

SOCIAL NETWORK ANALYSIS BERITA HOAX YANG TERSEBAR MELALUI MEDIA ONLINE DENGAN PEMBOBOTAN TF-IDF DAN PEMODELAN TOPIK MENGGUNAKAN LATENT DIRICHLET ALLOCATION

Muhammad Rizky¹, Puji Winar Cahyo²

INTISARI

Latar Belakang: Dewasa ini perkembangan teknologi menjadi salah satu faktor yang mempercepat persebaran berita yang dikategorikan sebagai berita palsu atau *hoax* yang dapat merugikan berbagai komponen masyarakat. Untuk itu akan dibangun sebuah model Social Network Analysis(SNA) sehingga masyarakat dapat menganalisis apakah setiap berita *hoax* yang beredar saling terkait dan dapat mengetahui tujuan adanya jaringan berita *hoax* tersebut.

Tujuan Penelitian: Menerapkan SNA pada berita *hoax* yang ada pada media online menggunakan Latent Dirichlet Allocation (LDA) *Topic Modeling* dan menampilkan grafik keterkaitan berita *hoax* pada platform Fact Checker Directory.

Metode Penelitian: Metode yang digunakan pada penelitian ini adalah SNA dan LDA *Topic Modeling* yang dikombinasikan dengan pembobotan TF-IDF

Hasil Penelitian: Pembentukan pemodelan topik pada sampel dengan jumlah data 500 data, topik optimal yang dihasilkan sebanyak 55 topik dengan nilai *Coherence Score* sebesar 0.506. Kemudian pada hasil analisis SNA kategori topik yang paling berpengaruh terhadap interaksi jaringan berita *hoax* Jokowi adalah kategori topik 1 “Media Masa Tentang Pemimpin Indonesia”. Dengan nilai *degree centrality* tertinggi, kategori topik 1 menjadi topik paling berpengaruh dari total jumlah interaksi yang dihasilkan dengan topik lain. Kategori topik 1 juga menjadi penghubung bagi interaksi topik lain di dalam jaringan dari nilai *betweenness centrality* yang didapat. Kategori topik 1 menjadi kategori topik yang paling banyak memiliki kedekatan dengan kategori topik lainnya dari nilai *closeness centrality* yang didapat.

Kesimpulan: Diketahui pada penelitian ini jumlah topik optimal yang dihasilkan berbanding lurus dengan jumlah data yang digunakan. Semakin banyak data yang digunakan maka semakin banyak jumlah topik yang dihasilkan secara optimal. Kemudian hasil analisis SNA kategori topik yang paling berpengaruh terhadap interaksi jaringan berita *hoax* Jokowi adalah kategori topik 1 “Media Masa Tentang Pemimpin Indonesia”.

Kata-kunci: SNA, TF-IDF, LDA, *Topic Modeling*

¹Mahasiswa Program Studi Informatika Universitas Jenderal Achmad Yani Yogyakarta

²Dosen Program Studi Informatika Universitas Jenderal Achmad Yani Yogyakarta

**SOCIAL NETWORK ANALYSIS OF HOAX NEWS DISTRIBUTED
THROUGH ONLINE MEDIA WITH TF-IDF WEIGHTING AND TOPIC
MODELING USING LATENT DIRICHLET ALLOCATION**

Muhammad Rizky¹, Puji Winar Cahyo²

ABSTRACT

Background: Today the development of technology is one of the factors that accelerates the spread of news that is categorized as fake news or hoaxes that can harm various components of society. For this reason, a Social Network Analysis (SNA) model will be built so that the public can analyze whether every hoax news circulating is related to each other and can find out the purpose of the hoax news network.

Objectives: Applying SNA to hoax news in online media using Latent Dirichlet Allocation (LDA) Topic Modeling and displaying a graph of the relationship between hoax news on the Fact Checker Directory platform.

Methods: The method used in this study is SNA and LDA Topic Modeling combined with TF-IDF weighting

Results: Formation of topic modeling in the sample with a total data of 500 data, the optimal topics generated are 55 topics with a Coherence Score value of 0.506. Then on the results of the SNA analysis, the topic category that has the most influence on the interaction of Jokowi's hoax news network is topic category 1 "Mass Media About Indonesian Leaders". With the highest degree centrality value, topic category 1 becomes the most influential topic out of the total number of interactions generated with other topics. Topic category 1 also becomes a link for the interaction of other topics in the network from the betweenness centrality value obtained. Topic category 1 is the topic category that has the most closeness to other topic categories from the closeness centrality value obtained.

Conclusion: It is known in this study that the number of optimal topics generated is directly proportional to the amount of data used. The more data used, the more the number of topics that are optimally generated. Then the results of the SNA analysis of the topic category that has the most influence on the interaction of Jokowi's hoax news network is topic category 1 "Mass Media About Indonesian Leaders".

Keywords: SNA, TF-IDF, LDA, Topic Modeling

¹Student Of Informatic Program Universitas Jenderal Achmad Yani Yogyakarta

²Lecturer Of Informatic Program Universitas Jenderal Achmad Yani Yogyakarta