

Memvisualisasikan Twit Kesehatan Berdasarkan Wilayah dan *Timestamp*

Bonpagna Kann¹, Sihem Amer-Yahia², Michael Ortega³, Jean-Louis Pépin⁴, Sébastien Bailly⁵

Intisari—Media sosial telah menjadi salah satu sumber data utama pada studi sosial, terutama melalui ekspresi pengguna seperti momen-momen penting dalam kehidupan sehari-hari atau perasaan dan persepsi pengguna terhadap topik diskusi tertentu. Dalam dunia kesehatan, media sosial dapat digunakan untuk mempelajari wacana orang tentang berbagai macam penyakit serta untuk memperoleh wawasan tentang dampak penyakit-penyakit tersebut terhadap kualitas hidup pasien. Saat ini, ketertarikan terhadap penerapan algoritme *machine learning* untuk meningkatkan prediksi penyakit melalui data yang diperoleh dari media sosial pengguna sudah makin tinggi. Pada penelitian ini, hampir 800 juta *post* diambil dari Twitter melalui prapemrosesan dan dengan menjalankan *time-aware ailment topic aspect model* (T-ATAM). Dengan langkah-langkah tersebut, diharapkan penyakit, gejala, dan pengobatan dua penyakit kronis, yaitu *sleep apnea* dan penyakit hati kronis, dapat diprediksi. Penelitian dilakukan terhadap twit berbahasa Inggris pada tahun 2018. Twit yang digunakan sebagai data penelitian ini sebagian besar berasal dari negara-negara di Eropa dan Amerika Serikat. Data yang diperoleh kemudian diolah menggunakan T-ATAM berdasarkan wilayah, *timestamp*, dan pengobatan, yaitu *continuous positive airway pressure* (CPAP), untuk melihat perbedaan sebaran teratas penyakit, gejala, serta pengobatan di berbagai wilayah; *timestamp*; dan sebelum, selama, serta setelah CPAP diperkenalkan. Berbagai visualisasi statistik ditampilkan pada penelitian ini, yang meliputi peta dunia, *word cloud*, dan histogram. Visualisasi-visualisasi tersebut dibuat berdasarkan sekitar 331.000 twit terkait penyakit hati serta satu juta twit terkait *sleep apnea*. Hasil penelitian menunjukkan bahwa depresi dan konsumsi alkohol adalah gejala utama penyakit hati. Sementara itu, kurangnya tidur malam dan terlalu banyak bekerja dianggap sebagai faktor utama penyebab *sleep apnea*.

Kata Kunci—Media Sosial, Pemodelan Topik, Visualisasi, Data Kesehatan, Analisis Data.

I. PENDAHULUAN

Media sosial memegang peranan penting dalam memberikan informasi mengenai kehidupan sehari-hari masyarakat pengguna internet melalui analisis *post* dalam skala besar sehingga dapat mewakili aktivitas dan persepsi pengguna terhadap berbagai macam topik. Banyak penelitian tentang media sosial yang telah dilakukan serta memiliki berbagai tujuan, termasuk mengeksplorasi tren dan membahas peristiwa

global, seperti perspektif politik [1]-[3], berita sehari-hari [4], [5], dan bencana alam [6], [7]. Twitter, yang merupakan salah satu contoh media sosial yang menarik dengan jutaan twit dikirimkan oleh pengguna setiap harinya, telah menjadi sumber data penting untuk mendeteksi dan memprediksi transisi [8], [9] dalam dunia kesehatan [10], [11]. Sebagian besar penelitian meneliti twit untuk mengkaji berbagai macam penyakit, termasuk kanker [12], [13] dan penyakit menular [7], [14].

Satu *post* di Twitter mungkin terlihat kurang bermanfaat, tetapi kombinasi dari semua *post* dapat diproses sehingga diperoleh pengetahuan yang berharga serta wawasan yang tak ternilai. Misalnya, pengguna Twitter dapat mengekspresikan keadaan mereka dengan membuat twit seperti “Saya masuk angin hari ini” atau “Punggung saya sakit. Apa yang harus saya lakukan?”. Informasi ini mungkin kurang menarik karena merupakan hal yang lumrah, tetapi gabungan *post-post* tersebut dapat memberikan informasi statistik penting dan membantu memantau gejala penyakit berdasarkan ruang dan waktu. Praktik ini telah diterapkan di Inggris [15], [16], dan Amerika Serikat [17], [18]. Mempelajari gejala penyakit melalui kiriman twit dapat membantu mengukur faktor risiko penyakit dan mempromosikan kampanye kesehatan di berbagai wilayah dan pada periode yang berbeda. Selain itu, mempelajari pengobatan penyakit yang disebutkan dalam twit dapat membantu mengukur efektivitas pengobatan yang diberikan, yaitu *continuous positive airway pressure* (CPAP) menyebabkan berubahnya wacana dalam twit.

Penelitian tentang penggunaan media sosial untuk memahami perspektif orang terhadap kondisi penyakit mereka dan mendeteksi topik diskusi serta gejala penyakit telah dilakukan. Akan tetapi, penelitian dan visualisasi twit berdasarkan wilayah dan *timestamp* belum pernah dilakukan sebelumnya, khususnya terkait *sleep apnea* dan penyakit hati kronis. Penelitian ini menerapkan model yang telah dikembangkan, yaitu *time-aware ailment topic aspect model* (T-ATAM) [8], untuk memprediksi penyakit, gejala, dan pengobatan *sleep apnea* dan penyakit hati kronis berdasarkan ruang dan waktu.

II. METODOLOGI

A. Dataset

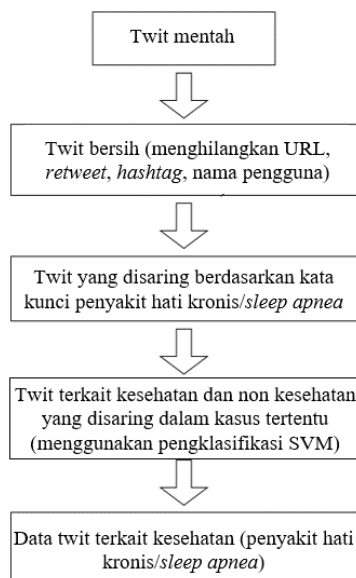
Sebuah *subset* Twit tahun 2018 diekstraksi menggunakan Twitter API. Penelitian ini berfokus pada twit yang berasal dari negara berbahasa Inggris dengan mayoritas merupakan negara-negara di benua Amerika dan Eropa. Konten yang dikumpulkan termasuk *timestamp* dan atribut geografis twit. Selain itu, *dataset* informasi kesehatan tentang penyakit hati dan *sleep apnea*, termasuk di dalamnya penyakit, gejala, dan pengobatannya, diekstraksi dari situs kesehatan dengan

^{1,2,3} CNRS, Univ. Grenoble Alpes, Bâtiment IMAG, 700 Av. Centrale, 38401, Saint-Martin-d'Hères, Grenoble, France (email: ¹bonpagnakann@gmail.com, ²sihem.Amer-Yahia@univ-grenoble-

³alpes.fr, michael.ortega@imag.fr)

^{4,5} INSERM, Univ. Grenoble Alpes, Av. des Maquis du Grésivaudan, 38700, La Tronche, Grenoble, France (email: ⁴jpepin@chu-grenoble.fr, ⁵sbailly@chu-grenoble.fr)

[Diterima: 9 Januari 2022, Revisi: 19 September 2022]



Gbr. 1 Diagram Pemrosesan Data.

TABEL I
JUMLAH PENYAKIT, GEJALA, DAN PENGobatan
PADA DATASET KATA KUNCI

Jenis Penyakit	Penyakit	Gejala	Pengobatan
Penyakit hati kronis	61	169	220
<i>Sleep apnea</i>	15	175	117

mengombinasikan metode manual dan otomatis menggunakan Beautiful Soup, yang merupakan paket untuk *scraping* situs dan *parsing* dokumen HTML dan XML di Python. Tabel I menunjukkan sejumlah penyakit, gejala, dan pengobatan penyakit hati kronis dan *sleep apnea* pada *dataset* kata kunci. Data yang digunakan pada penelitian ini diperoleh dari sumber yang dapat diakses publik, yaitu Twitter API. Twit dan data pengguna dari basis data dianonimkan untuk melindungi privasi pengguna.

B. Prapemrosesan Data

Prapemrosesan dilakukan terhadap twit yang telah dikumpulkan, dengan langkah sebagaimana ditunjukkan pada Gbr. 1. Twit mentah diekstraksi beserta atribut tempat, *timestamp*, dan teksnya sebelum diproses pada langkah selanjutnya untuk menghapus URL, *retweet*, *hashtag*, dan nama pengguna. Twit bersih kemudian disaring berdasarkan kata kunci penyakit hati kronis dan *sleep apnea*, termasuk penyakit, gejala, dan pengobatannya, untuk memastikan bahwa twit-twit tersebut secara khusus berkaitan dengan kedua penyakit yang diteliti. Selain itu, daftar *stopword* juga digunakan untuk menyaring kata-kata yang tidak perlu (seperti subjek, preposisi, artikel, atau kata umum) dari data twit. Berkas twit yang telah disaring kemudian digabungkan dengan berkas yang bersih dan tidak terkait dengan kesehatan untuk kemudian diubah ke dalam format LIBSVM. Dengan bantuan Octave, twit yang terkait dan yang tidak terkait dengan kesehatan diklasifikasi menggunakan MATLAB. Keluaran akhir dari prapemrosesan adalah berkas label yang terprediksi. Berkas tersebut dan berkas twit digunakan untuk menghasilkan

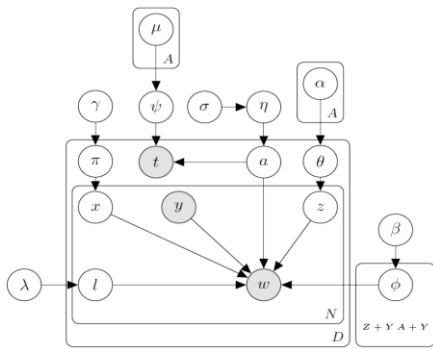
berkas kesehatan akhir yang kemudian dijadikan masukan model T-ATAM untuk penyakit hati dan *sleep apnea*.

C. Pemrosesan Data pada T-ATAM

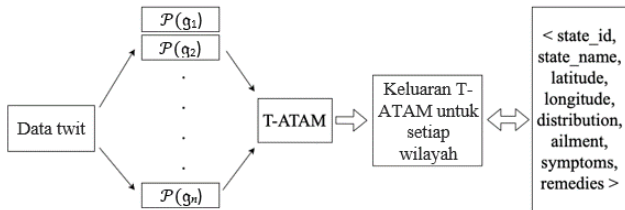
1) *Tinjauan Time-Aware Ailment Topic Aspect Model (T-ATAM)*: Pada penelitian ini, model *time-aware* yang diusulkan sebelumnya, yaitu T-ATAM [8], digunakan untuk memprediksi twit terkait kesehatan dengan menggunakan variabel yang diambil dari distribusi multinomial korpus. Proses dimulai dengan menyajikan data pemetaan twit ke dokumen dengan *timestamp* dan atribut geografis yang berbeda. Berdasarkan *dataset*, himpunan *post* $\mathcal{P} = \{\rho_1, \rho_2, \dots, \rho_n\}$ dianggap sebagai *post* yang memiliki pengertian sebagai unit pembaruan aktivitas pengguna di media sosial, salah satunya Twitter. Terdapat beberapa atribut pada sebuah *post*, di antaranya identitas tempat, *timestamp*, dan konten *post*. Selanjutnya, $\mathcal{G} = \{g_1, g_2, \dots, g_n\}$ merupakan atribut geografis setiap *post*. \mathcal{P}_g merupakan sekelompok *post* yang berasal dari wilayah $g \in \mathcal{G}$. Variabel $\mathcal{T} = \{t_1, t_2, \dots, t_n\}$ menyatakan atribut temporal untuk membagi *post* untuk menjalankan T-ATAM dan memperoleh hasil berdasarkan bulan. \mathcal{P}_g^t merujuk pada kumpulan *post* di \mathcal{P} yang berasal dari wilayah g pada periode t . Terakhir, $\mathcal{D} = \{\mathcal{D}_g^{t_1}, \mathcal{D}_g^{t_2}, \dots, \mathcal{D}_g^{t_n}\}$ digunakan untuk mendefinisikan kumpulan dokumen yang mewakili agregat twit dari wilayah g untuk *timestamp* t_i yang berbeda. Setiap variabel dijelaskan sebagai berikut.

- \mathcal{P} : *post*
- \mathcal{G} : wilayah
- \mathcal{T} : *timestamp*
- \mathcal{P}_g^t : *post* di wilayah g pada *timestamp* t
- \mathcal{D}_g^t : dokumen yang memetakan *post* $\rho \in \mathcal{P}_g^t$ ke dokumen.

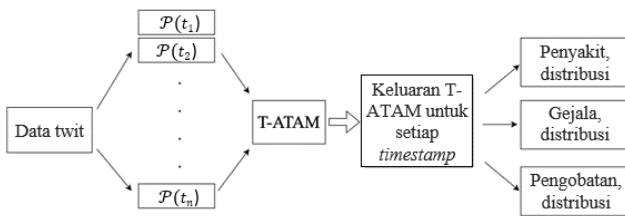
T-ATAM adalah model *time-aware* yang *timestamp* t setiap *post*-nya diperlakukan sebagai variabel acak [4]. Kata demi kata terkait kesehatan yang disaring berdasarkan penyakit unik, termasuk kanker hati, *central sleep apnea*, atau sirosis, dibangkitkan terlebih dahulu dalam dokumen sebelum dokumen twit dibuat. Kata-kata tersebut dapat menunjukkan gejala maupun pengobatan suatu penyakit; misalnya sakit punggung merupakan gejala dari kanker hati, sedangkan operasi adalah pengobatannya. Selama proses berlangsung, dua variabel acak, yaitu ℓ dan x , digunakan untuk mengidentifikasi jenis setiap kata, termasuk kata umum atau kata latar belakang. Selanjutnya, kata dapat ditarik dari distribusi kosakata yang umum ke seluruh korpus atau yang dihasilkan dari topik distribusi z . Ketika kata tersebut memiliki keterkaitan dengan kesehatan (kasus $\ell = 1$; $x = 1$), variabel acak y diizinkan untuk menentukan klasifikasi kata tersebut, termasuk *aspect-neutral* (kasus $y = 0$), gejala (kasus $y = 1$), atau pengobatan (kasus $y = 2$). Akhirnya, kata-kata disimpulkan berdasarkan penyakit yang terkait dengan dokumen. Namun, terlihat bahwa pada T-ATAM, *timestamp* setiap dokumen dihasilkan berdasarkan penyakit yang terkait dengan twit dari basis data. *Timestamp* merupakan variabel acak yang diamati, sehingga perlu diperoleh hasil satu distribusi waktu per penyakit $\{\psi, a \in A\}$ untuk mendapatkan *timestamp* twit tertentu. Selanjutnya,



Gbr. 2 Time-aware ailment topic aspect model (T-ATAM).



Gbr. 3 Pemrosesan data pada T-ATAM berdasarkan wilayah.



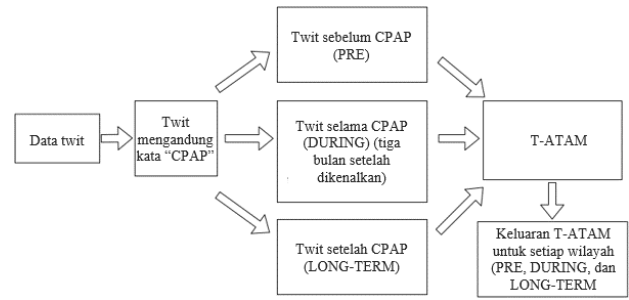
Gbr. 4 Pemrosesan data pada T-ATAM berdasarkan timestamp.

timestamp berdasarkan distribusi multinomial dengan parameter ψ_a , yang diperoleh dari distribusi Dirichlet dan ditentukan oleh vektor khusus untuk setiap penyakit, dihasilkan berdasarkan penyakit a yang terkait dengan dokumen. Proses dari model diilustrasikan pada Gbr. 2.

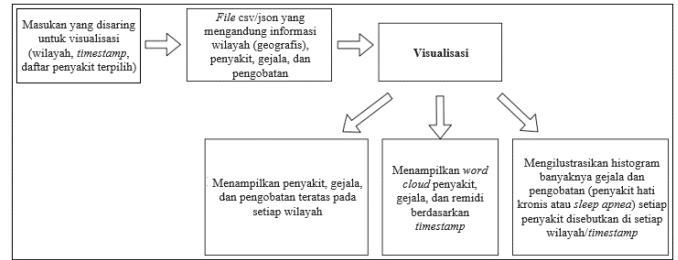
2) *Pemrosesan Data pada T-ATAM Berdasarkan Wilayah:* Bagian ini membahas tentang prediksi penyakit, gejala, dan pengobatan di setiap wilayah dataset twit. Pertama, twit p dikelompokkan dan dibagi berdasarkan wilayah g sebelum setiap data dimasukkan ke T-ATAM untuk memperoleh sebaran vektor penyakit di setiap wilayah. Dalam hal ini, penyakit teratas di setiap wilayah dengan sebaran tertinggi dipilih beserta informasi lainnya, seperti gejala dan pengobatan. Proses ini disajikan pada Gbr. 3.

3) *Pemrosesan Data pada T-ATAM Berdasarkan Timestamp:* Pada tahap ini, dilakukan pengamatan terhadap perubahan penyakit, gejala, dan pengobatan pada periode yang berbeda, yaitu periode Januari hingga Desember 2018. Twit p dikelompokkan dan dipisahkan berdasarkan timestamp t . T-ATAM dijalankan untuk memperoleh data penyakit, gejala, dan pengobatan yang terjadi pada setiap bulan. Ringkasan proses ini ditampilkan pada Gbr. 4.

4) *Pemrosesan Data pada T-ATAM Berdasarkan Penanganan:* Salah satu aspek penting dalam menganalisis twit terkait kesehatan adalah kemampuan untuk menjawab



Gbr. 5 Pemrosesan data pada T-ATAM berdasarkan pengobatan.



Gbr. 6 Proses visualisasi twit terkait kesehatan.

pertanyaan terkait kondisi sebelum dan sesudah perawatan yang diberikan. Misalnya, kondisi pasien setelah menerima pengobatan CPAP. CPAP adalah mode ventilasi pernapasan yang digunakan untuk mengobati *sleep apnea*. Dalam penelitian ini, hasil perubahan penyakit, gejala, dan pengobatan diamati sejak sebelum dan sesudah CPAP diperkenalkan. Untuk menjawab pertanyaan tersebut, dataset dibagi menjadi tiga kategori utama untuk setiap wilayah, yaitu PRE (sebelum CPAP disebutkan dalam twit), DURING (tiga bulan setelah CPAP pertama kali disebutkan), dan LONG-TERM (selambat-lambatnya tiga bulan setelah CPAP pertama kali disebutkan). Selanjutnya, T-ATAM diterapkan untuk memperoleh set-set sebagaimana ditunjukkan pada Gbr. 5.

D. Visualisasi Data

Setelah T-ATAM dijalankan, data keluaran divisualisasikan dengan cara yang berbeda-beda. Gbr. 6 menggambarkan proses visualisasi twit terkait kesehatan. Data keluaran yang dihasilkan oleh model digunakan untuk membuat visualisasi. Setelah data disaring, berkas keluaran diproses menurut wilayah dan timestamp yang berbeda. D3.js dapat dimanfaatkan untuk berbagai tujuan, termasuk menampilkan informasi penyakit di setiap wilayah pada peta, word cloud, dan histogram. Agar diperoleh data masukan yang tepat untuk pengimplementasian visualisasi, form dibuat pada front end untuk memastikan bahwa visualisasi bekerja sesuai permintaan dan menjadi lebih interaktif. Pembatasan dibagi menjadi spasial, temporal, sebelum/sesudah pengobatan, dan jenis penyakit yang diminta. Berbagai jenis tampilan dijelaskan di bawah ini.

1) *Visualisasi Data pada Peta:* Data keluaran dari T-ATAM menurut wilayah, termasuk penyakit teratas dengan sebaran, gejala, dan pengobatannya, ditampilkan pada peta.

2) *Word Cloud:* Data keluaran berdasarkan wilayah, timestamp, dan sebelum/sesudah pengobatan dapat dipilih untuk menghasilkan word cloud.

TABEL II
JUMLAH TWIT PADA SETIAP LANGKAH PRAPEMROSESAN

Langkah Prapemrosesan	Jumlah Twit	
	Penyakit Hati	Sleep Apnea
Data twit yang diperoleh	797.703.079	797.703.079
Data twit bersih	374.709.349	374.709.349
Data twit yang disaring berdasarkan <i>dataset</i> kata kunci	3.474.225	7.545.953
Data twit terkait penyakit (<i>support-vector machine classifier</i>)	331.479	1.063.445

TABEL III
PENYAKIT TERKAIT PENYAKIT HATI KRONIS DAN SLEEP APNEA

Sleep Apnea	Penyakit Hati Kronis
1. Bronkitis kronis	1. Alcoholic liver disease (ALD)
2. COPD	2. Hepatitis B
3. Mixed sleep apnea	3. Batu Empedu
4. Obstructive sleep apnea	4. Kanker hati
5. Positional sleep apnea	5. Hemokromatosis
6. Non-Cheyne-Stokes breathing central sleep apnea (CSA)	6. Kolangitis sklerosis primer
7. CSA	7. Ensefalopati Hepatik
8. Emfisema	8. Sirosis
	9. Karsinoma hepatoseluler
	10. Metastatic liver cancer
	11. Hipertensi portal
	12. Acute alcoholic hepatitis
	13. Nonalcoholic fatty liver

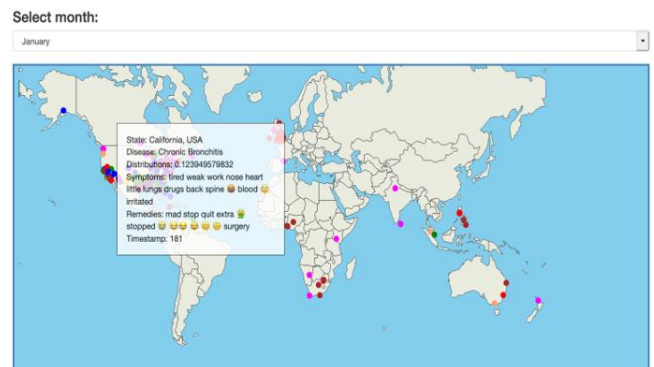
3) *Histogram*: Data keluaran berdasarkan granularitas spasial dan temporal dapat digunakan untuk menghasilkan histogram pada pemilihan pengguna, termasuk penyakit, gejala, dan pengobatan.

III. HASIL

Twit terkait penyakit hati kronis dan *sleep apnea* diperoleh setelah dilakukan prapemrosesan data twit. Berdasarkan jumlah twit pada setiap langkah yang ditunjukkan pada Tabel II, 800 juta twit berhasil terkumpul, dengan sekitar 331.000 twit terkait penyakit hati dan satu juta twit terkait *sleep apnea*. Selanjutnya, 13 macam penyakit terkait penyakit hati dan delapan macam terkait *sleep apnea* diperoleh dari hasil menjalankan T-ATAM. Macam-macam penyakit tersebut ditunjukkan pada Tabel III.

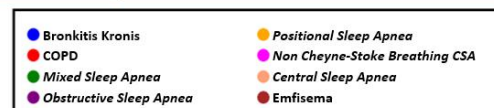
Pada *sleep apnea*, penelitian berfokus pada penggunaan istilah-istilah kunci. Diketahui bahwa pengguna memiliki cara yang berbeda-beda untuk mengekspresikan kata “snoring” dan “CPAP” dalam twit; jumlah penyebutan kata-kata tersebut tidak terlalu tinggi karena dalam kehidupan sehari-hari pengguna dimungkinkan menggunakan kata lain yang lebih sederhana. Berdasarkan *dataset*, kata “snoring” muncul di 1.084 twit di 395 wilayah, sedangkan “CPAP” muncul di 55 twit dan di 41 wilayah. Kajian dampak CPAP terhadap twit pengguna juga dilakukan untuk mengetahui perubahan penyakit, gejala, dan pengobatan *sleep apnea* di masing-masing wilayah. Istilah-istilah yang memiliki arti yang mirip dengan “snoring” adalah “snore”, “snores”, “snored”, “snoring”, “wheeze”, “wheezing”, dan “wheeze”. Di sisi lain, kata-kata yang memiliki arti yang sama dengan “CPAP” dan terapinya adalah “mask”, “masks”, “positive airway pressure (PAP)”, “Philips”, “compliance”, “humidifier”, dan “therapy”.

Data Display on World Map



(a)

Data Display on World Map



(b)

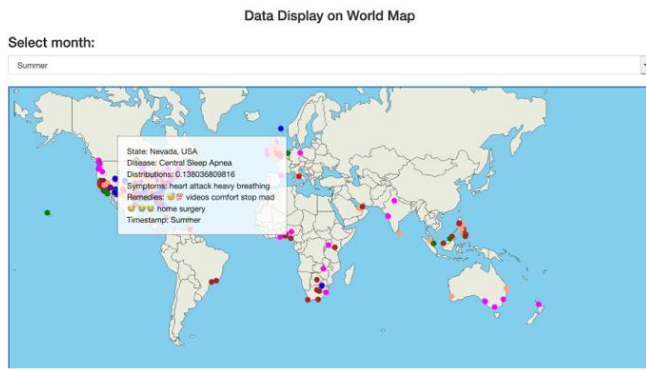
Gbr. 7 Visualisasi data pada peta untuk *sleep apnea* berdasarkan bulan, (a) Januari dan (b) Juli.

A. Tampilan Data pada Peta

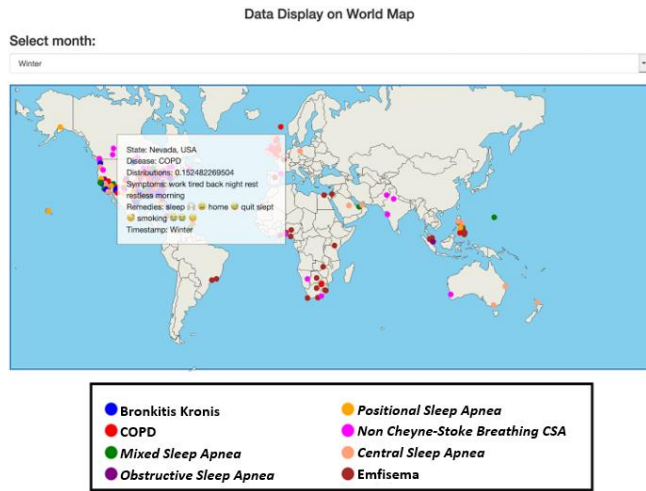
1) *Tampilan Data Berdasarkan Bulan*: Data keluaran perubahan *sleep apnea* diamati selama berbulan-bulan. Gbr. 7 menyajikan perubahan penyakit *sleep apnea* di California pada bulan Januari dan Juli. Berdasarkan gambar tersebut, terlihat jelas bahwa bronkitis kronis menjadi topik pembicaraan paling populer di Twitter pada bulan Januari. Sementara itu, diskusi beralih ke *mixed sleep apnea* enam bulan kemudian.

2) *Tampilan Data Berdasarkan Musim*: Studi data berdasarkan musim yang berbeda, termasuk musim semi, musim panas, musim gugur, dan musim dingin, dilakukan untuk meningkatkan pengamatan perubahan penyakit berdasarkan *timestamp*. Data masukan telah dipisahkan berdasarkan bulan, tetapi data dikelompokkan berdasarkan musim dengan menggunakan bulan berlangsungnya musim sebelum dimasukkan ke T-ATAM untuk menghasilkan keluaran.

Pada Gbr. 8, terlihat bahwa untuk *sleep apnea* di Nevada, *chronic obstructive pulmonary disease* (COPD) banyak disebutkan oleh pengguna dalam twit pada musim dingin. Sebaliknya, *central sleep apnea* lebih sering disebutkan selama musim panas, dengan banyak gejala serangan jantung dan



(a)



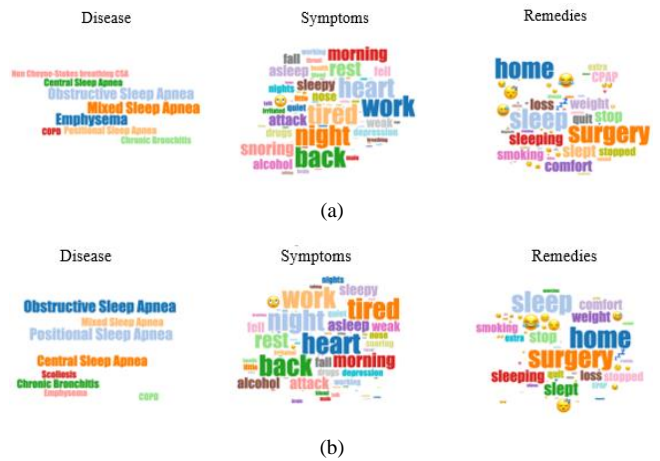
(b)

Gbr. 8 Visualisasi data pada peta untuk *sleep apnea* berdasarkan musim, (a) musim panas, (b) musim dingin.

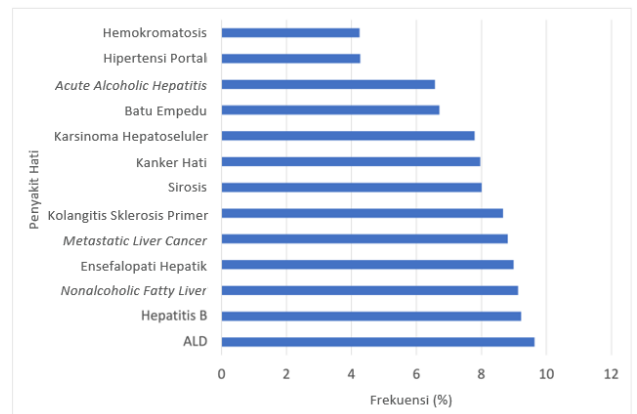
masalah pernapasan. Terkait penyakit hati, lebih banyak pengguna membuat twit tentang hepatitis B selama musim dingin di California. Sebagai perbandingan, lebih banyak pengguna membicarakan tentang *alcoholic liver disease* (ALD) selama musim panas 2018. Kebanyakan orang membicarakan tentang konsumsi alkohol di musim dingin, sehingga hasil di musim panas dapat merupakan konsekuensi dari gejala yang disebutkan di musim sebelumnya.

B. Tampilan pada Word Clouds dan Histograms

1) *Tampilan Data Berdasarkan Bulan*: Menjalankan T-ATAM berdasarkan bulan sangat penting dilakukan karena dapat memberikan informasi yang signifikan mengenai perubahan penyakit, baik penyakit hati kronis maupun *sleep apnea* pada bulan yang berbeda. Misalnya, pada Gbr. 9, *sleep apnea*, *obstructive sleep apnea*, *mixed sleep apnea*, dan *emfisema* adalah penyakit yang paling sering dibahas di Twitter pada bulan Januari. Sementara itu, *positional sleep apnea* dan *central sleep apnea* menunjukkan peningkatan yang signifikan enam bulan kemudian. Selain itu, kelelahan, kurang tidur malam, dan serangan jantung tetap menjadi gejala utama *sleep apnea* di bulan Januari dan Juli, dengan mendengkur lebih umum terjadi pada bulan Januari. Pembedahan, tidur di rumah, diet yang tepat, dan berhenti merokok tetap menjadi solusi



Gbr. 9 Word cloud untuk *sleep apnea* berdasarkan bulan, (a) Januari dan (b) Juli.



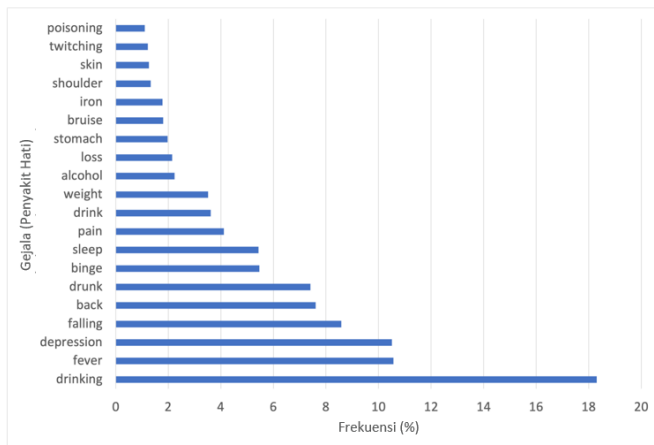
Gbr. 10 Histogram penyakit hati kronis pada musim semi.

utama yang disarankan pada media sosial, sementara CPAP lebih sering disebutkan di bulan Januari daripada di bulan Juli.

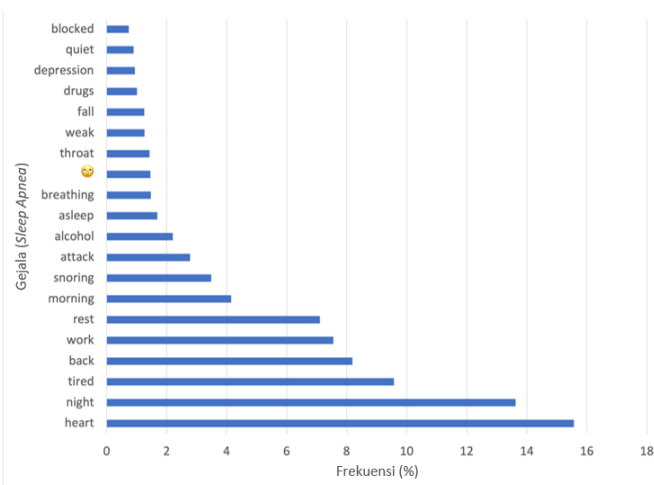
2) *Tampilan Data Berdasarkan Musim*: Selain melalui *word cloud*, hasil pengamatan juga divisualisasikan melalui histogram untuk menyajikan data keluaran kuantitatif yang lebih spesifik. Berdasarkan Gbr. 10, ALD, hepatitis B, dan *nonalcoholic fatty liver* adalah penyakit yang paling umum didiskusikan dalam percakapan di media sosial, dengan lebih dari 9% pada musim semi, sedangkan *nonalcoholic fatty liver* menggeser ALD di musim dingin.

Gbr. 11 menunjukkan bahwa konsumsi alkohol adalah gejala penyakit hati kronis yang paling banyak disebutkan di Twitter, kemudian diikuti oleh demam dan depresi. Tiga gejala teratas tersebut memiliki kombinasi hampir 40% dari total persentase semua gejala. Selanjutnya, *sleep apnea*, *obstructive sleep apnea*, *mixed sleep apnea*, dan *positional sleep apnea* merupakan diskusi yang paling populer di media sosial selama musim semi. Di sisi lain, terjadi peningkatan yang signifikan terhadap jumlah penyebutan bronkitis kronis di musim dingin.

Selain serangan jantung, kurang tidur malam, dan kelelahan, yang merupakan gejala paling umum di kedua musim, terlihat pula bahwa mendengkur dan masalah pernapasan atau masalah tenggorokan lebih banyak didiskusikan pada musim dingin. Sementara itu, sebagaimana dapat disimpulkan dari Gbr. 12,



Gbr. 11 Histogram gejala penyakit hati kronis pada musim dingin.



Gbr. 12 Histogram gejala sleep apnea pada musim dingin.



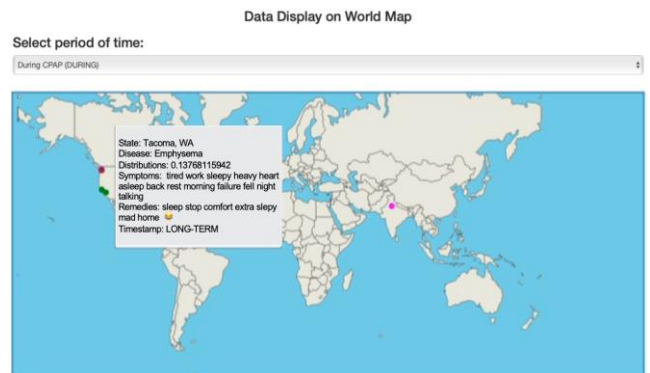
Gbr. 13 Tampilan data sebelum CPAP.

depresi dan jatuh tampaknya lebih banyak disebutkan pada twit pada musim panas.

3) *Tampilan Data Sebelum/Setelah CPAP:* Terdapat beberapa perubahan penyakit dengan gejala dan pengobatannya di setiap periode. Gbr. 13 sampai Gbr. 15 menunjukkan perubahan penyakit di Tacoma. Perubahan tersebut meliputi bronkitis kronis pada PRE dan DURING dengan gejala mengantuk, obat-obatan, gangguan hidung, dan gangguan tenggorokan, menjadi emfisema pada LONG-TERM dengan



Gbr. 14 Data displays during CPAP.



Gbr. 15 Tampilan data setelah CPAP.

gejala meliputi kelelahan, serangan jantung, dan mengantuk di pagi hari. Pada saat yang sama, tidak ditemukan gejala tentang masalah pernapasan, seperti masalah tenggorokan atau hidung. Oleh karena itu, perubahan sebaran penyakit di setiap wilayah setelah CPAP disebutkan dalam twit menarik untuk diamati karena dapat menunjukkan efektivitas pengobatan dan perkembangan penyakit setelah satu pengobatan diperkenalkan.

IV. DISKUSI DAN BATASAN

Media sosial dapat digunakan untuk mendiagnosis gejala penyakit karena pengguna dapat menggambarkan kondisi yang dialami dengan kata-kata sendiri. Namun, penting juga untuk menjadi pertimbangan bahwa informasi di media sosial berasal dari lima perspektif yang berbeda dan latar belakang yang terbatas. Oleh karena itu, kebenaran informasi yang dibagikan di media sosial tidak dapat dipastikan karena informasi tersebut singkat, tidak terverifikasi, dan tidak terstruktur.

Meskipun cakupan geografis terbatas karena penetrasi internet yang rendah di beberapa wilayah, media sosial memungkinkan pengguna untuk berbagi dan mengumpulkan informasi dengan lebih nyaman. Berdasarkan kumpulan twit, terlihat bahwa lebih banyak orang berbagi informasi tentang sleep apnea daripada penyakit hati. Penelitian sebelumnya menyebutkan bahwa pada tahap awal, orang dengan penyakit hati tidak merasakan keluhan apapun. Oleh karenanya, penderita mungkin tidak menyadari adanya gejala atau tanda penyakit [19]. Selain itu, sleep apnea adalah penyakit yang lebih umum terjadi dan secara langsung memengaruhi kehidupan sehari-hari. Misalnya, ketika merasa mengantuk atau lelah, pengguna internet mungkin membuat post di media

sosial untuk sekadar berbagi informasi tentang kondisi ataupun mengundang pengguna media sosial lainnya untuk mendiskusikan masalah yang dialami. Selain itu, meskipun konsumsi alkohol adalah penyebab penyakit hati yang paling umum, depresi adalah masalah yang lebih sering dibahas, sebagaimana ditunjukkan pada Gbr. 12. Dapat disimpulkan bahwa depresi merupakan akar dari permasalahan lain, seperti konsumsi alkohol, demam, dan kehilangan berat badan, yang juga turut berkontribusi terhadap penyakit hati. Lebih lanjut, penelitian menunjukkan bahwa kurang tidur malam secara aktif dibahas di media sosial sebagai salah satu gejala utama *sleep apnea* yang mengarah ke komplikasi lain, termasuk serangan jantung dan kantuk di pagi hari. Berdasarkan penelitian sebelumnya dengan subjek mahasiswa, mahasiswa memiliki durasi tidur paling pendek, bahkan pola tidur yang tidak teratur, dibandingkan dengan orang dewasa lainnya [20].

V. KESIMPULAN

Menggunakan T-ATAM untuk memprediksi penyakit, gejala, dan pengobatan dari tweet dapat memberikan informasi bermanfaat tentang penyakit hati dan *sleep apnea* berdasarkan wilayah dan *timestamp*, sehingga perubahan dapat diamati di area dan periode tertentu. Selain itu, T-ATAM juga digunakan untuk menguji pengaruh beberapa pengobatan dan perkembangan penyakit. Data dapat disaring dan dianalisis menurut periode sebelum/sesudah penggunaan pengobatan tersebut. Hasilnya menunjukkan bahwa informasi ini dapat digunakan untuk lebih memahami persepsi dan pengalaman orang terhadap penyakit dan pengobatannya. Meskipun terdapat keterbatasan, hasil penelitian ini dapat menjadi titik awal yang bermanfaat bagi penelitian selanjutnya, sehingga teknologi di bidang kesehatan dapat lebih dikembangkan.

KONFLIK KEPENTINGAN

Penulis menyatakan bahwa tidak ada konflik kepentingan dalam penelitian ini.

KONTRIBUSI PENULIS

Konseptualisasi, Bonpagna Kann dan Sihem Amer-Yahia; metode, Bonpagna Kann dan Sihem Amer-Yahia; perangkat lunak, Bonpagna Kann dan Michael Ortega; validasi, Bonpagna Kann, Sihem Amer-Yahia, dan Jean-Louis Pépin; analisis formal, Bonpagna Kann; investigasi, Bonpagna Kann dan Sihem Amer-Yahia; sumber daya, Bonpagna Kann, Sihem Amer-Yahia, Jean-Louis Pépin, dan Sébastien Bailly; kurasi data, Bonpagna Kann; penulisan—penyusunan draf asli, Bonpagna Kann; penulisan—peninjauan dan penyuntingan, Bonpagna Kann dan Sihem Amer-Yahia; visualisasi, Bonpagna Kann; pengawasan, Sihem Amer-Yahia, Jean-Louis Pépin, dan Sébastien Bailly; administrasi proyek, Jean-Louis Pépin, dan Sébastien Bailly; akuisisi pendanaan, Jean-Louis Pépin, dan Sébastien Bailly.

UCAPAN TERIMA KASIH

Penelitian ini didukung oleh Proyek CDP LIFE dengan kerjasama antara Laboratoire d'Informatique de Grenoble dan Rumah Sakit CHU Grenoble.

REFERENSI

- [1] S. Dorle dan N. Pise, "Political Sentiment Analysis through Social Media," *2018 Second Int. Confer. Comput. Methodol., Commun. (ICCMC)*, 2018, hal. 869–873.
- [2] C. Udanor, S. Aneke, dan B.O. Ogbuokiri, "Determining Social Media Impact on The Politics of Developing Countries Using Social Network Analytics," *Program*, Vol. 50, No. 4, hal. 481–507, Sep. 2016.
- [3] G. Wolfsfeld, E. Segev, dan T. Sheaffer, "Social Media and the Arab Spring," *Int. J. Press/Politics*, Vol. 18, No. 2, hal. 115–137, Jan. 2013.
- [4] N. Hoog dan P. Verboon, "Is the News Making Us Unhappy? The Influence of Daily News Exposure on Emotional States," *Brit. J. Psychol.*, Vol. 111, No. 2, hal. 157–173, Mei 2020.
- [5] S. Petrović, M. Osborne, dan V. Lavrenko, "Streaming First Story Detection with Application to Twitter: Human Language Technologies," *2010 Annu. Confer. North Amer. Chapter, Assoc. Comput. Linguist.*, 2020, hal. 181–189.
- [6] T. Sakaki, M. Okazaki, dan Y. Matsuo, "Earthquake Shakes Twitter Users," *Proc. 19th Int. Confer. World Wide Web (WWW '10)*, 2010, hal. 851–860.
- [7] S.K.T. Bhavaraju, C. Beyney, dan C. Nicholson, "Quantitative Analysis of Social Media Sensitivity to Natural Disasters," *Int. J. Disaster Risk Reduct.*, Vol. 39, hal. 1–10, Okt. 2019.
- [8] S. Sidana, dkk., "Health Monitoring on Social Media over Time," *Proc. 39th Int. ACM SIGIR Conf. Res., Develop. Inf. Retr.*, 2016, hal. 1–6.
- [9] S. Shrivastava dan P.S. Nair, "Mood Prediction on Tweets Using Classification Algorithm," *Int. J. Sci., Res. (IJSR)*, Vol. 4, No. 11, hal. 295–299, Nov. 2015.
- [10] N.S. Cook, dkk., "Patients' Perspectives on COPD: Findings from a Social Media Listening Study," *ERJ Open Res.*, Vol. 5, No. 1, hal. 1–10, Feb. 2019.
- [11] S. Gohil, S. Vuik, dan A. Darzi, "Sentiment Analysis of Health Care Tweets: Review of the Methods Used," *JMIR Public Health, Surveill.*, Vol. 4, No. 2, hal. 1–10, Apr. 2018.
- [12] M.P. Hamm, dkk., "Social Media Use Among Patients and Caregivers: A Scoping Review," *BMJ Open*, Vol. 3, No. 5, hal. 1–9, Mei 2013.
- [13] D. Murthy dan M. Eldredge, "Who Tweets About Cancer? An Analysis of Cancer-Related Tweets in the USA," *Digit. Health*, Vol. 2, hal. 1–16, Jul. 2016.
- [14] Y.T. Yang, M. Horneffer, dan N. DiLisio, "Mining Social Media and Web Searches for Disease Detection," *J. Pub. Health Res.*, Vol. 2, No. 1, hal. 17–21, Mei 2013.
- [15] V. Lampos dan N. Cristianini, "Tracking the Flu Pandemic by Monitoring the Social Web," *2010 2nd Int. Workshop Cogn. Inf. Process.*, 2010, hal. 411–416.
- [16] E.M. Glowacki, A.J. Lazard, dan G.B. Wilcox, "E-Cigarette Topics Shared by Medical Professionals: A Comparison of Tweets from the United States and United Kingdom," *Cyberpsychol., Behav., Soc. Netw.*, Vol. 20, No. 2, hal. 133–137, Feb. 2017.
- [17] A. Culotta, "Towards Detecting Influenza Epidemics by Analyzing Twitter Messages," *Proc. First Workshop Soc. Media Anal. (SOMA '10)*, 2010, hal. 115–122.
- [18] P.M. Polgreen, Y. Chen, D.M. Pennock, dan F.D. Nelson, "Using Internet Searches for Influenza Surveillance," *Clin. Infect. Dis.*, Vol. 47, No. 11, hal. 1443–1448, Des. 2008.
- [19] E.A. Tsochatzjis, J. Bosch, dan A.K. Burroughs, "Liver Cirrhosis," *Lancet*, Vol. 383, No. 9930, hal. 1749–1761, Mei 2014.
- [20] A. Schlarb, A. Friedrich, dan M. Claßen, "Sleep Problems in University Students – An Intervention," *Neuropsychiatr. Dis. Treat.*, Vol. 13, hal. 1989–2001, Jul. 2017.