

REFERENCE

- Akilandeswari, P., Bettala, S. D., Alankritha, P., Srimathi, H. and Krithik Sudhan, D., 2019. Dynamic Data Auditing Using MongoDB in Cloud Platform. *In: Smys, S., Bestak, R., Chen, J. I.-Z., and Kotuliak, I., eds. International Conference on Computer Networks and Communication Technologies* [online]. Singapore: Springer Singapore, 275–285. Available from: http://link.springer.com/10.1007/978-981-10-8681-6_26 [Accessed 8 Jul 2024].
- Albalawi, R., Yeap, T. H. and Benyoucef, M., 2019. Toward A Real-Time Social Recommendation System. *In: Proceedings of the 11th International Conference on Management of Digital EcoSystems* [online]. Presented at the MEDES '19: 11th International Conference on Management of Digital EcoSystems, Limassol Cyprus: ACM, 336–340. Available from: <https://dl.acm.org/doi/10.1145/3297662.3365789> [Accessed 17 Mar 2024].
- Anon., 100AD. *docker container stats* [online]. Docker Documentation. Available from: <https://docs.docker.com/reference/cli/docker/container/stats/> [Accessed 22 Jul 2024].
- Anon., 2024. *Reddit User Age, Gender, & Demographics (2024)* [online]. Exploding Topics. Available from: <https://explodingtopics.com/blog/reddit-users> [Accessed 6 Jul 2024].
- Anon., 2024a. *JSON* [online]. GitHub. Available from: <https://github.com/reddit-archive/reddit/wiki/JSON> [Accessed 5 Jul 2024].
- Anon., 2024b. *Quick Start - PRAW 7.7.2.dev0 documentation* [online]. Available from: https://praw.readthedocs.io/en/latest/getting_started/quick_start.html [Accessed 6 Jul 2024].
- Anon., 2024c. *Pusher Realtime Reddit API | Pusher blog* [online]. Available from: <https://pusher.com/blog/pusher-realtime-reddit-api/#use-cases-for-the-pusher-realtime-reddit-api> [Accessed 6 Jul 2024].
- Araujo, J. M. A., de Moura, A. C. E., da Silva, S. L. B., Holanda, M., Ribeiro, E. de O. and da Silva, G. L., 2021. Comparative Performance Analysis of NoSQL Cassandra and MongoDB Databases. *In: 2021 16th Iberian Conference on Information Systems and Technologies (CISTI)* [online]. Presented at the 2021 16th Iberian Conference on Information Systems and Technologies (CISTI), 1–6. Available from: <https://ieeexplore-ieee.org.ezproxy.ugm.ac.id/document/9476319/?arnumber=9476319> [Accessed 8 Jul 2024].
- Arshi Saloot, M. and Nghia Pham, D., 2021. Real-time Text Stream Processing: A Dynamic and Distributed NLP Pipeline. *In: 2021 International Symposium on Electrical, Electronics and Information Engineering* [online]. Presented at the ISEEIE 2021: 2021 International Symposium on Electrical, Electronics and Information Engineering, Seoul Republic of Korea: ACM, 575–584. Available from: <https://dl.acm.org/doi/10.1145/3459104.3459198> [Accessed 8 Jul 2024].
- Aslam, A., Qamar, U., Khan, R. A., Saqib, P., Ahmad, A. and Qadeer, A., 2019. Opinion Mining Using Live Twitter Data. *In: 2019 IEEE International*

- Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC)* [online]. Presented at the 2019 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC), New York, NY, USA: IEEE, 36–39. Available from: <https://ieeexplore.ieee.org/document/8919578/> [Accessed 15 May 2023].
- Baumgartner, J., Zannettou, S., Keegan, B., Squire, M. and Blackburn, J., 2020. *The Pushshift Reddit Dataset* [online]. [online]. Available from: <https://zenodo.org/records/3608135> [Accessed 5 Jul 2024].
- Belford, M., Mac Namee, B. and Greene, D., 2017. Stability of Topic Modeling via Matrix Factorization. [online]. Available from: <http://arxiv.org/abs/1702.07186> [Accessed 8 Jul 2024].
- Bellaouar, S., Bellaouar, M. M. and Ghada, I. E., 2021. Topic Modeling: Comparison of LSA and LDA on Scientific Publications. In: *2021 4th International Conference on Data Storage and Data Engineering* [online]. Presented at the DSDE '21: 2021 4th International Conference on Data Storage and Data Engineering, Barcelona Spain: ACM, 59–64. Available from: <https://dl.acm.org/doi/10.1145/3456146.3456156> [Accessed 22 Jul 2024].
- Bennawy, M. and El-Kafrawy, P., 2022. Contextual Data Stream Processing Overview, Architecture, and Frameworks Survey, 9 (1).
- Bergamaschi, S. and Po, L., 2015. Comparing LDA and LSA Topic Models for Content-Based Movie Recommendation Systems. In: . 247–263.
- Blair, S. J., Bi, Y. and Mulvenna, M. D., 2020. Aggregated topic models for increasing social media topic coherence. *Applied Intelligence*, 50 (1), 138–156.
- Chawla, H. and Khattar, P., 2020. *Data Lake Analytics on Microsoft Azure: A Practitioner's Guide to Big Data Engineering* [online]. Berkeley, CA: Apress. Available from: <https://link.springer.com/10.1007/978-1-4842-6252-8> [Accessed 7 Jul 2024].
- Habibabadi, S. K. and Haghighi, P. D., 2019. Topic Modelling for Identification of Vaccine Reactions in Twitter. In: *Proceedings of the Australasian Computer Science Week Multiconference* [online]. Presented at the ACSW 2019: Australasian Computer Science Week 2019, Sydney NSW Australia: ACM, 1–10. Available from: <https://dl.acm.org/doi/10.1145/3290688.3290735> [Accessed 2 Oct 2022].
- Hananto, V. R., Serdült, U. and Kryssanov, V., 2021. A Tourism Knowledge Model through Topic Modeling from Online Reviews. In: *2021 7th International Conference on Computing and Data Engineering* [online]. Presented at the ICCDE 2021: 2021 7th International Conference on Computing and Data Engineering, Phuket Thailand: ACM, 87–93. Available from: <https://dl.acm.org/doi/10.1145/3456172.3456211> [Accessed 2 Oct 2022].
- Harenslak, B., n.d. Data Pipelines with Apache Airflow.

- Hidayatullah, A. F., Pembrani, E. C., Kurniawan, W., Akbar, G. and Pranata, R., 2018. Twitter Topic Modeling on Football News. *In: 2018 3rd International Conference on Computer and Communication Systems (ICCCS)* [online]. Presented at the 2018 3rd International Conference on Computer and Communication Systems (ICCCS), Nagoya, Japan: IEEE, 467–471. Available from: <https://ieeexplore.ieee.org/document/8463231/> [Accessed 2 Oct 2022].
- Hlupić, T. and Puniš, J., 2021. An Overview of Current Trends in Data Ingestion and Integration. *In: 2021 44th International Convention on Information, Communication and Electronic Technology (MIPRO)* [online]. Presented at the 2021 44th International Convention on Information, Communication and Electronic Technology (MIPRO), 1265–1270. Available from: <https://ieeexplore-ieee-org.ezproxy.ugm.ac.id/document/9597149/?arnumber=9597149> [Accessed 7 Jul 2024].
- How, M., 2020. *The Modern Data Warehouse in Azure: Building with Speed and Agility on Microsoft's Cloud Platform* [online]. Berkeley, CA: Apress. Available from: <http://link.springer.com/10.1007/978-1-4842-5823-1> [Accessed 7 Jul 2024].
- Huang, C., Cahill, M., Fekete, A. and Rohm, U., 2020. Deciding When to Trade Data Freshness for Performance in MongoDB-as-a-Service. *In: 2020 IEEE 36th International Conference on Data Engineering (ICDE)* [online]. Presented at the 2020 IEEE 36th International Conference on Data Engineering (ICDE), 1934–1937. Available from: <https://ieeexplore-ieee-org.ezproxy.ugm.ac.id/document/9101730/?arnumber=9101730> [Accessed 8 Jul 2024].
- Huber, A., Schroeder, D. T., Pogorelov, K., Griwodz, C. and Langguth, J., 2022. A Streaming System for Large-scale Temporal Graph Mining of Reddit Data. *In: 2022 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW)* [online]. Presented at the 2022 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW), Lyon, France: IEEE, 1153–1162. Available from: <https://ieeexplore.ieee.org/document/9835250/> [Accessed 13 May 2024].
- Kolajo, T., Daramola, O. and Adebisi, A., 2019. Big data stream analysis: a systematic literature review. *Journal of Big Data*, 6 (1), 47.
- Kumar, J. and Garg, V., 2017. Security analysis of unstructured data in NOSQL MongoDB database. *In: 2017 International Conference on Computing and Communication Technologies for Smart Nation (IC3TSN)* [online]. Presented at the 2017 International Conference on Computing and Communication Technologies for Smart Nation (IC3TSN), 300–305. Available from: <https://ieeexplore-ieee-org.ezproxy.ugm.ac.id/document/8284495/?arnumber=8284495> [Accessed 8 Jul 2024].
- Miell, I. and Sayers, A. H., 2019. *Docker in practice*. Second edition. Shelter Island, New York: Manning Publications.

- Monish, P., Kumari, S. and Narendra Babu, C., 2018. Automated Topic Modeling and Sentiment Analysis of Tweets on SparkR. *In: 2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT)* [online]. Presented at the 2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Bangalore: IEEE, 1–7. Available from: <https://ieeexplore.ieee.org/document/8493973/> [Accessed 15 May 2023].
- Mulunda, C. K., Wagacha, P. W. and Muchemi, L., 2018. Review of Trends in Topic Modeling Techniques, Tools, Inference Algorithms and Applications. *In: 2018 5th International Conference on Soft Computing & Machine Intelligence (ISCMI)* [online]. Presented at the 2018 5th International Conference on Soft Computing & Machine Intelligence (ISCMI), Nairobi, Kenya: IEEE, 28–37. Available from: <https://ieeexplore.ieee.org/document/8703231/> [Accessed 15 May 2023].
- Poloskei, I., 2021. Continuous natural language processing pipeline strategy. *In: 2021 IEEE 15th International Symposium on Applied Computational Intelligence and Informatics (SACI)* [online]. Presented at the 2021 IEEE 15th International Symposium on Applied Computational Intelligence and Informatics (SACI), Timisoara, Romania: IEEE, 000221–000224. Available from: <https://ieeexplore.ieee.org/document/9465571/> [Accessed 7 Jul 2024].
- Rafi Imam, Md., Jyoti, O., Afrin, Z., Munawar Hossain, Md. and Hossain Mou, T., 2023. Suicidal Thought Detection Using NLP(Natural Language Processing) on Reddit Data. *In: 2023 26th International Conference on Computer and Information Technology (ICCIT)* [online]. Presented at the 2023 26th International Conference on Computer and Information Technology (ICCIT), Cox's Bazar, Bangladesh: IEEE, 1–6. Available from: <https://ieeexplore.ieee.org/document/10441163/> [Accessed 6 Jul 2024].
- Ramanan, B., Drabeck, L., Woo, T., Cauble, T. and Rana, A., 2020. ~PB&J~ - Easy Automation of Data Science/Machine Learning Workflows. *In: 2020 IEEE International Conference on Big Data (Big Data)* [online]. Presented at the 2020 IEEE International Conference on Big Data (Big Data), Atlanta, GA, USA: IEEE, 361–371. Available from: <https://ieeexplore.ieee.org/document/9378128/> [Accessed 8 Jul 2024].
- Raptis, T. P. and Passarella, A., 2023. A Survey on Networked Data Streaming With Apache Kafka. *IEEE Access*, 11, 85333–85350.
- Samanta, A. K. and Chaki, N., 2021. Performance Monitoring of MongoDB on Varied Cluster Configuration: An Experimental Approach. *In: 2021 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT)* [online]. Presented at the 2021 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT), Zallaq, Bahrain: IEEE, 525–530. Available from: <https://ieeexplore.ieee.org/document/9581673/> [Accessed 8 Jul 2024].

- Sekar, V. R., Kannan, T. K. R., N, S. and Vijay, P., 2023. Hybrid Perception Analysis of World Leaders in Reddit using Sentiment Analysis. *In: 2023 International Conference on Advances in Intelligent Computing and Applications (AICAPS)* [online]. Presented at the 2023 International Conference on Advances in Intelligent Computing and Applications (AICAPS), Kochi, India: IEEE, 1–5. Available from: <https://ieeexplore.ieee.org/document/10074005/> [Accessed 6 Jul 2024].
- Srinivasan, V. and K, C., 2021. Data Aggregation Of Tweets And Topic Modelling Based On The Twitter Dataset. *In: 2021 the 3rd International Conference on Big Data Engineering and Technology (BDET)* [online]. Presented at the BDET 2021: 2021 the 3rd International Conference on Big Data Engineering and Technology, Singapore Singapore: ACM, 15–20. Available from: <https://dl.acm.org/doi/10.1145/3474944.3474947> [Accessed 2 Oct 2022].
- Stoneman, E., n.d. Learn Docker in a Month of Lunches.
- Tun, M. T., Nyaung, D. E. and Phyu, M. P., 2019. Performance Evaluation of Intrusion Detection Streaming Transactions Using Apache Kafka and Spark Streaming. *In: 2019 International Conference on Advanced Information Technologies (ICAIT)* [online]. Presented at the 2019 International Conference on Advanced Information Technologies (ICAIT), Yangon, Myanmar: IEEE, 25–30. Available from: <https://ieeexplore.ieee.org/document/8920960/> [Accessed 8 Jul 2024].
- Wang, Z., Marella, S., Inada, M., Defino, P. R., Jagannath, A. and VijayaRenu, L., 2022. Latency vs Cost : Data Ingestion options at Twitter Scale. *In: 2022 IEEE International Conference on Big Data (Big Data)*. Presented at the 2022 IEEE International Conference on Big Data (Big Data), 223–227.
- Wingerath, W., Ritter, N. and Gessert, F., 2019. General-Purpose Stream Processing. *In: Real-Time & Stream Data Management* [online]. Cham: Springer International Publishing, 57–74. Available from: http://link.springer.com/10.1007/978-3-030-10555-6_5 [Accessed 7 Jul 2024].
- Wu, J. and Yang, T.-I., 2018. Dynamic CPU allocation for Docker containerized mixed-criticality real-time systems. *In: 2018 IEEE International Conference on Applied System Invention (ICASI)* [online]. Presented at the 2018 IEEE International Conference on Applied System Innovation (ICASI), Chiba: IEEE, 279–282. Available from: <https://ieeexplore.ieee.org/document/8394587/> [Accessed 8 Jul 2024].