ini antara lain sebagai berikut:

1. Algoritma yang digunakan adalah *Latent Dirichlet Allocation* (LDA) dengan mencari trending topik yang sedang hangat diperbincangkan di *Instagram* Universitas Multi Data Palembang.
2. Aplikasi android yang dilakukan dalam pemodelan topik adalah *Instagram*.
3. Pengambilan data dilakukan dengan melakukan *Scraping*.
4. Dataset yang digunakan sebanyak 500 data.
5. Library yang digunakan sebagai alat bantu penelitian ini adalah NLTK dan Sastrawi.

6. Proses analisis pemodelan topik menggunakan tools *Jupyter Notebook* dengan bahasa pemrograman Python.
7. Pelabelan data dilakukan oleh peneliti dan di validasi oleh bagian marketing Universitas Multi Data Palembang.
8. Topik yang digunakan ada 8 topik yaitu Wisuda, Tips, Alumni, Akademik, Promosi, Prestasi, Peringatan, dan Event

1.4 Tujuan dan Manfaat

Tujuan penelitian ini adalah untuk menerapkan metode *Latent Dirichlet Allocation* (LDA) dalam pemodelan topik terhadap pengunjung media sosial Universitas Multi Data Palembang dengan menggunakan aplikasi *Instagram*.

Manfaat dari penelitian ini adalah :

1. Memahami cara menerapkan algoritma *Latent Dirichlet Allocation* (LDA) terhadap analisis pemodelan topik dari aplikasi.
2. Memahami cara *Scraping* data dengan menggunakan *Jupyter Notebook*.
3. Mengetahui trending topik yang sedang hangat diperbincangkan di *Instagram* Universitas Multi Data Palembang.

1.5 Metode Penelitian

Pada tahap ini dibahas mengenai langkah-langkah yang dilakukan untuk Analisis engagement pengunjung sosial media universitas multi data Palembang berdasarkan topik dengan menggunakan *Latent Dirichlet Algorithm* (LDA), yaitu sebagai berikut:

1. Identifikasi Masalah

Pada tahap ini, akan dilakukan identifikasi masalah mengenai pemodelan topik pada aplikasi instagram Universitas Multi Data Palembang dengan metode Algoritma *Latent Dirichlet Algorithm* (LDA). Dengan memperoleh masalah utama yang ingin diatasi, menjadikan penelitian ini sangat penting untuk dilakukan.

2. Studi Literatur

Pada tahap ini dilakukan studi pustaka dari beberapa sumber, misalnya jurnal dan buku-buku yang masih berhubungan dengan topik penelitian. Hal ini bertujuan untuk mendapatkan informasi dan data yang dapat digunakan untuk melakukan penelitian ini.

3. Pengumpulan Dataset

Pengumpulan data pada situs Instagram dilakukan dengan proses *Scraping* menggunakan *Jupyter Notebook*. Dataset yang diambil dari tahun 2021 sampai 2023. Fitur data yang digunakan dalam sistem diperoleh dari akun Instagram Universitas Multi Data Palembang dan data yang terkumpul sekitar 500 data.

4. Pre-Processing Data

Pada tahap Pre-Processing, dataset yang didapatkan dari instagram akan melakukan tahapan Case Folding, Tokenize, Stop Word, dan Stemming. Case Folding dalam Pre-Processing adalah suatu tahap yang bertujuan untuk mengubah semua karakter dalam teks menjadi huruf kecil atau huruf besar, Tokenize dalam Pre-Processing adalah suatu tahap yang

dapat membantu dalam menghilangkan tanda baca dan karakter-karakter yang tidak diperlukan dalam analisis data teks, Stop word dalam Pre-Processing adalah suatu tahap yang dapat mengurangi kompleksitas data dan mempercepat proses pemrosesan data teks, dan Stemming dalam Pre-Processing adalah suatu tahap yang dapat menghasilkan bentuk dasar (root word) dari kata-kata dalam teks yang diolah. Dengan melakukan Stemming, kata-kata yang memiliki varian atau bentuk yang sama dapat digabungkan menjadi satu entitas, sehingga dapat memperkecil kompleksitas data dan meningkatkan efektivitas analisis data teks.

5. Pelabelan Data

Pada tahap pelabelan data, banyaknya data untuk diberi label ialah 500 data. Pelabelan data akan terbagi menjadi 8 kategori topik yaitu Prestasi, Akademik, Alumni, Tips, Wisuda, Promosi, Event, dan Peringatan. Pada tahap pelabelan data dilakukan oleh peneliti dan divalidasi oleh bagian marketing Universitas Multi Data Palembang.

6. Perancangan Sistem

Pada tahap ini metode Decision Tree dan Random Forest digunakan untuk memprediksi topik pada postingan Instagram, yang merupakan jumlah topik yang ada di Instagram Universitas Multi Data Palembang.

7. Validasi

Pada tahap ini dilakukan validasi model, validasi yang ditargetkan adalah fase di mana hasil yang diperoleh sistem diperiksa kebenarannya dan berjalan sesuai dengan apa yang diinginkan.

8. Evaluasi

Pada tahap terakhir, sistem yang dibangun dievaluasi, yang tujuannya adalah untuk memastikan kebenaran hasil yang diperoleh dan meminimalkan kesalahan. Pada fase ini, metode *confusion matrix* digunakan untuk menghitung *precision*, *recall*, *f1-score*, dan *accuracy*.

1.6 Sistematika Penulisan

Bab-bab laporan ini diuraikan dalam sistematika penulisan. Ada lima bab yang menyusun tugas akhir ini, beberapa di antaranya merupakan sub-bab. Berikut ini adalah struktur penulisan yang sistematis:

BAB 1 PENDAHULUAN

Sebagai pedoman penulisan karya ilmiah, bab ini menjelaskan tentang latar belakang, rumusan pertanyaan, ruang lingkup, tujuan dan manfaat, metodologi penelitian, dan sistematika penulisan.

BAB 2 LANDASAN TEORI

Bab ini membahas penjelasan teori yang berhubungan dengan judul yang digunakan dalam penelitian berupa pemodelan topik, Algoritma *Latent Dirichlet Algorithm* (LDA), dan penelitian terkait.

BAB 3 RANCANGAN SISTEM

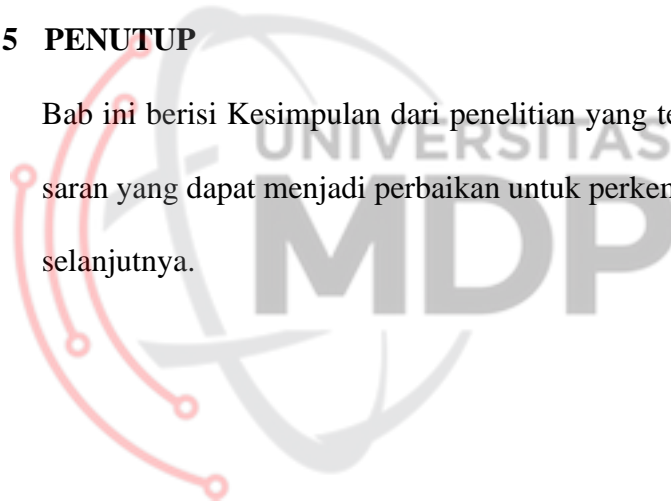
Bab ini menjelaskan tentang identifikasi masalah, studi literatur, pengumpulan dataset, pre-processing data, pelabelan data, perancangan sistem, evaluasi, validasi.

BAB 4 HASIL DAN PEMBAHASAN

Bab ini berisi pembahasan yang menjelaskan langka-langkah yang dilakukan untuk mendapatkan hasil.

BAB 5 PENUTUP

Bab ini berisi Kesimpulan dari penelitian yang telah dilakukan dan saran yang dapat menjadi perbaikan untuk perkembangan penelitian selanjutnya.



DAFTAR PUSTAKA

- Alfanzar, A. I., Khalid, K., & Rozas, I. S. (2020). Topic modelling skripsi menggunakan metode latent dirichlet allocation. *JSil (Jurnal Sistem Informasi)*, 7(1), 7–13.
- Arianto, B. W., & Anuraga, G. (2020). Pemodelan topik pengguna twitter mengenai aplikasi “ruangguru.” *Jurnal ILMU DASAR*, 21(2), 149–154.
- Ashari, R. M. H., & Sitorus, O. F. (2023). Pengaruh content marketing terhadap customer engagement kopi kenangan. *Jurnal EMT KITA*, 7(1), 38–46.
- Enos. (2022, April 22). *What is search engine scraping?* <https://www.enostech.com/what-is-search-engine-scraping/>.
- Instagram. (2023). *We bring you closer to the people.* <https://about.instagram.com/about-us>.
- Jupyter. (2023, May 10). *Jupyter.* <https://jupyter.org/>.
- Kabiru, I. N., & Sari, P. K. (2019). Analisa Konten Media Sosial E-commerce Pada Instagram Menggunakan Metode Sentiment Analysis Dan Lda-based Topic Modeling (studi Kasus: Shopee Indonesia). *EProceedings of Management*, 6(1).
- Kannitha, D. Z. T., Mustafid, M., & Kartikasari, P. (2022). Pemodelan topik pada keluhan pelanggan menggunakan algoritma latent dirichlet allocation dalam media sosial twitter. *Jurnal Gaussian*, 11(2), 266–277.
- Nurlayli, A., & Nasichuddin, M. A. (2019). Topik modeling penelitian dosen jpte uny pada google scholar menggunakan latent dirichlet allocation. *Elinvo (Electronics, Informatics, and Vocational Education)*, 4(2), 154–161.
- Patmawati, P., & Yusuf, M. (2021). Analisis topik modelling terhadap penggunaan sosial media twitter oleh pejabat negara. *Building of Informatics, Technology and Science (BITS)*, 3(3), 122–129.
- Putu, N. L. P. M., & Amrullah, A. Z. (2021). Analisis sentimen dan pemodelan topik pariwisata lombok menggunakan algoritma naive bayes dan latent dirichlet allocation. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(1), 123–131.
- Python. (2023, May 8). *The python tutorial.* <https://docs.python.org/3/tutorial/>.
- Raihannisa Fitriah. (2021, October 6). *11 Universitas di sumatera selatan, nomor 9 view jembatan ampera.* <https://sumsel.inews.id/Berita/Universitas-Di-Sumatera-Selatan/All>.

- Rangga Rizky A. (2018, August 17). *Topic modelling menggunakan latent dirchlect allocation*. <https://Medium.Com/@ranggaantok/Topic-Modelling-Menggunakan-Latent-Dirchlect-Allocation-3fdf979ffd05>.
- Rohadian, S., & Amir, M. T. (2019). Upaya membangun customer engagement melalui media sosial instagram. *Journal of Entrepreneurship, Management and Industry (JEMI)*, 2(4), 179–186.
- Rudi Dian Arifin. (2023, February 28). *Pengertian instagram beserta sejarah, fungsi, tujuan, manfaat, dll*. <https://Dianisa.Com/Pengertian-Instagram/>.
- Saifullah, S., Zarlis, M., Zakaria, Z., & Sembiring, R. W. (2017). Analisa terhadap perbandingan algoritma decision tree dengan algoritma random tree untuk pre-processing data. *J-SAKTI (Jurnal Sains Komputer Dan Informatika)*, 1(2), 180–185.
- Setijohatmo, U. T., Rachmat, S., Susilawati, T., & Rahman, Y. (2020). Analisis metoda latent dirichlet allocation untuk klasifikasi dokumen laporan tugas akhir berdasarkan pemodelan topik. *Prosiding Industrial Research Workshop and National Seminar*, 11(1), 402–408.
- Siringoringo, R., Jamaluddin, J., & Perangin-Angin, R. (2020). Pemodelan topik berita menggunakan latent dirichlet allocation dan k-means clustering. *Jurnal Informatika Kaputama (JIK)*, 4(2), 216–222.
- Syaifuddin, A., Harianto, R. A., & Santoso, J. (2020). Analisis trending topik untuk percakapan media sosial dengan menggunakan topic modelling berbasis algoritme lda. *INSYST: Journal of Intelligent System and Computation*, 2(1), 12–19.
- Syed, S., & Spruit, M. (2017). Full-text or A\abstract? examining topic coherence scores using latent dirichlet allocation. *2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, 165–174. <https://doi.org/10.1109/DSAA.2017.61>
- Umam. (2021). *Pengertian media sosial, sejarah, fungsi, jenis, manfaat, dan perkembangannya*. <https://Www.Gramedia.Com/Literasi/Pengertian-Media-Sosial/>.
- Universitas MDP. (2021). *Universitas multi data palembang*. <https://Mdp.Ac.Id/>.
- Universitas MDP. (2023). *Instagram*. <https://Www.Instagram.Com/Universitasmdp/>.
- Zhao, B. (2017). Web scraping. In L. A. Schintler & C. L. McNeely (Eds.), *Encyclopedia of Big Data* (pp. 1–3). Springer International Publishing. https://doi.org/10.1007/978-3-319-32001-4_483-1