

Abstrak

Perkembangan teknologi informasi dan komunikasi memberikan dorongan perubahan sektor pendidikan menuju pembelajaran berbasis digital melalui platform teknologi pendidikan (*edutech*). Salah satu platform edutech terbesar yang ada di Indonesia adalah Ruangguru, yang telah digunakan oleh jutaan pengguna dan memperoleh banyak ulasan dari penggunanya di media sosial, khususnya X (twitter). Ulasan tersebut menggambarkan pengalaman dan persepsi pengguna terhadap kualitas layanan, namun bersifat tidak terstruktur dan sulit untuk dianalisis secara manual karena memiliki jumlah yang besar. Permasalahan penelitian ini adalah bagaimana mengidentifikasi topik utama yang dibahas dalam ulasan pengguna aplikasi Ruangguru secara objektif dan efisien. Penelitian ini memiliki tujuan untuk menentukan jumlah topik yang muncul dalam ulasan pengguna aplikasi Ruangguru di media sosial (twitter). Metode yang digunakan adalah *Latent Dirichlet Allocation* (LDA) sebagai pendekatan pemodelan topik berbasis *machine learning* yang bersifat tidak terawasi (*unsupervised learning*). Data penelitian berjumlah 3.282 ulasan pengguna berbahasa Indonesia yang dikumpulkan dari bulan Januari hingga Oktober 2025. Hasil penelitian diharapkan dapat memberikan gambaran mengenai tema yang dominan dalam ulasan pengguna sehingga dapat menjadi bahan evaluasi bagi pengembang aplikasi serta kontribusi akademis dalam bidang *text mining*.

Kata kunci: pemodelan topik, *Latent Dirichlet Allocation*, ulasan pengguna, Ruangguru, media sosial X.

Abstract

The development of information and communication technology has driven change in the education sector towards digital-based learning through educational technology (edutech) platforms. One of the largest edutech platforms in Indonesia is Ruangguru, which has been used by millions of users and has received numerous reviews from its users on social media, particularly X (Twitter). These reviews describe users' experiences and perceptions of service quality, but they are unstructured and difficult to analyze manually due to their large number. The research problem is how to identify the main topics discussed in Ruangguru app user reviews objectively and efficiently. This study aims to determine the number of topics that appear in Ruangguru app user reviews on social media (Twitter). The method used is Latent Dirichlet Allocation (LDA) as an unsupervised machine learning-based topic modeling approach. The research data consists of 3,282 Indonesian-language user reviews collected from January to October 2025. The results of this study are expected to provide an overview of the dominant themes in user reviews, which can be used as evaluation material for application developers and as an academic contribution in the field of text mining.

Keywords: topic modeling, *Latent Dirichlet Allocation*, user reviews, Ruangguru, social media X.

Daftar Isi

HALAMAN JUDUL	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN ORISINALITAS	iv
MOTTO DAN PERSEMBAHAN	v
KATA PENGANTAR	vi
Abstrak	viii
Abstract.....	ix
Daftar Isi	x
Daftar Gambar	xiii
Daftar Tabel	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.2.1 Permasalahan	2
1.2.2 Metode Usulan.....	3
1.2.3 Pertanyaan Penelitian	3
1.3 Tujuan dan Manfaat	3
1.3.1 Tujuan Penelitian	3
1.3.2 Manfaat Penelitian.....	3
1.4 Batasan Masalah	4
1.5 Sistematika Penulisan	4
BAB II KAJIAN PUSTAKA	6
2.1 Text Mining.....	6
2.1.1 Scrapping Data	6
2.1.2 Preprocessing.....	6
2.2 Pemodelan Topik.....	7
2.3 Metode LDA (<i>Latent Dirichlet Allocation</i>)	7
2.4 Aplikasi Ruangguru	8
2.5 Evaluasi Model	9
2.6 Penelitian Terkait	9
BAB III METODE PENELITIAN	18

3.1 Alur Penelitian	18
3.2 Pengumpulan Data	19
3.3 Preprocessing	19
3.3.1 Case Folding	19
3.3.2 Cleaning	20
3.3.3 Normalisasi	20
3.3.4 Tokenizing	21
3.3.5 Stopword Removal	22
3.3.6 Stemming	22
3.4 Proses <i>Bigram</i> dan <i>Trigram</i>	23
3.5 Ekstraksi Fitur <i>Bag of Word</i> (BoW)	23
3.6 Pemodelan Topik Menggunakan <i>Latent Dirichlet Allocation</i> (LDA)	24
3.7 Skenario Pengujian	24
3.8 Evaluasi Menggunakan <i>Coherence Score</i>	26
3.9 Interpretasi Topik	26
BAB IV HASIL DAN PEMBAHASAN	27
4.1 Deskripsi Data	27
4.2 Preprocessing Data	27
4.2.1 Case Folding	27
4.2.2 Cleaning	28
4.2.3 Normalisasi	29
4.2.4 Tokenizing	29
4.2.5 Stopword	30
4.2.6 Stemming	30
4.3 Pembentukan N-gram (Unigram, Bigram, dan Trigram)	31
4.3.1 Unigram	31
4.3.2 Bigram	32
4.3.3 Trigram	32
4.4 Bag of Word	33
4.4.1 BoW unigram	33
4.4.2 BoW bigram	34
4.4.3 BoW trigram	35
4.5 Skenario Pengujian	36
4.5.1 Skenario 1	36

4.5.2 Skenario 2	37
4.5.3 Skenario 3	38
4.5.4 Skenario 4	39
4.5.5 Skenario 5	41
4.5.6 Perbandingan Skenario Pengujian	42
4.6 Hasil Pemodelan Topik dengan Latent Dirichlet Allocation	42
4.7 Evaluasi Menggunakan <i>Coherence Score</i>	43
4.8 Visualisasi	44
4.9 Interpretasi Topik	45
4.10 Tampilan Sistem	64
BAB V PENUTUP	70
5.1 Kesimpulan	70
5.2 Saran	71
REFERENSI	72
LAMPIRAN	75

Daftar Gambar

Gambar 2. 1 Metode LDA.....	8
Gambar 3. 1 Alur Penelitian	18
Gambar 4.2 1 Hasil Case Folding.....	28
Gambar 4.2 2 Hasil Case Folding.....	28
Gambar 4.2 3 Hasil Normalisasi	29
Gambar 4.2 4 Hasil Tokenizing	30
Gambar 4.2 5 Hasil Stopword	30
Gambar 4.2 6 Hasil Stemming	31
Gambar 4.3 1 Hasil Unigram.....	31
Gambar 4.3 2 Hasil Bigram.....	32
Gambar 4.3 3 Hasil Trigram	32
Gambar 4.4 1 Diagram BoW Unigram.....	33
Gambar 4.4 2 Diagram BoW Bigram	34
Gambar 4.4 3 Diagram BoW Trigram	35
Gambar 4.8 1 Visualisasi Menggunakan pyLDAvis.....	44