

MELAMPAUI TRADISI: ANALISIS GALIAN TEKS UNTUK DATA KUALITATIF BERSKALA BESAR

Omar K. Burhan

Departemen Psikologi, Universitas Brawijaya, Malang, Indonesia

Korespondensi: burhan.ok@ub.ac.id

BEYOND TRADITION: TEXT MINING ANALYSIS FOR LARGE-SCALE QUALITATIVE DATA

Manuscript type: Editorial Note

Abstract

Recent advances in digital technology have made it possible to access and analyze large amounts of data, providing scholars in psychology with new ways to gain insights from qualitative sources such as social media, news articles, and interviews. Traditional methods, however, struggle to handle such large data volumes effectively. Text mining offers an efficient solution, using algorithms to find patterns and themes in large text datasets. This article introduces text mining techniques for analyzing large-scale qualitative data, using two issues of Jurnal Psikologi Ulayat (JPU) as a case study. Using techniques such as word frequency analysis, word clouds, bi-grams, and pairwise correlations, five main themes were identified: (1) Indonesian-based psychology, (2) mental health, (3) interpersonal and family relationships, (4) psychometric assessments, and (5) social phenomena related to fandom and pop culture. This study demonstrates the potential of text mining as an efficient approach to uncover patterns and themes in large qualitative datasets.

Article history:

Received 14 October 2024
Received in revised form 14 November 2024
Accepted 14 November 2024
Available online 1 December 2024

Keywords:

Jurnal Psikologi Ulayat
qualitative data
research methodology
text mining

Abstrak

Kemajuan teknologi digital telah memungkinkan akses dan analisis data skala besar, memberikan peluang bagi peneliti psikologi untuk menggali wawasan baru dari data kualitatif seperti media sosial dan artikel berita. Namun, metode tradisional memiliki keterbatasan dalam menangani volume data yang besar. Galian teks (text mining) hadir sebagai solusi yang efisien, menggunakan algoritma untuk menemukan pola dan tema dalam kumpulan teks besar. Artikel ini memperkenalkan teknik galian teks untuk menganalisis data kualitatif skala besar dengan studi kasus dua edisi Jurnal Psikologi Ulayat (JPU). Melalui analisis frekuensi kata, awan kata, bi-gram, dan pairwise correlation, lima tema utama teridentifikasi: (1) psikologi berbasis Indonesia, (2) kesehatan mental, (3) hubungan antar individu dan keluarga, (4) pengukuran psikometris, dan (5) fenomena sosial terkait fandom dan budaya pop. Ilustrasi ini menunjukkan potensi galian teks sebagai pendekatan efisien untuk mengidentifikasi pola dan tema dalam data kualitatif berskala besar.

Kata Kunci: data kualitatif, galian teks, metodologi penelitian

Dalam beberapa tahun terakhir, perkembangan teknologi digital telah memungkinkan para peneliti psikologi mengakses dan mengolah data dalam jumlah yang jauh lebih besar dibanding sebelumnya. Kemajuan ini memberikan peluang besar untuk menggali wawasan baru dari data kualitatif yang berlimpah, seperti media sosial, artikel berita, atau teks wawancara. Meski demikian, metode tradisional yang masih digunakan oleh banyak peneliti kualitatif di Indonesia sering tidak efisien untuk menangani volume data yang besar.

Metode tradisional, seperti wawancara mendalam dan analisis manual dari transkrip, memang memberikan hasil yang kaya secara kontekstual, namun memiliki keterbatasan ketika harus menangani ratusan atau ribuan teks. Proses ini memerlukan waktu, tenaga, dan sumber daya manusia

yang besar, dan sering kali sulit untuk diterapkan pada data berskala besar seperti komentar di media sosial atau ribuan artikel berita daring.

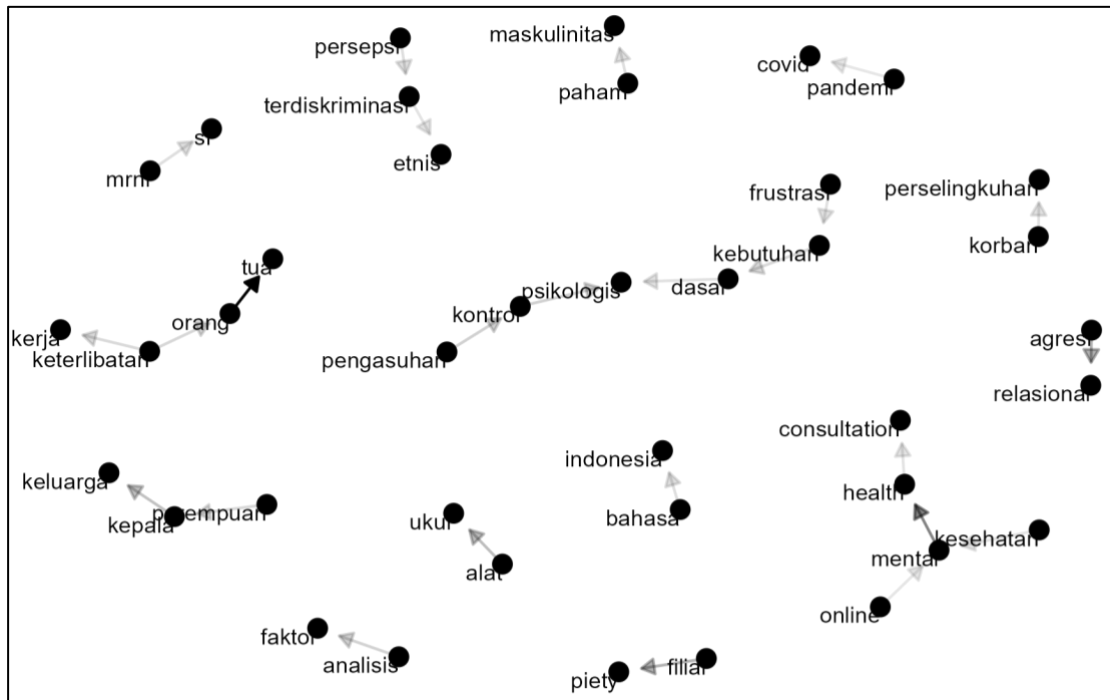
Galian teks, atau yang dalam bahasa Inggris dikenal sebagai *text mining*, adalah metode untuk mengekstrak informasi bermakna dari kumpulan teks yang besar (Thakur & Kumar, 2021). Metode ini menggunakan algoritma komputer untuk menemukan pola, tema, dan hubungan antara kata-kata dalam data teks. Galian teks menawarkan efisiensi yang lebih tinggi dibanding analisis manual, sekaligus memberikan hasil yang lebih terukur dan dapat direplikasi.

Catatan editorial ini memperkenalkan teknik galian teks untuk menganalisis data kualitatif berskala besar. Sebagai contoh, galian teks dapat membantu peneliti Indonesia dalam memahami bagaimana masyarakat memandang fenomena psikologis melalui teks daring, seperti ulasan kesehatan mental di media sosial atau opini publik tentang isu-isu politik yang diekspresikan di artikel berita. Teknik ini juga dapat digunakan untuk menganalisis hasil wawancara atau survei kualitatif dalam skala besar, memberikan hasil yang lebih cepat tanpa mengorbankan kedalaman analisis. Dalam konteks terapi klinis, psikolog dapat mengeksplorasi teks catatan terapi, transkrip, atau diari tentang klien dengan menciptakan otomatisasi untuk menelusuri perkembangan dalam proses terapi (Smink dkk., 2019).

Sejalan dengan visi dan misi Jurnal Psikologi Ulayat (JPU), pemanfaatan galian teks dapat membantu menemukan pola-pola psikologis yang unik dalam masyarakat Indonesia. Selama lebih dari 12 tahun, JPU telah berkomitmen untuk menerbitkan hasil penelitian yang mencerminkan kekhasan psikologi Indonesia. Analisis galian teks terhadap artikel-artikel yang dipublikasikan JPU berpotensi menciptakan peta fenomena psikologis yang khas bagi Indonesia.

Sebagai ilustrasi galian teks, artikel ini akan menyajikan analisis sederhana menggunakan perangkat lunak R. Ada banyak pendekatan untuk melakukan galian teks dengan R. Dalam ilustrasi ini, penulis menggunakan pendekatan “*tidytext*” dari Silge dan Robinson (2017). Secara spesifik, penulis menganalisa terbitan JPU Volume 11 No. 1 tahun 2024 dan Volume 10 No. 2 tahun 2023, dengan total 17 artikel dan data kotor sebanyak 104.133 kata. Sebagai catatan, analisis ini memang terbatas dan tidak lengkap mengingat tujuannya sebagai demonstrasi.

Untuk mendapatkan gambaran konten dan tema-tema di dua terbitan tersebut, penulis membuat analisis frekuensi kata (Figur 1), awan-kata (Figur 2), *bi-graph* (Figur 3), dan *pairwise-correlation* (Figur 4). Detail langkah kode-kode R yang digunakan dapat pembaca lihat di laman *Open Science Framework* (OSF) penulis (<https://osf.io/u25z3/>). Berikut adalah hasil analisis galian teks sederhana terhadap dua terbitan JPU yang dimaksud.

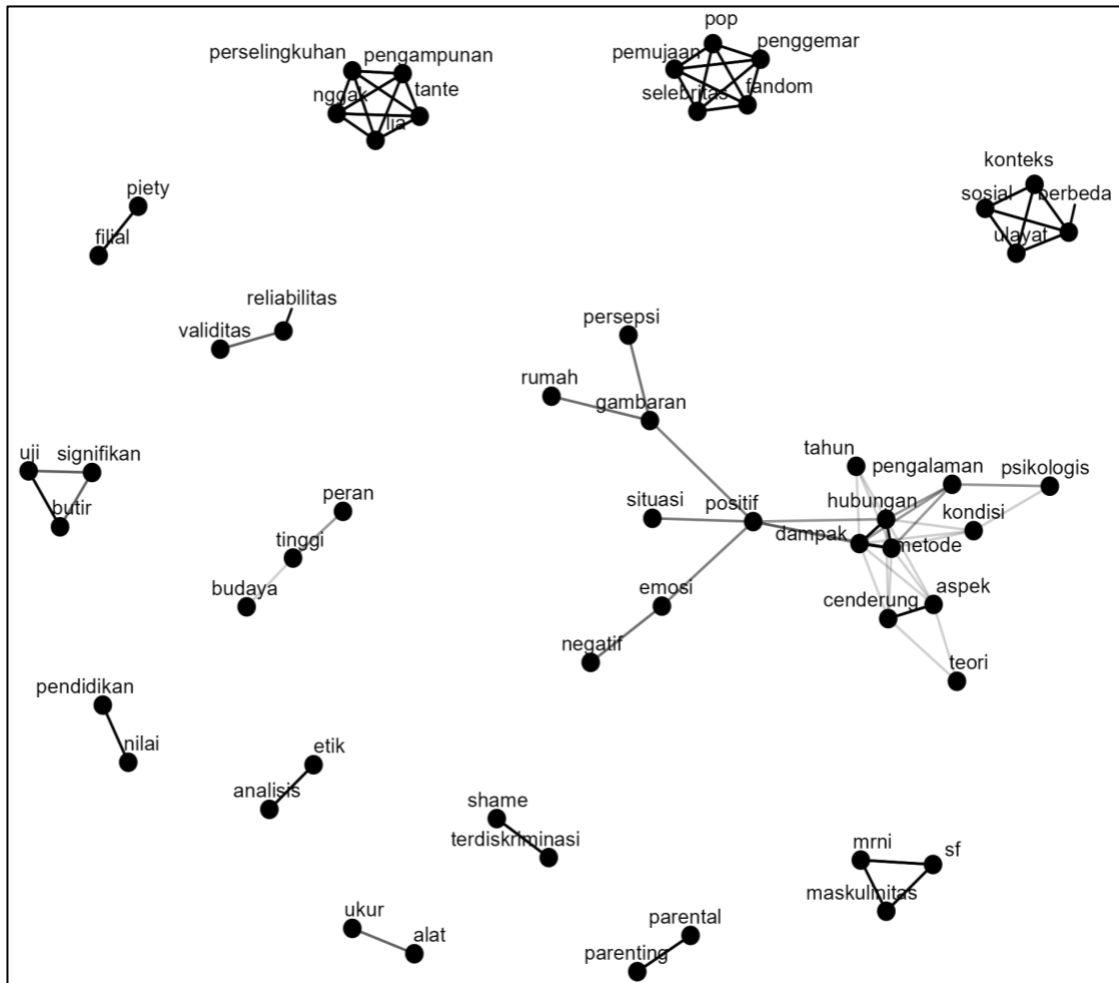


Figur 3. Jaringan Bi-gram

Dari *bi-gram* (Figur 3), kombinasi seperti “kesehatan mental”, “*online mental*”, “*mental health*”, dan “*health consultation*” menunjukkan fokus pada isu-isu kesehatan mental, termasuk konsultasi daring yang merupakan isu kontemporer yang cukup hangat di bidang psikologi modern. Frasa “pandemi Covid” menunjukkan perhatian pada dampak Covid-19 terhadap aspek psikologis dan sosial. Tema “pengasuhan kontrol”, “kontrol psikologis”, “psikologis dasar”, “kebutuhan dasar”, dan “frustrasi kebutuhan” menyoroti gaya pengasuhan dan kebutuhan psikologis dasar. Kata “*filial piety*” mengindikasikan adanya perhatian terhadap nilai-nilai kesalehan atau kewajiban anak terhadap orang tua, yang relevan dalam konteks budaya keluarga Indonesia. Frasa seperti “keterlibatan orang”, “keterlibatan kerja”, “korban perselingkuhan”, dan “agresi relasional” mencerminkan dinamika sosial dan konflik dalam hubungan. Analisis ini juga kembali memaparkan fokus JPU pada isu gender, sebagaimana tampak pada frasa “perempuan kepala”, “kepala keluarga”, dan “paham maskulinitas”, yang mungkin terkait dengan pembahasan akan perempuan sebagai kepala keluarga. Secara umum, jaringan *bi-gram* ini menegaskan fokus JPU pada isu-isu kesehatan mental, keluarga, pengasuhan, dan nilai-nilai keluarga dalam konteks sosial dan budaya Indonesia.

Terakhir adalah analisis *pairwise correlation*. Bila *bi-gram* menunjukkan kemunculan dua kata secara berurutan, *pairwise correlation* menghitung korelasi antara dua kata dari berapa kali dua kata tersebut muncul bersama-sama dalam sebuah teks, tetapi tidak harus berurutan (misalnya dalam paragraf atau kalimat yang sama). Figur 4 menunjukkan *pairwise correlation* kata-kata. Sebagai

catatan, untuk menyederhanakan, hanya padanan kata yang muncul lebih dari 50 kali dan yang berkorelasi sangat kuat yang disajikan di sini ($\Phi > .80$).



Figur 4. Pairwise Correlation

Figur 4 menunjukkan beberapa pola yang berulang dari analisis sebelumnya. Misalnya, kata “*filial*” berkorelasi kuat dengan kata “*piety*”, hal ini tentunya tidak mengejutkan mengingat dua kata tersebut membentuk sebutan sebuah konstruk. Selain itu, kata “*validitas*” berasosiasi kuat dengan “*reliabilitas*”, korelasi kedua kata ini ditandem dengan korelasi lainnya (misalnya: “*alat*” dan “*ukur*”; “*uji*”, “*butir*”, “*signifikan*”) mengindikasikan fokus JPU terhadap pengukuran yang valid dan reliabel dalam setiap artikel penelitian yang dipublikasikan. Interelasi antara kata-kata “*konteks*”, “*sosial*”, “*berbeda*”, dan “*ulayat*” mengindikasikan perhatian analitis yang besar terhadap budaya dan konteks. Selain itu, korelasi kata-kata seperti “*fandom*”, “*selebritas*”, “*pemujaan*”, dan “*pop*” mengindikasikan adanya bahasan pada isu kontemporer pada kaula muda Indonesia, mungkin terkait dengan *K-Pop*. Terakhir, korelasi antara “*perselingkuhan*”, “*pengampunan*” dan “*korban*” menunjukkan adanya perhatian terhadap dampak negatif dalam dinamika relasi, khususnya terkait

konflik dalam hubungan. Figur ini secara keseluruhan mencerminkan pendekatan komprehensif JPU terhadap isu-isu penting dalam konteks psikologi, sosial dan budaya Indonesia.

Berdasarkan analisis yang telah dibuat, tema-tema yang muncul dari JPU Volume 11, No.1 tahun 2024 dan Volume 10 No. 2 tahun 2023 disajikan di Tabel 1.

Tabel 1.

Tema-tema dalam JPU Volume 11 No. 1 Tahun 2024 dan Volume 10 No. 2 Tahun 2023

Tema	Sumber Analisis	Deskripsi
Psikologi berbasis Indonesia	Frekuensi kata (Figur 1), Awan kata (Figur 2), <i>Pairwise correlation</i> (Figur 4)	Kata “penelitian” dan “Indonesia” dalam Figur 1 menunjukkan fokus utama JPU pada penelitian berbasis konteks lokal Indonesia. Kata “etnis” dan “budaya” dalam Figur 2 memperkuat perhatian JPU pada isu ke-Indonesiaan, sementara korelasi kata-kata seperti “konteks”, “sosial”, dan “ulayat” dalam Figur 4 menunjukkan analisis terhadap budaya dan konteks sosial Indonesia.
Kesehatan mental	Frekuensi kata (Figur 1), Awan kata (Figur 2), <i>Bi-gram</i> (Figur 3), <i>Pairwise correlation</i> (Figur 4)	Kombinasi kata seperti “kesehatan mental”, “ <i>online mental</i> ”, “ <i>mental health</i> ”, dan “ <i>health consultation</i> ” menyoroti perhatian JPU terhadap kesehatan mental, termasuk layanan konsultasi daring yang menjadi isu kontemporer. Istilah terkait Covid-19, seperti “Pandemi covid”, menunjukkan perhatian pada dampak pandemi terhadap kesehatan mental dan sosial.
Hubungan antarindividu dan keluarga	Frekuensi kata (Figur 1), Awan kata (Figur 2), <i>Bi-gram</i> (Figur 3), <i>Pairwise correlation</i> (Figur 4)	Kata-kata seperti “orang”, “anak”, “keluarga”, dan “pengasuhan” dalam Figur 1 dan 2 mencerminkan fokus JPU pada dinamika keluarga dan hubungan antarindividu. Frasa seperti “ <i>filial piety</i> ”, “pengasuhan kontrol”, dan “kontrol psikologis” dalam Figur 3 mengindikasikan perhatian terhadap nilai-nilai keluarga, dinamika pengasuhan, dan kewajiban anak kepada orang tua. Korelasi antara kata “perselingkuhan”, “pengampunan”, dan “korban” dalam Figur 4 menunjukkan perhatian terhadap dinamika konflik dalam hubungan.
Pengukuran psikometris	<i>Bi-gram</i> (Figur 3), <i>Pairwise correlation</i> (Figur 4)	Korelasi kuat antara kata “validitas” dan “reliabilitas”, serta kata-kata terkait seperti “alat”, “ukur”, “uji”, dan “butir”, menekankan fokus JPU pada pentingnya pengukuran yang valid dan reliabel dalam penelitian psikologi yang diterbitkan.
Fandom dan budaya pop	<i>Bi-gram</i> (Figur 3), <i>Pairwise correlation</i> (Figur 4)	Korelasi antara kata “fandom”, “selebritas”, “pemujaan”, dan “pop” menunjukkan minat JPU pada fenomena sosial terkait idolasi selebritas dan budaya pop, mungkin khususnya dalam konteks kaula muda Indonesia dan tren <i>K-Pop</i> .

Penutup

Penulis telah memperkenalkan secara singkat teknik galian teks. Penulis menganalisis tema-tema yang muncul di dua isu terbitan JPU untuk mengilustrasikan kebermanfaatan teknik ini. Penulis juga melampirkan teknis analisisnya (termasuk persiapan, data, dan kode-kode R) di laman OSF (<https://osf.io/u25z3/>). Dua isu JPU yang menjadi perhatian mengemukakan lima tema yang sesuai dengan komitmen JPU sebagai jurnal yang berciri ke-Indonesia-an, yaitu: (1) psikologi berbasis Indonesia, (2) kesehatan mental, (3) hubungan antar individu dan keluarga, (4) pengukuran psikometris, dan (5) fandom dan budaya pop. Terakhir, penulis menegaskan bahwa tujuan pengenalan teknik galian teks ini bukan untuk menggantikan teknik analisis kualitatif tradisional, melainkan untuk memberikan tambahan alat yang dapat meningkatkan kualitas dan keandalan penelitian kualitatif. Akhir kata, penulis berharap semakin banyak peneliti psikologi Indonesia yang memanfaatkan teknik ini.

REFERENSI

- Silge, J., & Robinson, D. (2017). *Text mining with R: A tidy approach*. O'Reilly Media. <https://www.tidytextmining.com/>
- Smink W., Sools, A. M., van der Zwaan, J. M., Wiegiersma, S., Veldkamp, B. P., & Westerhof, G. J. (2019) Towards text mining therapeutic change: A systematic review of text-based methods for Therapeutic Change Process Research. *PLoS ONE*, 14(12), e0225703. <https://doi.org/10.1371/journal.pone.0225703>
- Thakur, K., & Kumar, V. (2021). Application of text mining techniques on scholarly research articles: Methods and tools. *New Review of Academic Librarianship*, 27(3–4), 302–326. <https://doi.org/10.1080/13614533.2021.1918190>