KLASIFIKASI TOPIK PADA WEB BERITA *ONLINE* DENGAN PENDEKATAN TEKS *MINING*.

SKRIPSI

M. Edo Noprian

161410264

PROGRAM STUDI SISTEM INFORMASI

FAKULTAS ILMU KOPMUTER

UNIVERSITA BINADARMA

PALEMBANG

2020

KLASIFIKASI TOPIK PADA WEB BERITA *ONLINE* DENGAN PENDEKATAN TEKS *MINING*.

M. Edo Noprian

161410264

Skripsi ini diajukan sebagai syarat memperoleh gelar Sarjana Komputer

PROGRAM STUDI SISTEM INFORMASI

FAKULTAS ILMU KOPMUTER

UNIVERSITA BINADARMA

PALEMBANG

2020

HALAMAN PENGESAHAN

KLASIFIKASI TOPIK WEB PADA BERITA ONLINE DENGAN PENDEKATAN TEKS MINING

M. EDO NOPRIAN
161410264

Telah diterima sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer pada Program Studi Sistem Informasi

Palembang, 08 September 2020
Fakultas Ilmu Komputer
Universitas Bina Darma
Dekan,

Pembimbing

Dr. Edi Surya Negara, M.Kom.

Dedy Syamsuar, S.Kom., M.I.T., Ph.D.

HALAMAN PERSETUJUAN

Skripsi Berjudul **"Klasifikasi Topik pada Web Berita *Online* dengan Pendekatan Teks *Mining*."** Oleh **M. Edo Noprian (161410264)** telah dipertahankan didepan komisi penguji pada **Selasa** tanggal **08 September 2020.**

Komisi Penguji

1. Ketua : Dr. Edi Surya Negara, M.Kom. ( )
2. Anggota : Maria Ulfa, M.Kom. ( )
3. Anggota : Firamon Syakti, M.M., M. Kom. ( )

Mengetahui,

Program Studi Sistem Informasi

Fakultas Ilmu Komputer

Universitas Bina Darma

Ketua,

Dr. Edi Surya Negara, M.Kom.

SURAT PERNYATAAN

Saya yang bertanda tangan dibawah ini :

Nama : M. Edo Noprian

NIM : 161410264

Dengan ini menyatakan bahwa :

1. Karya tulis saya (Skripsi) adalah asli dan belum pernah diajukan untukmendapatkan gelar akademik (Sarjana) di Universitas Bina Darma atauperguruan tinggi lainnya ;
2. Karya tulis ini murni gagasan, rumusan dan penelitian saya dengan arahandari tim pembimbing ;
3. Di dalam karya tulis ini tidak terdapat karya atau pendapat yang telah ditulis atau di publikasikan orang lain, kecuali secara tertulis dengan jelas dikutipdengan mencantumkan nama pengarang dan memasukkan ke dalam daftarrujukan ;
4. Saya bersedia tugas skripsi, di cek keasliannya menggunakan plagiarismchecker serta di unggah ke internet, sehingga dapat diakses secara daring ;
5. Surat pernyataan ini saya tulis dengan sungguh-sungguh dan apabila terbuktimelakukan penyimpangan atau ketidakbenaran dalam pernyataan ini makasaya bersedia menerima sanksi dengan peraturan dan perundang-undanganyang berlaku ;

Demikian surat pernyataan ini saya buat agar dapat dipergunakan sebagaimana mestinya.

Pelambang, 08 September 2020

Yang membuat pernyataan,

M. Edo Noprian

161410264

ABSTRAK

Pesatnya perkembangan dalam dunia informasi digital menyebabkan peningkatan volume informasi yang berbentuk teks seperti dokumen berita. Di sosial media dokumen berita sangat banyak di unggah dalam rentan waktu yang begitu cepat dan salah satunya adalah *Twitter*. *Twitter* adalah layanan sosial media yang sudah sangat banyak melayani pengguna sehingga menjadikannya sebagai salah satu sosial media yang memiliki data yang sangat besar. Dari data yang sangat besar tersebut dapat dimanfaatkan sebagai sumber dokumen berita untuk web berita *online*. Akan tetapi dengan banyaknya topik yang di ektraksi pada data *Twitter* membuat data yang masuk memiliki beragam topik yang menyebabkan kesulitan dalam mengidentifikasi topik dari kumpulan data yang diambil dan akan membutuhkan waktu yang banyak jika harus dilakukan secara manual oleh manusia. Sedangkan, data tersebut berpotensi dibutuhkan untuk memberikan informasi secepat mungkin.

Dengan dilakukannya penelitian ini, bertujuan untuk mengklasifikasikan topik pada data yang diambil dari *Twitter* secara otomatis sehingga dapat membuat klasifikasi pada dokumen berita yang diambil, dapat lebih efektiv dan efesien dan tidak membutuhkan waktu sebanyak dilakukan manual oleh manusia. Penelitian dilakukan dengan metode *Latent Dirichlet Allocation* (*LDA*). Dokumen berita yang akan di klasifikasi adalah dokumen berita bahasa Indonesia dan akan di klasifikasikan kedalam topik yang akan ditentukan. Eksperimen pemodelan topik dengan metode LDA menyimpulkan bahwa jumlah topik yang didbentuk dari 9094 data tweet adalah 10 topik. Hasil eksperimen ini *Latent Dirichlet Allocaion* dapat digunakan untuk *text mining*, tetapi data tweet yang dihasilkan belum layak untuk dijadikan informasi dalam pengambilan topik berita. Karena hasil dari metode ini masih membutuhkan peninjauan kembali oleh manusia untuk mengetahui fakta dari *topic* yang dihasilkan.

Kata Kunci : Klasifikasi Dokumen Berita, *Latent Dirichlet Allocation* (*LDA*)

ABSTRACT

The rapid development in the world of digital information has led to an increase in the volume of information in the form of *text* such as news documents. On social media, news documents are very much uploaded in a very vulnerable time and one of them is Twitter. Twitter is a social media service that has served very many users, making it one of the social media that has huge data. From this very large data, it can be used as a *source* of news documents for *online* news websites. However, with the many topics extracted from the Twitter data, the incoming data has a variety of topics which causes difficulty in identifying topics from the data collection taken and will require a lot of time if it has to be done manually by humans. Meanwhile, the data is potentially needed to provide information as quickly as possible.

By doing this research, it aims to classify topics on data taken from Twitter automatically so that it can classify the news documents that are taken, can be more effective and efficient and do not require as much time as done manually by humans. The research was conducted using the Latent Dirichlet Allocation (LDA) method. News documents to be classified are Indonesian news documents and will be classified into topics to be determined. The topic modeling experiment with the LDA method concluded that the number of topics that were formed from 9094 tweet data was 10 topics. The results of this experiment, Latent Dirichlet Allocaion, can be used for text mining, but the resulting tweet data is not suitable for information in taking news topics. Because the results of this method still require human review to find out the facts of the resulting topic.

Keywords: News Document Classification, *Latent Dirichlet Allocation* (*LDA*)

KATA PENGANTAR

بِسْمِ اللَّهِ الرَّحْمَٰنِ الرَّحِيمِ

Puji syukur kehadiran Tuhan Yang Maha Esa karena berkat rahmat dan karunia-nya skripsi yang berjudul "**Klasifikasi Topik Pada Web Berita Online Dengan Pendekatan Teks Mining"** dapat diselesaikan dengan baik untuk memenuhi salah satu syarat mendapatkan gelar Sarjana Komputer di Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Bina Darma.

Dalam penulisan skripsi ini, tentunya masih jauh dari sempurna.Hal ini dikarenakan keterbatasnya pengetahuan yang dimiliki.Oleh karena itu dalam rangka melengkapi kesempurnaan dari penulisan skripsi ini diharapkan adanya saran dan kritik yang diberikan bersifat membangun.

Pada kesempatan yang baik ini, tak lupa penulis menghaturkan terima kasih kepada semua pihak yang telah memberikan bimbingan, pengarahan, nasehat dan pemikiran dalam penulisan skripsi penelitian ini, terutama kepada :

1. Dr. Sunda Ariana, M.Pd., M.M. selaku Rektor Universitas Bina Darma Palembang.
2. Dedy Syamsuar, Ph.D. selaku Dekan Fakultas Ilmu Komputer.
3. Dr. Edi Surya Negara, M.Kom. selaku Ketua Program Studi Sistem Informasi, sekaligus pembimbing dalam menyelesaikan skripsi ini.
4. Kepada Ibu Maria Ulfa, M.Kom dan Firamon Syakti, M.M., M. Kom. sebagai penguji.
5. Kepada seluruh dosen dan mahasiswa Universitas Bina Darma yang telah membantu atas terlaksananya skripsi tersebut.
6. Kepada orang tua yang selalu memberikan semangat dan do'a sehingga dapat menyelesaikan skripsi ini.
7. Kepada kedua adik-adikku **"Syahrul"** dan **"Wulan"** yang selalu memberikan dukungan.

8. Kepada teman satu bimbingan yaitu Fajar, Tiara, Aldo, Fitra, Romadon, Siska, Yaya.
9. Kepada **Fajar** yang tida henti memberikan semangat.
10. Kepada Rekan-rekan kepengurusan HIMSIF Universitas BinaDarma.

Palembang, Agustus 2020

Penulis

DAFTAR ISI

DATAR GAMBAR

DAFTAR *SOURCE CODE*