DAFTAR PUSTAKA

Alash, H. M., & Al-Sultany, G. A. (2020). Improve topic modeling algorithms based on Twitter hashtags. *Journal of Physics: Conference Series*, *1660*(1), 012100. https://doi.org/10.1088/1742-6596/1660/1/012100

Annur, C. M. (2023). *Meski Trennya Turun, Media Online Tetap Jadi Sumber Berita Utama Masyarakat Indonesia*. https://databoks.katadata.co.id/datapublish/2023/06/16/meski-trennya-turun-media-online-tetap-jadi-sumber-berita-utama-masyarakat-indonesia

Anwar, K. (2022). Analisa sentimen Pengguna Instagram Di Indonesia Pada Review Smartphone Menggunakan Naive Bayes. *KLIK: Kajian Ilmiah Informatika Dan Komputer*, *2*(4), 148–155. https://doi.org/10.30865/klik.v2i4.315

Apriani, A., Zakiyudin, H., & Marzuki, K. (2021). Penerapan Algoritma Cosine Similarity dan Pembobotan TF-IDF System Penerimaan Mahasiswa Baru pada Kampus Swasta. *Jurnal Bumigora Information Technology (BITe)*, *3*(1), 19–27. https://doi.org/10.30812/bite.v3i1.1110

Blei, D. M. (2003). Latent Dirichlet Allocation. *The Journal of Machine Learning Research*, *3*, 993–1022.

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, *55*(4), 77–84. https://doi.org/10.1145/2133806.2133826

Borg, I., & Groenen, P. J. F. (2005). *Modern multidimensional scaling: Theory and applications* (2nd ed). Springer.

Cendana, M., & Permana, S. D. H. (2019). Pra-Pemrosesan Teks Pada Grup Whatsapp Untuk Pemodelan Topik. *Junal Mantik Penusa*, *3*(3), 107–116.

Chuang, J., Manning, C. D., & Heer, J. (2012). Termite: Visualization techniques for assessing textual topic models. *Proceedings of the International Working Conference on Advanced Visual Interfaces*, 74–77. https://doi.org/10.1145/2254556.2254572

Ellis, C. H., & Evans, C. E. L. (2022). Nutrition Communication in Public Health and the Media. In C. E. L. Evans, *Transforming Food Environments* (1st ed., pp. 173–185). CRC Press. https://doi.org/10.1201/9781003043720-12

Fahlevvi, M. R., & Sn, A. (2022). Topic Modeling on Online News.Portal Using Latent Dirichlet Allocation (LDA). *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, *16*(4), 335. https://doi.org/10.22146/ijccs.74383

Fauzi, A., Setiawan, E. B., & Baizal, Z. K. A. (2019). Hoax News Detection on Twitter using Term Frequency Inverse Document Frequency and Support Vector Machine Method. *Journal of Physics: Conference Series*, *1192*, 012025. https://doi.org/10.1088/1742-6596/1192/1/012025

Gadri, S., & Moussaoui, A. (2015). Information retrieval: A new multilingual stemmer based on a statistical approach. *2015 3rd International Conference on Control, Engineering & Information Technology (CEIT)*, 1–6. https://doi.org/10.1109/CEIT.2015.7233113

Gangadharan, V., & Gupta, D. (2020). Recognizing Named Entities in Agriculture Documents using LDA based Topic Modelling Techniques. *Procedia Computer Science*, *171*, 1337–1345. https://doi.org/10.1016/j.procs.2020.04.143

George, S., & Srividhya, V. (2020). *Comparison of LDA and NMF Topic Modeling Techniques for Restaurant Reviews*. *62*.

Griffiths, T. L., & Steyvers, M. (2002). A probabilistic approach to semantic representation. In W. D. Gray & C. D. Schunn (Eds.), *Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Society* (1st ed., pp. 381–386). Routledge. https://doi.org/10.4324/9781315782379-102

Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences*, *101*(suppl_1), 5228–5235. https://doi.org/10.1073/pnas.0307752101

Gunawan, B., Pratiwi, H. S., & Pratama, E. E. (2018). Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, *4*(2), 113. https://doi.org/10.26418/jp.v4i2.27526

Hadi, A. F., C. W., D. B., & Hasan, Moh. (2017). Text MIning Pada Media Sosial Twitter (Studi Kasus: Masa Tenang Pilkada 2017 Putaran 2). *Universitas Jember*.

Haldar, P., Viswanath, L., & Kumar Srivastava, A. (2022). Nutrition: A Boon To Healthy Early Childhood. *International Journal of Advanced Research*, *10*(02), 1124–1128. https://doi.org/10.21474/IJAR01/14320

Hanafi, A. (2009). *Pengenalan Bahasa Suku Bangsa Indonesia Berbasis Teks Menggunakan Metode N-gram* [Universitas Telkom]. https://repository.telkomuniversity.ac.id/pustaka/94404/pengenalan-bahasa-suku-bangsa-indonesia-berbasis-teks-menggunakan-metode-n-gram.html

Hardiyanti, L., Anggraini, D., & Kurniawati, A. (2023). Identify Reviews of Pedulilindungi Applications using Topic Modeling with Latent Dirichlet Allocation Method. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, *17*(4), 441. https://doi.org/10.22146/ijccs.86025

Hudaya, C. S., Fakhrurroja, H., & Alamsyah, A. (2019). Analisis Persepsi Konsumen Terhadap Brand GO-JEK Pada Media Sosial Twitter Menggunakan Metode Sentiment Aanalysis Dan Topic Modelling. *Jurnal Mitra Manajemen*, *3*(6), 664–673. https://doi.org/10.52160/ejmm.v3i6.244

Irianto, D. P. (2006). *Panduan Gizi Lengkap Keluarga Dan Olahragawan* (I). CV. Andi Offset.

Jurafsky, D., & Matrin, J. H. (2023). *Speech and Language Processing*. Stanford.

Kabiru, I. N., & Sari, P. K. (2019). Analisa Konten Media Sosial E-Commerce Pada Instagram Menggunakan Metode Sentimen Analysis Dan LDA-Based Topic Modeling (Studi Kasus: Shopee Indonesia). *eProceedings of Management*, *6*(1), 12–19.

Kaur, P., & Buttar, P. K. (2018). Review On Stemming Techniques. *International Journal of Advanced Research in Computer Science*, *9*(5), 64–68. https://doi.org/10.26483/ijarcs.v9i5.6308

Klohe, K., Prazeres Da Costa, C., Lien, N., Holmboe-Ottesen, G., Rychlik, M., Haavardsson, I., Stordalen, G., Singh, S., Engebretsen, I., Iversen, P. O., & Winkler, A. S. (2017). Nutrition – A global challenge for health. *Tidsskrift for Den norske legeforening*. https://doi.org/10.4045/tidsskr.17.0679

Laxmi, M. D., Indriati, I., & Fauzi, M. A. (2018). Query Expansion Pada Sistem Temu Kembali Informasi Berbahasa Indonesia Dengan Metode

Pembobotan TF-IDF Dan Algoritme Cosine Similarity Berbasis Wordnet. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *3*(1), 823–830.

Lin, J. (1991). Divergence measures based on the Shannon entropy. *IEEE Transactions on Information Theory*, *37*(1), 145–151. https://doi.org/10.1109/18.61115

Liu, J., Nie, H., Li, S., Chen, X., Cao, H., Ren, J., Lee, I., & Xia, F. (2021). Tracing the Pace of COVID-19 Research: Topic Modeling and Evolution. *Big Data Research*, *25*, 100236. https://doi.org/10.1016/j.bdr.2021.100236

M C, S. P., Reddy, B. R., Tharun Reddy, D. S., & Gupta, D. (2022). Comparative Analysis of Research Papers Categorization using LDA and NMF Approaches. *2022 IEEE North Karnataka Subsection Flagship International Conference (NKCon)*, 1–7. https://doi.org/10.1109/NKCon56289.2022.10127059

Natalia, C., Suprata, F., Surbakti, F. P. S., & Clarence, S. (2021). Penentuan Standar Spesifikasi Kerja di Café Berdasarkan Big Data dengan Metode LDA dan AHP. *Jurnal Rekayasa Sistem Industri*, *10*(2), 211–226. https://doi.org/10.26593/jrsi.v10i2.5228.211-226

Nugroho, D. D. A., & Alamsyah, A. (2018). Analisis Konten Pelanggan Airbnb Pada Network Sosial Media Twitter. *eProceedings of Management Current Archives About*, *5*(2), 1623–1626.

Prasanna, P. L., & Rao, D. R. (2019). A Text Mining Research Based on Topic Modeling using Latent Dritchlent Allocation. *International Journal of Recent Technology and Engineering*, *7*(5), 308–317.

Prihatini, P. M., Suryawan, I. K., & Mandia, I. (2017). Feature extraction for document text using Latent Dirichlet Allocation. *Journal of Physics: Conference Series*, *953*, 012047. https://doi.org/10.1088/1742-6596/953/1/012047

Putra, K. B., & Kusumawardani, R. P. (2017). Analisis Topik Informasi Publik Media Sosial di Surabaya Menggunakan Pemodelan Latent Dirichlet Allocation (LDA). *Jurnal Teknik ITS*, *6*(2), A446-450. https://doi.org/10.12962/j23373539.v6i2.23205

Pypi.org. (2023). *pyLDAvis*. https://pypi.org/project/pyLDAvis/

Rahmawati, T. (2020). Aplikasi Principal Component Analysis (PCA) Untuk Mereduksi Faktor-Faktor Yang Berpengaruh Dalam Peramalan Konsumsi Listrik. *Teknomatika: Jurnal Informatika Dan Komputer*, *7*(1), 21–32.

Rizkia, S. (2019). *Analisis Sentimen Kepuasan Pelanggan Terhadap Internet Provider Indihome di Twitter Menggunakan Metode Decision Tree dan Pembobotan TF-IDF* [Universitas Telkom]. https://repository.telkomuniversity.ac.id/pustaka/152195/analisis-sentimen-kepuasan-pelanggan-terhadap-internet-provider-indihome-di-twitter-menggunakan-metode-decision-tree-dan-pembobotan-tf-idf.html

Robert, Delir Haghighi, P., Burstein, F., Urquhart, D., & Cicuttini, F. (2021). Investigating Individuals' Perceptions Regarding the Context Around the Low Back Pain Experience: Topic Modeling Analysis of Twitter Data. *Journal of Medical Internet Research*, *23*(12), e26093. https://doi.org/10.2196/26093

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

Sahria, Y., & Fudholi, D. H. (2020). Analysis of Health Research Topics in Indonesia Using the LDA (Latent Dirichlet Allocation) Topic Modeling Method. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *4*(2), 336–344. https://doi.org/10.29207/resti.v4i2.1821

Setiawan, J. H., & Stellarosa, Y. (2021). Analisis Isi Pemberitaan COVID-19 Pada Media Online Di Indonesia Maret 2020 – Februari 2021. *LSPR Communication And Business Institute*.

Setijohatmo, U. T., Rachmat, S., Susilawati, T., & Rahman, Y. (2020). Analisis Metoda Latent Dirichlet Allocation untuk Klasifikasi Dokumen Laporan Tugas Akhir Berdasarkan Pemodelan Topik. *Prosiding The 11th Industrial Research Workhop and National Seminar*, *11*(1).

Sievert, C., & Shirley, K. (2014). LDAvis: A method for visualizing and interpreting topics. *Proceedings of the Workshop on Interactive Language*

Learning, Visualization, and Interfaces, 63–70. https://doi.org/10.3115/v1/W14-3110

Siregar, R. R. A., Sinaga, F. A., & Arianto, R. (2017). Aplikasi Penentuan Dosen Penguji Skripsi Menggunakan Metode TF-IDF dan Vector Space Model. *Computatio : Journal of Computer Science and Information Systems*, *1*(2), 171. https://doi.org/10.24912/computatio.v1i2.1014

Socrates, I. G. A., Akbar, A. L., Akbar, M. S., Arifin, A. Z., & Herumurti, D. (2016). Optimasi Naive Bayes Dengan Pemilihan Fitur Dan Pembobotan Gain Ratio. *Lontar Komputer : Jurnal Ilmiah Teknologi Informasi*, 22. https://doi.org/10.24843/LKJITI.2016.v07.i01.p03

Sumadiria, H. (2005). *Jurnalistik Indonesia menulis berita dan feature: Panduan praktis jurnalis profesional* (Cet. 1). Simbiosa Rekatama Media.

Syaifuddin, A., Harianto, R. A., & Santoso, J. (2020). Analisis Trending Topik untuk Percakapan Media Sosial dengan Menggunakan Topic Modelling Berbasis Algoritme LDA. *Journal of Intelligent System and Computation*, *2*(1), 12–19. https://doi.org/10.52985/insyst.v2i1.150

Utami, M. T., Tulili, T. R., & Topadang, A. (2017). Implementasi Metode City Block Distance pada Identifikasi Citra Tanda Tangan. *JTT (Jurnal Teknologi Terpadu)*, *5*(2), 134. https://doi.org/10.32487/jtt.v5i2.273

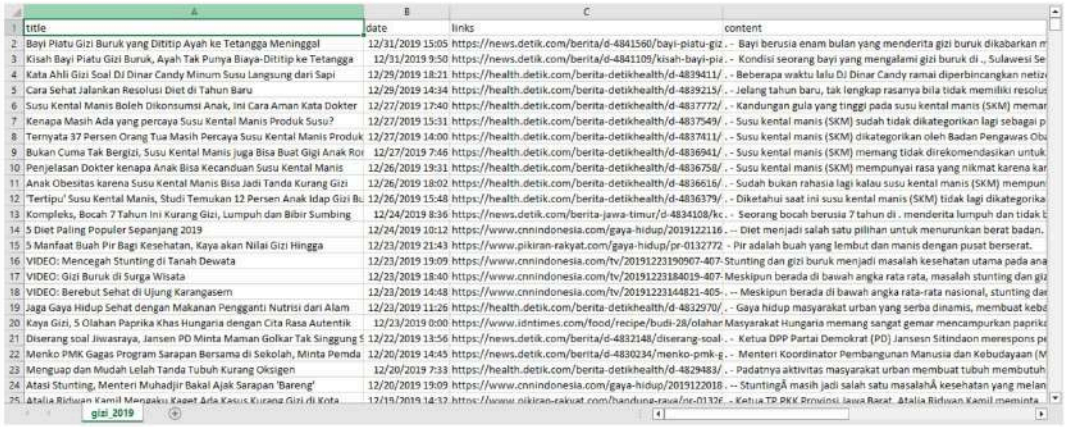
Utomo, P. E. P., Manaar, M., Khaira, U., & Suratno, T. (2021). Analisis Sentimen Online Review Pengguna Bukalapak Menggunakan Metode Algoritma TF-IDF. *JUSS (Jurnal Sains Dan Sistem Informasi)*, *2*(2), 35–39. https://doi.org/10.22437/juss.v2i2.8469

Wahyu, P. G. G., Setiawan, I., & Saputri, R. I. (2023). Gambaran Perilaku Masyarakat Dalam Mencari Informasi Kesehatan Melalui Internet (Studi pada Kecamatan Pasirjambu, Kabupaten Bandung). *Padjadjaran Journal of Dental Researchers and Students*, *7*(1), 81. https://doi.org/10.24198/pjdrs.v7i1.40474

Y, C., Kiran, P., & P B, M. (2022). The Novel Method for Data Preprocessing CLI. *Advances in Intelligent Systems and Technologies*, 117–120. https://doi.org/10.53759/aist/978-9914-9946-1-2_21

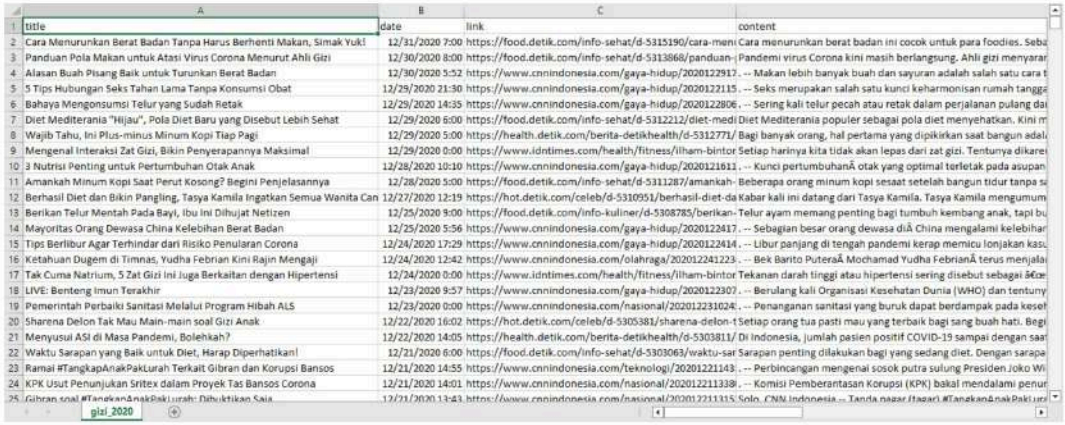
Yoren. (2018). *Perbandingan Raw TF dan Binary TF pada pencarian di situs Museum Wayang Kekayon Yogyakarta* [Universitas Sanata Dharma Yogyakarta]. http://repository.usd.ac.id/id/eprint/32223

Lampiran 1 Hasil pengambilan data



	A	B	C	
1	title	date	links	content
2	Bayi Piatu Gizi Buruk yang Dititip Ayah ke Tetangga Meninggal	12/31/2019 15:05	https://news.detik.com/berita/d-4841560/bayi-piatu-giz	. - Bayi berusia enam bulan yang menderita gizi buruk dikabarkan m
3	Kisah Bayi Piatu Gizi Buruk, Ayah Tak Punya Biaya-Dititip ke Tetangga	12/31/2019 9:50	https://news.detik.com/berita/d-4841109/kisah-bayi-pia	. - Kondisi seorang bayi yang mengalami gizi buruk di ., Sulawesi Se
4	Kata Ahli Gizi Soal DJ Dinar Candy Minum Susu Langsung dari Sapi	12/29/2019 18:21	https://health.detik.com/berita-detikhealth/d-4839411/	. - Beberapa waktu lalu DJ Dinar Candy ramai diperbincangkan netiz
5	Cara Sehat Jalankan Resolusi Diet di Tahun Baru	12/29/2019 14:34	https://health.detik.com/berita-detikhealth/d-4839215/	. - Jelang tahun baru, tak lengkap rasanya bila tidak memiliki resolus
6	Susu Kental Manis Boleh Dikonsumsi Anak, Ini Cara Aman Kata Dokter	12/27/2019 17:40	https://health.detik.com/berita-detikhealth/d-4837772/	. - Kandungan gula yang tinggi pada susu kental manis (SKM) memar
7	Kenapa Masih Ada yang percaya Susu Kental Manis Produk Susu?	12/27/2019 15:31	https://health.detik.com/berita-detikhealth/d-4837549/	. - Susu kental manis (SKM) sudah tidak dikategorikan lagi sebagai p
8	Ternyata 37 Persen Orang Tua Masih Percaya Susu Kental Manis Produk	12/27/2019 14:00	https://health.detik.com/berita-detikhealth/d-4837411/	. - Susu kental manis (SKM) dikategorikan oleh Badan Pengawas Obı
9	Bukan Cuma Tak Bergizi, Susu Kental Manis juga Bisa Buat Gigi Anak Ror	12/27/2019 7:46	https://health.detik.com/berita-detikhealth/d-4836941/	. - Susu kental manis (SKM) memang tidak direkomendasikan untuk
10	Penjelasan Dokter kenapa Anak Bisa Kecanduan Susu Kental Manis	12/26/2019 19:31	https://health.detik.com/berita-detikhealth/d-4836758/	. - Susu kental manis (SKM) mempunyai rasa yang nikmat karena kar
11	Anak Obesitas karena Susu Kental Manis Bisa Jadi Tanda Kurang Gizi	12/26/2019 18:02	https://health.detik.com/berita-detikhealth/d-4836616/	. - Sudah bukan rahasia lagi kalau susu kental manis (SKM) mempun
12	'Tertipu' Susu Kental Manis, Studi Temukan 12 Persen Anak Idap Gizi Bu	12/26/2019 15:48	https://health.detik.com/berita-detikhealth/d-4836379/	. - Diketahui saat ini susu kental manis (SKM) tidak lagi dikategorika
13	Kompleks, Bocah 7 Tahun Ini Kurang Gizi, Lumpuh dan Bibir Sumbing	12/24/2019 8:36	https://news.detik.com/berita-jawa-timur/d-4834108/kc	. - Seorang bocah berusia 7 tahun di . menderita lumpuh dan tidak b
14	5 Diet Paling Populer Sepanjang 2019	12/24/2019 10:12	https://www.cnnindonesia.com/gaya-hidup/2019122116	. -- Diet menjadi salah satu pilihan untuk menurunkan berat badan.
15	5 Manfaat Buah Pir Bagi Kesehatan, Kaya akan Nilai Gizi Hingga	12/23/2019 21:43	https://www.pikiran-rakyat.com/gaya-hidup/pr-0132772	- Pir adalah buah yang lembut dan manis dengan pusat berserat.
16	VIDEO: Mencegah Stunting di Tanah Dewata	12/23/2019 19:09	https://www.cnnindonesia.com/tv/20191223190907-407-	Stunting dan gizi buruk menjadi masalah kesehatan utama pada ana
17	VIDEO: Gizi Buruk di Surga Wisata	12/23/2019 18:40	https://www.cnnindonesia.com/tv/20191223184019-407-	Meskipun berada di bawah angka rata rata, masalah stunting dan giz
18	VIDEO: Berebut Sehat di Ujung Karangasem	12/23/2019 14:48	https://www.cnnindonesia.com/tv/20191223144821-405-	. -- Meskipun berada di bawah angka rata-rata nasional, stunting dar
19	Jaga Gaya Hidup Sehat dengan Makanan Pengganti Nutrisi dari Alam	12/23/2019 11:26	https://health.detik.com/berita-detikhealth/d-4832970/	. - Gaya hidup masyarakat urban yang serba dinamis, membuat keba
20	Kaya Gizi, 5 Olahan Paprika Khas Hungaria dengan Cita Rasa Autentik	12/23/2019 0:00	https://www.idntimes.com/food/recipe/budi-28/olahar	Masyarakat Hungaria memang sangat gemar mencampurkan paprik
21	Diserang soal Jiwasraya, Jansen PD Minta Maman Golkar Tak Singgung S	12/22/2019 13:56	https://news.detik.com/berita/d-4832148/diserang-soal	. - Ketua DPP Partai Demokrat (PD) Jansesn Sitindaon merespons pe
22	Menko PMK Gagas Program Sarapan Bersama di Sekolah, Minta Pemda	12/20/2019 14:45	https://news.detik.com/berita/d-4830234/menko-pmk-g	. - Menteri Koordinator Pembangunan Manusia dan Kebudayaan (M
23	Menguap dan Mudah Lelah Tanda Tubuh Kurang Oksigen	12/20/2019 7:33	https://health.detik.com/berita-detikhealth/d-4829483/	. - Padatnya aktivitas masyarakat urban membuat tubuh membutuh
24	Atasi Stunting, Menteri Muhadjir Bakal Ajak Sarapan 'Bareng'	12/20/2019 19:09	https://www.cnnindonesia.com/gaya-hidup/2019122018	-- StuntingÃ masih jadi salah satu masalahÃ kesehatan yang melan
25	Atalia Ridwan Kamil Mengaku Kaget Ada Kasus Kurang Gizi di Kota	12/19/2019 14:12	https://www.pikiran-rakyat.com/bandung-raya/pr-0132€	. - Ketua TP PKK Provinsi Jawa Barat, Atalia Ridwan Kamil meminta

gizi_2019



	A	B	C	
1	title	date	link	content
2	Cara Menurunkan Berat Badan Tanpa Harus Berhenti Makan, Simak Yuk!	12/31/2020 7:00	https://food.detik.com/info-sehat/d-5315190/cara-men	Cara menurunkan berat badan ini cocok untuk para foodies. Seba
3	Panduan Pola Makan untuk Atasi Virus Corona Menurut Ahli Gizi	12/30/2020 8:00	https://food.detik.com/info-sehat/d-5313868/panduan-	Pandemi virus Corona kini masih berlangsung. Ahli gizi menyarar
4	Alasan Buah Pisang Baik untuk Turunkan Berat Badan	12/30/2020 5:52	https://www.cnnindonesia.com/gaya-hidup/2020122917	. -- Makan lebih banyak buah dan sayuran adalah salah satu cara t
5	5 Tips Hubungan Seks Tahan Lama Tanpa Konsumsi Obat	12/29/2020 21:30	https://www.cnnindonesia.com/gaya-hidup/2020122115	. -- Seks merupakan salah satu kunci keharmonisan rumah tangga
6	Bahaya Mengonsumsi Telur yang Sudah Retak	12/29/2020 14:35	https://www.cnnindonesia.com/gaya-hidup/2020122806	. -- Sering kali telur pecah atau retak dalam perjalanan pulang dai
7	Diet Mediterania "Hijau", Pola Diet Baru yang Disebut Lebih Sehat	12/29/2020 6:00	https://food.detik.com/info-sehat/d-5312212/diet-medi	Diet Mediterania populer sebagai pola diet menyehatkan. Kini m
8	Wajib Tahu, Ini Plus-minus Minum Kopi Tiap Pagi	12/29/2020 5:00	https://health.detik.com/berita-detikhealth/d-5312771/	Bagi banyak orang, hal pertama yang dipikirkan saat bangun adal
9	Mengenal Interaksi Zat Gizi, Bikin Penyerapannya Maksimal	12/29/2020 0:00	https://www.idntimes.com/health/fitness/ilham-bintor	Setiap harinya kita tidak akan lepas dari zat gizi. Tentunya dikare
10	3 Nutrisi Penting untuk Pertumbuhan Otak Anak	12/28/2020 10:10	https://www.cnnindonesia.com/gaya-hidup/2020121611	. -- Kunci pertumbuhanÃ otak yang optimal terletak pada asupan
11	Amankah Minum Kopi Saat Perut Kosong? Begini Penjelasannya	12/28/2020 5:00	https://food.detik.com/info-sehat/d-5311287/amankah-	Beberapa orang minum kopi sesaat setelah bangun tidur tanpa s
12	Berhasil Diet dan Bikin Pangling, Tasya Kamila Ingatkan Semua Wanita Can	12/27/2020 12:19	https://hot.detik.com/celeb/d-5310951/berhasil-diet-da	Kabar kali ini datang dari Tasya Kamila. Tasya Kamila mengumum
13	Berikan Telur Mentah Pada Bayi, Ibu Ini Dihujat Netizen	12/25/2020 9:00	https://food.detik.com/info-kuliner/d-5308785/berikan-	Telur ayam memang penting bagi tumbuh kembang anak, tapi bu
14	Mayoritas Orang Dewasa China Kelebihan Berat Badan	12/25/2020 5:56	https://www.cnnindonesia.com/gaya-hidup/2020122417	. -- Sebagian besar orang dewasa diÃ China mengalami kelebihar
15	Tips Berlibur Agar Terhindar dari Risiko Penularan Corona	12/24/2020 17:29	https://www.cnnindonesia.com/gaya-hidup/2020122414	. -- Libur panjang di tengah pandemi kerap memicu lonjakan kas
16	Ketahuan Dugem di Timnas, Yudha Febrian Kini Rajin Mengaji	12/24/2020 12:42	https://www.cnnindonesia.com/olahraga/20201224122	. -- Bek Barito PuteraÃ Mochamad Yudha FebrianÃ terus menjalai
17	Tak Cuma Natrium, 5 Zat Gizi Ini Juga Berkaitan dengan Hipertensi	12/24/2020 0:00	https://www.idntimes.com/health/fitness/ilham-bintor	Tekanan darah tinggi atau hipertensi sering disebut sebagai â€œe
18	LIVE: Benteng Imun Terakhir	12/23/2020 9:57	https://www.cnnindonesia.com/gaya-hidup/2020122307	. -- Berulang kali Organisasi Kesehatan Dunia (WHO) dan tentuny
19	Pemerintah Perbaiki Sanitasi Melalui Program Hibah ALS	12/23/2020 0:00	https://www.cnnindonesia.com/nasional/20201223024	. -- Penanganan sanitasi yang buruk dapat berdampak pada keset
20	Sharena Delon Tak Mau Main-main soal Gizi Anak	12/22/2020 16:02	https://hot.detik.com/celeb/d-5305381/sharena-delon-t	Setiap orang tua pasti mau yang terbaik bagi sang buah hati. Begi
21	Menyusui ASI di Masa Pandemi, Bolehkah?	12/22/2020 14:05	https://health.detik.com/berita-detikhealth/d-5303811/	Di Indonesia, jumlah pasien positif COVID-19 sampai dengan saat
22	Waktu Sarapan yang Baik untuk Diet, Harap Diperhatikan!	12/21/2020 6:00	https://food.detik.com/info-sehat/d-5303063/waktu-sar	Sarapan penting dilakukan bagi yang sedang diet. Dengan sarapa
23	Ramai #TangkapAnakPakLurah Terkait Gibran dan Korupsi Bansos	12/21/2020 14:55	https://www.cnnindonesia.com/teknologi/20201221143	. -- Perbincangan mengenai sosok putra sulung Presiden Joko Wi
24	KPK Usut Penunjukan Sritex dalam Proyek Tas Bansos Corona	12/21/2020 14:01	https://www.cnnindonesia.com/nasional/20201221133€	. -- Komisi Pemberantasan Korupsi (KPK) bakal mendalami penur
25	Gibran soal #TangkapAnakPakLurah: Dibuktikan Saja	12/21/2020 13:43	https://www.cnnindonesia.com/nasional/20201221131	Solo, CNN Indonesia -- Tanda pagar (tagar) #TangkapAnakPakLura

gizi_2020

Link: Dataset lengkap

Lampiran 2 Contoh berita

. -- Calon wakil presiden nomor urut 3 Mahfud MD menyindir program makan siang dan susu gratis yang dicanangkan pasangan calon nomor urut 2 Prabowo Subianto dan Gibran Rakabuming Raka.Mahfud membandingkan program itu dengan program yang ia dan Ganjar Pranowo tawarkan yang dinamakan Gastronomi."Makan siang gratis susu dan sebagainya itu kan impor, kira-kira barang-barang impor. Kalau Gastronomi dari bumi-bumi kita dan laut laut kita," kata Mahfud MD saat jumpa pers di Djakarta Theatre, Jakarta, Sabtu (30/12). Lihat Juga :Mahfud Pamer 21 Program Unggulan: Lebih dari Sekadar Makan SiangMahfud mempertanyakan prospek jangka panjang program makan siang dan susu gratis yang ditawarkan oleh Prabowo dan Gibran.Ia mengatakan rakyat ibaratnya harus diberi kail pancing, bukan cuma ikan. Di sisi lain, pasangan Ganjar-Mahfud merancang program jangka panjang soal pemberian gizi bagi masyarakat. Mereka ingin memberi makanan yang betul-betul bernilai gizi."Bukan hanya makan siang, tapi makanannya juga sehat," ujar Mahfud MD.Ganjar-Mahfud menghadapi dua pasangan calon di Pilpres 2024. Dua pasangan itu adalah Anies Baswedan-Muhaimin Iskandar dan Prabowo Subianto-Gibran Rakabuming Raka.Salah satu program yang populer di pilpres ini adalah program makan siang gratis dari pasangan Prabowo-Gibran.Sebelumnya, Prabowo-Gibran gencar mengkampanyekan program makan siang dan susu gratis. Program itu mereka tawarkan untuk menjawab persoalan stunting dan tantangan generasi emas Indonesia.Lihat Juga :Ganjar Desak DPR Usut KPU soal Surat Suara di Taipei: Kalau Lalai LucuMega proyek itu diprediksi butuh Rp400 triliun per tahun. Dewan Pakar TKN Prabowo-Gibran Panji Irawan yakin anggaran Indonesia cukup untuk mendanai program itu."Kami sudah menghitung. Jadi memang angkanya bisa mencapai mungkin ratusan triliun, tetapi kita juga sudah menghitung bahwasanya di dalam kita punya koleksi dari tax (pajak) masih banyak kebocoran," ungkap Panji. (dhf/pra)

Lampiran 3 *Source code preprocessing* data

```
#case folding
data['content'] = data['content'].astype(str)
data['content'] = data['content'].apply(lambda x:" ".
join(x.lower() for x in x.split()))
data['content'].head()

#punctuation removal
data['content'] = data['content'].str.replace('[^\w\s]',''
',regex = True)
data['content']

#number removal
data['content'] = data['content'].str.replace(r'[\d+]',' ',
regex = True)
data['content']

#stopword
additional_stopwords = ['baca', 'gr', 'gram', 'seks',
'lainnya', 'tujuannya', 'pemeriksaan', 'iye', 'ndag', 'sih',
'per', 'gairah', 'simak', 'video', 'gambas', 'wdw', 'fds',
'persen', 'persennya', 'tak', 'ngga', 'cm','mmhg', 'foto',
'ega', 'scroll', 'to', 'continue', 'with', 'content',
'advertisement', 'prf', 'lihat','fey', 'chs', 'daftar', 'mmu',
'adr', 'odi', 'up', 'ain', 'ard',  'adv', 'rs', 'rsud',
'miligram','tips', 'hut', 'advertorial', 'resep', 'halaman',
'membaca', 'ya', 'lus', 'ribu', 'kg', 'kilogram', 'survei',
'deh', 'ujung', 'ujungnya', 'sang', 'wib', 'senin', 'selasa',
'rabu', 'kamis', 'jumat', 'sabtu', 'minggu', 'satu', 'dua',
'tiga', 'empat', 'lima', 'enam', 'tujuh', 'delapan',
'sembilan', 'triliun', 'kna', 'bbn', 'juta', 'kilometer',
'cnn', 'simak', 'lho', 'com', 'kkal', 'm', 'trik', 'detik',
'ton', 'agt', 'mengutip', 'bercinta', 'libido', 'foreplay',
'tahan', 'yook', 'mah', 'miliar', 'live', 'pt', 'saksikan',
'psp', 'agn', 'januari', 'februari', 'maret', 'april', 'mei',
'juni', 'juli', 'agustus', 'september', 'oktober', 'november',
'desember', 'asr', 'papar', 'rt', 'rw', 'rp', 'g', 'menit',
'avd', 'fjr', 'klik', 'berikut', 'instagram', 'pixabay',
'istockphoto', 'ilustrasi', 'ekor', 'kilo', 'vs', 'nma',
'selamat', 'mencoba', 'els', 'review', 'youtube', 'tiktok',
'tiktoknya', 'tv', 'bmw', 'pikiran', 'rakyat', 'cnnindonesia',
'si', 'update', 'sdm', 'sdt', 'tim', 'tayang', 'catat', 'rea',
'laper', 'seksual', 'photo', 'fat', 'sbmptn', 'prodi',
'jurusan', 'sendok', 'fef', 'nomor', 'bernomor', 'jam',
'galau', 'derajat', 'celcius', 'perguruan', 'ltmpt', 'wis',
'dirumahaja', 'ard', 'iwd', 'ptj', 'wita', 'infografis', 'fl',
'oz', 'cerita', 'ita', 'tokoh', 'figur', 'year', 'review',
'xxl', 'm', 'piala', 'kilas', 'dosen', 'dilansir', 'penulis',
'rds', 'idn', 'times', 'aff', 'dl', 'akun', 'osc',
```

```
'mengunjungi', 'whatsapp', 'livestrong', 'kuliner', 'intip',
'saintek', 'soshum', 'artikel', 'iwd', 'youtubenya', 'bumbu',
'gambaran', 'teruskan', 'lo', 'bir', 'inflasi', 'ekonomi',
'perekonomian', 'mel', 'aud', 'cup', 'mel', 'melansir',
'cangkir', 'imb', 'twitter', 'dihubungi', 'detikfood',
'snmptn', 'imbuhnya', 'box', 'pmg', 'isa', 'eks', 'dl', 'oh',
'ya', 'iya', 'so', 'asyik', 'download', 'lth', 'del', 't',
'detikhealth', 'tbk', 'semoga', 'detikcom', 'vip', 'vvip',
'cc', 'inh', 'tst', 'saksikan', 'info', 'piala', 'u', 'har',
'yuk', 'unggahan', 'please', 'pop', 'us', 'dzu', 'rzr', 'fby',
'rasulullah', 'saw', 'nabi', 'page', 'dhf', 'ain', 'food',
'ndtv', 'aor', 'ceo', 'xl', 's', 'mab', 'pua', 'next', 'inh',
'juh', 'tsa', 'rir', 'ory', 'sur', 'mnf', 'daftar', 'isi',
'frd', 'dal', 'sfr', 'dzu', 'put', 'surat',
'pixabayilustrasi', 'nan', 'sarap', 'tinggal', 'baiknya',
'berkali', 'kali', 'kurangnya', 'mata', 'olah', 'sekurang',
'setidak', 'tama', 'tidaknya']

with open('/content/drive/MyDrive/SKRIPSI/Topic
Modeling/2019/stopword.txt') as file:
    stopwords = file.read().split()

all_stopwords = stopwords + additional_stopwords

data['content'] = data['content'].apply(lambda x:
                                        " ".join(word for word
in x.split()
                                                 if word not
in all_stopwords))

print(data['content'])

#Tokenizing
import re
# Function to Tokenize words
def tokenize(text):
    tokens = re.split('\W+', text) #W+ means that either a
word character (A-Za-z0-9_) or a dash (-) can go there.
    return tokens
def convertToString(term):
  if type(term) is str:
    return term
  else:
    return str(term)
data['content'] = data['content'].apply(lambda x:
tokenize(x.lower()))
#We convert to lower as Python is case-sensitive.
data.head()
```

```
#stemming
#create stemmer
factory = StemmerFactory()
stemmer = factory.create_stemmer()
#stemming process
def kata_stem(teks):
    stem_teks = " ".join([stemmer.stem(i) for i in teks])
    return stem_teks
data['content'] = data['content'].apply(lambda x:
kata_stem(x))
data.head()

#tokenizing
def tokenize(text):
    tokens = re.split('\W+', text) #W+ means that either a
word character (A-Za-z0-9_) or a dash (-) can go there.
    return tokens
data['content'] = data['content'].apply(lambda x:
tokenize(x.lower()))
#We convert to lower as Python is case-sensitive.
data.head()
```

Lampiran 4 *Source code feature extraction*

```
# Baca file CSV (ganti 'nama_file.csv' dengan nama file yang
sesuai)
df = pd.read_csv('/content/drive/MyDrive/SKRIPSI/Topic
Modeling/2019/preprocessinggizi2019.csv')

# Kolom yang ingin digunakan (ganti 'content' dengan nama
kolom yang sesuai)
content_column = 'content'

# Tokenisasi (misalnya, dengan memisahkan berdasarkan spasi)
def preprocess(text):
    return text.lower().split()

# Hitung TF (Term Frequency)
def calculate_tf(term, document):
    tokens = preprocess(document)
    term_count = tokens.count(term.lower())
    total_words = len(tokens)
    return term_count / total_words

# Hitung DF (Document Frequency)
def calculate_df(term, documents):
    doc_count = sum(1 for doc in documents for item in
preprocess(doc) if term.lower() == item.lower())
    return doc_count

# Hitung IDF (Inverse Document Frequency)
def calculate_idf(term, documents):
    N = len(documents)
    df = calculate_df(term, documents)
    idf = math.log(N / (df + 1))
    return idf

# Hitung TF-IDF untuk semua kata
all_documents = df[content_column]
tfidf_scores = {}
for doc_content in all_documents:
    tokens = preprocess(doc_content)
    for token in tokens:
        tf = calculate_tf(token, doc_content)
        df = calculate_df(token, all_documents)
        idf = calculate_idf(token, all_documents)
        tfidf = tf * idf
        tfidf_scores[token] = (tf, df, idf, tfidf)

# # Tampilkan hasil
# print(f"{'Kata':<25} {'TF':<10} {'DF':<8} {'IDF':<12} {'TF-
IDF':<10}")
```

```
# print("-" * 65)
# for word, (tf, df, idf, tfidf) in tfidf_scores.items():
#     print(f"{word:<25} {tf:.5f}    {df:<8} {idf:.5f}
 {tfidf:.5f}")


# Membuat DataFrame baru dari hasil perhitungan TF-IDF
tfidf_df = pd.DataFrame(tfidf_scores).T
tfidf_df.columns = ['TF', 'DF', 'IDF', 'TF-IDF']
tfidf_df.index.name = 'Kata'

# Menyimpan DataFrame ke dalam file CSV
tfidf_df.to_csv('/content/drive/MyDrive/SKRIPSI/Topic
Modeling/2019/hasil_tfidf_2019.csv')

# Menampilkan DataFrame (opsional)
print(tfidf_df)
```

```
#Create Bigram & Trigram
#bigram & trigram
df = pd.read_csv('/content/drive/MyDrive/SKRIPSI/Topic
Modeling/2019/preprocessinggizi2019.csv') #create data frame

text = df['content']
text_list = []
for i in range(len(text)) :
    bbb = text[i].replace('[', '')
    bbb = bbb.replace(']', '')
    bbb = bbb.replace("'", "")
    bbb = bbb.replace(",", "")
    temp = []
    for j in bbb.split() :
        temp.append(j)
    text_list.append(temp)

print(len(text_list))

df.head()

print(text_list)

# Add bigrams and trigrams to docs,minimum count 25 means only
that appear 25 times or more.
bigram = Phrases(text_list, min_count=20)
trigram = Phrases(bigram[text_list])

for idx in range(len(text_list)):
    for token in bigram[text_list[idx]]:
        if '_' in token:
```

```
            # Token is a bigram, add to document.
            text_list[idx].append(token)
    for token in trigram[text_list[idx]]:
        if '_' in token:
            # Token is a bigram, add to document.
            text_list[idx].append(token)

# Menampilkan bigram dan trigram
print("Bigrams: ", bigram)
print("Trigrams: ", trigram)
```

Lampiran 5 *Source code topic modeling*

```
# Create a dictionary representation of the documents.
dictionary = corpora.Dictionary(text_list)

dictionary.filter_extremes(no_below=1, no_above=0.1)
#no_below (int, optional) - Keep tokens which are contained in
at least no_below documents.
#no_above (float, optional) - Keep tokens which are contained
in no more than no_above documents (fraction of total corpus
size, not an absolute number).

print(dictionary)

#build corpus
doc_term_matrix = [dictionary.doc2bow(doc) for doc in
text_list]

print(len(doc_term_matrix))
print(doc_term_matrix[100])

tfidf = models.TfidfModel(doc_term_matrix) #build TF-IDF model
corpus_tfidf = tfidf[doc_term_matrix]
```

```
import numpy as np
from tabulate import tabulate

class LDA:
    def __init__(self, num_topics, alpha, beta, num_iters,
random_state=None):
        self.num_topics = num_topics
        self.alpha = alpha
        self.beta = beta
        self.num_iters = num_iters
        self.random_state = random_state

    def fit(self, corpus_tfidf, dictionary,
random_state=None):
        if self.random_state is not None:
            np.random.seed(self.random_state)
        # Initialize
        self.corpus = corpus_tfidf
        self.dictionary = dictionary
        self.num_docs = len(corpus_tfidf)
        self.vocab_size = len(dictionary)
        self.doc_lengths = [sum(freq for _, freq in doc) for
doc in corpus_tfidf]
        # Initialize topic assignments randomly
        self.doc_topic_counts = np.zeros((self.num_docs,
self.num_topics))
```

```
        self.topic_word_counts = np.zeros((self.num_topics,
self.vocab_size))
        self.topic_counts = np.zeros(self.num_topics)
        self.dist_topics = np.zeros((self.num_docs,
self.num_topics))

        for doc_idx, doc in enumerate(corpus_tfidf):
            for word_id, freq in doc:
                topic = np.random.randint(self.num_topics)
                self.doc_topic_counts[doc_idx, topic] += freq
                self.topic_word_counts[topic, word_id] += freq
                self.topic_counts[topic] += freq

        # Gibbs sampling
        for _ in range(self.num_iters):
            for doc_idx, doc in enumerate(corpus_tfidf):
                for word_id, freq in doc:
                    topic = self._sample_topic(doc_idx,
word_id)
                    self.doc_topic_counts[doc_idx, topic] +=
freq
                    self.topic_word_counts[topic, word_id] +=
freq
                    self.topic_counts[topic] += freq

        # Compute topic distributions
        self.dist_topics = (self.doc_topic_counts +
self.alpha) / (
            np.sum(self.doc_topic_counts, axis=1)[:,
np.newaxis] + self.num_topics * self.alpha
        )

        self.dist_words = (self.topic_word_counts + self.beta)
/ (
            np.sum(self.topic_word_counts, axis=1)[:,
np.newaxis] + self.vocab_size * self.beta
        )

    def _sample_topic(self, doc_idx, word_id):
        probs = (
            (self.topic_word_counts[:, word_id] + self.beta) /
            (self.topic_counts + self.vocab_size * self.beta)
*
            (self.doc_topic_counts[doc_idx, :] + self.alpha)
        )
        probs /= np.sum(probs)
        return np.random.choice(self.num_topics, p=probs)

    def get_topics(self, num_words=10):
```

```
        topics = []
        for topic_idx in range(self.num_topics):
            topic_dist = self.dist_words[topic_idx, :]
            top_word_ids = np.argsort(topic_dist)[::-
1][:num_words]
            top_words = [self.dictionary[word_id] for word_id
in top_word_ids]
            topics.append(top_words)
        return topics

    def get_document_topics(self):
        document_topics = []
        for doc_idx in range(self.num_docs):
            topic_probs = self.dist_topics[doc_idx, :]
            top_topic_idx = np.argmax(topic_probs)
            top_topic_prob = topic_probs[top_topic_idx]
            document_topics.append((top_topic_idx,
top_topic_prob))
        return document_topics

    def inference(self, corpus_tfidf):
        doc_topic_dists = []
        for doc in corpus_tfidf:
            doc_topic_dist = np.zeros(self.num_topics)
            for word_id, freq in doc:
                word_topic_dist = (
                    (self.topic_word_counts[:, word_id] +
self.beta) /
                    (self.topic_counts + self.vocab_size *
self.beta)
                )
                word_topic_dist /= np.sum(word_topic_dist)
                doc_topic_dist += word_topic_dist * freq  #
Mengalikan dengan frekuensi
            if np.sum(doc_topic_dist) == 0:
                doc_topic_dist = np.ones(self.num_topics) /
self.num_topics  # Handle kasus di mana semua distribusi topik
adalah nol
            else:
                doc_topic_dist /= np.sum(doc_topic_dist)  #
Normalisasi distribusi topik dokumen
            doc_topic_dists.append(doc_topic_dist)
        return np.array(doc_topic_dists), None  # Menambahkan
nilai kedua yang diharapkan oleh pyLDAvis

# Panggil fungsi untuk membuat model LDA
num_topics = 9
alpha = 0.1
beta = 0.01
```

```
num_iters = 100

random_state = 42  # Atur sesuai kebutuhan Anda
lda = LDA(num_topics, alpha, beta, num_iters,
random_state=random_state)
lda.fit(corpus_tfidf, dictionary)

# Menampilkan representasi topik dengan kata-kata dan bobotnya
print("\nRepresentasi Topik:")
topics = lda.get_topics(num_words=10)  # Mengambil 10 kata
teratas untuk setiap topik
for topic_idx, topic in enumerate(topics):
    topic_repr = " + ".join([f"{weight:.5f}*\"{word}\"" for
word, weight in zip(topic, lda.dist_words[topic_idx])])
    print(f"Topik {topic_idx}:")
    print(f" ({topic_repr})")

# Mendapatkan representasi dokumen dengan topik terbesar
document_topics = lda.get_document_topics()

# Menyiapkan data untuk ditampilkan
table_data = []
for topic_idx in range(lda.num_topics):
    topic_words = lda.get_topics()[topic_idx]
    top_words_str = ", ".join(topic_words)  # Perubahan di
sini
    topic_docs_count = sum(1 for doc_topic in document_topics
if doc_topic[0] == topic_idx)
    table_data.append([topic_idx, top_words_str,
topic_docs_count])

# Menampilkan tabel
print("\n", tabulate(table_data, headers=["Topic", "Kata
Penyusun (Top 10 kata)", "Jumlah Dokumen"], tablefmt="grid"))
```

```
def compute_coherence_values(dictionary, corpus, texts, limit,
start, step):
        coherence_values = []
        model_list = []

        for num_topics in range(start, limit, step):
            lda = LDA(num_topics, alpha, beta, num_iters,
random_state=random_state)
            lda.fit(corpus_tfidf, dictionary)
            model_list.append(lda)

            topics = lda.get_topics()
```

```
            coherence_model_lda =
CoherenceModel(topics=topics, texts=texts,
dictionary=dictionary, coherence='c_v')

coherence_values.append(coherence_model_lda.get_coherence())

        return model_list, coherence_values

start=1
limit=11
step=1
model_list, coherence_values =
compute_coherence_values(dictionary, corpus=corpus_tfidf,

texts=text_list, start=start, limit=limit, step=step)
#show graphs
import matplotlib.pyplot as plt
x = range(start, limit, step)
plt.plot(x, coherence_values)
plt.xlabel("Num Topics")
plt.ylabel("Coherence score")
plt.legend(("coherence_values"), loc='best')
plt.show()

# Print the coherence scores
for m, cv in zip(x, coherence_values):
    print("Num Topics =", m, " has Coherence Value of",
round(cv, 5))
```

Lampiran 6 Tautan *source code* lengkap *topic modeling*

Topic Modeling