Jurnal Elektronik Ilmu Komputer Udayana
Volume 12, No 4. May 2024

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Klusterisasi Fitur Tanya Dokter Pada Website Alodokter Menggunakan Metode Latent Dirichlet Allocation

I Putu Fajar Tapa Mahendra[a1], I Gusti Ngurah Anom Cahyadi Putra [a2]

[a]Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Udayana
Kuta Selatan, Badung, Bali, Indonesia
[1]ftapamahendra@gmail.com
[2] anom.cp@unud.ac.id

Abstract

Digital advancements have change information-seeking behaviors, particularly in health inquiries for the people. The Alodokter website's "Tanya Dokter" feature facilitates an easy connections with medical experts to ask question regarding health. The posed questions tend to mirror evolving health trends and public misunderstandings regarding health issues. Manual analysis of data in "Tanya Dokter" features proves challenging, prompting the use of Latent Dirichlet Allocation (LDA) topic modeling. This research categorizes Alodokter topics, unveiling common health issues. The optimal model reveals 8 clusters with diverse topic distributions. Validation metrics using coherence score with 0.258481 as the highest value affirm the model's efficacy. Optimal outcomes stem from combination of parameter such as 8 topics, alpha 0.02, and beta 0.02. This study may offers Alodokter and healthcare providers an informed perspective on an accessible approach to categorize health questions effectively using Topic Modelling Latent Dirichlet Allocation.

Keywords: *Alodokter, Tanya Dokter, Topic Modelling, Latent Dirichlet Allocation, Health*

1. Pendahuluan

Perkembangan teknologi digital dimasa saat ini telah mengubah cara orang-orang dalam mencari informasi. Informasi yang dulunya hanya didapatkan ketika membeli dan membaca suatu buku kini dapat diakses dengan mudah melalui internet. Begitu juga dalam pencarian informasi yang membutuhkan pendapat seorang ahli. Dengan bantuan teknologi digital, informasi dari pakar yang dulunya membutuhkan kita untuk bertemu langsung dengan pakarnya, kini dapat dilakukan secara online melalui suatu platform. Salah satu contoh platform yang dimana masyarakat dapat bertanya kepada seorang pakar atau ahli adalah Alodokter dengan fiturnya yaitu Tanya Dokter.

Melalui Tanya Dokter, masyarakat dapat dengan mudah bertanya mengenai masalah kesehatan yang mereka alami dengan pakar terkait. Banyaknya pertanyaan yang dilontarkan pada website ini dapat mencerminkan dinamika kesehatan masyarakat, khususnya masyarakat Indonesia. Tidak berhenti sampai disitu, pertanyaan yang dilontarkan juga dapat mencerminkan miskonsepsi tentang kesehatan pada masyarakat serta kesenjangan pengetahun terkait kesehatan. Kemampuan dalam menganalisa data pada Tanya Dokter dapat memberikan pemahaman lebih mendalam mengenai tren kesehatan yang dialami masyarakat serta membantu dalam menyesuaikan kontern terkait yang dibutuhkan masyarakat.

Namun mengingat banyaknya data yang ada pada fitur Tanya Dokter teserbut, menganalisa data secara manual akan menjadi tantangan yang sangat sulit. Metode analisis secara tradisional, sering kali tidak cocok dan tidak efisien untuk mendapatkan informasi yang berarti dari analisa data pada fitur Tanya Dokter. Untuk membantu dalam mengatasi masalah tersebut, kita dapat menggunakan metode *topic modelling* untuk membantu dalam mengekstrak topik inti pada data Tanya Dokter. Salah satu metode *topic modelling* yang dapat digunakan adalah *Latent Dirichlet Allocation*.

Topic modelling secara statistic bekerja dengan cara mengeksplorasi dokumen yang diberikan dan merepresentasikan mereka sebagai kumpulan istilah yang sering muncul bersamaan dalam dokumen [1]. Menurut Gurcan, LDA pada topic modelling adalah suatu metode pendekatan yang generatif yang digunakan untuk menemukan pola semantic yang ada pada suatu korpus dokumen yang relatif tidak terstruktur [2]. Menggunakan LDA, pendekatan secara sistematik dapat dilakukan untuk menyaring teks

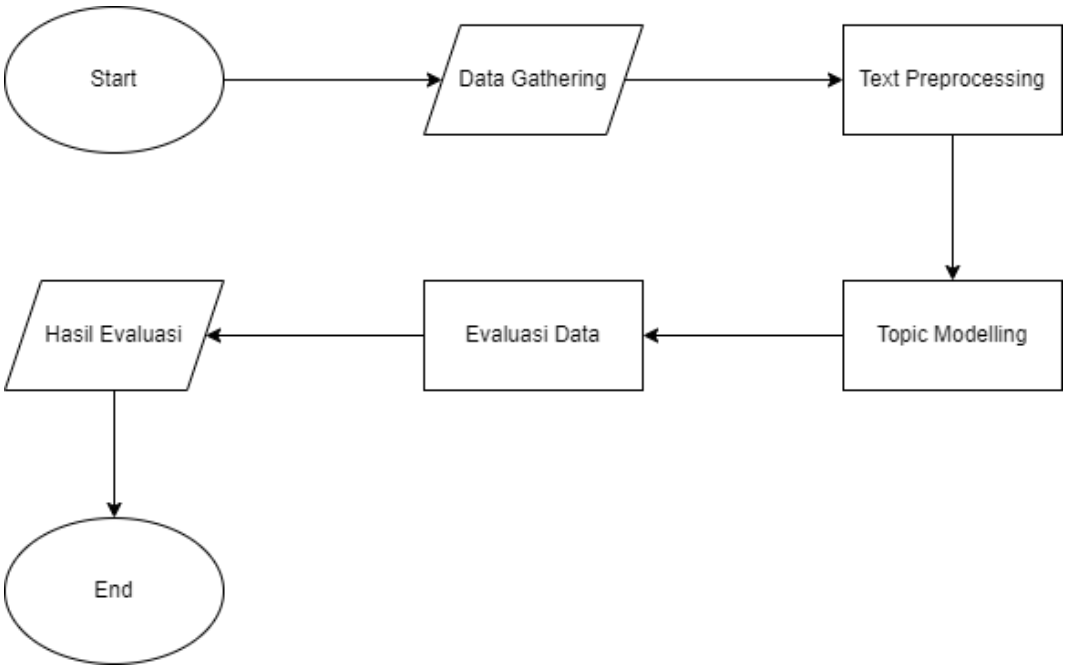
menjadi topik yang dapat dikenali, memungkinkan dilakukannya identifikasi tren kesehatan yang ada pada fitur Tanya Dokter. Dengan menggunakan LDA, akan diasumsikan terdapat distribusi berbagai macam topik pada suatu kumpulan dokumen, dimana setiap dokumen direpresentasikan sebagai sebaran dari topik dan setiap topik sebagai distribusi kata-kata di dalam dokumen [3]. Mengunakan sebaran topik yang dihasilkan, akan direpresentasikan dalam bentuk probabilitas dengan probabilitas tertinggi yang akan menjadi faktor utama dalam menentukan termasuk klsuter mana suatu data teks tersebut.

Melalui penelitian ini, peneliti berharap dengan menggunakan metode *topic modelling Latent Dirichlet Allocation* ini dapat membantu dalam membuat suatu model yang dapat menunjukan masalah kesehatan yang sering muncul menggunakan klasterisasi pada data dari website Alodokter pada fitur Tanya Dokter. Dengan melakukan hal tersebut peneliti berharap dapat menyediakan Alodokter serta komunitas penyedia layanan kesehatan dengan pandangan yang medalam untuk membentuk pendekatan yang lebih terinformasi dan ramah kepada pengguna aplikasi atau platform kesehatan digital atau membantu pengguna yang masih bingung dalam menentukan topik yang ingin ditanyakan pada website Alodokter dalam mengkategorikan pertanyaan mereka sehingga dapat ditangani oleh ahli yang tepat.

2. Metode Penelitian

2.1. Alur Penelitian

Berikut adalah alur dari penelitian yang dilakukan:



Gambar 1. Alur Penelitian

2.2. Pengumpulan Data

Pengumpulan data teks pada fitur Tanya dokter pada website Alodokter dilakukan dengan menggunakan teknik *web scrapping*. Dalam proses pengumpulan datanya, digunakan bantun *library selenium* yang mana *library selenium* ini dapat diakses melalui bahasa pemrograman *Python*. Pengambilan data dilakukan dalam 10 tahap pengulangan dimana setiap tahap mengambil sekitar 120-135 data dan langsung disimpan dalam bentuk *.tsv (tab separated value)*. Hal ini dilakukan untuk mengurangi resiko kehilangan data yang sudah terambil pada saat pengambilan yang diakibatkan baik oleh koneksi internet maupun masalah pada perangkat keras.dari data yang telah didapatkan tersebut, data yang sebelumnya terbagi menjadi 10 bagian kemudian disatukan dan kemudian akan dilakukan pembersihan data duplikat. Setelah seluruh proses dilakukan, didapatkan sisa data yang berjumlah

p-ISSN: 2301-5373
e-ISSN: 2654-5101

1099. Berikut contoh data yang didapatkan menggunakan teknik *web scrapping* pada fitur Tanya Dokter yang ada pada website Alodokter dapat dilihat pada tabel 1:

Table 1 Contoh Data

No	Question	Answer
1	"Teman saya ada yg abis dikecup lehernya sama cowoknya sampe masih ada bekasnya dok, dia malu, cara apa untuk menghilangkan bekas kecupan di lehernya dok?"	['Alo, terimakasih atas pertanyaannya.\nKecupan yang cukup kuat di area leher memang bisa menyisakan bekas yang cukup lama hilangnya, misalnya berupa ruam kemerahan, memar, lecet, dan sebagainya. Tergantung keparahannya, bekas kecupan ini bisa memerlukan waktu beragam untuk sembuh, bisa singkat, bisa juga lambat. Saran kami, coba Anda arahkan rekan Anda untuk:\nMandi yang rajin dan bersihkan lehernya dengan baik\nKompres dingin bekas kecupan di lehernya\nOleskan pelembab ke area kulit lehernya\nJangan berlebihan menggosok atau memanipulasi bekas kecupan di leher\nMinum banyak air putih, makan makanan yang mengandung kaya vitamin C serta antioksidan lainnya\nApabila ia merasa sangat terganggu dengan bekas kecupan di lehernya tersebut, Anda bisa arahkan ia untuk periksa langsung ke dokter ya..\nSemoga membantu.']
2	"Dok, apakah disini saya bisa minta resep dokter untuk obat batuk pilek anak? Ini untuk anak saya umur 4 tahun yg batuk pilek sudah seminggu belum sembuh dok"	['Alo, selamat siang\nbatuk pilek ialah merupakan keluhan yang sering menimpa anak-anak, dimana keluhan batuk pilek umumnya disebabkan oleh infeksi virus atau bakteri, namun paling sering akibat infeksi virus yang dimana dapat sembuh dengan sendirinya dalam waktu 7-10 hari jadi tidak perlu penanganan khusus. namun memang keluhan batuk pilek ini mengganggu aktivitas dan tidur anak. untuk membuat anak terasa nyaman dan meredakan keluhannya ada beberapa tips yang bisa bunda lakukan seperti :\npastikan anak tidur cukup\nsaat anak tidur posisikan kepala lebih tinggi jadi gunakan bantal dikepala saat tidur\nberikan anak lebih banyak air putih bila perlu hnagat\nberikan madu pada anak setengah sendok dicampur dengan teh hangat sebelum tidur\noleskan balsem khusus anak di bagian dada, leher dan punggungnya\nkonsumsi obat batuk pilek sirup khusus anak yang dapat bunda beli bebas di apotik\njika lebih dari 10 hari keluhan

		batuk pilek tidak sembuh, atau adanya kondisi demam tinggi diatas 38 derajat serta adanya sesak napas maka segera temui dokter secara langsung untuk mendapatkan penanganan lebih lanjut dengan tepat.\nsemoga dapat membantu']

2.3. Text Preprocessing

Text preprocessing adalah tahap dimana teks asli diubah dengan menghilangkan *unsure textual* yang tidak diperlukan dalam pengolah yang lebih lanjut [4]. Pada tahap ini, data teks yang dikumpulkan akan diolah agar dapat diproses pada tahap pelatihan model. Berikut adalah tahap-tahap dalam *text preprocessing*:

a. Case Folding
Case folding adalah tahap *preprocessing* dimana setiap kata pada data teks akan disetarakan format hurufnya menjadi huruf besar atau huruf kecil. Pada penelitian ini seluruh huruf akan diubah menjadi huruf kecil

b. Penghapusan tanda baca
Pada tahap ini seluruh tanda baca yang terdapat pada data teks akan dihapus. Tanda baca dihapus karena dianggap tidak memiliki nilai makna dalam ektraksi informasi pada data teks

c. Tokenization
Pada tahap *tokenization*, data teks yang ada akan dipenggal menjadi perkata dan kemudian disimpan dalam bentuk array. Hal ini dilakukan untuk mempermudah pengolahan data teks

d. Stopword removal
Pada tahap *stopword removal*, beberapa kata yang dianggap tidak memiliki makna akan dihapus. Kata-kata yang dihapus biasanya berupa kata sambung seperti “dan”, “lalu” dan “setelah”.

e. Stemming
Stemming adalah tahap dimana kata yang ada pada data teks, diubah bentuknya menjadi kata dasar. Hal ini dilakukan untuk agar sistem yang dibuat dapat memahami makna dokumen dengan lebih baik

f. Bag of word
Bag of word adalah tahap dimana kata-kata yang ada pada seluruh korpus akan dihitung frekuensi kemunculannya. *Bag of word* membantu dalam merepresentasikan data teks menjadi data angka yang lebih mudah dipahami oleh komputer

g. Term Weighting
Term weighting adalah tahap dimana setiap kata yang ada pada data teks pada seluruh korpus akan diberikan bobot. Pada pemberian bobot pada kata ini, akan digunakan metode *TF-IDF*. *TF-IDF* adalah metode pemberian bobot yang menggunakan nilai *TF (Term frequency)* yang berdasarkan frekuensi kemunculan kata pada dokumen dengan *IDF (Inverse Document Frequency)* yang bersarkan pada kemunculan kata pada suatu kumpulan dokumen. Berikut formula *TF-IDF*:

$$TF - IDF = TF * IDF \quad [1]$$

$$TF = \frac{jumlah\ kemunculan\ suatu\ kata\ (x)}{jumlah\ kata\ dalam\ dokumen} \quad [2]$$

$$IDF = \log \frac{jumlah\ dokumen}{jumlah\ dokumen\ dengan\ kata\ (x)} \quad [3]$$

2.4. Topic Modelling

Dalam *machine learning*, *topic* modelling adalah salah satu bentuk *unsupervised machine learning* yang menyaring data teks agar dapat mengidentifikasikan pola kemunculan kata yang menandakan topik yang mendasari data text tersebut [5]. Dalam topic modelling ini, setiap dokumen dalam suatu

korpus direpresentasikan sebagai kombinasi yang terdiri dari beberapa topik, sedangakan topik itu sendiri direpresentasikan sebagai kombinasi yang terdiri dari beberapa kata.

Tujuan representasi ini dilakukan untuk dapat menentukan distribusi penyebaran topik pada suatu dokumen dan distribusi kata pada suatu topik. Hal ini dilakukan karena umumnya topic modelling menggunakan pendekatan probabilistik, yang artinya saat model dari topic modelling mengatakan bahwa suatu dokumen terdiri 20% topik X, 30% topic Y dan 50% topik Z, ini mencerminkan probabilitas berdasarkan kata yang ada pada dokumen. Salah satu metode yang termasuk ke dalam kategori *topic modelling* adalah *LDA (Latent Dirichlet Allocation)*.

Latent Dirichlet Allocation adalah sebuah pendekatan yang didasarkan pada teorema definetti, dimana metode ini digunakan untuk menangkap beberapa topik yeng tersebar diantara beberapa kumpulan dokumen [5]. Konsep utama dari metode *Latent Dirichlet Allocation* ini dalam *melakukan topic modelling* adalah dengan merepresentasikan topik sebagai campuran dari beberapa topik yang berbeda, dimana topik itu sendiri direpresentasikan oleh distribusi kata pada dokumen. Berikut cara kerja dari LDA:

a. Inisiasi
   Jumlah topik yang akan diektrak pada saat topic modelling akan ditentukan pada tahap ini
b. Random Assignment
   Setelah itu, setiap kata akan dimasukan ke salah satu topik ada. Jumlah topik yang ada sesuai dengan inisiasi yang dilakukan
c. Iterative Reassignment
   Lalu akan dilakukan iterasi berkali-kali dengan aturan
   - Iterasi untuk setiap dokumen *d*
     - Iterasi kata *w* pada document *d*
     - Iterasi untuk setiap topik *t*
       - Hitung kedua probabilitas:
         - *P(t/d)*: proporsi kata pada suatu dokumen yang dimasukan pada topic *t*
         - *P(w/t)*: proporsi suatu topik *t* diberikan terhadap seluruh dokumen *d* yang berasal dari kata *w*
       - Pengkategorian ulang kata *w* kedalam topic *t* bedasarkan probabilitas yang didapatkan dari hasil *P(t/d) * P(w/t)*

d. Convergence
   Tahap dimana algoritma sudah mulai stabil dalam memberikan topik pada suatu kata setiap kali iterasi dilakukan
e. Output
   Setiap dokumen memiliki distribusi berdasarkan topik, dan setiap topik memiliki distribusi kata.

2.5. Parameter Tuning

Parameter Tuning adalah tahap dimana dilakukan beberapa modifikasi pada variabel model sehingga menghasilkan hasil yang berbeda. Tujuan dari parameter tuning ini adalah untuk membantu dalam menemukan model yang paling baik dalam melakukan *topic modelling*. Pada penelitian ini model akan dilatih ulang dengan menggunakan kombinasi parameter tuning yang berbeda. Adapun parameter yang akan dijadikan sebagai parameter tuning:

a. Jumlah topik, dari nilai 5 sampai 10
b. Nilai *alpha* yang merepresentasikan penyebaran topik dalam suatu dokumen dengan rentang nilai 0.1, 0.2, 0.3, 0.4 dan 0.5
c. Nilai *beta* yang merepresentasikan penyebaran kata yang merepresentasikan suatu topik dengan rentang nilai 0.1, 0.2, 0.3, 0.4 dan 0.5

2.6. Evaluasi Model dengan *Coherence Score*

Coherence score adalah salah satu metode evaluasi yang umum digunakan untuk mengevaluasi model topic modelling. Metrik ini menghitung konsistensi dari kata pada suatu topik untuk mengevaluasi apakah suatu topik dapat di interpretasi dengan cara menghitung kemiripan semantik dari kata yang ada pada suat topik [6]. Suatu statement atau kalimat dapat dikatakan koheren, apabila setiap katanya

saling mendukung satu sama lain [7]. Dalam evaluasi menggunakan metode *coherence score*, ada beberapa tahapan yang harus dilalui. Adapun tahapan yang dilalui:

a. Segmentasi
   Pemenggalan kumpulan kata atau kalimat menjadi kata tunggal
b. Estimasi Probabilitas
   Estimasi probabilitas dari subset menggunakan data korpus yang besar
c. Confirmation Measure
   Menghitung nilai menggunakan probabilitas untuk mengetahui indikasi seberepa mungkin suatu subset kata dilihat bersama.
d. Aggregation
   Kemudian nilai dari *confirmation measure* akan di melalui proses agregasi untuk mendapatkan satu nilai *coherence score*.

Formula *coherence score*:

$$Coherence_{NPMI} = \frac{1}{N}\Sigma_{i=1}^{N} \frac{\log\left(\frac{p(w_i, w_j) + \epsilon}{p(w_i) \times p(w_j)}\right)}{-\log(p(w_i, w_j) + \epsilon)} \quad [4]$$

$p(w_i)$ = probabilitas kata w_i mucnul dalam korpus
$p(w_j)$ = probabilitas kata w_j muncul dalam korpus
$p(w_i, w_j)$ = probabilitas 2 kata muncul bersamaan dalam suatu topik atau konteks

3. Hasil dan Pembahasan

3.1. Hasil *Topic Modelling*

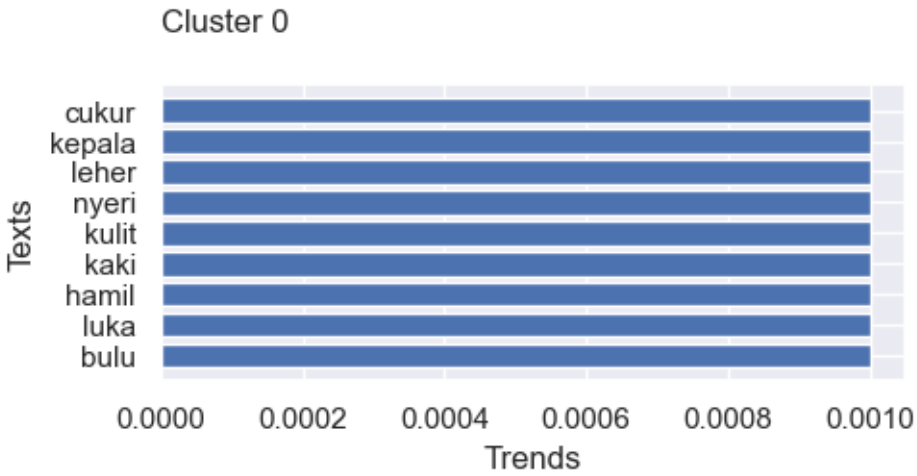
Model LDA dilatih dengan kombinasi paremeter tuning yang disebutkan diatas dengan jumlah 150 kombinasi. Dari seluruh kombinasi yang ada, didapatkan kombinasi 8 topik, 0.2 *alpha* dan 0.2 *beta* sebagai kombinasi terbaik dengan nilai *coherence score* 0.258481. Berikut distibusi topik pada kluster LDA dari model terbaik dapat dilihat pada tabel 2:

Tabel 2. Hasil Kluster

Kluster Topik	Kata
Kluster 0	bulu, luka, hamil, kaki, kulit, nyeri, leher, kepala, cukur, kolesterol
Kluster 1	hamil, kulit, anak, mata, gatal, kaki, tidur, guna, ginjal, wajah
Kluster 2	kulit, hamil, rambut, wajah, vagina, makan, anak, infeksi, gatal, luka
Kluster 3	gigi, luka, nyeri, sakit, batuk, demam, makan, kulit, mata, obat
Kluster 4	rambut, telinga, makan, darah, konsumsi, kulit, obat, batuk, sakit, ganggu
Kluster 5	mata, kulit, urat, hamil, makan, asam, darah, menstruasi, nyeri, sakit
Kluster 6	payudara, kaki, hamil, gigi, kulit, ibu, makan, jerawat, bayi, nyeri

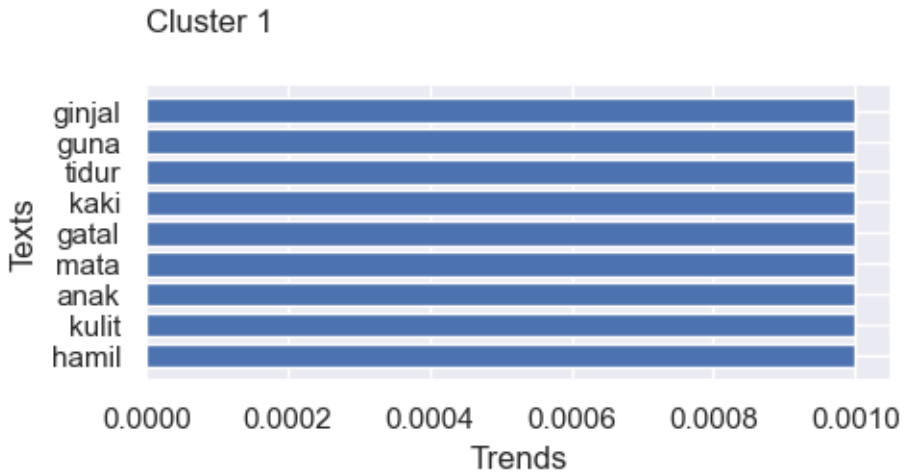
p-ISSN: 2301-5373
e-ISSN: 2654-5101

Berikut visualisasi dari hasil klusterisasi beserta bobot dari katanya:



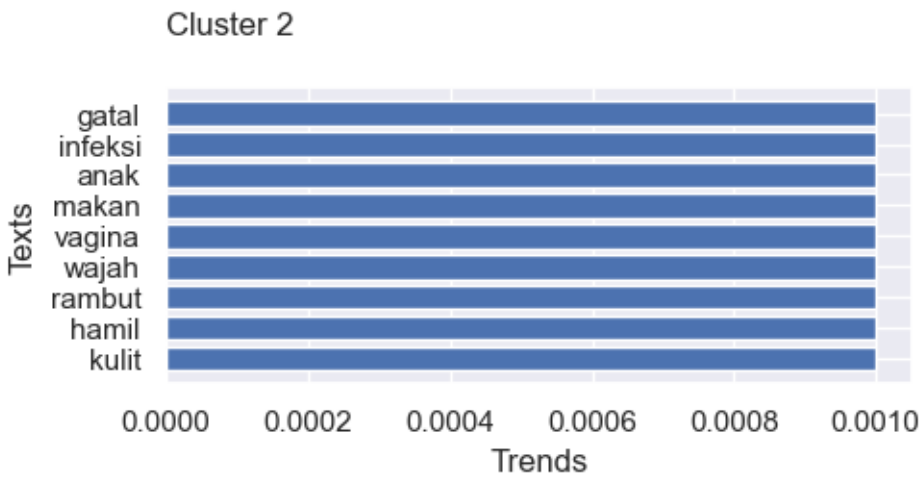
Gambar 2. Kluster 0

Pada gambar 2 dapat dilihat kluster 0 terbentuk dari kata “cukur”, “kepala”, “leher”, “nyeri”, “kulit”, “kaki”, “hamil”, “luka”, dan “bulu”. Pada kluster ini, seluruh kata memiliki bobot trend yang sama.



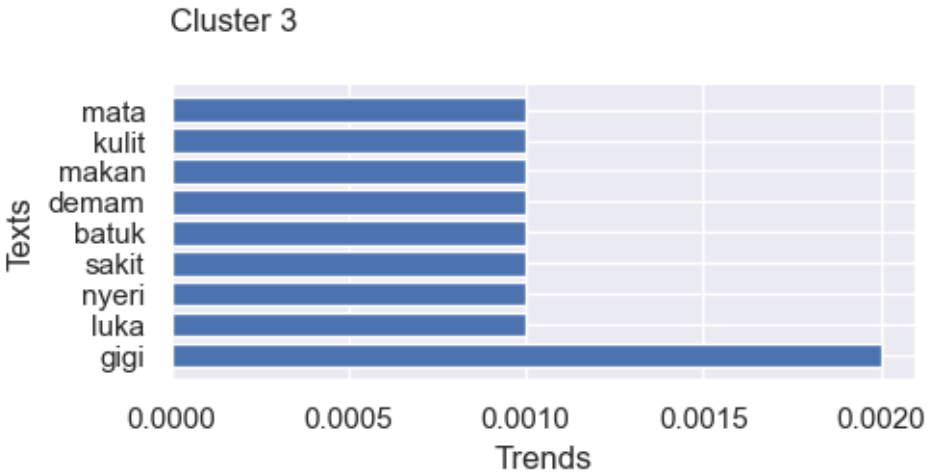
Gambar 3. Kluster 1

Pada gambar 3 dapat dilihat kluster 1 terbentuk dari kata “ginjal”, “guna”, “tidur”, “kaki”, “gatal”, “mata”, “anak”, “kulit” dan “hamil. Pada kluster ini, seluruh kata pada kluster memiliki bobot trend yang sama.



Gambar 4. Kluster 2

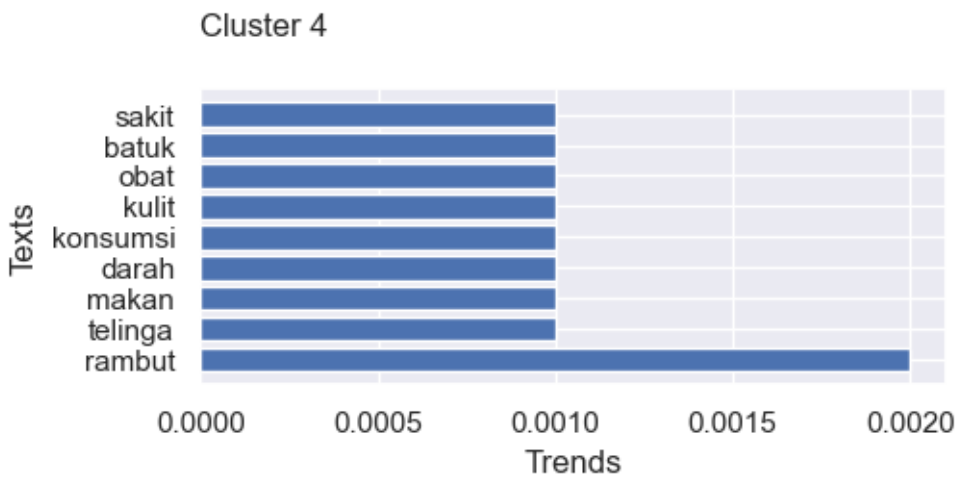
Pada gambar 4 dapat dilihat kluster 2 terbentuk dari kata “gatal”, “infeksi”, “anak”, “makan”, “vagina”, “wajah”, “rambut”, “hamil” dan “kulit”. Pada kluster 2, setiap kata pembentuknya memiliki bobot yang sama



Gambar 5. Kluster 3

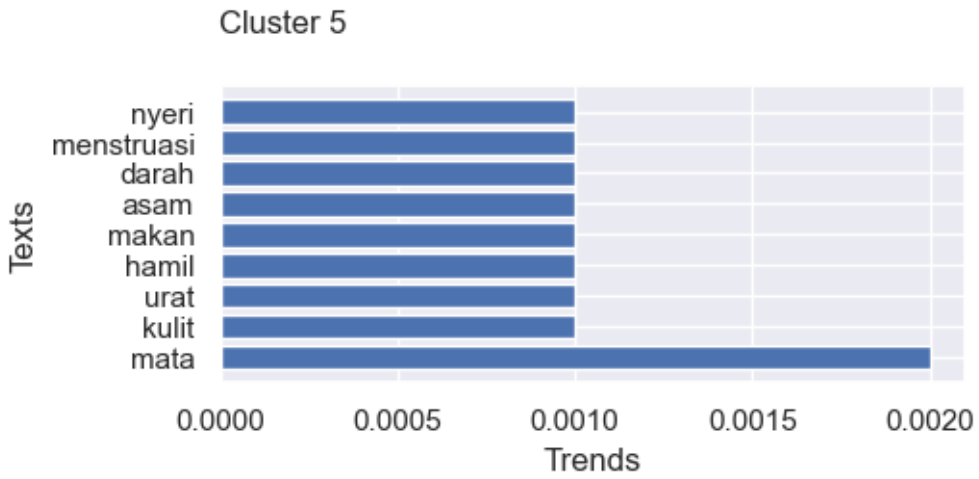
Pada gambar 5 dapat dilihat kata pembentuk dari kluster 3, yaitu “mata”, “kulit”, “makan”, “demam”, “batuk”, “sakit”, “nyeri”, “luka” dan “gigi”. Pada kluster 3 ini, kata “gigi” memiliki bobot yang paling tinggi

p-ISSN: 2301-5373
e-ISSN: 2654-5101



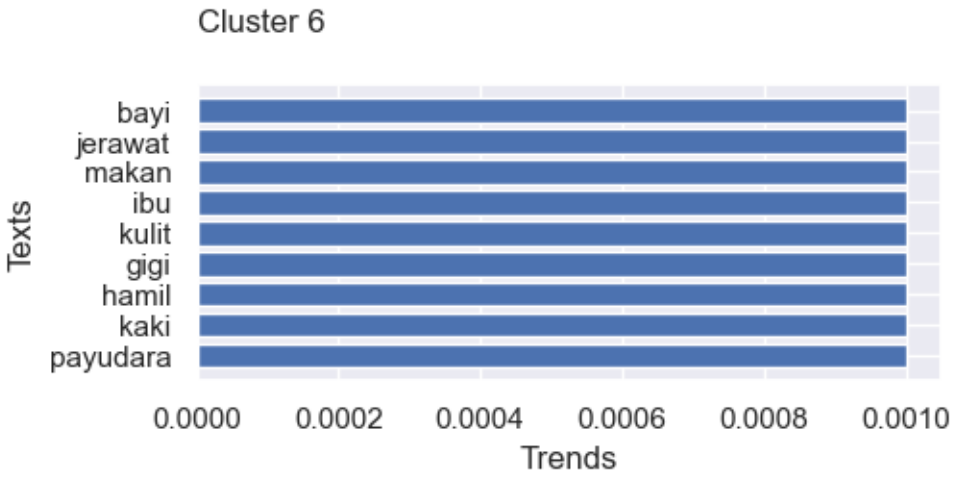
Gambar 6. Kluster 4

Pada gambar 6 dapat dilihat kata pembentuk dari kluster 4, yaitu “sakit”, “batuk”, “obat”, “kulit”, “konsumsi”, “darah”, “makan”, “telinga” dan “rambut”. Pada kluster 4 ini, kata “rambut” memiliki bobot yang paling tinggi



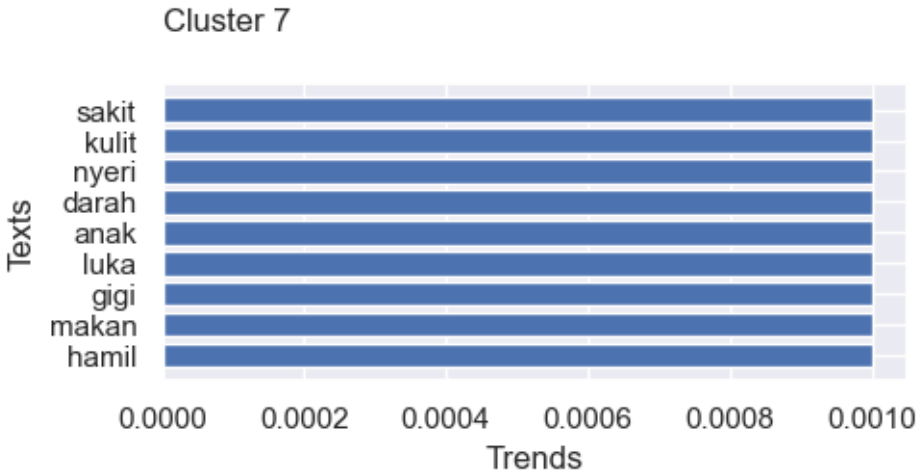
Gambar 7. Kluster 5

Pada gambar 7 dapat dilihat kata pembentuk dari kluster 5, yaitu “nyeri”, “menstruasi”, “darah”, “asam”, “makan”, “hamil”, “urat”, “kulit” dan “mata”. Pada kluster 5 ini, kata “mata” memiliki bobot yang paling tinggi



Gambar 8. Kluster 6

Pada gambar 8 dapat dilihat kata pembentuk dari kluster 6, yaitu “bayi”, “jerawat”, “makan”, “ibu”, “kulit”, “gigi”, “hamil”, “kaki” dan “payudara”. Pada kluster 6 ini seluruh kata pembentuk kluster 6 ini memiliki bobot yang sama.



Gambar 9. Kluster 7

Pada gambar 9 dapat dilihat kata pembentuk dari kluster 7, yaitu “sakit”, “kulit”, “nyeri”, “darah”, “anak”, “luka”, “gigi”, “makan” dan “hamil”. Pada kluster 7 ini seluruh kata pembentuknya memiliki bobot yang sama.

4. Conclusion

Berdasarkan penelitian yang telah dilakukan didapatkan hasil bahwa, dari 150 model yang dilatih berdasarkan kombinasi antara jumlah topik nilai *alpha* dan nilai *beta*, model terbaik adalah model dengan jumlah topik 8, nilai *alpha* 0.2 dan nilai *beta* 0.2. Model terbaik ditentukan dengan mencari nilai *coherence score* terbaik diantara model yang lain. Pada kombinasi yang disebutkan tadi, didapatkan nilai *coherence score* berupa 0.258481. Dari nilai coherence score tersebut dapat dikatakan bahwa topic modelling menggunakan metode Latent Dirichlet Allocation masih memiliki kekurangan. Hal tersebut juga dapat dilihat pada visualisasi data pada setiap kluster yang ada dimana pada beberapa kluster bobot kata pembentuk kluster memeliki bobot yang sama. Peneliti menduga hal ini disebabkan oleh kurangnya dataset serta luasnya persebaran topik pada fitur Tanya Dokter dari website Alodokter

p-ISSN: 2301-5373
e-ISSN: 2654-5101

Refrensi

[1] N. A. Tresnasari, T. B. Adji dan A. E. Permanasari, “Social-Child-Case Document Clustering based on Topic Modeling using Latent Dirichlet Allocation” *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 14, no. 2, p. 179-188, 2020
[2] F. Gurcan, O. Ozyurt, dan N. Cagitay, “Investigation of Emerging Trends in the E-Learning Field Using Latent Dirichlet Allocation” *International Review of Research in Open and Distributed Learning*, vol. 22, no. 2, p. 1–18, 2021
[3] K. Porter, “Analyzing the DarkNetMarkets subreddit for evolutions of tools and trends using LDA topic modeling” *Digital Investigation*, vol. 26, p. S87- S97, 2018
[4] H. Najjichah, A. Syukur, dan H. Subagyo, “Pengaruh Text Preprocessing dan Kombinasinya pada Peringkas Dokumen Otomatis Teks Berbahasa Indonesia” *Jurnal Teknologi Informasi*. vol. 15, no. 1, p. 1-11. 2019
[5] P. Kherwa dan P. Bansal, “Topic Modeling: A Comprehensive Review” *ICST Transactions on Scalable Information Systems*, vol. 7, no. 24, 2018
[6] H. Rahimi, J. L. Hoover, D. Mimno, H. Naacke. C. Constantin dan B. Amann, “Contextualized Topic Coherence Metrics”, *arXiv*:2305.14587, 2023
[7] S. K. Ray, A. Ahmad, dan C. A. Kumar, “Review and Implementation of Topic Modeling in Hindi” *Applied Artificial Intelligence*, vol. 33, no. 11, p. 979-1007, 2019

This page is intentionally left blank.