

Pemodelan Topik Risiko Bunuh Diri berdasarkan Konten Media Sosial dengan Latent Dirichlet Allocation

Khairul Walady Putra¹, Putra Pandu Adikara², Budi Darma Setiawan³

Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya
Email: ¹k.walady@student.ub.ac.id, ²adikara.putra@ub.ac.id, ³s.budidarma@ub.ac.id

Abstrak

Bunuh diri merupakan masalah kesehatan yang serius. Tidak hanya menyebabkan hilangnya nyawa secara sia-sia, bunuh diri juga dapat meninggalkan dampak yang berkepanjangan bagi mereka yang ditinggalkan. Meskipun begitu, stigma dan kekhawatiran akan perlakuan diskriminatif masih menjadi penghambat dalam upaya pencegahan bunuh diri. Mereka yang memiliki pemikiran bunuh diri cenderung memilih media sosial sebagai tempat bercerita. Pemahaman terhadap topik yang mereka bicarakan dapat menjadi salah satu langkah dalam peningkatan upaya pencegahan bunuh diri. Oleh karena itu, pada penelitian ini dilakukan pemodelan topik menggunakan metode Latent Dirichlet Allocation untuk mendapatkan gambaran mengenai topik yang dibicarakan dalam *subreddit r/SuicideWatch*. Pengujian terhadap pemodelan topik yang dilakukan menghasilkan nilai *coherence* tertinggi sebesar 0,2947. Nilai tersebut diperoleh menggunakan parameter $\alpha = 1/T$, $\beta = 1/T$, dan $T = 9$. Walaupun memiliki nilai *coherence* tertinggi dibanding pengujian lain, pengujian tersebut menghasilkan topik yang sulit diinterpretasi karena banyaknya kata umum yang muncul. Pengujian lain yang menggunakan parameter $\alpha = 50/T$, $\beta = 1/T$, dan $T = 5$ memberikan nilai *coherence* yang lebih rendah, tetapi topik yang dihasilkan lebih mudah untuk diinterpretasi. Beberapa topik yang dihasilkan, antara lain, rasa ketidakberdayaan, kondisi emosional, hubungan sosial, serta pemikiran atau perencanaan bunuh diri.

Kata kunci: pemodelan topik, risiko bunuh diri, latent dirichlet allocation, topic coherence, reddit

Abstract

Suicide is a serious health problem. Not only does it cause the loss of life in vain, suicide can also leave a lasting impact on those left behind. However, stigma and concerns about discriminatory treatment are still obstacles in suicide prevention efforts. Those who have thoughts of suicide tend to choose social media as a place to tell their stories. Understanding the topics they discuss can be one step in improving suicide prevention efforts. Therefore, in this research topic modeling was carried out using the Latent Dirichlet Allocation method to get an overview of the topics discussed in the r/SuicideWatch subreddit. Highest coherence score achieved on the evaluation was 0,2947. This value was obtained using the parameters $\alpha = 1/T$, $\beta = 1/T$, and $T = 9$. Even though it has the highest coherence value compared to other tests, this test produces topics that are difficult to interpret because of the many common words that appear. Another test using the parameters $\alpha = 50/T$, $\beta = 1/T$, and $T = 5$ gives lower coherence values, but the resulting topics are easier to interpret. Identified topics includes feelings of helplessness, emotional conditions, social relationships, and suicidal thoughts or plan.

Keywords: topic modelling, suicide risk, latent dirichlet allocation, topic coherence, reddit

1. PENDAHULUAN

Bunuh diri merupakan masalah kesehatan yang serius. Berdasarkan data WHO (2023), bunuh diri menempati peringkat keempat penyebab kematian pada rentang umur 15-29 tahun. Setidaknya ada 703 ribu orang meninggal tiap tahunnya karena bunuh diri. Tidak hanya

menyebabkan hilangnya nyawa secara sia-sia, bunuh diri juga dapat meninggalkan dampak yang berkepanjangan bagi mereka yang ditinggalkan. Keluarga dan teman yang ditinggalkan mungkin akan merasakan kesedihan yang berkepanjangan, rasa bersalah, rasa gelisah, depresi, hingga munculnya pemikiran bunuh diri (Hamdan, et al., 2019;

Wagner, Hofmann, & Grafiadeli, 2021).

Terdapat beberapa faktor risiko yang dapat mendorong seseorang untuk mengakhiri hidupnya sendiri. Faktor risiko tersebut meliputi gangguan psikologis, seperti depresi, bipolar, dan kecemasan, atau masalah personal, seperti perundungan, kehilangan pekerjaan, dan kehilangan orang tersayang (Rabani, Khan, & Khanday, 2020). Faktor risiko yang dibiarkan begitu saja dapat memicu munculnya pemikiran bunuh diri, yang kemudian berlanjut ke tahap perilaku bunuh diri, hingga akhirnya dilakukan percobaan bunuh diri.

Menurut (WHO, 2023), kegiatan pencegahan bunuh diri di banyak negara masih terbilang belum efektif. Stigma yang ada di masyarakat menyebabkan mereka yang memiliki masalah psikologis sering tidak mendapat pertolongan yang mereka butuhkan (Kim, et al., 2015). Hal ini dikarenakan adanya rasa khawatir terhadap perlakuan diskriminatif yang mungkin diterima dari orang-orang di sekitar. Mereka yang memiliki masalah psikologis pun cenderung memilih media sosial sebagai tempat bercerita. Oleh karena itu, data yang bersumber dari media sosial dapat digunakan untuk membantu dalam memperdalam pemahaman terhadap mereka yang memiliki masalah psikologis, termasuk pemikiran bunuh diri. Pemahaman terhadap topik yang mereka bicarakan dapat menjadi salah satu langkah dalam peningkatan upaya pencegahan bunuh diri.

Media sosial Reddit cukup sering digunakan sebagai sumber data dalam melakukan identifikasi terhadap kondisi psikologis. Beberapa kondisi tersebut antara lain depresi (Song, et al., 2018; Tadesse, et al., 2019), pemikiran bunuh diri (Chatterjee, et al., 2022), atau stres (Turcan, et al., 2021). Reddit memiliki batasan karakter yang lebih banyak dibanding Twitter. Hal ini memungkinkan pengguna untuk mendeskripsikan keadaan emosionalnya dengan lebih detail. Reddit juga memiliki beberapa forum atau *subreddit* yang didedikasikan untuk berbagai topik kesehatan mental, salah satunya *subreddit* r/SuicideWatch. Subreddit ini didedikasikan untuk topik yang berkaitan dengan bunuh diri.

Beberapa peneliti telah mencoba untuk melakukan pemodelan topik pada kasus kesehatan mental dengan data yang bersumber dari media sosial. Salah satunya pada penelitian (Kamarudin, Beigi, & Liu, 2020), pemodelan topik terhadap konten yang berhubungan dengan

kesehatan mental dilakukan menggunakan Latent Dirichlet Allocation pada data yang bersumber dari Reddit. Pada penelitian lainnya, (Sik, Németh, & Katona, 2021) melakukan pemodelan topik terhadap konten daring yang berhubungan dengan depresi. Penelitian ini menggunakan data yang bersumber dari beberapa forum daring, salah satunya Reddit.

Pada penelitian ini akan dilakukan percobaan untuk melakukan pemodelan topik pada konten yang berhubungan dengan risiko bunuh diri. Data yang digunakan bersumber dari media sosial Reddit. Pemodelan topik akan dilakukan dengan menggunakan metode Latent Dirichlet Allocation. Harapannya, pemodelan topik yang dilakukan dapat memberikan gambaran mengenai topik yang dibicarakan dalam *subreddit* r/SuicideWatch.

2. DATA DAN METODOLOGI

2.1 Pengumpulan Data

Pengumpulan data dilakukan dengan menggunakan Python Reddit API Wrapper. Data diperoleh dari media sosial Reddit pada bagian *subreddit* r/SuicideWatch. Pengumpulan data dilakukan mulai tanggal 10 Oktober sampai dengan 31 Oktober 2023. Data yang diperoleh berjumlah 2465 dokumen. Setiap dokumen terdiri dari judul dan isi.

2.2 Text Preprocessing

Preprocessing merupakan serangkaian proses yang dilakukan untuk mengubah data teks yang tidak terstruktur menjadi data terstruktur. Setiap langkah dalam *text preprocessing* bertujuan untuk menghilangkan informasi yang dianggap kurang penting dalam analisis teks (Anandarajan, Hill, & Nolan, 2019). *Text preprocessing* juga dapat mengurangi dimensi dari data teks yang digunakan, sehingga dapat mengurangi kompleksitas proses analisis. Rangkaian *text preprocessing* yang dilakukan pada penelitian ini, meliputi *case folding*, *cleaning*, *tokenization*, *stopword removal*, dan *lemmatization*.

2.3 Latent Dirichlet Allocation

Latent Dirichlet Allocation (LDA) merupakan salah satu metode pemodelan topik. Prinsip dasar dari LDA adalah setiap kumpulan dokumen dapat direpresentasikan oleh sekumpulan topik laten (Blei, Ng, & Jordan, 2003). Setiap topik pada dokumen

direpresentasikan oleh kumpulan peluang kata. Setiap dokumen dalam suatu kumpulan dokumen terdiri dari kumpulan topik yang sama, tetapi dengan proporsi yang berbeda antar dokumen.

Perhitungan peluang suatu topik pada suatu dokumen dilakukan dengan menggunakan parameter α yang berupa peluang prior topik tersebut. Semakin kecil nilai parameter α , semakin sedikit jumlah topik pada setiap dokumen (George & Doss, 2018). Beberapa cara yang digunakan untuk menentukan nilai α berdasarkan jumlah topik T yang digunakan, antara lain, $\alpha = 1/T$ (Mardones-Segovia, et al., 2023) dan $\alpha = 50/T$ (Putri & Kusumaningum, 2017). Persamaan peluang topik pada suatu dokumen ditunjukkan pada Persamaan (1).

$$\theta_{t,d} = \frac{n_{t,d} + \alpha}{\sum_{t=1}^K n_{t,d} + K * \alpha} \quad (1)$$

Keterangan:

- $n_{t,d}$ = Jumlah kata dengan topik t pada dokumen d
- α = Peluang prior topik
- K = Jumlah topik

Perhitungan peluang suatu kata pada suatu topik dilakukan dengan menggunakan parameter β yang berupa peluang prior kata tersebut. Semakin kecil nilai parameter β , jumlah kata dengan peluang yang signifikan dalam suatu topik akan lebih sedikit (George & Doss, 2018). Persamaan peluang kata pada suatu topik ditunjukkan pada Persamaan (2).

$$\varphi_{w,t} = \frac{n_{w,t} + \beta}{\sum_{w=1}^W n_{w,t} + V * \beta} \quad (2)$$

Keterangan:

- $n_{w,t}$ = Jumlah kata w dengan topik t
- β = Peluang prior kata
- W = Jumlah kata pada seluruh dokumen
- V = Jumlah kata unik pada seluruh dokumen

Hasil dari Persamaan (1) dan Persamaan (2) digunakan untuk menentukan topik dari setiap kata dalam kumpulan dokumen. Persamaan peluang topik pada suatu kata ditunjukkan pada Persamaan (3).

$$\zeta_{t,w,d} = \theta_{t,d} * \varphi_{w,t} \quad (3)$$

Keterangan:

- $\zeta_{t,w,d}$ = Peluang topik t pada kata w dalam

dokumen d

$\theta_{t,d}$ = Peluang topik t pada dokumen d

$\varphi_{w,t}$ = Peluang kata w pada topik t

Algoritma dari Latent Dirichlet Allocation sebagai berikut:

1. Tentukan jumlah topik, jumlah iterasi, serta nilai parameter α dan β .
2. Inisialisasi setiap kata pada setiap dokumen dengan suatu topik secara acak.
3. Untuk setiap kata w pada setiap dokumen d , hapus topik pada kata w , kemudian tentukan topik baru untuk kata w berdasarkan Persamaan (4).

$$topik_{w,d} = arg \max_{t \in T} \zeta_{t,w,d} \quad (4)$$

Keterangan:

T = Himpunan seluruh topik

4. Ulangi langkah 3 hingga mencapai jumlah iterasi yang ditentukan.
5. Hitung peluang setiap topik pada setiap dokumen menggunakan Persamaan (1).
6. Hitung peluang setiap kata pada setiap topik menggunakan Persamaan (2).

2.4 C_v Coherence

C_v Coherence merupakan salah satu metode yang digunakan dalam mengukur *topic coherence*. Metode ini diusulkan oleh (Röder, Both, & Hinneburg, 2015) dan berdasarkan penelitian yang dilakukan, C_v Coherence memiliki kinerja yang lebih baik dibanding metode pengukuran *topic coherence* lain. C_v Coherence terdiri dari empat tahapan: (i) *segmentation*, (ii) *probability calculation*, (iii) *confirmation measure*, dan (iv) *aggregation*.

- i. *Segmentation* dilakukan dengan memasangkan setiap kata pada himpunan kata teratas dari setiap topik dengan himpunan kata teratas itu sendiri berdasarkan Persamaan (5).

$$S_{set}^{one} = \{(W', W^*) | W' \in W; W^* = W\} \quad (5)$$

Keterangan:

(W', W^*) = Pasangan himpunan bagian W' dan W^*

W = Himpunan kata dengan peluang tertinggi pada suatu topik

w_i = Kata dalam himpunan W

- ii. *Probability calculation* dilakukan dengan menghitung peluang kata pada dokumen virtual yang diperoleh menggunakan

Boolean Sliding Window. *Window* yang digunakan akan bergerak satu kata pada setiap langkah dan menghasilkan satu dokumen virtual baru.

- iii. Nilai *confirmation measure* dari suatu pasangan himpunan bagian diperoleh dengan menghitung *cosine similarity* antara vektor yang merepresentasikan W' dan W^* . Vektor dari himpunan bagian W' diperoleh menggunakan Persamaan (6). Vektor dari himpunan bagian W^* diperoleh menggunakan Persamaan (7). Setiap elemen dari kedua vektor tersebut diperoleh menggunakan persamaan (8). *Confirmation measure* antara kedua vektor diperoleh dengan menggunakan Persamaan (9).

$$\vec{v}(W') = \{NPMI(W', w_j)^\gamma\}_{j=1, \dots, |W'|} \quad (6)$$

$$\vec{v}(W^*) = \sum_{w_i \in W^*} \vec{v}(W') \quad (7)$$

$$NPMI(W', w_j)^\gamma = \left(\frac{\log \frac{P(w_i, w_j) + \epsilon}{P(w_i) * P(w_j)}}{-\log(P(w_i, w_j) + \epsilon)} \right)^\gamma \quad (8)$$

$$s_{cos}(\vec{u}, \vec{w}) = \frac{\sum_{i=1}^{|W'|} u_i * w_i}{\|\vec{u}\| * \|\vec{w}\|} \quad (9)$$

Keterangan:

- γ = Parameter bobot nilai NPMI
- w_j = Kata dalam himpunan W
- ϵ = Konstanta untuk mencegah logaritma nol
- \vec{u} = Representasi vektor dari himpunan bagian W'
- \vec{w} = Representasi vektor dari himpunan bagian W^*

- iv. Nilai *topic coherence* suatu topik diperoleh dengan melakukan perhitungan *arithmetic mean* terhadap nilai *confirmation measure* dari setiap pasangan himpunan bagian pada topik tersebut.

$$coherence(t) = \frac{1}{n} \sum_{i=1}^{n_t} s_{cos}(\vec{u}, \vec{w}) \quad (10)$$

Keterangan:

- n_t = Jumlah pasangan himpunan bagian pada topik t

2.5 Relevance

Relevance merupakan salah satu metode alternatif untuk melakukan pemeringkatan kata dalam suatu topik (Sievert & Shirley, 2014). Nilai *relevance* suatu kata diperoleh dari nilai peluang dan nilai *lift* kata tersebut. Nilai *lift*

dapat digunakan untuk mengurangi kemunculan kata dengan frekuensi tinggi pada topik. Persamaan nilai *lift* ditunjukkan pada Persamaan (11). Pengukuran nilai *relevance* dilakukan dengan menggunakan satu parameter bobot λ yang memiliki rentang nilai 0 hingga 1. Persamaan nilai *relevance* ditunjukkan pada Persamaan (12).

$$lift(w, t) = \log \left(\frac{\varphi_{w,t}}{P(w)} \right) \quad (11)$$

$$rel(w, t) = \lambda \log(\varphi_{w,t}) + (1 - \lambda) lift(w, t) \quad (12)$$

Keterangan:

$\varphi_{w,t}$ = Peluang kata w pada topik t

$P(w)$ = Peluang kata w

λ = Parameter bobot

3. HASIL DAN PEMBAHASAN

3.1. Hasil Pengujian Parameter α , β , dan Jumlah Topik

Pengujian dilakukan dengan menggunakan jumlah topik serta nilai parameter α dan β yang berbeda. Jumlah topik T yang digunakan adalah sebanyak lima hingga sepuluh topik. Pengujian ini dilakukan dengan menggunakan nilai parameter $\alpha = 50/T$ dan $\alpha = 1/T$. Nilai parameter β yang digunakan sebesar $1/T$. Jumlah iterasi yang dilakukan sebanyak 10 iterasi. Perhitungan nilai *topic coherence* dilakukan menggunakan 20 kata teratas yang diurutkan berdasarkan peluang, *window_size* = 110, $\gamma = 1$, dan $\epsilon = 1 \times 10^{-12}$. Hasil Pengujian tersebut ditunjukkan pada Tabel 1.

Tabel 1. Hasil Pengujian Parameter α , β , dan Jumlah Topik

α	β	Jumlah Topik (T)	Coherence
50/T	1/T	5	0,2617
		6	0,2540
		7	0,2543
		8	0,2348
		9	0,2379
1/T	1/T	10	0,2335
		5	0,2937
		6	0,2920
		7	0,2926
		8	0,2933
		9	0,2947
		10	0,2937

Berdasarkan hasil dari pengujian yang dilakukan, nilai *coherence* yang diperoleh dari seluruh pengujian terbilang rendah. Nilai *coherence* tertinggi diperoleh pada pengujian α

= 1/T, $\beta = 1/T$, dan $T = 9$. Nilai *coherence* yang diperoleh dari pengujian tersebut sebesar 0,2947. Pengujian menggunakan nilai $\alpha = 1/T$ menghasilkan nilai *coherence* yang lebih tinggi dibanding pengujian $\alpha = 50/T$. Daftar peringkat 20 kata dengan peluang tertinggi dari setiap topik pada pengujian $\alpha = 1/T$, $\beta = 1/T$, dan $T = 9$ ditunjukkan pada Tabel 2.

Tabel 2. Daftar Peringkat 20 Kata dengan Peluang Tertinggi dari Setiap Topik pada Pengujian $\alpha = 1/T$, $\beta = 1/T$, dan $T = 9$

Topik	Peringkat Kata
0	<i>feel, want, go, like, get, know, life, think, even, one, make, time, try, people, never, live, thing, friend, really, die</i>
1	<i>want, feel, go, get, know, life, like, think, make, even, try, one, time, live, people, would, die, thing, take, well</i>
2	<i>get, want, feel, go, like, life, know, time, think, make, would, try, thing, even, one, well, year, really, live, people</i>
3	<i>feel, want, go, get, like, know, life, think, even, try, time, one, people, make, tell, well, thing, friend, would, help</i>
4	<i>feel, want, like, get, go, know, life, think, even, time, people, make, try, never, one, live, really, year, thing, friend</i>
5	<i>go, want, feel, get, life, like, know, think, even, time, one, make, die, try, live, thing, never, people, bad, would</i>
6	<i>like, feel, get, want, go, life, know, think, even, make, thing, friend, time, people, really, anymore, would, year, bad, good</i>
7	<i>get, feel, want, like, go, life, know, people, time, think, one, try, make, even, work, live, really, friend, year, thing</i>
8	<i>want, go, feel, like, get, life, know, people, die, think, time, make, would, even, live, one, friend, day, try, year</i>

Daftar peringkat kata yang diperoleh pengujian $\alpha = 1/T$, $\beta = 1/T$, dan $T = 9$ menunjukkan banyak kata yang sama di seluruh topik. Beberapa kata, seperti *feel, want, go, like*, dan *get* mendominasi peringkat kata dari setiap topik. Munculnya banyak kata yang sama pada daftar peringkat kata setiap topik dapat disebabkan karena nilai α yang rendah. Nilai α yang rendah menyebabkan setiap dokumen cenderung didominasi oleh satu topik tertentu dan kata yang sering muncul di banyak dokumen memiliki peluang yang cukup tinggi di setiap topik.

Pengujian menggunakan nilai $\alpha = 50/T$ menghasilkan nilai *coherence* yang lebih rendah.

Nilai *coherence* tertinggi dari pengujian tersebut diperoleh dengan jumlah topik sebanyak lima. Nilai *coherence* yang diperoleh bernilai 0,2617. Daftar peringkat 20 kata dengan peluang tertinggi dari setiap topik pada pengujian $\alpha = 50/T$, $\beta = 1/T$, dan $T = 5$ ditunjukkan pada Tabel 3.

Tabel 3. Daftar Peringkat 20 Kata dengan Peluang Tertinggi dari Setiap Topik pada Pengujian $\alpha = 50/T$, $\beta = 1/T$, dan $T = 5$

Topik	Peringkat Kata
0	<i>really, never, well, bad, see, love, keep, nothing, talk, long, last, ever, pain, world, happy, shit, lot, actually, suicide, sleep</i>
1	<i>like, anymore, good, day, everything, f***ing, care, always, leave, tired, point, everyone, lose, month, alone, win, without, away, seem, first</i>
2	<i>feel, want, go, life, know, think, even, try, thing, live, much, end, say, work, help, way, anything, start, give, job</i>
3	<i>one, people, friend, family, hate, someone, back, person, enough, many, break, matter, money, sure, post, stay, past, understand, honestly, experience</i>
4	<i>get, time, make, would, year, die, take, tell, kill, still, need, find, could, since, something, right, parent, anyone, hope, hurt</i>

Daftar peringkat kata yang diperoleh dari pengujian $\alpha = 50/T$, $\beta = 1/T$, dan $T = 5$ terdiri dari kata yang lebih beragam dibanding hasil pada Tabel 2. Nilai α yang lebih tinggi menyebabkan distribusi topik yang dihasilkan dari pengujian ini lebih merata. Nilai *coherence* yang lebih tinggi pada pengujian menggunakan $\alpha = 1/T$ dapat disebabkan oleh topik yang terdiri dari banyak kata yang bersifat umum. Hal ini dapat terjadi karena *coherence* tidak dapat membedakan kata dengan frekuensi tinggi dengan kata yang lebih deskriptif terhadap suatu topik. Berdasarkan daftar peringkat kata pada Tabel 3, terdapat beberapa topik yang dapat diinterpretasikan.

Topik 0 berisi kata *last, suicide*, dan *sleep* yang dapat dikaitkan dengan keinginan bunuh diri. Selain itu, isi dari beberapa dokumen dengan peluang topik 0 tertinggi menunjukkan pemikiran dan percobaan bunuh diri. Oleh karena itu, dapat disimpulkan bahwa topik 0 berkaitan dengan pemikiran atau perencanaan bunuh diri. Penggalan dokumen dari topik 0 ditunjukkan pada Tabel 4.

Tabel 4. Penggalan dokumen dari Topik 0

Peluang	Penggalan Dokumen
0,3125	<i>Hello, and yeah, the title says it all. Against my better judgement, I decided to do yet another attempt by ingesting large amount of my meds. ... A lot of bad things happened lately, it all pushed me to my limit, and here I am playing what would probably be my last Factorio session while waiting for my ultimate demise.</i>

Topik 1 berisi kata *anymore*, *tired*, dan *point* yang dapat dikaitkan dengan perasaan ingin menyerah. Selain itu, isi dari beberapa dokumen dengan peluang topik 1 tertinggi menunjukkan pernyataan ingin menyerah. Oleh karena itu, dapat disimpulkan bahwa topik 1 berkaitan dengan rasa ketidakberdayaan. Penggalan dokumen dari topik 1 ditunjukkan pada Tabel 5.

Tabel 5. Penggalan dokumen dari Topik 1

Peluang	Penggalan Dokumen
0,3651	<i>I am ready to give up. I'm not smart enough to get a PhD or go to medical school. Hell, I'm probably not even smart enough to finish my neuroscience degree (currently a freshman).</i>

Topik 2 berisi kata *feel* yang digunakan untuk mengungkapkan kondisi emosional. Selain itu, isi dari beberapa dokumen dengan peluang topik 2 tertinggi menunjukkan cerita mengenai kondisi emosional. Oleh karena itu, dapat disimpulkan bahwa topik 2 berkaitan dengan kondisi emosional. Penggalan dokumen dari topik 2 ditunjukkan pada Tabel 6.

Tabel 6. Penggalan dokumen dari Topik 2

Peluang	Penggalan Dokumen
0,3258	<i>I feel completely fatigued and exhausted. My relationship just ended. I don't want to do it anymore. Everyone sees a different picture of me as successful happy strong. But I feel the most unsafe it's ever been. I just want to stop it all and jump. I have been thinking seriously about it and working out details. How to wrap up loose ends</i>

Topik 3 berisi kata yang berkaitan dengan orang lain, seperti *people*, *friend*, *family*, *someone*, dan *person*. Selain itu, isi dari beberapa dokumen dengan peluang topik 3 tertinggi menunjukkan kondisi hubungan dengan teman dan keluarga. Oleh karena itu, dapat disimpulkan bahwa topik 3 berkaitan dengan hubungan sosial. Penggalan dokumen dari topik 3 ditunjukkan pada Tabel 7.

Tabel 7. Penggalan dokumen dari Topik 3

Peluang	Penggalan Dokumen
0,3651	<i>I am ready to give up. I'm not smart enough to get a PhD or go to medical school. Hell, I'm probably not even smart enough to finish my neuroscience degree (currently a freshman).</i>

Topik 4 berisi kata *die*, *kill*, dan *hurt* yang dapat dikaitkan dengan perbuatan bunuh diri. Selain itu, isi dari beberapa dokumen dengan peluang topik 4 tertinggi menunjukkan pernyataan keinginan untuk bunuh diri. Oleh karena itu, topik 4 dapat dikaitkan dengan pemikiran atau perencanaan bunuh diri. Penggalan dokumen dari topik 4 ditunjukkan pada Tabel 8.

Tabel 8. Penggalan dokumen dari Topik 4

Peluang	Penggalan Dokumen
0,3850	<i>I do not remember when I became depressed, or much of anything for that matter, but in 4th grade my mom walked in on me trying to kill myself, weapon to neck- and she walked out of the room and never said a word about it. I later also told her of my many attempts and she did nothing.</i>

3.2. Hasil Pengujian Pengaruh Penggunaan Nilai Relevance

Pengujian dilakukan berdasarkan nilai coherence tertinggi pada pengujian sebelumnya. Pada pengujian ini, pemeringkatan kata dilakukan berdasarkan nilai *relevance* dengan nilai parameter $\lambda = 0,6$. Hasil pengujian ditunjukkan pada Tabel 9.

Tabel 9. Hasil Pengujian Penggunaan Nilai Relevance

α	β	Jumlah Topik (T)	Coherence
50/T	1/T	5	0,2617
1/T	1/T	10	0,2937

Berdasarkan hasil pengujian pada Tabel 7, nilai *coherence* pada pengujian $\alpha = 1/T$, $\beta = 1/T$, dan $T = 9$ mengalami penurunan sebesar 0,0052. Nilai *coherence* pada pengujian $\alpha = 50/T$, $\beta = 1/T$, dan $T = 5$ tidak mengalami perubahan. Daftar peringkat 20 kata dengan *relevance* tertinggi dari setiap topik pada pengujian $\alpha = 1/T$, $\beta = 1/T$, dan $T = 9$ ditunjukkan pada Tabel 10.

Tabel 10. Daftar Peringkat 20 Kata dengan

Relevance Tertinggi dari Setiap Topik pada Pengujian $\alpha = 1/T, \beta = 1/T, \text{ dan } T = 9$

Topik	Peringkat Kata
0	<i>feel, want, go, like, know, life, get, even, one, think, make, time, never, try, people, live, say, friend, thing, tell</i>
1	<i>want, feel, go, get, know, life, like, think, try, make, even, live, one, take, people, die, would, time, work, bad</i>
2	<i>get, go, want, feel, like, life, time, know, would, think, thing, make, try, well, year, even, one, really, work, end</i>
3	<i>feel, go, want, get, know, like, think, try, life, tell, even, one, time, people, help, well, friend, thing, end, make</i>
4	<i>feel, like, want, know, get, life, go, think, even, never, year, people, really, live, try, make, time, one, thing, bad</i>
5	<i>go, want, feel, get, life, like, think, know, time, even, die, one, try, make, never, live, bad, thing, tired, good</i>
6	<i>like, get, feel, go, want, life, know, think, make, even, thing, friend, anymore, time, really, good, people, bad, year, would</i>
7	<i>get, feel, want, like, go, life, people, know, time, work, one, think, try, make, live, even, keep, really, year, see</i>
8	<i>want, go, feel, like, life, get, die, people, know, would, live, make, time, think, one, even, help, day, friend, year</i>

Daftar peringkat kata pada Tabel 10 menunjukkan beberapa perbedaan pada urutan kata jika dibandingkan dengan kata pada Tabel 2. Akan tetapi, peringkat kata teratas masih didominasi oleh kata yang umum, seperti *feel, want, go, like, dan get*. Hal ini menunjukkan bahwa, pada pengujian ini, nilai parameter $\lambda = 0,6$ belum efektif untuk mengurangi kemunculan kata yang bersifat umum pada daftar peringkat kata dalam topik.

Daftar peringkat kata pada Tabel 11 tidak memiliki perbedaan sama sekali jika dibandingkan dengan Tabel 3. Pada pengujian ini, nilai parameter $\lambda = 0,6$ tidak memberikan dampak apapun terhadap interpretabilitas topik yang dihasilkan.

Tabel 11. Daftar Peringkat 20 Kata dengan Relevance Tertinggi dari Setiap Topik pada Pengujian $\alpha = 50/T, \beta = 1/T, \text{ dan } T = 5$

Topik	Peringkat Kata
0	<i>really, never, well, bad, see, love, keep, nothing, talk, long, last, ever, pain, world, happy, shit, lot, actually, suicide, sleep</i>
1	<i>like, anymore, good, day, everything, f***ing, care, always, leave, tired, point, everyone, lose, month, alone, win, without, away, seem, first</i>

Topik	Peringkat Kata
2	<i>feel, want, go, life, know, think, even, try, thing, live, much, end, say, work, help, way, anything, start, give, job</i>
3	<i>one, people, friend, family, hate, someone, back, person, enough, many, break, matter, money, sure, post, stay, past, understand, honestly, experience</i>
4	<i>get, time, make, would, year, die, take, tell, kill, still, need, find, could, since, something, right, parent, anyone, hope, hurt</i>

4. PENUTUP

Berdasarkan hasil pengujian dan analisis yang telah dilakukan, pemodelan topik risiko bunuh diri berdasarkan konten media sosial menggunakan Latent Dirichlet Allocation menghasilkan nilai *coherence* tertinggi sebesar 0,2947. Nilai tersebut diperoleh menggunakan parameter $\alpha = 1/T, \beta = 1/T, \text{ dan } T = 9$. Walaupun memiliki nilai *coherence* yang lebih tinggi dibanding pengujian lain, pengujian tersebut menghasilkan topik yang sulit diinterpretasi karena banyaknya kata umum yang muncul. Pengujian lain yang menggunakan parameter $\alpha = 50/T, \beta = 1/T, \text{ dan } T = 5$ memberikan nilai *coherence* yang lebih rendah, tetapi topik yang dihasilkan lebih mudah untuk diinterpretasi. Beberapa topik yang diidentifikasi, antara lain, topik mengenai rasa ketidakberdayaan, hubungan sosial, serta pemikiran atau perencanaan bunuh diri.

Saran untuk penelitian berikutnya yaitu penggunaan nilai parameter λ yang lebih kecil guna mengurangi kemunculan kata yang bersifat umum dari daftar peringkat kata. Selain itu, pembentukan *n-gram* berdasarkan kata umum yang diperoleh dapat memperjelas makna dari kata tersebut. Penghapusan kata yang dianggap umum juga dapat dilakukan pada hasil akhir dari pemodelan topik.

DAFTAR PUSTAKA

Anandarajan, M., Hill, C., & Nolan, T., 2019. *Practical Text Analytics: Maximizing the Value of Text Data*. Cham: Springer Nature.

Blei, D.M., Ng, A.Y. & Jordan, M.I., 2003. Latent Dirichlet Allocation. *Journal of Machine Learning Research*, [online] Tersedia di: <<https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf>>

- Chatterjee, M., Kumar, P., Samanta, P. & Sarkar, D., 2022. Suicide Ideation Detection from Online Social Media: A Multi-Modal Feature Based Technique. *International Journal of Information Management Data Insights*, 2(2), p.100103.
- George, C.P. & Doss, H., 2018. Principled Selection of Hyperparameters in the Latent Dirichlet Allocation Model. *Journal of Machine Learning Research*, [online] Tersedia di: <<https://www.jmlr.org/papers/v18/15-595.html>>
- Hamdan, S., Berkman, N., Lavi, N., Levy, S. & Brent, D., 2019. The Effect of Sudden Death Bereavement on the Risk for Suicide. *Crisis*.
- Kamarudin, N.S., Beigi, G. & Liu, H., 2021. A Study on Mental Health Discussion through Reddit. In: *2021 International Conference on Software Engineering & Computer Systems and 4th International Conference on Computational Science and Information Management (ICSECS-ICOCSIM)*. Pekan, Malaysia, 24-26 August 2021. IEEE.
- Kim, W.J., Song, Y.J., Ryu, H.S., Ryu, V., Kim, J.M., Ha, R.Y., Lee, S.J., Namkoong, K., Ha, K. & Cho, H.S., 2015. Internalized Stigma and Its Psychosocial Correlates in Korean Patients with Serious Mental Illness. *Psychiatry research*, 225(3), pp.433-439.
- Mardones-Segovia, C., Wheeler, J.M., Choi, H.J., Wang, S. & Cohen, A.S., 2023. Model Selection for Latent Dirichlet Allocation in Assessment Data. *Psychological Test and Assessment Modeling*, 65(1), pp.3-35.
- Putri, I.R. & Kusumaningrum, R., 2017. Latent Dirichlet allocation (LDA) for Sentiment Analysis toward Tourism Review in Indonesia. *Journal of Physics: Conference Series*, [online] Tersedia di: <<https://iopscience.iop.org/article/10.1088/1742-6596/801/1/012073/meta>>
- Rabani, S.T., Khan, Q.R. & Khanday, A.M.U.D., 2020. Detection of Suicidal Ideation on Twitter using Machine Learning & Ensemble Approaches. *Baghdad science journal*, 17(4), pp.1328-1328.
- Röder, M., Both, A. & Hinneburg, A., 2015. Exploring the Space of Topic Coherence Measures. In: *Proceedings of the eighth ACM international conference on Web search and data mining*. Shanghai, China, 2-6 February 2015. New York: ACM.
- Sievert, C. & Shirley, K., 2014. LDAvis: A Method for Visualizing and Interpreting Topics. In: *Proceedings of the workshop on interactive language learning, visualization, and interfaces*. Maryland, USA, 27 June 2014. ACL.
- Sik, D., Németh, R. & Katona, E., 2021. Topic Modelling Online Depression Forums: Beyond Narratives of Self-Objectification and Self-Blaming. *Journal of Mental Health*, 32(2), pp.386-395.
- Song, H., You, J., Chung, J.W. & Park, J.C., 2018. Feature Attention Network: Interpretable Depression Detection from social Media. *Proceedings of the 32nd Pacific Asia conference on language, information and computation*. Hong Kong, 1-3 December 2018. ACL.
- Syed, S. & Spruit, M., 2017. Full-Text or Abstract? Examining Topic Coherence Scores using Latent Dirichlet Allocation. In: *2017 IEEE International conference on data science and advanced analytics (DSAA)*. Tokyo, Japan, 19-21 October 2017. IEEE.
- Tadesse, M.M., Lin, H., Xu, B. & Yang, L., 2019. Detection of Depression-Related Posts in Reddit Social Media Forum. *IEEE Access*, 7, pp.44883-44893.
- Turcan, E., Muresan, S. & McKeown, K., 2021. Emotion-infused models for explainable psychological stress detection. In: *Proceedings of the 2021 conference of the North American Chapter of the Association for Computational Linguistics: human language technologies*. 6-11 June 2021. ACL.
- Wagner, B., Hofmann, L. & Grafiadeli, R., 2021. The Relationship Between Guilt, Depression, Prolonged grief, and Posttraumatic Stress Symptoms After Suicide Bereavement. *Journal of Clinical Psychology*, 77(11), pp.2545-2558.
- WHO, 2023. *Suicide*. [online] Tersedia di: <<https://www.who.int/news-room/fact->

sheets/detail/suicide>