

INTISARI

ANALISIS *TOPIC MODELING* MENGGUNAKAN *LATENT DIRICHLET ALLOCATION* DAN *LATENT SEMANTIC ANALYSIS* DARI DATA BERITA UNTUK *KNOWLEDGE MANAGEMENT* COVID-19

Oleh

Locita Kumara Sugiyana

22/495110/PPA/06286

Pandemi COVID-19 memberikan banyak sekali data yang dapat diolah menjadi informasi dan pengetahuan, salah satunya adalah data berita, dengan adanya pengetahuan ini diharapkan penanganan dan pencegahan COVID-19 atau pandemi selanjutnya dapat dilakukan dengan lebih optimal, pengetahuan ini dapat diperoleh dengan memanfaatkan *knowledge management* dan *big data analytic*. *Topic modeling* merupakan metode untuk mengekstraksi pengetahuan, pemodelan topik berasal dari LSA yang menentukan struktur sematik pada korpus, namun LSA memiliki kelemahan ketika dihadapkan pada kata dengan arti atau makna yang mirip menyebabkan topik yang dihasilkan terkadang mirip, LDA merupakan algoritma yang bersifat probabilistik yang dapat digunakan untuk mengidentifikasi topik tersembunyi dan lebih spesifik.

Penelitian ini bertujuan untuk mengeksplorasi algoritma pemodelan topik dalam mengidentifikasi topik pada berita yang mungkin mempengaruhi perubahan kasus COVID-19, kinerja LSA mendapatkan rata-rata *coherence score* 0,487345, waktu eksekusi 153ms, dan *jaccard similarity* 0,089841, LDA mendapatkan rata-rata *coherence score* 0,458974, waktu eksekusi 5097ms, dan *jaccard similarity* 0,0039, pada penelitian ini menunjukkan bahwa LDA lebih baik dalam memberikan informasi yang lebih spesifik terkait topik pada berita dibandingkan dengan LSA, penelitian ini menunjukkan bahwa terdapat perbedaan topik yang sangat terlihat ketika kasus baru COVID-19 meningkat maupun menurun, ketika kasus baru meningkat berdasarkan pemodelan topik didapatkan bahwa minggu tersebut terdapat kumpulan masa ataupun varian virus baru, sedangkan pada kasus baru menurun didapatkan bahwa minggu tersebut sedang terjadi Pemberlakuan Pembatasan Kegiatan Masyarakat.

Kata Kunci: *Knowledge Management, COVID-19, Topic Modeling, Latent Sematic Analysis, Latent Dirichlet Allocation.*

ABSTRACT

***TOPIC MODELING ANALYSIS USING LATENT DIRICHLET
ALLOCATION AND LATENT SEMANTIC ANALYSIS FROM NEWS DATA
FOR KNOWLEDGE MANAGEMENT COVID-19***

By

Locita Kumara Sugiyana

22/495110/PPA/06286

The COVID-19 pandemic provides a lot of data that can be processed into information and knowledge, one of which is news data, with this knowledge it is hoped that the handling and prevention of COVID-19 or the next pandemic can be done more optimally, this knowledge can be obtained by utilising knowledge management and big data analytics. Topic modeling is a method for extracting knowledge, topic modeling comes from LSA which determines the semantic structure of the corpus, but LSA has a weakness when faced with words with similar meanings or meanings causing the resulting topics to sometimes be similar, LDA is a probabilistic algorithm that can be used to identify hidden topics and more specific topic.

This research aims to explore topic modeling algorithms in identifying topics in news that might affect changes in COVID-19 cases, LSA performance gets an average coherence score of 0.487345, 153ms execution time, and jaccard similarity 0.089841, LDA gets an average coherence score of 0.458974, 5097ms execution time, and jaccard similarity 0.0039, this study shows that LDA is better at providing more specific information related to topics in news compared to LSA. This research shows that there are differences in topics that are very visible when new cases of COVID-19 increase or decrease, when new cases increase based on topic modelling it is found that that week there is a mass collection or new virus variants, while in new cases decreasing it is found that that week there is Community Activity Restrictions Enforcement.

Keywords: Knowledge Management, COVID-19, Topic Modeling, Latent Semantic Analysis, Latent Dirichlet Allocation.