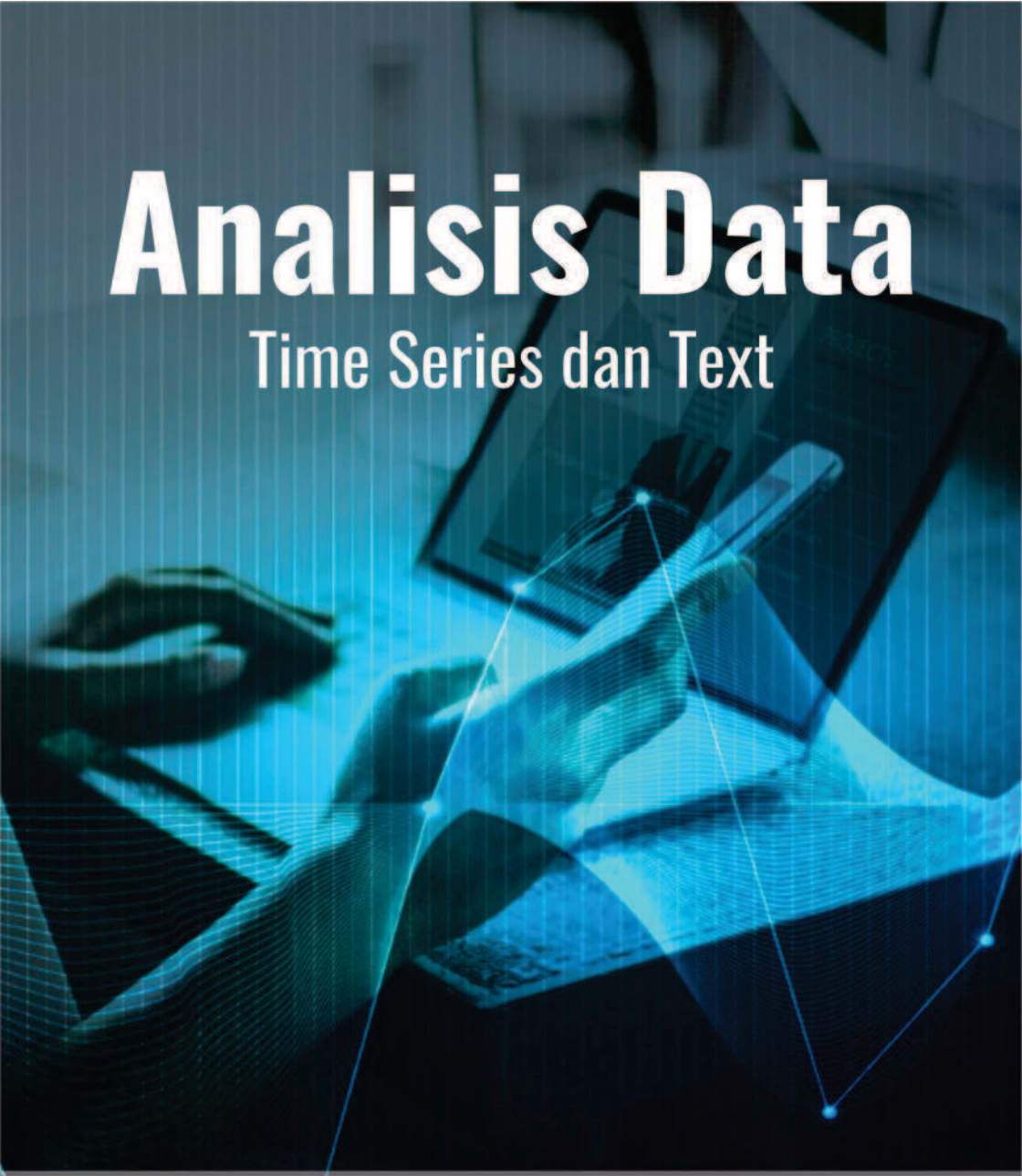
Analisis Data

Time Series dan Text

Dr. Ir. Wahyudi, S.T., M.T.

Analisis Data

Time Series dan Text



eureka
media aksara
Anggota IKAPI
No. 225/JTE/2021

0858 5343 1992
eurekamediaaksara@gmail.com
Jl. Banjaran RT.20 RW.10
Bojongsari - Purbalingga 53362

ISBN 978-623-120-663-3

ANALISIS DATA: TIME SERIES DAN TEXT

Dr. Ir. Wahyudi, S.T., M.T.

PENERBIT CV.EUREKA MEDIA AKSARA

ANALISIS DATA: TIME SERIES DAN TEXT

Penulis : Dr. Ir. Wahyudi, S.T., M.T.

Desain Sampul : Ardyan Arya Hayuwaskita

Tata Letak : Rizki Rose Mardiana

ISBN : 978-623-120-663-3

Diterbitkan oleh : **EUREKA MEDIA AKSARA, MEI 2024**
ANGGOTA IKAPI JAWA TENGAH
NO. 225/JTE/2021

Redaksi:
Jalan Banjaran, Desa Banjaran RT 20 RW 10 Kecamatan Bojongsari Kabupaten Purbalingga Telp. 0858-5343-1992
Surel : eurekamediaaksara@gmail.com
Cetakan Pertama : 2024

KATA PENGANTAR

Puji syukur selalu terucap kepada Allah SWT yang sampai saat ini telah memberikan nikmat sehat, sehingga penulis bisa menyelesaikan buku ini walaupun masih terdapat kendala yang masih dapat diselesaikan. Terima kasih juga penulis ucapkan kepada semua yang berkontribusi atas selesainya tulisan ini. Keterbatasan waktu menjadi salah satu hal yang menjadi kesulitan dalam penulisan buku ini. Namun berkat dukungan dari semua pihak, akhirnya tulisan ini dapat selesai tepat waktu. Penulis menyadari masih banyak kekurangan dalam tulisan ini. Oleh karena itu penulis memohon maaf atas kesalahan yang mungkin ada pada buku ini.

Penulis berharap buku yang berjudul "Analisis Data: Time Series dan Text" bisa bermanfaat bagi pembaca. Mohon untuk memaklumi jika terdapat penjelasan yang sulit untuk dimengerti. Untuk itu penulis mengharapkan kritik maupun saran, sehingga penulis bisa memperbaikinya dikemudian hari. Terimakasih atas ketertarikan Anda untuk membaca buku yang penulis buat.

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

ANALISIS DATA: TIME SERIES DAN TEXT

Dr. Ir. Wahyudi, S.T., M.T.

BAB 1 ANALISIS TIME SERIES

Konsep Utama ACF ARIMA Autoregresif Rata-rata bergerak PACF Stasioner Rangkaian waktu. Bab ini membahas topik analisis *Time Series / Deret Waktu* dan aplikasinya. Penekanan diberikan pada identifikasi struktur yang mendasari *Time Series* dan menentukan model *Autoregressive Integrated Moving Average* (ARIMA) yang sesuai.

A. Gambaran Umum Analisis Time Series

Analisis *Time Series* mencoba memodelkan struktur yang mendasari pengamatan diambil dari waktu ke waktu. Sebuah *Time Series*, dilambangkan $Y = a + bX$, adalah urutan nilai yang sama dari waktu ke waktu. Sebagai contoh, Gambar 1.1 menunjukkan plot jumlah penumpang bulanan maskapai penerbangan internasional selama periode 12 tahun.

BAB 2 ANALISIS TEKS

Analisis teks, kadang-kadang disebut analitik teks, mengacu pada representasi, pemrosesan, dan pemodelan data tekstual untuk mendapatkan wawasan yang berguna. Komponen penting dari analisis teks adalah penambangan teks, proses menemukan hubungan dan pola yang menarik dalam koleksi teks yang besar.

Analisis teks menderita kutukan dimensi tinggi. Ambil contoh buku anak-anak yang populer, Green Eggs dan Ham. Penulis Theodor Geisel (Dr. Seuss) ditantang untuk menulis seluruh buku dengan hanya 50 kata yang berbeda. Dia menjawabnya dengan buku Green Eggs and Ham, yang berisi 804 kata, hanya 50 di antaranya yang berbeda. Ke-50 kata tersebut adalah:

a, am, and, anywhere, are, be, boat, box, car, could, dark, do, eat, eggs, fox, goat, good, green, ham, here, house, I, if, in, let, like, ay, me, mouse, not, on, or, rain, Sam, say, see, so, thank, that, the,hem, there, they, train, tree, try, will, with, would, you.

Green Eggs and Ham adalah sebuah buku yang sederhana. Analisis teks sering kali berurusan dengan data tekstual yang jauh lebih kompleks. Sebuah *corpus* (jamak:

DAFTAR PUSTAKA

A. Agarwal, B. Xie, I. Vovsha, O. Rambow and R. Passonneau, "Sentiment Analysis of Twitter Data," In Proceedings of the Workshop on Languages in Social Media, pp. 30–38, 2011.

A. Agarwal, F. Biadsy, and K. R. Mckeown, "Contextual Phrase-Level Polarity Analysis Using Lexical Affect Scoring and Syntactic N-Grams," Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics, pp. 24–32, 2009.

A. Go, R. Bhayani, and L. Huang, "Twitter Sentiment Classification Using Distant Supervision," CS224N Project Report, Stanford, pp. 1–12, 2009.

A. McCallum, K. Nigam, J. Rennie, and K. Seymore, "A Machine Learning Approach to Building Domain-Specific Search Engines," IJCAI, vol. 99, 1999.

"Amazon Mechanical Turk" [Online]. Available: http://www.mturk.com/. [Accessed 7 April 2014].

A. Pak and P. Paroubek, "Twitter as a Corpus for Sentiment Analysis and Opinion Mining," Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10), pp. 19–21, 2010.

B. Liu, X. Li, W. S. Lee, and P. S. Yu, "Text Classification by Labeling Words," AAAI, vol. 4, pp. 425–430, 2004.

B. O'Connor, R. Balasubramanyan, B. R. Routledge, and N. A. Smith, "From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series," Proceedings of the Fourth

International Conference on Weblogs and Social Media, ICWSM '10, pp. 122–129, 2010.

B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis," Foundations and Trends in Information Retrieval, vol. 2, no. 1–2, pp. 1–135, 2008.

B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs Up? Sentiment Classification Using Machine Learning Techniques," Proceedings of EMNLP, pp. 79–86, 2002.

C. D. Manning, P. Raghavan, and H. Schütze, "Document and Query Weighting Schemes," in Introduction to Information Retrieval, Cambridge, United Kingdom, Cambridge University Press,2008, p. 128.

"Critical Assessment of Information Extraction in Biology (BioCreative)" [Online]. Available: http://www.biocreative.org/. [Accessed 2 April 2014].

"Curl and libcurl Tools" [Online]. Available: http://curl.haxx.se/. [Accessed 27 March 2014].

"DataSift: Power Decisions with Social Data," DataSift [Online]. Available: http://datasift.com/. [Accessed 12 June 2014].

D. M. Blei, "Probabilistic Topic Models," Communications of the ACM, vol. 55, no. 4, pp. 77–84,2012.

D. M. Blei, "Topic Modeling Software" [Online]. Available: http://www.cs.princeton.edu/~blei/ topicmodeling.html. [Accessed 11 June 2014].

D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet Allocation," Journal of Machine Learning Research, vol. 3, pp. 993–1022, 2003.

Dr. Seuss, "Green Eggs and Ham," New York, NY, USA, Random House, 1960.

"Gnip: The Source for Social Data," GNIP [Online]. Available: http://gnip.com/. [Accessed 12 June 2014].

G. K. Zipf, Human Behavior and the Principle of Least Effort, Reading, MA: Addison-Wesley, 1949.

G. Salton and C. Buckley, "Term-Weighting Approaches in Automatic Text Retrieval," in Information Processing and Management, 1988, pp. 513–523.

H. Saif, Y. He, and H. Alani, "Semantic Sentiment Analysis of Twitter," Proceedings of the 11th International Conference on The Semantic Web (ISWC'12), pp. 508–524, 2012.

J. Chang, "lda: Collapsed Gibbs Sampling Methods for Topic Models," CRAN, 14 October 2012. [Online]. Available: http://cran.r-project.org/web/packages/lda/. [Accessed 3April 2014].

J. J. Godfrey and E. Holliman, "Switchboard-1 Release 2," Linguistic Data Consortium, Philadelphia, 1997. [Online]. Available: http://catalog.ldc.upenn.edu/LDC97S62. [Accessed 2 April 2014].Bibliography ADVANCED ANALYTICAL THEORY AND METHODS: TEXT ANALYSIS

M. E. Newman, "Power Laws, Pareto Distributions, and Zipf's Law," Contemporary Physics, vol. 46, no. 5, pp. 323–351, 2005.

M. Hu and B. Liu, "Mining and Summarizing Customer Reviews," Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 168–177, 2004.

M. Porter, "Porter's English Stop Word List," 12 February 2007. [Online]. Available: http://snowball.tartarus.org/algorithms/english/stop.txt. [Accessed 2 April 2014].

M. Steinbach, G. Karypis, and V. Kumar, "A Comparison of Document Clustering Techniques," KDD workshop on text mining, vol. 400, no. 1, 2000.

N. Seco, T. Veale, and J. Hayes, "An Intrinsic Information Content Metric for Semantic Similarity in WordNet," ECAI, vol. 16, pp. 1089–1090, 2004.

P. D. Turney, "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews," Proceedings of the Association for Computational Linguistics, pp. 417–424, 2002.

P. Koehn, "Europarl: A Parallel Corpus for Statistical Machine Translation," MT Summit, 2005.

P. Resnik, "Using Information Content to Evaluate Semantic Similarity in a Taxonomy," In Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI'95), vol. 1,pp. 448–453, 1995.

P. Soucy and G. W. Mineau, "A Simple KNN Algorithm for Text Categorization," ICDM, pp. 647–648, 2001.

ProgrammableWeb, “API Directory” [Online]. Available: http://www.programmableweb.com/apis/directory. [Accessed 27 March 2014].

R. Rehurek, “Python Gensim Library” [Online]. Available: http://radimrehurek.com/gensim/. [Accessed 8 April 2014].

“The Penn Treebank Project,” University of Pennsylvania [Online]. Available: ttp://www.cis.upenn.edu/~treebank/home.html. [Accessed 26 March 2014].

T. Joachims, “Transductive Inference for Text Classification Using Support Vector Machines,” ICML, vol. 99, pp. 200–209, 1999.

T. Minka, “Estimating a Dirichlet Distribution,” 2000.

T. Pedersen, “Information Content Measures of Semantic Similarity Perform Better Without Sense-Tagged Text,” Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, pp. 329–332, June 2010.

Twitter, “Twitter Developers Site” [Online]. Available: https://dev.twitter.com/. [Accessed 27 March 2014].

W. N. Francis and H. Kucera, “Brown Corpus Manual,” 1979. [Online]. Available: http://icame. uib.no/brown/bcm.html.

Wikipedia, “List of Open APIs” [Online]. Available: http://en.wikipedia.org/wiki/List_of_open_APIs. [Accessed 27 March 2014].

"XML Path Language (XPath) 2.0," World Wide Web Consortium, 14 December 2010. [Online]. Available: http://www.w3.org/TR/xpath20/. [Accessed 27 March 2014].

Y. Li, D. McLean, Z. A. Bandar, J. D. O'Shea, and K. Crockett, "Sentence Similarity Based on Semantic Nets and Corpus Statistics," IEEE Transactions on Knowledge and Data Engineering, vol. 18, no. 8, pp. 1138–1150, 2006.