

KOMBINASI ANALISIS BIBLIOMETRIK DENGAN *LATENT DIRICHLET ALLOCATION* SEBAGAI PEMODELAN TOPIK *CASHLESS SOCIETY*

Hanifa Sepriadi^{*1}, Cindy Rudianta², Atiek Iriany³, Solimun⁴, Adji Achmad Rinaldo Fernandes⁵

^{1,2,3,4,5}Universitas Brawijaya, Malang

Email: ¹sepriadi1412@student.ub.ac.id, ²rudiantacindy@student.ub.ac.id, ³atiek.irianny@gmail.com,

⁴solimun@ub.ac.id, ⁵fernandes@ub.ac.id

*Penulis Korespondensi

(Naskah masuk: 16 September 2024, diterima untuk diterbitkan: 14 April 2025)

Abstrak

Era digitalisasi dan komputasi telah dimulai, ditandai dengan munculnya teknologi digital yang merasuk ke berbagai aspek kehidupan, sementara data juga terus berkembang menjadi *big data*. Setelah era covid 19, metode pembayaran *non-tunai* berkembang sangat pesat, sehingga banyak penelitian mengenai *cashless society*. Tujuan dari penelitian ini adalah memodelkan topik-topik yang berkaitan dengan *cashless society* untuk mendapatkan variabel dan indikator yang terkait dengan menggunakan analisis bibliometrik dan *latent dirichlet allocation*. Data penelitian ini berasal dari artikel publikasi ilmiah dan hasil *web scrapping* di *twitter* yang bertemakan *cashless society*. Hasil penelitian menunjukkan bahwa terdapat 5 variabel dan 21 indikator yang berhubungan dengan *cashless society*.

Kata kunci: *Big Data, Cashless Society, Bibliometric, Latent Dirichlet Allocation, Topic Modeling*

COMBINATION OF BIBLIOMETRIC ANALYSIS WITH LATENT DIRICHLET ALLOCATION AS TOPIC MODELING OF CASHLESS SOCIETY BEHAVIOR

Abstract

The era of digitalization and computing has begun, marked by the emergence of digital technology which permeates various aspects of life, while data also continues to develop into big data. After the covid 19 era, non-cash payment methods developed very rapidly, so there were many studies on the cashless society. The purpose of this research is to model topics related to the cashless society to obtain related variables and indicators using bibliometric analysis and latent dirichlet allocation. This research data comes from scientific publication articles and web scrapping results on twitter with the theme of cashless society. The results showed that there are 5 variables and 21 indicators related to cashless society.

Keywords: *Big Data, Cashless Society, Bibliometric, Latent Dirichlet Allocation, Topic Modeling*

1. PENDAHULUAN

Era digitalisasi dan komputasi telah dimulai, ditandai dengan munculnya teknologi digital yang merasuk ke berbagai aspek kehidupan, sementara data juga terus berkembang menjadi *big data*. *Big data* adalah istilah yang digunakan untuk menggambarkan data dalam jumlah besar, kompleks, dan bervariasi yang diperoleh dari berbagai sumber, termasuk penggunaan teknologi digital dan sensor. Terdapat potensi besar dalam *big data*, yaitu wawasan inovatif, memahami masalah, bahkan memprediksi sesuatu berdasarkan data (Han dkk., 2022).

Terdapat beberapa cara untuk mengumpulkan data untuk analisis *big data*, salah satunya adalah

web scrapping. Metode *web scrapping* akan mengumpulkan data dari sebuah *website* dengan cara mengekstrak elemen teks pada HTML (*Hypertext Markup Language*). HTML mendefinisikan dan menampung informasi atau data yang ada di dalam sebuah *website* (Chapagain, 2019). Mengekstrak data teks disebut juga dengan *text mining*. Manfaat dari *text mining* adalah menemukan informasi berharga yang tersembunyi di dalam data teks yang terstruktur (Feinerer dkk., 2008).

Hasil dari *text mining* dilanjutkan dengan pemodelan topik. *Topic modeling* adalah menganalisa kata-kata yang terdapat pada sebuah teks dan memodelkannya ke dalam bahasa alami. *Topic modeling* sering digunakan pada berbagai data

teks seperti pemilihan email, artikel, koran, *twitter*, dan masih banyak lainnya. *Topic modeling* memiliki 2 metode yang umum digunakan, yaitu *Latent Semantic Analysis* (LSA) dan *Latent Dirichlet Allocation* (LDA). LDA bekerja lebih stabil untuk memproses data dalam jumlah besar jika dibandingkan dengan LSA karena LDA merupakan model topik dengan probabilistic generative (Blei dkk.,2003).

Garbhapu & Bodapati (2020) membandingkan LDA dengan LSA untuk memodelkan topik dari Alkitab bahasa Inggris. Hasil dari penelitian ini menunjukkan bahwa LDA menunjukkan nilai koherensi yang lebih baik yaitu 0.58018 dibandingkan dengan LSA dengan nilai koherensi 0.50395. Penelitian yang dilakukan oleh [5] juga menunjukkan bahwa LDA memiliki performa yang lebih baik dibandingkan dengan LSA dengan memodelkan topik-topik pada dataset berita BBC. Penelitian yang dilakukan oleh Kalapeli dkk. (2020) dengan menggunakan data teks publikasi ilmiah juga menyimpulkan bahwa LDA bekerja lebih baik daripada LSA.

Kebiasaan masyarakat dalam bertransaksi perlahan berubah sejak era COVID-19, masyarakat lebih sering menggunakan pembayaran non-tunai dengan menggunakan dompet digital atau e-wallet untuk memudahkan transaksi dan membuat transaksi lebih aman. *World Health Organization* (WHO) menghimbau masyarakat untuk menerapkan pembayaran nirsentuh atau yang dikenal dengan pembayaran *non-tunai*. Tujuan dari pembayaran tanpa uang tunai adalah untuk mengurangi kontak fisik dengan pembayaran tunai dengan menggunakan dompet digital atau transaksi elektronik (Van & Wolf, 2019). Oleh karena itu, tujuan dari penelitian ini adalah memodelkan topik-topik yang berkaitan dengan perilaku *cashless* menggunakan kombinasi analisis bibliometrik dan *latent dirichlet allocation* untuk meningkatkan penggunaan *cashless payment* sehingga memudahkan masyarakat dalam bertransaksi.

2. TINJAUAN PUSTAKA

2.1 *Systematic Literature Review* (SLR)

SLR adalah tinjauan untuk mengetahui apa yang sudah diketahui dari penelitian yang sudah ada sebelumnya tentang fenomena, subjek atau topik tertentu (Yang dkk., 2013). Terdapat beberapa tujuan dalam penelitian menggunakan SLR, diantaranya adalah mengidentifikasi, mereview, mengevaluasi dan menginterpretasikan semua penelitian yang tersedia dengan area topik fenomena yang menarik dan pertanyaan penelitian tertentu yang relevan. Penggunaan metode SLR dalam menentukan variabel didukung oleh sumber data yang *up-to-date*. Dalam prosesnya, salah satu aplikasi yang sering digunakan adalah *VOSviewer*.

VOSviewer merupakan program yang dikembangkan untuk membuat dan melihat peta bibliometrik. Program ini tersedia secara gratis untuk komunitas penelitian bibliometrik (lihat www.vosviewer.com). Dalam penelitian, *VOSviewer* digunakan untuk analisis bibliometrik, pemetaan topik untuk penelitian terbaru, mencari referensi yang paling banyak digunakan dalam bidang tertentu dan lain-lain. *VOSviewer* digunakan untuk membuat peta atau jurnal berdasarkan data kutipan bersama atau membuat peta kata kunci berdasarkan data kemunculan bersama. Program ini menawarkan penampil yang memungkinkan peta bibliometrik diperiksa secara lengkap. *VOSviewer* dapat menyajikan dan merepresentasikan informasi khusus tentang peta grafis bibliometrik. *VOSviewer* dapat menampilkan peta bibliometrik yang besar dengan cara yang mudah untuk menginterpretasikan hubungan. *VOSviewer* dapat membaca dataset dari berbagai situs jurnal online seperti *Scopus*, *Google Scholar*, *Web of Science*, *Dimension*, dan *Pubmed*.

2.2 Analisis Bibliometrik

Analisis bibliometrik menyediakan metode yang akurat dan objektif untuk mengevaluasi kontribusi sebuah artikel terhadap kemajuan pengetahuan, yang sering digunakan sebagai alat untuk menganalisis tren dan kinerja dalam bidang tertentu (Enzama, 2018). Pendekatan ini digunakan untuk memahami perkembangan dalam studi masyarakat nontunai, dengan tujuan untuk mengidentifikasi kesenjangan penelitian dan inovasi sebagai dasar untuk penelitian di masa depan.

Studi bibliometrik dapat memberikan evaluasi dan penilaian yang relevan terhadap relevansi sosial dan ilmiah dari disiplin ilmu atau bidang penelitian tertentu. Bibliometrik digunakan untuk mengidentifikasi pola publikasi ilmiah (Roder dkk.,2015), terutama publikasi ilmiah yang dikendalikan oleh Elsevier, Wiley, Taylor & Francis, Springer Nature, dan Sage. Bibliometrika efektif dalam menyediakan kumpulan data yang dapat digunakan untuk meningkatkan kualitas penelitian. Analisis bibliometrik digunakan untuk mencari tren, pola, dan hubungan variabel.

2.3 *Topic Coherence*

Topic modeling merupakan salah satu aplikasi statistik dalam mengolah data teks. Topic modeling adalah teknik yang digunakan untuk menemukan topik implisit dalam koleksi dokumen. Topic modeling sering digunakan untuk mengorganisir, meringkas, dan membantu peneliti dalam mengeksplorasi corpora yang besar. Algoritma topic modeling dapat digunakan pada koleksi dokumen teks yang luas seperti email, Twitter, koran, publikasi ilmiah, dan koleksi dokumen lainnya (Blei dkk.,2003).

Menentukan jumlah topik dalam pemodelan topik merupakan aspek penting yang harus diperhatikan. Dengan jumlah topik yang tepat, isu yang dihasilkan dapat lebih optimal, sehingga mudah diinterpretasikan oleh manusia. Koherensi topik mengasumsikan bahwa topik yang baik adalah topik yang memiliki M kata terbaik di setiap topik yang memiliki peluang tinggi untuk muncul bersamaan di semua dokumen (Rosen dkk., 2012). Menghitung koherensi topik dimulai dengan menghitung NPMI setiap kata untuk kata terbaik pada setiap topik, yang dirumuskan sebagai persamaan berikut.

$$NPMI = \frac{\log\left(\frac{P(w_i, w_j) + \varepsilon}{P(w_i)P(w_j)}\right)}{-\log(P(w_i, w_j) + \varepsilon)}$$

Keterangan:

$P(w_i, w_j)$: Peluang kemunculan kata ke- i dan ke- j secara bersama
 ε : Nilai konstanta

Setelah menghitung nilai NPMI setiap kata dibentuk vektor kata menggunakan persamaan berikut.

$\mathbf{v}_{w_i} = [NPMI(w_i, w_j), \dots, NPMI(w_i, w_M)]$
 dengan:

$NPMI(w_i, w_M)$: Nilai NPMI kata ke- i dan kata ke- j
 i, j : indeks kata ($i, j = 1, 2, 3, \dots, M$)

Kemudian dapat dibentuk sebuah vektor topik (\mathbf{v}_t) dengan menjumlahkan M vektor kata terbaik. Perhitungan *topic coherence* dilakukan dengan menghitung rata-rata dari *cosine similarity* setiap vektor kata terhadap vektor topik dengan rumus sebagai persamaan berikut.

$$C_V = \frac{1}{M} \sum_{i=1}^M \cos(\mathbf{v}_{w_i}, \mathbf{v}_t)$$

Keterangan:

$\cos(\mathbf{v}_{w_i}, \mathbf{v}_t)$: *cosine similarity* vektor kata ke- i terhadap vektor topik
 i : indeks kata ($i = 1, 2, 3, \dots, M$)

Nilai akhir dari *topic coherence* merupakan rata-rata dari seluruh hasil perhitungan pada seluruh topik.

2.4 Latent Dirichlet Allocation

Latent Dirichlet Allocation (LDA) adalah sebuah metode pembangunan model untuk data teks. Metode Latent Dirichlet Allocation diperkenalkan oleh [4] dengan ide utama setiap dokumen direpresentasikan sebagai campuran topik laten dan setiap topik direpresentasikan sebagai distribusi kata. Sebagai contoh, topik seni memiliki kata “musik”, “lagu”, dan “gambar” dengan probabilitas

yang tinggi. Topik komputer memiliki kata “data”, “internet”, dan “listrik” dengan probabilitas tinggi.

Ada dua metode yang populer dalam pemodelan topik, yaitu pemodelan topik linier dan pemodelan topik probabilistik. Pemodelan topik linier sering digunakan dalam analisis semantik laten (LSA). Pemodelan topik probabilistik merupakan pengembangan dari metode pemodelan topik linier. Pada metode pemodelan topik probabilistik, topik didefinisikan sebagai distribusi kata, dan dokumen didefinisikan sebagai kumpulan beberapa topik dengan bobot tertentu. Metode yang umum digunakan adalah Latent Dirichlet Allocation (LDA) (Blei dkk., 2003).

Salah satu metode yang dapat digunakan dalam LDA adalah dengan menggunakan Gibbs sampling. Langkah-langkah perhitungan Latent Dirichlet Allocation menggunakan Gibbs sampling adalah sebagai berikut:

- 1) Menentukan banyak topik (T)
- 2) Memberi label setiap kata pada setiap dokumen dengan topik secara acak.
- 3) Menghapus label topik pada kata ke- i dalam dokumen ke- d .
- 4) Melakukan *gibbs sampling* pada kata ke- i menggunakan persamaan dibawah ini terhadap setiap topik (Rosen-Zvi dkk., 2012).

$$P(z_i = j | w_i = m, \mathbf{y}, \mathbf{w}) \propto \frac{C_{mj} + \beta}{\sum_m C_{mj} + V\beta} * \frac{C_{dj} + \alpha}{\sum_d C_{dj} + T\alpha}$$

Keterangan:

$z_i = j$: penempatan kata ke- i dengan label topik ke- j pada sebuah dokumen ke- d .
 $w_i = m$: representasi kata ke- i adalah kata m pada dokumen ke- d .
 \mathbf{y} : vektor label topik seluruh dokumen tanpa mencantumkan observasi label kata ke- i
 \mathbf{w} : vektor kata dalam seluruh dokumen tanpa mencantumkan observasi kata ke- i
 C_{mj} : jumlah kemunculan kata m dengan label topik ke- j pada dokumen ke- d
 C_{dj} : jumlah kemunculan topik ke- j pada dokumen ke- d
 $\sum_m C_{mj}$: jumlah kemunculan kata m dengan label topik ke- j pada seluruh dokumen
 $\sum_d C_{dj}$: jumlah kemunculan topik ke- j pada seluruh dokumen
 α : prior sebaran topik terhadap dokumen
 β : prior sebaran kata terhadap topik
 V : banyak kata
 T : banyak topik

- 5) Memberi label topik baru kata ke- i berdasarkan hasil perhitungan persamaan diatas.
- 6) Ulangi langkah tiga sampai dengan lima pada setiap kata dalam seluruh dokumen.
- 7) Menghitung bobot dokumen terhadap topik menggunakan persamaan (Rosen-Zvi dkk., 2012).

$$\theta_{dj} = \frac{C_{dj} + \alpha}{\sum_d C_{dj} + T\alpha}$$

Keterangan:

θ_{dj} : bobot dokumen ke- d terhadap topik ke- j

C_{dj} : jumlah kemunculan topik ke- j pada dokumen ke- d

$\sum_d C_{dj}$: jumlah kemunculan topik ke- j pada seluruh dokumen

α : prior sebaran topik terhadap dokumen

T : banyak topik

3. METODE PENELITIAN

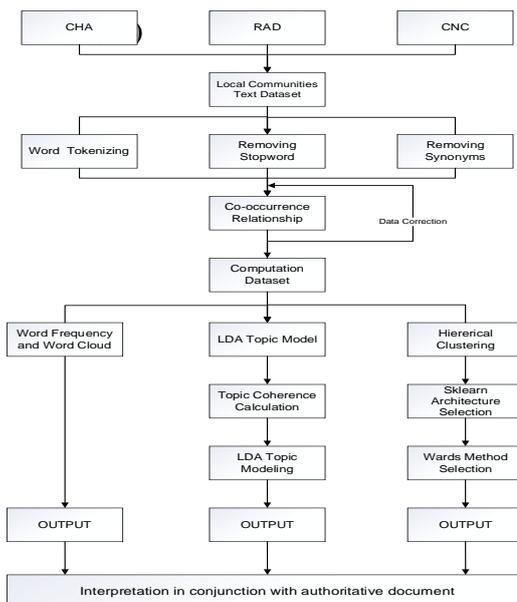
Penelitian ini menggunakan kombinasi analisis bibliometrik dan *Latent Dirichlet Allocation* (LDA) untuk memodelkan topik-topik yang berkaitan dengan masyarakat tanpa uang tunai. Analisis bibliometrik dilakukan dengan menggunakan artikel ilmiah dari *Scopus* terkait penelitian mengenai *cashless society*. Hasil dari bibliometrik akan menghasilkan tren publikasi penelitian mengenai *cashless society* dan mendeskripsikan variabel-variabel yang terkait dengan *cashless society*. Sementara itu, LDA menggunakan data dari *web scrapping* melalui *YouTube* dan *Instagram*.

Hasil LDA akan menjadi indikator untuk semua variabel endogen dan eksogen. Setelah mendapatkan indikator-indikator tersebut, kuesioner disusun dan disebarakan kepada 250 responden di Kota Bekasi. Hasil dari kuesioner digunakan untuk menghubungkan indikator hasil LDA dengan variabel hasil Bibliometrik, hubungan tersebut akan diukur dengan menggunakan analisis faktor eksplanatori.

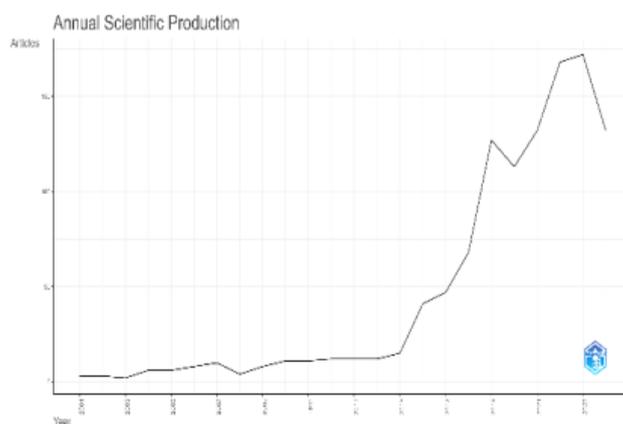
4. HASIL DAN PEMBAHASAN

4.1. Analisis Bibliometrik

Hasil penelusuran awal terhadap dokumen Scopus dengan tema *cashless society* menghasilkan 1265 dokumen. Pada tahap pencarian awal, tahun publikasi tidak dibatasi karena bertujuan untuk memetakan perkembangan penelitian dengan tema *cashless society*. Kemudian dilakukan penyaringan dokumen berdasarkan jenis dokumen yaitu artikel, sehingga ditemukan 945 dokumen. Dari hasil dokumen yang ditemukan, dilakukan analisis bibliometrik yang memuat beberapa informasi antara lain nama pengarang, kata kunci, bahasa, tahun terbit, judul artikel, jumlah sitasi, nama jurnal penerbit, abstrak, afiliasi, dan DOI. Berdasarkan hasil tersebut, dapat disimpulkan bahwa *cashless society* berkembang begitu pesat. Menurut data 5 tahun terakhir yaitu tahun 2019 sampai dengan 2023, frekuensi penelitian terkait *cashless society* mengalami peningkatan setiap tahunnya, hal ini dapat dilihat pada Gambar 2.

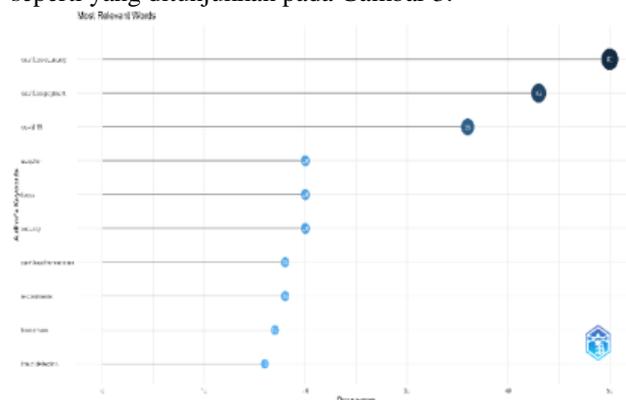


Gambar 1. Peta Jalan LDA



Gambar 2. Frekuensi Penelitian Setiap Tahun

Berdasarkan hasil bibliometrik, diperoleh kata kunci yang relevan dengan masyarakat nontunai seperti yang ditunjukkan pada Gambar 3.



Gambar 3. Most Relevant Keyword

Topic	Economy	Payment	Adoption	Consumer Focus	Intention To Use Cashless Society	Frequent Behavior To Use Cashless Society	Policy
Digitalization	0.639	-0.057	0.064	0.067	0.087	-0.049	0.106
Ease of Payment	-0.159	0.517	-0.012	-0.003	-0.062	-0.027	-0.092
Readiness	-0.186	0.146	0.623	0.152	-0.121	0.009	-0.197
Accuracy	-0.166	-0.002	0.670	0.024	-0.079	-0.014	0.181
Social	-0.101	0.194	0.054	-0.193	0.641	0.107	0.177
Culture	0.042	0.175	-0.084	-0.157	0.504	0.092	0.042
Transaction	0.522	-0.047	0.075	-0.107	-0.121	-0.130	-0.034
Card	-0.058	0.551	-0.100	-0.049	0.191	0.147	-0.057
Ease of Use	-0.080	0.125	-0.167	0.031	-0.088	0.603	0.061
Payment Method	-0.038	-0.089	-0.020	-0.193	-0.173	0.661	0.181
Habit	0.192	0.163	0.068	0.178	0.149	-0.089	0.607
System	0.548	0.180	-0.101	0.008	-0.035	0.118	-0.053
Believe	-0.040	-0.093	-0.146	0.686	0.159	-0.088	-0.118
Commitment	0.105	-0.097	-0.157	0.616	-0.188	0.006	-0.115
Collaborative	0.018	0.693	0.047	-0.163	0.019	0.126	0.082
Acceptance	-0.143	-0.130	0.677	-0.002	0.139	-0.008	-0.142
Personal	0.044	-0.116	-0.137	0.004	0.603	0.078	-0.016
Security	-0.197	-0.159	0.082	0.538	0.134	-0.104	0.017
Frequency of Use	0.068	-0.046	-0.161	-0.167	-0.090	0.560	-0.038
Need	0.156	0.193	0.152	-0.192	0.096	0.178	0.642
Contextual Factors	-0.090	0.000	0.179	-0.144	-0.027	-0.064	0.588

4.4. Factor Analysis (EFA)

Berdasarkan hasil LDA, kuesioner disebarkan guna mendapatkan data untuk menghubungkan variabel dan indikator dari hasil LDA. Hasil tersebut digunakan untuk EFA dengan menghitung nilai factor loading dari semua indikator terhadap variabelnya, hasil EFA dapat dilihat pada Tabel 2.

5. KESIMPULAN

Berdasarkan hasil analisis dan pengujian yang telah dilakukan penelitian ini, diketahui bahwa tren penelitian mengenai *cashless society* berkembang pesat dari tahun 2019 hingga 2023. Variabel yang sering digunakan adalah *economy, payment, adoption, policy, and consumer focus*.

Hasil *topic coherence* menunjukkan bahwa topik optimal yang dapat dibentuk adalah 21 topik. Topik-topik tersebut digunakan sebagai indikator dari variabel-variabel yang mempengaruhi *cashless society*.

DAFTAR PUSTAKA

- BLEI, D. M., NG, A. Y., & JORDAN, M. I. 2003. Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3, 993–1022.
- CHAPAGAIN, A. 2019. Hands-On Web Scraping with Python: Perform advanced scraping operations using various Python libraries

and tools such as Selenium, Regex, and others. Packt Publishing Ltd.

- EZAMA-NICOLÁS, R. 2018. A Bibliometric Method for Assessing Technological Maturity: The Case of Additive Manufacturing. *Scientometrics*, 117(3), 1425–1452.
- FEINERER, I., HORNIK, K., & MEYER, D. 2008. Text Mining Infrastructure in R. *Journal of Statistical Software*, 25(5). <https://doi.org/10.18637/jss.v025.i05>
- GARBHAPU, V. K., & BODAPATI, P. 2020. A comparative analysis of Latent Semantic analysis and Latent Dirichlet allocation topic modeling methods using Bible data. *Indian Journal of Science and Technology*, 13(44), 4474–4482.
- HAN, J., TONG, H., & TONG, H. 2022. Data mining: concepts and techniques. Cambridge: Morgan kaufmann.
- KALEPALLI, Y., TASNEEM, S., TEJA, P. D. P., & MANNE, S. 2020. Effective comparison of LDA with LSA for topic modelling. In 2020 4th International conference on intelligent computing and control systems (ICICCS) (pp. 1245-1250). IEEE.
- RÖDER, M., BOTH, A., & HINNEBURG, A. 2015. Exploring the Space of Topic Coherence Measures. *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, 399–408. <https://doi.org/10.1145/2684822.2685324>

- ROSEN-ZVI, M., GRIFFITHS, T., STEYVERS, M., & SMYTH, P. 2012. The Author-Topic Model for Authors and Documents.
- VAN KLAVEREN, C., & DE WOLF, I. 2019. Systematic Reviews in Education Research: In Contemporary Economic Perspectives in Education. <https://doi.org/10.2307//j.ctt14jxsqg.4>
- YANG, L., SUN, Y., & LIU, Z. 2013. A Bibliometric Investigation of Research Trends in Computational Simulation for Engineering. *Scientometrics*, 95(2), 529-546. doi:10.1007/s11192-012-0872-z
- YULIANTI, N., ULPAWATI, U., & SUSANTI, S. 2022. Analisis Bibliometrik Determinan Kejadian Stunting Pada Balita. *Jurnal Riset Rumpun Ilmu Kesehatan*, 1(2), 120-129.

Halaman ini sengaja dikosongkan