

INTISARI

PEMODELAN TOPIK PADA MEDIA SOSIAL TWITTER
MENGGUNAKAN METODE LDA2VEC
(Studi kasus: Pembicaraan mengenai belanja *online*)

Oleh
Dinda Asyifa K
18/424285/PA/18390

Pesatnya perkembangan teknologi dan meningkatnya pengguna internet telah membuat aktivitas pada seluruh aspek kehidupan beralih ke digital, tidak terkecuali dalam aspek ekonomi. Meningkatnya aktivitas belanja secara daring membuat para pelaku usaha terus bersaing dalam mencakup konsumen dengan memerhatikan kebutuhan dan juga perilakunya yang dapat tercermin pada sosial media. Untuk menemukan topik tersembunyi pada pembicaraan konsumen di sosial media tersebut, tugas akhir ini membahas pemodelan topik mengenai pembicaraan belanja daring pada *Twitter* menggunakan metode LDA2Vec. LDA2vec merupakan metode hibrida antara *Latent Dirichlet Allocation* (LDA) dan *word2vec*. Metode penyematan kata *word2vec* mempelajari vektor kata yang memberikan informasi mengenai bagaimana kata-kata berkorelasi secara semantik ditambahkan bersamaan dengan vektor dokumen yang dipelajari dalam LDA sehingga terbentuk vektor konteks dan kemudian digunakan dalam memprediksi kata. Hasil dari pemodelan topik merupakan topik pembicaraan mengenai belanja daring pada media sosial *Twitter*, nilai koherensi yang mengukur kesamaan semantik antara kata dalam hasil pemodelan topik, proporsi topik pada korpus, probabilitas kata pada topik, serta proporsi topik dalam dokumen.

Kata kunci: pemodelan topik, LDA2vec, word2vec, LDA



ABSTRACT

TOPIC MODELING ON SOCIAL MEDIA USING LDA2VEC

(Case study: Conversations about online shopping on Twitter)

By

Dinda Asyifa K

18/424285/PA/18390

The rapid development of technology and the increasing number of internet users have made activities in all aspects of life shift to digital, including the economic aspect. The increase in online shopping activities makes business actors continue to compete in covering consumers by paying attention to their needs and behavior which can be reflected on social media. To find hidden topics in consumer conversations on social media, this final project discusses topic modeling regarding online shopping conversations on Twitter using the LDA2Vec. LDA2vec is a hybrid method between Latent Dirichlet Allocation (LDA) and word2vec. The word embedding method word2vec which learns word vectors that provide information about how semantically correlated words are added together with the document vectors studied in LDA to form context vectors and then used to predict words. The results of the topic modeling are topics of conversation about online shopping on Twitter social media, coherence score that measure the semantic similarity between words in the topic modeling results, the proportion of topics in the corpus, the probability of words on the topic, and the proportion of topics in the document.

Keyword: topic modeling, LDA2Vec, word2vec, LDA