IMPLEMENTASI ALGORITMA LATENT DIRICHLET ALLOCATION UNTUK TOPIC MODELING TERHADAP DATA TWITTER TERKAIT PANDEMI COVID-19

SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh gelar

Sarjana Komputer (S.Kom.)

Antonius Wisnu Setyawan

00000025802

PROGRAM STUDI INFORMATIKA

FAKULTAS TEKNIK DAN INFORMATIKA

UNIVERSITAS MULTIMEDIA NUSANTARA

TANGERANG

2021

LEMBAR PENGESAHAN

IMPLEMENTASI ALGORITMA LATENT DIRICHLET ALLOCATION UNTUK TOPIC MODELING TERHADAP DATA TWITTER TERKAIT PANDEMI COVID-19

Oleh

Nama	: Antonius Wisnu Setyawan
NIM	: 00000025802
Program Studi	: Informatika
Fakultas	: Teknik dan Informatika

Tangerang, 25 Juni 2021

Ketua Sidang	Dosen Penguji
Digitally signed by Dennis Gunawan DN: CN=Dennis Gunawan, C=ID, OU=Department of Informatics, O=Universitas Multimedia Nusantara, E=dennis.gunawan@umn.ac.id Reason: I am approving this document Location: Jakarta Date: 2021-06-28 10:24:08	26 Juni 2021
Dennis Gunawan, S.Kom., M.Sc.	Andre Rusli, S.Kom., M.Sc.
Dosen Pembimbing I	Dosen Pembimbing II
Julio Christian Young, S.Kom., M.Kom.	Alethea Suryadibrata, S.Kom., M.Eng.

Mengetahui,

Ketua Program Studi Informatika

Digitally signed by Marlinda Vasty Overbeek
Date: 2021.06.28 11:39:30 +07'00'

Marlinda Vasty Overbeek, S.Kom., M.Kom.

PERNYATAAN TIDAK MELAKUKAN PLAGIAT

Dengan ini saya:

Nama : Antonius Wisnu Setyawan

NIM : 00000025802

Program Studi : Informatika

Fakultas : Teknik dan Informatika

menyatakan bahwa Skripsi yang berjudul "**IMPLEMENTASI ALGORITMA LATENT DIRICHLET ALLOCATION UNTUK TOPIC MODELING TERGADAP DATA TWITTER TERKAIT PANDEMI COVID-19**" ini adalah karya ilmiah saya sendiri, bukan plagiat dari karya ilmiah yang ditulis oleh orang lain atau lembaga lain, dan semua karya ilmiah orang lain atau lembaga lain yang dirujuk dalam Skripsi ini telah disebutkan sumber kutipannya serta dicantumkan di Daftar Pustaka. Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan Skripsi maupun dalam penulisan laporan Skripsi, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk mata kuliah Skripsi yang telah saya tempuh.

Tangerang, 5 Juni 2021

(Antonius Wisnu Setyawan)

PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Demi perkembangan ilmu pengetahuan, menyetujui dan memberikan izin kepada **Universitas Multimedia Nusantara** hak Bebas Royaliti Non-eksklusif (*Non-exclusive Royalty-Free Right*) atas karya ilmiah saya yang berjudul:

IMPLEMENTASI ALGORITMA LATENT DIRICHLET ALLOCATION UNTUK TOPIC MODELING TERHADAP DATA TWITTER TERKAIT PANDEMI COVID-19

beserta perangkat yang diperlukan.

Dengan Hak Bebas Royalti Non-eksklusif ini, pihak **Universitas Multimedia Nusantara** berhak menyimpan, mengalihmedia atau *format*-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mendistribusi dan menampilkan atau mempublikasikan karya ilmiah saya di internet atau media lain untuk kepentingan akademis, tanpa perlu meminta izin dari saya maupun memberikan royalti kepada saya, selama tetap mencantumkan nama saya sebagai penulis karya ilmiah tersebut. Demikian pernyataan ini saya buat dengan sebenarnya untuk dipergunakan sebagaimana mestinya.

Tangerang, 5 Juni 2021

(Antonius Wisnu Setyawan)

HALAMAN PERSEMBAHAN / MOTO

"*Neque porro quisquam est qui dolorem ipsum quia dolor sit amet, consectetur, adipisci velit, sed quia nonnumquam eiusmodi tempora incidunt ut labore et dolore magnam aliquam quaerat voluptatem.*"

Cicero

KATA PENGANTAR

Puji syukur kepada Tuhan Yang Maha Esa atas segala karunia dan rahmat-Nya sehingga Skripsi yang berjudul "**IMPLEMENTASI ALGORITMA LATENT DIRICHLET ALLOCATION UNTUK TOPIC MODELING TERHADAP DATA TWITTER TERKAIT PANDEMI COVID-19**" dapat diselesaikan dengan baik dan tepat waktu.

Dalam penyusunan Skripsi ini, banyak pihak telah membantu, maka dari itu pada kesempatan ini terima kasih diucapkan kepada:

1. Bapak Dr. Ninok Leksono selaku Rektor Universitas Multimedia Nusantara,
2. Bapak Dr. Eng. Niki Prastomo, S.T., M.Sc., Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara,
3. Ibu Marlinda Vasty Overbeek, S.Kom., M.Kom., Ketua Program Studi Informatika Universitas Multimedia Nusantara,
4. Bapak Julio Christian Young, S.Kom., M.Kom., dan Ibu Alethea Suryadibrata, S.Kom., M.Eng. yang membimbing, memberikan arahan dan masukan dalam pembuatan Skripsi dengan baik,
5. Bapak dan Ibu Dosen Program Studi Informatika Universitas Multimedia,
6. Keluarga khsususnya untuk Mama yang telah memberi motivasi dan nasihat-nasihat, serta Almarhum Papa yang selalu memberi inspirasi untuk terus maju, dan juga sanak saudara yang telah memberi semangat,
7. Teman terdekat dan saudara sepupu khususnya untuk Billy, Audy, Bramantyo, Chrisanta, Carissa, Benny, Samuel, Neven, Michael, Grace, serta Kak Stella yang selalu membantu dalam menyegarkan pikiran dan memberikan hiburan serta memberi semangat dalam proses penyusunan Skripsi,
8. Teman-teman seperjuangan lainnya,

9. Semua pihak yang tidak dapat disebutkan satu per satu yang telah membantu memberikan dukungan serta doa hingga penyusunan laporan Skripsi ini dapat terselesaikan dengan baik.

Dalam penulisan Skripsi ini masih jauh dari sempurna, oleh karena itu segala bentuk kritik dan saran yang bersifat membangun akan sangat membantu dalam menyempurnakan penulisan Skripsi ini. Semoga Skripsi ini dapat bermanfaat, baik sebagai sumber informasi maupun sumber inspirasi.

Tangerang, 5 Juni 2021

Antonius Wisnu Setyawan

IMPLEMENTASI ALGORITMA LATENT DIRICHLET ALLOCATION UNTUK TOPIC MODELING TERHADAP DATA TWITTER TERKAIT PANDEMI COVID-19

ABSTRAK

Pada tahun 2019, penemuan virus penyakit baru mengejutkan masyarakat dunia, yang tidak lama kemudian penyebaran virus penyakit ini berkembang menjadi pandemi. Pandemi ini disebabkan oleh virus *severe acute respiratory syndrome coronavirus 2* (SARS-CoV-2) atau yang dikenal sebagai virus Corona. Penyakit yang ditimbulkan akibat virus ini disebut *Coronavirus Disease 2019* atau yang lebih dikenal dengan COVID-19. Indonesia merupakan salah satu negara yang terdampak pandemi ini dan berjuang untuk menghadapi pandemi virus COVID-19. Untuk menekan penyebaran virus ini, telah dilakukan oleh pemerintah Indonesia berbagai cara, salah satunya adalah dengan memperjuangkan vaksinasi yang sudah mulai dilakukan pemerintah. Namun vaksinasi juga menimbulkan tanggapan positif maupun negatif bagi masyarakat Indonesia serta menimbulkan berbagai macam isu atau kontroversi maupun teori konspirasi yang beredar pada sosial media, dan salah satunya melalui Twitter. Pada penelitian ini dilakukan perancangan sistem untuk melakukan *topic modeling* dengan metode *Latent Dirichlet Allocation* (LDA) dilakukan untuk melihat ragam topik yang beredar pada masyarakat melalui kicauan atau *tweet* pada sosial media Twitter. Set data pada penelitian ini berupa *tweet* dari sosial media Twitter. Pengumpulan set data dilakukan mulai tanggal 20 Maret 2021 sampai dengan 29 Maret 2021 dengan kata kunci "#Covid-19" berbahasa Indonesia dan berlokasi Indonesia. Percobaan yang dilakukan dengan skenario perbandingan set data sebelum dilakukan penyaringan dan set data setelah dilakukan penyaringan beserta uji coba perbandingan *hyperparameter alpha* dengan *hyperparameter* jumlah topik. Hasil dari penelitian yang dilakukan menunjukkan implementasi metode LDA untuk melakukan *topic modeling* pada *tweet* dari sosial media Twitter berhasil dikembangkan. Dari hasil uji coba didapatkan performa metode LDA untuk melakukan *topic modeling* pada set data sebelum dilakukan penyaringan memiliki nilai *coherence score* terbaik yaitu 0.5232 dan untuk set data setelah dilakukan penyaringan memiliki nilai *coherence score* terbaik yaitu 0.4484.

Kata Kunci: COVID-19, *latent dirichlet allocation*, *topic coherence*, *topic modeling*, *tweet*

IMPLEMENTATION OF LATENT DIRICHLET ALLOCATION ALGORITHM FOR TOPIC MODELING ON TWITTER DATA RELATED TO COVID-19 PANDEMIC

ABSTRACT

In 2019, the discovery of a new disease virus shocked the world community, which soon afterwards the spread of this disease virus developed into a pandemic. This pandemic was caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) or what is known as the Corona virus. The disease caused by this virus is called Coronavirus Disease 2019 or better known as COVID-19. Indonesia is one of the countries affected by this pandemic and is struggling to deal with the COVID-19 virus pandemic. To suppress the spread of this virus, the Indonesian government has done various ways, one of which is fighting for vaccinations that the government has started to do. However, vaccination also generated positive and negative responses to the Indonesian people and raised various issues or controversies as well as conspiracy theories circulating on social media, and one of them is via Twitter. In this study, a system design for topic modeling with the Latent Dirichlet Allocation (LDA) method was carried out to see the various topics that were circulating in the community through tweets on Twitter social media. The data set in this study was acquired in the form of tweets from social media Twitter. The data set was collected from March 20, 2021 to March 29, 2021 with the keyword "# Covid-19" in Indonesian and located in Indonesia. Experiments were carried out with a data set comparison scenario before filtering and a data set after filtering along with a comparison trial between the alpha hyperparameter and the number of topics hyperparameter. The results of the research conducted show that the implementation of the LDA method for topic modeling on tweets from social media Twitter has been successfully developed. From the test results, it was found that the performance of the LDA method for conducting topic modeling on the data set before screening had the best coherence score, namely 0.5232, and the best coherence score for the data set after filtering was 0.4484.

Keywords: COVID-19, *latent dirichlet allocation*, *topic coherence*, *topic modeling*, *tweet*

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR RUMUS

DAFTAR LAMPIRAN