PEMODELAN TOPIK CUITAN MEDIA ONLINE PADA TWITTER MENGGUNAKAN LATENT DIRICHLET ALLOCATION

SKRIPSI

Diajukan untuk memenuhi salah satu syarat memperoleh gelar Sarjana
Komputer pada Program Studi S1 Informatika

Disusun Oleh

Anisykurli Faza Ramadhani

18.11.2438

PROGRAM SARJANA

PROGRAM STUDI INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

2023

PEMODELAN TOPIK CUITAN MEDIA ONLINE PADA TWITTER MENGGUNAKAN LATENT DIRICHLET ALLOCATION

SKRIPSI

Diajukan untuk memenuhi salah satu syarat memperoleh gelar Sarjana
Komputer pada Program Studi S1 Informatika

Disusun Oleh

Anisykurli Faza Ramadhani

18.11.2438

PROGRAM SARJANA

PROGRAM STUDI INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

2023

HALAMAN PERSETUJUAN

SKRIPSI

PEMODELAN TOPIK CUITAN MEDIA ONLINE PADA TWITTER MENGGUNAKAN LATENT DIRICHLET ALLOCATION

yang disusun dan diajukan oleh

Anisykurli Faza Ramadhani

18.11.2438

telah disetujui oleh Dosen Pembimbing Skripsi

pada tanggal 7 Juni 2023

Dosen Pembimbing,

Anggit Dwi Hartanto, M.Kom

NIK. 190302163

HALAMAN PENGESAHAN

SKRIPSI

PEMODELAN TOPIK CUITAN MEDIA ONLINE PADA TWITTER MENGGUNAKAN LATENT DIRICHLET ALLOCATION

yang disusun dan diajukan oleh

Anisykurli Faza Ramadhani

18.11.2438

Telah dipertahankan di depan Dewan Penguji

pada tanggal 26 Juni 2023

Susunan Dewan Penguji

Nama Penguji	Tanda Tangan
Arif Dwi Laksito, M.kom **NIK. 190302150**	
Andi Sunyoto, M.Kom., Dr. **NIK. 190302052**	
Anggit Dwi Hartanto, M.Kom **NIK. 190302163**	

Skripsi ini telah diterima sebagai salah satu persyaratan

untuk memperoleh gelar Sarjana Komputer

Tanggal 26 Juni 2023

DEKAN FAKULTAS ILMU KOMPUTER

Hanif Al Fatta, S.Kom., M.Kom.

NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama Mahasiswa : Anisykurli Faza Ramadhani
NIM : 18.11.2438

Menyatakan bahwa skripsi dengan judul berikut:

PEMODELAN TOPIK CUITAN MEDIA ONLINE PADA TWITTER MENGGUNAKAN LATENT DIRICHLET ALLOCATION

Dosen Pembimbing : Anggit Dwi Hartanto, M.Kom

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 26 Juni 2023

Yang Menyatakan,

Anisykurli Faza Ramadhani

HALAMAN PERSEMBAHAN

Dengan telah diselesaikannya skripsi ini, penulis mempersembahkan karya tulis ini kepada:

1. Kedua orang tua saya yang selalu memberi motivasi dan doa untuk saya.
2. Bapak Anggit Dwi Hartanto, M.Kom. yang telah membimbing saya dalam mengerjakan skripsi.
3. Teman-teman terdekat saya yang selalu memberikan semangat dalam mengerjakan skripsi.
4. Teman-teman kelas 18-IF09 yang menemani saya dari awal hingga akhir perkuliahan di Universitas Amikom Yogyakarta.

KATA PENGANTAR

Puji dan syukur penulis ucapkan kehadirat Allah SWT. Yang telah memberikan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi ini dengan judul **"PEMODELAN TOPIK CUITAN MEDIA ONLINE PADA TWITTER MENGGUNAKAN LATENT DIRICHLET ALLOCATION"** yang menjadi salah satu syarat untuk menyelesaikan masa studi program sarjana di Universitas AMIKOM Yogyakarta.

Dalam menyusun skripsi ini, penulis banyak mendapat dukungan, bimbingan dan kemudahan dari berbagai pihak. Pada kesempatan ini penulis ingin mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. M. Suyanto, M.M. selaku Rektor Universitas AMIKOM Yogyakarta.
2. Bapak Hanif Al Fatta, M.Kom. selaku Dekan Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta.
3. Ibu Windha Mega Pradnya D., M.Kom. selaku Ketua Program Studi S1 Informatika Universitas AMIKOM Yogyakarta.
4. Bapak Anggit Dwi Hartanto, M.Kom. selaku Dosen Pembimbing yang selalu memberikan bimbingan, saran dan masukan dalam penulisan skripsi.

Penulis menyadari bahwa skripsi ini masih memiliki banyak kekurangan, maka dari itu kritik dan saran yang membangun sangat dibutuhkan penulis untuk perbaikan karya selanjutnya. Akhir kata semoga penelitian ini dapat bermanfaat bagi pembaca dan menambah wawasan khususnya dalam bidang *natural language processing*.

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR KODE

INTISARI

Media sosial twitter merupakan salah satu *platform* yang wajib digunakan oleh media-media online di Indonesia untuk menyebarkan berita dalam bentuk *tweet*. Banyaknya *tweet* yang dibuat oleh media online menyebabkan sulit untuk menganalisis topik-topik apa saja yang dibahas dalam pemberitaan khususnya berita tentang covid-19. Tujuan dari penelitian ini adalah untuk menemukan topik-topik tersembunyi dari kumpulan *tweet* yang dibuat oleh beberapa media online berbahasa Indonesia dengan teknik pemodelan topik.

Metode LDA merupakan metode pemodelan topik populer yang banyak digunakan dan juga merupakan metode penyempurnaan dari metode-metode sebelumnya seperti *Latent Semantic Analysis* (LSA) dan *Probabilistic Latent Semantic Analysis* (PLSA). Untuk mendapatkan model yang maksimal, dilakukan tahapan *preprocessing* pada data teks yang akan digunakan dan juga evaluasi model menggunakan *topic coherence* untuk menentukan banyak topik yang akan dihasilkan, kemudian topik yang dihasilkan akan dilakukan analisis untuk setiap topiknya menggunakan visualisasi hasil, sehingga akan diketahui makna atau pembahasan apa yang terkandung pada masing-masing topik.

Pemodelan topik yang dilakukan menghasilkan 12 topik dengan *coherence score* sebesar 0,403044. Kata-kata yang muncul pada setiap topik dapat diinterpretasi dengan baik sehingga setiap topik memiliki makna yang dapat dipahami. Hasil menunjukkan bahwa topik tentang *update* kasus harian covid-19 muncul paling banyak di antara topik lainnya, sedangkan topik tentang perkembangan kasus positif di Jawa-Bali menjadi topik yang paling sedikit dibahas oleh media online.

Kata kunci: *Latent Dirichlet Allocation*, Pemodelan Topik, Berita, Media Online

ABSTRACT

Twitter social media is one of the platforms used by online media in Indonesia to spread news in the form of tweets. The large number of tweets made by online media makes it difficult to analyze what topics are discussed in the news, especially news about covid-19. The purpose of this research is to find hidden topics from a collection of tweets made by several Indonesian-language online media with topic modeling techniques.

The LDA method is a popular topic modeling method that is widely used and is also a refinement of previous methods such as Latent Semantic Analysis (LSA) and Probabilistic Latent Semantic Analysis (PLSA). To get the maximum model, the preprocessing stage is carried out on the text data to be used and also evaluates the model using topic coherence to determine the number of topics that will be generated, then the resulting topics will be analyzed for each topic using visualization of the results, so that it will be known what meaning or discussion is contained in each topic.

Topic modeling resulted in 12 topics with a coherence score of 0.403044. The words that appear in each topic can be interpreted well so that each topic has a meaning that can be understood. The results show that the topic of daily covid-19 case updates appears the most among other topics, while the topic of positive case developments in Java-Bali is the topic least discussed by online media.

Keywords: *Latent Dirichlet Allocation, Topic Modeling, News, Online Media*