

Pemodelan Topik Pengguna Twitter Mengenai Aplikasi “Ruangguru”

Topic Modeling for Twitter Users Regarding the "Ruangguru" Application

Bagus Wicaksono Arianto^{*}, Gangga Anuraga
 Department of Statistics, Faculty of Mathematics and Natural Sciences,
 PGRI Adi Buana Surabaya
^{*}E-mail: baguswa@gmail.com

ABSTRACT

PT Ruang Raya Indonesia ("Ruangguru") is the largest and most comprehensive technology company in Indonesia that focuses on education-based services. In 2019 there were 15 million Ruangguru users and 300.00 teachers who had joined and were present in 32 provinces in Indonesia. It prepared a number of expansion strategies to become a company valued at more than US \$ 1 billion in the next year or two. The purpose of this research is to classify the opinions of Ruangguru users about the services provided so that it can be an evaluation material in improving their services using the latent dirichlet allocation method. The data used comes from a collection of tweets of Twitter users in Indonesia using the Twitter API. The Twitter account used in this study is @ruangguru. The results of the analysis showed that the public perception of Twitter users by using latent dirichlet allocation was formed into 28 topics.

Keywords: latent dirichlet allocation, ruangguru, twitter.

PENDAHULUAN

Ruangguru merupakan aplikasi digital asli Indonesia yang berbasis pendidikan. Berdiri sejak tahun 2014 hingga tahun 2019 telah mendapat 15 juta pengguna serta lebih dari 300 ribu mitra guru. Ruangguru hadir di hampir seluruh provinsi di Indonesia. Berbagai penghargaan telah diraih seperti UNICEF Young Innovation forum 2015 dan bubu awards 2016 kategori Indonesia Best education web. Ruangguru berkeinginan untuk menjadi mitra bagi pemerintah demi memberikan pendidikan berkualitas serta berkolaborasi dengan berbagai pihak guna mencapai tujuan meningkatnya kualitas guru, maka mutu pendidikan di Indonesia. Ruangguru menyiapkan sejumlah strategi ekspansi untuk menjadi unicorn atau perusahaan bernilai lebih dari US\$ 1 miliar dalam setahun-dua tahun mendatang. Dengan semakin banyak pengguna serta mitra guru peneliti ingin mengelompokkan persepsi pengguna ruangguru yang berada di media sosial twitter mengenai aplikasi ruangguru menggunakan metode *Latent Dirichlet Allocation* (LDA).

Topic modeling atau pemodelan topik merupakan metode *clustering* yang termasuk dalam *unsupervised learning*. Dalam *unsupervised learning* tidak ada label untuk suatu objek. Terdapat 3 tipe *clustering* yaitu *hard clustering*, *hierarchical clustering*, dan *soft/fuzzy clustering*. Pemodelan topik termasuk dalam *soft/fuzzy clustering* yang

mana setiap objek dapat dimiliki lebih dari satu *cluster* dengan tingkat tertentu (Doig, 2015). Terdapat beberapa teknik yang dapat digunakan untuk pemodelan topik, salah satunya adalah model LDA. LDA adalah metode yang digunakan untuk melakukan analisis pada dokumen yang sangat besar. LDA juga digunakan untuk meringkas, melakukan pengelompokan, menghubungkan maupun memproses data.

Penelitian yang pernah dilakukan mengenai *topic modeling* oleh Utami (2017) analisis topik data media sosial twitter menggunakan model topik *latent dirichlet allocation* memperoleh hasil bahwa pemodelan topik LDA pada kelima lokasi *tweet* di Kota Bogor dan rentang waktu tertentu berhasil membentuk topik dengan informasi atau deskripsi topik untuk setiap lokasi *tweet*. Penelitian lain oleh Al-khairi, *et al.*, (2018) mengenai deteksi topik fashion pada twitter dengan latent dirichlet allocation diperoleh hasil jumlah topik yang paling optimal adalah 20 topik. Penelitian lain oleh Kurniawan (2018) mengenai sistem monitoring percakapan pada toko online menggunakan metode LDA studi kasus: toko online “berrybenka.com” diperoleh hasil jumlah topik yang paling optimal adalah 10 topik.

Penelitian ini ingin mengelompokkan *tweet* masyarakat mengenai layanan ruangguru dengan LDA. Perbedaan penelitian ini dengan sebelumnya adalah iterasi yang digunakan dan penentuan kelompok yang optimum



Gambar 2. Hasil *Latent Dirichlet Allocation* (LDA)

Gambar 2 menunjukkan bahwa hasil *topic modeling* menggunakan algoritma *latent dirichlet allocation* didapatkan sebanyak 28 topik yang terbentuk. Gambar 2 menunjukkan bahwa beberapa topik saling beririsan berarti topik-topik tersebut memiliki beberapa kata yang sama. Topik 5 dan topik 8 yang memiliki kata yang sama seperti kata ruangguru dan bimbel daring. Selanjutnya akan dibahas lima topik terbesar yang sudah terbentuk dari hasil analisis pemodelan topik.

1. Topik 0 LDA

Hasil analisis topik pemodelan menggunakan algoritma LDA pada topik 0 ditunjukkan pada Tabel 3.

Tabel 3. Topik 0

Persentase	Kata	Peluang	
Topik 0	8.5	dapat_diskon_beli_pakai	0.026
		Terus	0.023
		hi_guna_kode_berikut	0.022
		Lagi	0.021
		Bersama	0.018
		Belajar	0.017
		Setelah	0.017
		dapat_diskon_ambassador	0.017
		Ambassador	0.016
		beli_pakai	0.016

Tabel 3 menunjukkan bahwa topik 0 mewakili sekitar 8,5 persen dari keseluruhan

topik yang terbentuk. Topik 0 terdapat beberapa kata seperti “dapat_diskon”, “beli pakai”, “belajar”. Kata yang memiliki peluang muncul terbesar pada topik 0 adalah *dapat_diskon_beli_pakai* dengan nilai sebesar 0.026. Topik 0 berisi tentang informasi mengenai diskon ruangguru dengan menggunakan kode yang tersedia. Topik 0 juga berisi tentang informasi mengenai pendapat masyarakat mengenai rekomendasi ambassador baru bagi ruangguru.

2. Topik 1 LDA

Hasil analisis topik pemodelan menggunakan algoritma LDA pada topik 1 ditunjukkan pada Tabel 4.

Tabel 4. Topik 1

Persentase	Kata	Peluang	
Topik 1	4.8	Baru	0.029
		Brand	0.026
		Begitu	0.023
		Tidak	0.021
		Kelas	0.02
		Mau	0.02
		jadi	0.02
		ruang_guru	0.018
		zenius	0.017
		edc	0.017

Tabel 4 menunjukkan bahwa topik 1 mewakili sekitar 4,8 persen dari keseluruhan topik yang terbentuk. Kata “baru” memiliki

peluang terbesar dalam topik 1 yaitu sebesar 0.029. Topik 1 juga terdapat beberapa kata seperti “brand”, “mau”, “ruangguru”. Topik 1 menunjukkan persepsi masyarakat yang masih bingung memilih aplikasi ruangguru atau zenius khususnya yang belum pernah menggunakan bimbingan belajar *online*.

3. Topik 2 LDA

Hasil analisis topik pemodelan menggunakan algoritma LDA pada topik 2 ditunjukkan pada Tabel 5.

Tabel 5. Topik 2

Persentase	Kata	Peluang	
Topik 2	4.6	ayo	0.041
		kenapa	0.032
		sama	0.018
		melalui_dm	0.018
		guru	0.017
		promo	0.015
		banyak	0.015
		tv	0.014
		harus	0.014
		baru	0.014

Tabel 5 menunjukkan bahwa topik 2 mewakili sekitar 4,6 persen dari keseluruhan topik yang terbentuk. Kata “ayo” memiliki peluang terbesar dalam topik 5 yaitu sebesar 0.041. Topik 2 juga terdapat beberapa kata seperti “kenapa”, “melalui_dm”, “promo”. Topik 2 menunjukkan promo ruangguru di berbagai media massa terutama media televisi.

4. Topik 3 LDA

Hasil analisis topik pemodelan menggunakan algoritma LDA pada topik 3 ditunjukkan pada Tabel 6.

Tabel 6. Topik 3

Persentase	Kata	Peluang	
Topik 3	4.4	Sangat	0.063
		Ya	0.05
		Share	0.038
		Sbmptn	0.031
		Atau	0.03
		Quipper	0.028
		Tolong	0.025
		Saja	0.024
		Jadi	0.024
		mohon_	0.022
		maaf	

Tabel 6 menunjukkan topik 3 mewakili sekitar 4,4 persen dari keseluruhan topik yang

terbentuk. Topik ke 4 terdapat beberapa kata seperti kata “sangat”, “sbmptn”, “quipper” dan lain-lain. Topik 3 menunjukkan permintaan masyarakat untuk saling berbagi akun bimbingan belajar daring.

5. Topik 4 LDA

Hasil analisis topik pemodelan menggunakan algoritma LDA pada topik 4 ditunjukkan pada Tabel 7.

Tabel 7. Topik 4

Persentase	Kata	Peluang	
Topik 4	4.4	Dari	0.037
		Coba	0.037
		sama	0.031
		juga_bisa	0.027
		kalau	0.022
		tidak	0.021
		Bisa	0.019
		Edc	0.018
		Pendidikan	0.018
		Ya	0.017

Tabel 7 menunjukkan bahwa topik 4 mewakili sekitar 4,4 persen dari keseluruhan topik yang terbentuk. Topik ke 4 terdapat beberapa kata seperti kata “coba”, “juga_bisa”, “edc” dan lain-lain. Topik 4 merupakan pertanyaan apakah bisa 1 akun digunakan dalam 2 *handphone* secara bersamaan.

KESIMPULAN

Metode pengklasteran LDA pada aplikasi ruangguru mengelompokkan data twitter menjadi 28 buah topik dengan topik yang sering diperbincangkan adalah diskon ruangguru.

DAFTAR PUSTAKA

- Al-khairi, Y U. Yudi W., & Budi L. P. 2018. *Deteksi Topik Fashion Pada Twitter Dengan Latent Dirichlet Allocation. Jurnal Aplikasi dan Teori Ilmu Komputer*. 1(1): 1-8.
- Doig C. 2015. *Introduction to topic modeling in python*. <http://chdoig.github.io/pytexas2015-topic-modeling> [1 November 2019].
- Kurniawan, W. 2018. *Sistem Monitoring Percakapan Pada Toko Online Menggunakan Metode Latent Dirichlet Allocation Studi Kasus: Toko Online “BERRYBENKA.COM”* [Skripsi]. Sleman:

- Universitas Islam Indonesia.
Ruangguru. (2017). *Ruangguru*.
https://ruangguru.com/general/about?utm_source=bimbelrg&utm_medium=referral&utm_campaign=footer [25 Oktober 2019]
- Utami, KP. 2017. *Analisis Topik Data Media Sosial Twitter Menggunakan Model Topik Latent Dirichlet Allocation* [Skripsi]. Bogor: Institut Pertanian Bogor.

