

Analysis on Scopus Articles Padang State University Based on SINTA Website

Kerin Hagia Aidillah*, Dodi Vionanda, Dony Permana dan Wita Resfi Ananta

Departemen Statistika, Universitas Negeri Padang, Padang, Indonesia

*Corresponding author: kerinaidillah@gmail.com

Submitted : 10 Februari 2025

Revised : 12 Februari 2025

Accepted : 28 Februari 2025

ABSTRACT

Universities have the responsibility to carry out education, research, and community service as mandated by Law Number 20 of 2003 on the National Education System in Article 20. The flagship research theme set by Universitas Negeri Padang (UNP) for the period 2020-2024 is "Development of Digital Learning Services and Development of Minangkabau Cuisine based on Local Potential." The focus of the flagship research activities at Padang State University encompasses two main research areas: 1) Digital Learning Services; and 2) Minangkabau Cuisine. The objective of this research is to compare the flagship research theme with the Scopus articles from Universitas Negeri Padang on the SINTA website. By analyzing the trends of Scopus article topics on the SINTA website using web scraping techniques and wordcloud visualization, it is concluded that there is a match between the trending topics of UNP's Scopus articles and UNP's flagship research theme, particularly in the field of Digital Learning Services. From the wordcloud results, which show keywords such as Learning, Development, Student, and Model. This research allows us to easily observe from the wordcloud visualization the trend of research topics in Scopus articles on SINTA at Universitas Negeri Padang, reflecting the realization of Universitas Negeri Padang flagship research theme for the period 2020-2024.

Keywords: *Web Scraping, Wordcloud, SINTA*



This is an open access article under the Creative Commons 4.0 Attribution License, which permits unrestricted use, distribution, and reproduction in medium, provided the original work is properly cited. ©2022 by author and Universitas Negeri Padang.

I. PENDAHULUAN

Perguruan tinggi dan universitas mempunyai tanggung jawab untuk melakukan pendidikan, riset atau penelitian, serta dedikasi kepada masyarakat berdasarkan amanat oleh Hukum Undang-Undang Nomor 20 Tahun 2003 mengenai Sistem Pendidikan Nasional atas Pasal 20. Riset di perguruan tinggi bertujuan untuk memperluas ilmu wawasan serta teknologi dan meningkatkan keselamatan masyarakat serta kegigihan bangsa, begitu juga dipaparkan dalam Hukum Undang-Undang Nomor 12 Tahun 2012 mengenai Pendidikan Tinggi atas Pasal 45 hingga 46. Hasil riset diwajibkan disebarluaskan ke berbagai sumber seperti seminar, publikasi, serta paten. Untuk mewujudkan tujuan riset, Universitas Negeri Padang (UNP) merumuskan strategi serta kebijakan riset yang dimasukkan dalam Rencana Strategis Penelitian UNP Tahun 2020– 2024 (Renstra UNP, 2020).

Pokok bahasan riset unggulan yang diresmikan UNP dalam periode 2020- 2024 yakni “*Pengembangan Layanan Digital Pembelajaran dan Pengembangan Kuliner Minangkabau berdasarkan atas potensi lokal*”. Tema riset unggulan UNP melingkupi 2 aspek, ialah: 1) Layanan Pembelajaran Digital; serta 2) Kuliner Minangkabau. Hasil formulasi riset unggulan disusun dengan adanya denah jalur (*road map*) dengan terperinci untuk periode 5 tahun (2020- 2024). Diharapkan aspek yang tercantum dalam Rencana Induk Penelitian UNP ini menjadi pedoman serta mempunyai arah riset yang jelas. Pokok bahasan serta aspek dalam riset unggulan perguruan tinggi UNP dapat digolongkan ditingkat nasional ataupun global (Renstra UNP, 2020).

Perbandingan ini diperlukan guna mengevaluasi efektivitas implementasi fokus strategis universitas dalam bidang penelitian, sekaligus mengidentifikasi potensi kesenjangan yang dapat menjadi dasar perencanaan strategis untuk meningkatkan kualitas dan relevansi penelitian di masa mendatang. Sehingga penelitian ini dilakukan untuk menganalisis dan memahami sejauh mana pokok bahasan riset unggulan yang ditetapkan oleh universitas selaras dengan output penelitian yang tercatat dalam SINTA. Berlandasan dari paparan diatas, tujuan riset ini yakni untuk membandingkan pokok bahasan riset unggulan yang diresmikan UNP dalam periode 2020-2024 ialah “*Pengembangan Layanan Digital Pembelajaran serta Pengembangan Kuliner Minangkabau berlandasan atas potensi lokal*” dengan Artikel terindeks Scopus yang telah Dosen serta mahasiswa UNP terbitkan. Dengan menganalisis tren topik Artikel Scopus di website SINTA menggunakan metode *web scraping* dan visualisasi *wordcloud*.

II. METODE PENELITIAN

A. Sumber Data dan Variabel Penelitian

Data yang digunakan dalam penelitian ini yakni data sekunder dengan variabel penelitian berupa judul artikel berafiliasi Universitas Negeri Padang yang terindeks *scopus* dengan rentang periode publikasi tahun 2020-2024. Data dalam penelitian ini didapat dengan cara mengakses website resmi SINTA <https://sinta.kemdikbud.go.id> yang terafiliasi Universitas Negeri Padang. Teknik pengumpulan data dengan menerapkan teknik web scraping menggunakan bantuan *Google Colaboratoty* dan software R studio. Data diperoleh sebanyak 2.489 judul artikel Scopus dalam periode tahun 2020-2024.

B. Teknik Analisis Data

Adapun gambaran metode penelitian sebagai berikut :



Gambar 1. Metode Penelitian

Dalam mengelola dataset pada riset ini, terdapat 3 prosedur yang dilalui yaitu (Pradana, 2020):

1. Mengumpulkan Data

Mengumpulkan data merupakan proses pengambilan data yang dilakukan untuk memenuhi kebutuhan penelitian, analisis, atau pengambilan keputusan. Dalam penelitian ini, proses dilakukan dengan menggunakan metode *web scraping* dengan software atau *tools google collaboratory* dengan data yang di scraping berupa judul beserta tahun terbit artikel scopus kemudian disimpan dalam bentuk file csv atau excel. *Web scraping* adalah teknik pengambilan data dari website dengan cara otomatis tanpa harus menyalinnya secara manual, berpusat dalam memperoleh informasi/data secara otomatis dan ekstraksi (Deviacita et al., 2019). Proses *web scraping* dilakukan dengan menggunakan *Google Colab* dengan bahasa pemrograman *Python* dan *library beautifulsoup* (Adila, 2022).

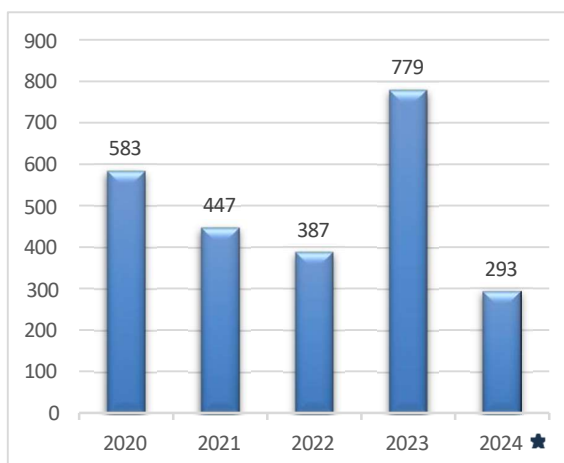
2. Pembersihan Data

Pembersihan data atau *data cleaning* merupakan proses pernghapusan dan eliminasi data yang tidak berarti (noise) seperti tanda baca dan karakter lain (Sudianto et al., 2022).

3. Analisis

Analisis dalam penelitian ini dengan menggunakan salah satu metode untuk memvisualisasikan data teks secara visual yaitu *wordcloud*. Visualisasi *wordcloud* menggambarkan frekuensi kata-kata dalam bentuk teks dengan ukuran yang proporsional terhadap kemunculannya. Semakin sering sebuah kata muncul dalam data, maka semakin besar ukuran kata tersebut dalam visualisasi *wordcloud*, menjadikannya informatif sekaligus estetik (Sahria, 2020).

Prosedur pertama dalam penelitian ini adalah mengakses jurnal SINTA yang hendak dijadikan sebagai sumber informasi data. Pada tahap ini dilakukan eksplorasi artikel scopus yang terafiliasi dari Universitas Negeri Padang dari tahun 2020-2024. Dari hasil eksplorasi pada jurnal SINTA, terkumpul 2.489 data dari Universitas Negeri Padang yang secara rinci menyajikan publikasi yang terindeks SINTA pada situs <https://sinta.kemdi kbud .go. id /affiliations/profile/497/?view= scopus>. Prosedur kedua yaitu melaksanakan pengambilan data dengan metode *web scraping*. *Web Scraping* adalah teknik pengumpulan data dari website, khususnya dari halaman web berbasis HTML atau XHTML. Dalam hal ini, data dikumpulkan dari halaman artikel SINTA berdasarkan judul artikel yang terindeks Scopus dan terafiliasi dengan Universitas Negeri Padang. Daftar jumlah artikel yang terkumpul dalam penelitian ini diperlihatkan oleh Gambar 2.



Gambar 2. Visualisasi Jumlah Judul Artikel Scopus UNP 2020-2024

Berdasarkan gambar diatas, menunjukkan jumlah Universitas Negeri Padang selama periode 2020-2024. Dimana terjadinya penurunan dalam 3 tahun berturut-turut dari tahun 2020 hingga 2022. Namun, pada tahun 2023 terjadi peningkatan yang signifikan dari tahun sebelumnya. Sementara, pada tahun 2024 ditandai simbol bintang (*), yang mengindikasikan bahwa tahun tersebut masih berjalan atau data belum lengkap. Sehingga secara keseluruhan, jumlah artikel mengalami fluktuasi yang dapat terjadi karena berbagai faktor.

Prosedur ketiga adalah pembersihan data yang diperoleh dari pengambilan data dengan web scraping. Sehingga, tidak ada data yang tidak tertata alhasil pengolahan data dapat dilakukan. Prosedur keempat ialah analisis data. Hasil analisis penelitian ini diharapkan dapat memberikan gambaran kepada peneliti yang ingin melihat tren topik artikel dari Universitas Negeri Padang. Hasil analisa atas riset ini menghasilkan visualisasi *wordcloud* yang dapat memberikan visualisasi informasi untuk menggambarkan data secara nyata dan jelas serta berdaya guna untuk setelah itu dilakukan analisis lanjutan sesuai tujuan data akan dipilih (Ulfah & Najiah, 2023).

III. HASIL DAN PEMBAHASAN

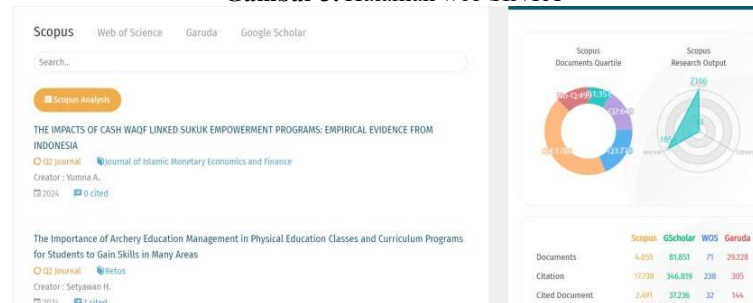
Hasil penelitian ini yakni menerapkan metode *web scraping* pada jurnal SINTA dengan bahasa pemrograman *Python* serta hasil scraping ditampung di excel dengan format csv. Setelah itu data dianalisis dengan software R untuk mengetahui tren topik penelitian artikel Scopus UNP.

A. Akses Jurnal SINTA

Berikut ini adalah halaman web SINTA yang digunakan untuk proses *scraping* berdasarkan judul artikel Scopus dari Universitas Negeri Padang. Halaman web jurnal SINTA yang di-scraping dapat dilihat pada Gambar 3 dan 4.



Gambar 3. Halaman web SINTA



Gambar 4. Artikel Scopus UNP

B. Scraping Data

Dalam melakukan metode *web scraping*, diawali dengan mengetahui bentuk HTML atas halaman website untuk memastikan dalam tag HTML mana data inti direpresentasikan. Selanjutnya struktur HTML yang akan digunakan pada jurnal SINTA dapat dilihat pada Gambar 5

```

    <div class="profile-article">
      <div class="stat-chart mt-1 mb-4 text-center"> </div>
      <ul class="article-tab mt-1 mb-3"> </ul>
      <div class="row"> </div>
      <br>
      <a href="https://sinta.kemdikbud.go.id/affiliations/scopusanalysis/497" class="btn btn-warning text-white mb-3"> </a>
      <div class="ar-list-item mb-5">
        <div class="ar-title"> == $0
          <a href="https://www.scopus.com/record/display.uri?eid=2-s2.0-85183375040&origin=resultslist" target="_blank"> </a>
        </div>
        <div class="ar-meta"> </div>
        <div class="ar-meta"> </div>
      </div>
    </div>
  
```

Gambar 5. HTML Artikel SINTA

Setelah mendapat tag laman web yang akan diambil. Informasi yang diperoleh dari penggunaan teknik web scraping berbasis Python untuk mempermudah proses pengambilan data. Salah satu URL target scraping yang digunakan dalam penelitian ini berasal dari laman SINTA yaitu <https://sinta.kemdikbud.go.id/affiliations/profile/497/?view=scopus>. Data yang diambil dari target web scraping ini kemudian diproses menggunakan Google Colaboratory dengan Python, dan hasilnya ditampilkan pada gambar 6.

Judul Artikel
Empowering Students: Examining the Effectiveness of Ta'lim Muta'allim Group Guidance
Social Capital of the Harau Community in Preserving Rendang Daun Kayu: Maintaining the Sustainability of Traditional Minangkabau Cuisine
A spotlight on gamma-mangostin: exploring its potential as antiviral agents
A community group decomposition of the willingness to accept: evidence from sand, stone and gravel mining activities in Padang Pariaman, West Sumatra, Indonesia
Strategies for Optimising Rice Management in New Rice Expansion Areas in West Sumatra Province, Indonesia
Effectiveness of Interactive Content with H5P for Moodle-Learning Management System in Blended Learning
Catalyzing Kaempferol from Broccoli Fruit (<i>Brassica oleracea</i> var. <i>italica</i>) on Tyrosinase Protein for Regulation of Melanin Synthesis in-Slico
Qatar 2022 World Cup Scorer Analysis
Development Food Agriculture Area Basic Land Capability in Dharmasraya Regency, West Sumatra Province
Citrus Limon Has the Potential to Reduce Oxidative Stress and Inflammation After Physical Activity/Exercise: Systematic Review

Gambar 6. Hasil Scraping

Dari *scraping* data tersebut diperoleh sebanyak 2.489 judul artikel Scopus dalam periode tahun 2020-2024. Artikel-artikel tersebut dipublikasikan melalui situs resmi SINTA dan berasal dari institusi yang berafiliasi dengan Universitas Negeri Padang.

C. Wordcloud





Gambar 7. Visualisasi Wordcloud Artikel Scopus UNP

Berdasarkan Gambar 7 diatas, dapat diamati bahwa terdapat sejumlah topik utama yang menonjol dan memiliki kata kunci yang mirip/berulang (redundansi). Oleh karena itu, untuk meningkatkan akurasi serta efisiensi dalam pengelompokan kategori, topik-topik yang memiliki kesamaan akan dikelompokkan dalam satu klaster tertentu. Proses pengelompokan ini bertujuan untuk menyederhanakan analisis dan meningkatkan pemahaman terhadap tren penelitian. Informasi lebih lanjut mengenai klasterisasi ini dapat ditemukan dalam Tabel 1.

Tabel 1. Tren Topik periode tahun 2020-2024

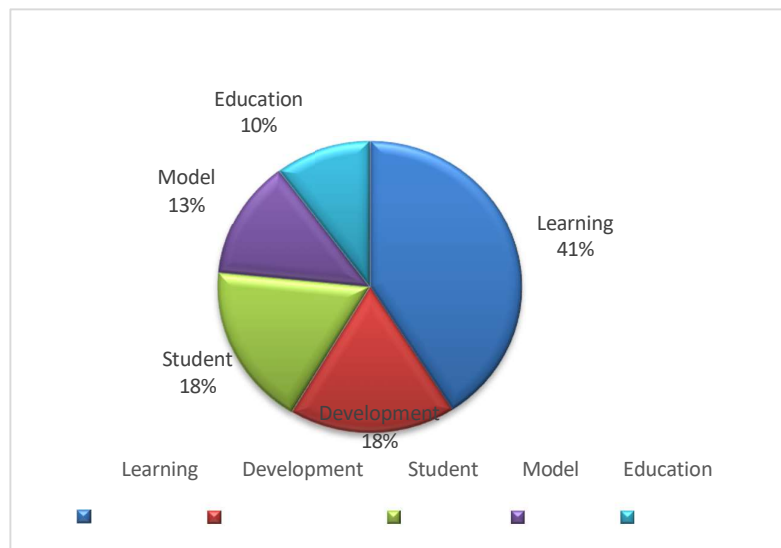
Tahun Tren Topik				
2024	2023	2022	2021	2020
1. Development	1. Learning	1. Learning	1. Learning	1. Learning
2. Learning	2. Student	2. Model	2. Students	2. Development
3. Indonesia	3. Development	3. Development	3. Development	3. School
4. Review	4. Study	4. Students	4. School	4. Student
5. Education	5. Indonesia	5. School	5. High	5. High
6. Students	6. Model	6. Study	6. Model	6. Model
7. West	7. Skills	7. Education	7. Effect	7. Indonesia
8. Performance	8. Education	8. Indonesia	8. Study	8. Mathematics
9. Systematic	9. High	9. Effect	9. Grade	9. Problem
10. Sumatra	10. School	10. Physical	10. Validity	10. Education
11. Environmental	11. Effect	11. Skills	11. Education	11. Improve
12. Model	12. Improve	12. Application	12. Improve	12. Preliminary
13. Study	13. Mathematics	13. Covid-19	13. Physics	13. Senior
14. Sustainable	14. Media	14. Design	14. Inquiry	14. Integrated
15. Management	15. West	15. Teaching	15. West	15. Solving

Adapun klaster yang disesuaikan dari kemiripan arti/inti dari kata-kata kunci yang ada dari setiap topik, kemudian topik tersebut dikelompokkan ke dalam klaster dapat dilihat pada tabel 2.

Tabel 2. Klasterisasi Topik

Klaster	Tren Tahunan	Tren Topik
Klaster 1	5 tahun	Learning, Development, Student, Education, Model
Klaster 2	4 tahun	School, Indonesia, Study
Klaster 3	3 tahun	High, Improve, Effect, West
Klaster 4	2 tahun	Mathematics, Skill
Klaster 5	Tidak ada hubungan dalam tahun	Review, Performance, Systematic, Sumatra, Enviromental, Sustainable, Management, Alat, Physical, Applications, Covid- 19, Design, Teaching, Grade, Validity, Physics, Inquiry, Dilema, Preliminary, Tua, Integrated, Solving

Berdasarkan klusterisasi tren topik penelitian menghasilkan sejumlah topik yang menonjol dengan 5 kata kunci yang ada di semua periode 2020 hingga 2024 yaitu *Learning, Development, Student, Education, Model*.). Faktor khusus yang memengaruhi banyaknya penelitian bertemakan *Pengembangan Layanan Digital Pembelajaran* adalah dikarenakan mayoritas publisher artikel di website SINTA yaitu dari Fakultas Ilmu Pendidikan dan Jurusan Pendidikan serta diwajibkannya publish SINTA untuk Sarjana S2 di Universitas Negeri Padang. Sehingga tema pengembangan pembelajaran mendominasi judul artikel di Universitas Negeri Padang. Adapun hasil visualisasi tren topik dari klaster 1 dengan bentuk diagram dapat dilihat sebagai berikut :



Gambar 8. Diagram Tren Frekuensi Kata

IV. KESIMPULAN

Berdasarkan hasil analisis yang telah dilakukan, terdapat beberapa output yang didapat yaitu terlihat adanya kecocokan tren topik artikel scopus UNP dengan pokok bahasan riset unggulan UNP. Fokus topik kegiatan menghasilkan sejumlah topik yang menonjol dengan 5 kata kunci yang ada di semua periode 2020 hingga 2024, yaitu pada aspek Layanan Digital Pembelajaran; terlihat dari hasil *wordcloud* tren topik menghasilkan kata kunci *Learning* (Pembelajaran), *Development* (Pengembangan), *Student* (Siswa), *Education* (Pendidikan), *Model* (Pemodelan). Berdasarkan hasil dari penelitian ini, dapat disimpulkan bahwa pokok bahasan riset unggulan yang ditetapkan oleh universitas selaras dengan output penelitian yang tercatat dalam artikel SINTA berafiliasi Universitas Negeri Padang yang terindeks *scopus*.

DAFTAR PUSTAKA

- Adila, N. (2022). Implementation of Web Scraping for Journal Data Collection on the SINTA Website. *Sinkron*, 7(4), 2478–2485. <https://doi.org/10.33395/sinkron.v7i4.11576>
- Busro, Mailana, A., & Sarifudin, A. (2021). Pendidikan Islam dalam Publikasi Internasional: Analisis Bibliometrik pada Database Scopus. *Edukasi Islami: Jurnal Pendidikan Islam*, 10(01), 413–426. <http://jurnal.staialhidayahbogor.ac.id/index.php/ei/article/view/1591/800>
- Galih Pradana, M. (2020). Penggunaan Fitur Wordcloud Dan Document Term Matrix Dalam Text Mining. *Jurnal Ilmiah Infromatika (JIF)*, 08(01), 38–43.
- Hasan, T. (2023). Analisis Publikasi Ilmiah Dosen Terindeks Scopus Pada Fakultas Keperawatan Universitas Riau Tahun 2011-2022. *Media Pustakawan*, 29(3), 228–240. <https://doi.org/10.37014/medpus.v29i3.3236>
- Imam Cholissodin, E. R. (2017). *Analisis Big Data - ICCA*. <https://icca.co.id/analisis-big-data/>
- Josi, A., Abdillah, L. A., & Suryayusra. (2014). *Penerapan teknik web scraping pada mesin pencari artikel ilmiah*. <http://arxiv.org/abs/1410.5777>
- Komalasari, R. (2021). Analisis karya tulis ilmiah dosen Institut Pertanian Bogor terindeks Scopus sampai tahun 2021. *Tahun*, 10(1), 17–32. <https://www.perpustakaan.ipb.ac.id/>
- LP2M UNP. (2016). *Rencana Induk Penelitian (Rip)*. 031, 4–5. <http://lpp.uad.ac.id/wp-content/uploads/2016/12/RIP-UAD-2016-2017-KOMPLIT.pdf>
- Sahria, Y. (2020). Implementasi Teknik Web Scraping pada Jurnal SINTA Untuk Analisis Topik Penelitian Kesehatan Indonesia. *URECOL (Unversity Research Colloquium)*, 297–306. <http://repository.urecol.org/index.php/proceeding/article/view/1079>
- Silge, J., & Robinson, D. (2017). *Text Mining with R Text Mining with R Revision History for the First Edition*. <http://oreilly.com/catalog/errata.csp?isbn=9781491981658>
- Ulfah, A., & Najiah, I. (2023). Implementasi Web Scraping Pada Situs Jurnal Sinta Menggunakan Framework Selenium Webdriver Python. *JIKA (Jurnal Informatika)*, 7(1), 29. <https://doi.org/10.31000/jika.v7i1.7037>