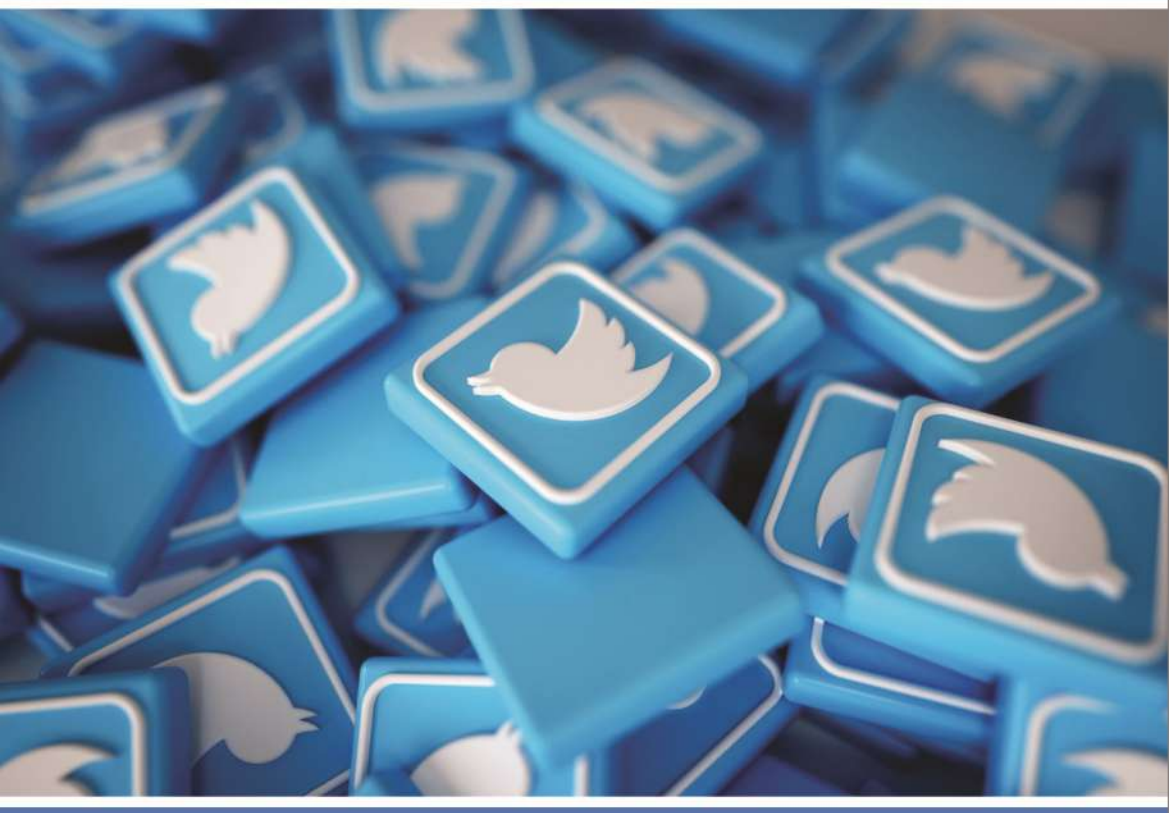
MENGUBAH CUITAN TWITTER MENJADI DATA BERHARGA

MENGGUNAKAN METODE *WORD2VEC* DAN *CONVOLUTIONAL NEURAL NETWORK*



Tim Penulis:

Prof. Dr.rer.nat. Wahyu Hardyanto, M.Si., Aji Purwinarko, S.Si., M.Cs., Nila Prasetya Aryani, S.Si., M.Si., Sugiyanto, S.Pd., M.Si., Muhamad Anbiya Nur Isiam, Nala Adina, Defin Andestian, Wahyu Setyaningrum, M Fadil Mardiansyah.

MENGUBAH CUITAN TWITTER MENJADI DATA BERHARGA

MENGGUNAKAN METODE *WORD2VEC* DAN *CONVOLUTIONAL NEURAL NETWORK*

Tim Penulis:
Prof. Dr.rer.nat. Wahyu Hardyanto, M.Si., Aji Purwinarko, S.Si., M.Cs., Nila Prasetya Aryani, S.Si., M.Si., Sugiyanto, S.Pd., M.Si., Muhamad Anbiya Nur Islam, Nala Adina, Defin Andestian, Wahyu Setyaningrum, M Fadil Mardiansyah.

Penerbit Widina

MENGUBAH CUITAN TWITTER MENJADI DATA BERHARGA MENGGUNAKAN METODE WORD2VEC DAN *CONVOLUTIONAL NEURAL NETWORK*

Penulis:
Wahyu Hardyanto, Aji Purwinarko, Nila Prasetya Aryani, Sugiyanto, Muhamad Anbiya Nur Islam, Nala Adina, Defin Andestian, Wahyu Setyaningrum, M Fadil Mardiansyah.

Desain Cover:
Septian Maulana

Sumber Ilustrasi:
www.freepik.com

Tata Letak:
Handarini Rohana

Editor:
Aji Purwinarko, S.Si., M.Cs.

ISBN:
978-623-500-205-7

Cetakan Pertama:
Juni, 2024

Hak Cipta Dilindungi Oleh Undang-Undang
by Penerbit Widina Media Utama

Dilarang keras menerjemahkan, memfotokopi, atau memperbanyak sebagian atau seluruh isi buku ini tanpa izin tertulis dari Penerbit.

PENERBIT:
WIDINA MEDIA UTAMA
Komplek Puri Melia Asri Blok C3 No. 17 Desa Bojong Emas
Kec. Solokan Jeruk Kabupaten Bandung, Provinsi Jawa Barat

Anggota IKAPI No. 360/JBA/2020
Website: www.penerbitwidina.com
Instagram: @penerbitwidina
Telepon (022) 87355370

PRAKATA

Puji dan syukur kami panjatkan kepada Tuhan Yang Maha Kuasa atas berkat dan rahmat-Nya sehingga buku dengan judul "Mengubah Cuitan Twitter Menjadi Data Berharga, Menggunakan Metode Word2vec dan *Convolutional Neural Network*," dapat terbit.

Opini merupakan ekspresi yang subyektif yang dapat menggambarkan sentimen seseorang, pendapat, atau perasaan tentang suatu kejadian dan sifat. Twitter menjadi salah satu media yang paling sering digunakan masyarakat untuk beropini. Opini publik atau masyarakat dapat dimanfaatkan dalam proses evaluasi guna meningkatkan mutu pendidikan. Akan tetapi bagaimana merangkum data mentah berupa opini publik di media sosial ini, lalu memanfaatkannya menjadi data berharga? Untuk dapat melakukannya kita perlu bantuan *software* tertentu. Oleh karena itu, maka buku ini pun hadir. Buku ini mendokumentasikan proses bagaimana data mentah tersebut diperoleh, kemudian mengolahnya menjadi data "siap saji".

Ucapan terimakasih yang sebesar-besarnya kami haturkan kepada semua pihak yang sudah membantu hingga buku ini dapat terbit. Semoga Tuhan Yang Maha Kuasa memberikan balasan yang berlipatganda. Semoga dengan hadirnya buku ini dapat memberikan sumbangan bagi perkembangan ilmu pengetahuan serta diharapkan dapat memberikan wawasan berharga bagi siapa saja yang berminat dan tertarik dengan bagaimana algoritma dunia maya bekerja.

DAFTAR ISI

1

MAKNA PENTING SEBUAH DATA

A. DEFINISI DATA

Data adalah istilah yang merujuk pada fakta-fakta yang disusun dalam bentuk yang dapat diukur, dihitung, atau diolah secara komputerisasi. Definisi data telah mengalami evolusi seiring dengan perkembangan teknologi informasi dan komunikasi. Pada dasarnya, data merupakan kumpulan informasi yang direkam, disimpan, dan diproses untuk tujuan tertentu. Data dapat berupa angka, teks, gambar, suara, atau bentuk lainnya yang dapat direpresentasikan dalam format digital atau analog.

Dalam konteks teknologi informasi, data sering dianggap sebagai bahan mentah yang menjadi dasar bagi proses analisis dan pengambilan keputusan. Data yang dikelola dengan baik memiliki beberapa karakteristik penting. Pertama, data harus akurat, artinya data tersebut harus tepat dan tidak mengandung kesalahan yang dapat memengaruhi keputusan yang diambil. Kedua, data harus lengkap, yaitu mencakup semua informasi yang diperlukan untuk tujuan tertentu. Ketiga, data harus relevan, yang berarti data tersebut memiliki kaitan atau relevansi dengan konteks atau permasalahan yang

2

WAWASAN SEPUTAR *TEXT MINING* TWITTER

A. TWITTER

Twitter adalah sebuah situs jejaring sosial yang sedang berkembang pesat saat ini karena pengguna dapat berinteraksi dengan pengguna lainnya dari komputer ataupun perangkat *mobile* mereka dari manapun dan kapanpun. Beragam macam mengenai berita dunia, gosip hiburan tentang selebriti, dan diskusi tentang produk yang baru saja dirilis semuanya dikumpulkan di Twitter dengan jelas. Selain hanya menampilkan berita dan laporan, Twitter sendiri juga merupakan platform besar tempat berbagai pendapat disajikan dan dipertukarkan (Liao *et al.,* 2017).

Dapat dilihat pada Gambar 2.1, tahun 2014 sampai dengan 2019 akun pengguna Twitter di Indonesia mengalami kenaikan yang signifikan, yaitu diperkirakan sekitar 12 juta pengguna di Indonesia (Statista, 2019a).

3

RANCANG BANGUN SISTEM

A. *PREPROCESSING*

Pada tahap awal yang dilakukan yaitu *preprocessing*. *Preprocessing* dilakukan untuk membersihkan *dataset* dari *noise* data, kemudian tahap selanjutnya yaitu *encode* data. Tahap *encoding* data dilakukan untuk menyederhanakan *dataset* dengan mengkonversi teks ke bentuk angka. Hasil dari *encoding* data berupa angka yang mewakili setiap kata dari kalimat dari *dataset.* Kemudian proses berlanjut ke tahap pembagian data, dimana *dataset* akan dibagi menjadi dua yaitu data *training* dan data *testing.* Dan tahap penting selanjutnya yaitu *word embedding* yang merupakan bagian dari *feature extraction*. Tahap ini memudahkan pada proses pelatihan *machine learning* yang membutuhkan jumlah *input* yang besar. Model *word embedding* sebelumnya akan digunakan untuk proses perhitungan dengan metode CNN untuk proses klasifikasi.

Tahapan berikutnya merupakan evaluasi model yang digunakan untuk mempresentasikan hasil dari *modelling* yaitu *confusion matrix* untuk menghitung nilai akurasi dari hasil yang didapatkan. Tahap akhir yang dilakukan dalam riset yaitu proses pembangunan *system* atau *system modelling* yang dimaksudkan untuk memvisualisasikan riset

4

MENAMBANG DATA

A. MENAMBANG DATA DARI TWITTER

Data di dapatkan dengan *crawling* Twitter. Twitter menyediakan *Free Open API* untuk memperbolehkan kita menambang data pada *website*nya. Sebelum melakukan implementasi *crawling* pada twitter kita memerlukan *API Key* yang teregistrasi untuk dapat berinteraksi dengan twitter.

Untuk dapat berkomunikasi dengan API twitter kita akan menggunakan *library python* yaitu *Tweepy*. Selain *Tweepy* masih ada beberapa *library python* lainnya seperti (*Twython, TwitterSeacrh*, dan lain-lain). Pada tulisan ini kita akan menggunakan API *Search* dan API *Stream* untuk proses *crawling* twitter.

Langkah awal instal *library tweepy* dengan menggunakan perintah ***pip install tweepy***. Hasil dari proses instalasi ditunjukkan oleh Gambar 4.1.

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Requirement already satisfied: tweepy in /usr/local/lib/python3.7/dist-packages (3.10.0)
Requirement already satisfied: requests-oauthlib>=0.7.0 in /usr/local/lib/python3.7/dist-packages (from tweepy) (1.3.1)
Requirement already satisfied: six>=1.10.0 in /usr/local/lib/python3.7/dist-packages (from tweepy) (1.15.0)
Requirement already satisfied: requests[socks]>=2.11.1 in /usr/local/lib/python3.7/dist-packages (from tweepy) (2.23.0)
Requirement already satisfied: oauthlib>=3.0.0 in /usr/local/lib/python3.7/dist-packages (from requests-oauthlib>=0.7.0->tweepy) (3.2.2)
Requirement already satisfied: urllib3!=1.25.0,!=1.25.1,<1.26,>=1.21.1 in /usr/local/lib/python3.7/dist-packages (from requests[socks]>=2.11.1->tweep
Requirement already satisfied: chardet<4,>=3.0.2 in /usr/local/lib/python3.7/dist-packages (from requests[socks]>=2.11.1->tweepy) (3.0.4)
Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.7/dist-packages (from requests[socks]>=2.11.1->tweepy) (2.10)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.7/dist-packages (from requests[socks]>=2.11.1->tweepy) (2022.9.24)
Requirement already satisfied: PySocks!=1.5.7,>=1.5.6 in /usr/local/lib/python3.7/dist-packages (from requests[socks]>=2.11.1->tweepy) (1.7.1)
```

Gambar 4.1 Hasil *install twepy*

DAFTAR PUSTAKA

Cai, G., & Xia, B. (2015). Convolutional neural networks for multimedia sentiment analysis. In *Natural language processing and chinese computing* (pp. 159–167). Springer.

Church, K. W. (2017). Emerging trends: Word2Vec. *Natural Language Engineering*, *23*(1), 155–162. https://doi.org/10.1017/S1351324916000334

Dey, L., Chakraborty, S., Biswas, A., Bose, B., & Tiwari, S. (2016). Sentiment analysis of review datasets using naive bayes and k-nn classifier. *ArXiv Preprint ArXiv:1610.09982*.

Ferdiana, R., Jatmiko, F., Purwanti, D. D., Ayu, A. S. T., & Dicka, W. F. (2019). Dataset Indonesia untuk analisis sentimen. *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi (JNTETI)*, *8*(4), 334–339.

Gupta, R. R., & Ranga, V. (2021). Comparative study of different reduced precision techniques in deep neural network. *In Proceedings of International Conference on Big Data, Machine Learning and Their Applications*, 123–136.

Hakim, I. R. (2018). Metode penulisan ilmiah. In *Surakarta: CV. Dwija Amarta Press* (Vol. 1, Issue 1).

Hamida, U. (2014). Penggunaan artificial neural network (ANN) untuk memodelkan kebutuhan energi untuk transportasi. *J. Teknol. Dan Manaj*, *12*(2), 57–65.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, *160*(1), 106–154.

Jianqiang, Z., Xiaolin, G., & Xuejun, Z. (2018). Deep convolution neural networks for twitter sentiment analysis. *IEEE Access*, *6*, 23253–23260.

Kotu, V., & Deshpande, B. (2014). *Predictive analytics and data mining: concepts and practice with rapidminer*. Morgan Kaufmann.

Kotu, V., & Deshpande, B. (2015). Data Exploration. In *Predictive Analytics and Data Mining*. https://doi.org/10.1016/b978-0-12-801460-8.00003-3

Liao, S., Wang, J., Yu, R., Sato, K., & Cheng, Z. (2017). CNN for Situations Understanding Based on Sentiment Analysis of Twitter Data. *Procedia Computer Science*, *111*(2015), 376–381. https://doi.org/10.1016/j.procs.2017.06.037

Lin, Y. (2021). *10 Twitter Statistics Every Marketer Should Know in 2019*. https://www.oberlo.com/blog/Twitter-statistics

Lutfi, A. A., Permanasari, A. E., & Fauziati, S. (2018). Sentiment analysis in the sales review of Indonesian marketplace by utilizing Support Vector Machine. *Journal of Information Systems Engineering and Business Intelligence*, *4*(1), 57–64.

Ma, L., & Zhang, Y. (2015a). Using word2Vec to process big text data. *Proceedings - 2015 IEEE International Conference on Big Data, IEEE Big Data 2015*, 2895–2897. https://doi.org/10.1109/BigData.2015.7364114

Ma, L., & Zhang, Y. (2015b). Using word2Vec to process big text data. *Proceedings - 2015 IEEE International Conference on Big Data, IEEE Big Data 2015*, 2895–2897. https://doi.org/10.1109/BigData.2015.7364114

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013a). Efficient estimation of word representations in vector space. *1st International Conference on Learning Representations, ICLR 2013 - Workshop Track Proceedings*, 1–12.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013b). Efficient estimation of word representations in vector space. *1st International Conference on Learning Representations, ICLR 2013 - Workshop Track Proceedings*, 1–12.

Nawangsari, R. P., Kusumaningrum, R., & Wibowo, A. (2019). Word2vec for indonesian sentiment analysis towards hotel reviews: An evaluation study. *Procedia Computer Science*, *157*, 360–366. https://doi.org/10.1016/j.procs.2019.08.178

Nurzahputra, A., & Muslim, M. A. (2016). Analisis Sentimen pada Opini Mahasiswa Menggunakan Natural Language Processing. *Seminar Nasional Ilmu Komputer (SNIK 2016)*, 114–118.

Pournaki, A., Gaisbauer, F., Banisch, S., & Olbrich, E. (2020). The twitter explorer: A framework for observing twitter through interactive networks. In *arXiv preprint arXiv*.

Putri, W. S. R., Nurwati, N., & Budiarti, M. (2016). Pengaruh media sosial terhadap perilaku remaja. *Prosiding Penelitian Dan Pengabdian Kepada Masyarakat*, *3*(1).

Riyanto, G. P. (2021). *Jumlah Pengguna Internet Indonesia 2021 Tembus 202 Juta*. https://tekno.kompas.com/read/2021/02/23/16100057/jumlah-pengguna-internet-indonesia-2021-tembus-202-juta

Statista. (2019a). Number of monthly active twitter users worldwide from 1st quarter 2010 to 1st quarter 2019. In *Statista*. Statista.

Statista. (2019b). Number of monthly active twitter users worldwide from 1st quarter 2010 to 1st quarter 2019. In *Statista*. Statista.

Tripathi, D. K., Chadha, S., & Tripathi, A. (2023). Metaheuristic enabled intelligent model for stock market prediction via integrating volatility spillover: India and its Asian and European counterparts. *Data & Knowledge Engineering*, *144*, 102127. https://doi.org/10.1016/J.DATAK.2022.102127

Wang, J.-H., Liu, T.-W., Luo, X., & Wang, L. (2018). An LSTM approach to short text sentiment classification with word embeddings. *Proceedings of the 30th Conference on Computational Linguistics and Speech Processing (ROCLING 2018)*, 214–223.

Zhang, Y., & Wallace, B. (2015). *A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification*.

Zhang, Z., Ye, Q., Zhang, Z., & Li, Y. (2011). Sentiment classification of Internet restaurant reviews written in Cantonese. *Expert Systems with Applications*, *38*(6), 7674–7682. https://doi.org/10.1016/j.eswa.2010.12.147

MENGUBAH CUITAN TWITTER MENJADI DATA BERHARGA

MENGGUNAKAN METODE *WORD2VEC* DAN *CONVOLUTIONAL NEURAL NETWORK*

Tercatat pengguna aktif internet Indonesia di tahun 2021 telah menembus 202 juta, ini merupakan jumlah yang sangat besar. Internet telah menyediakan berbagai layanan *online* salah satunya adalah Twitter, yaitu microblog menjadi salah satu media yang paling sering digunakan masyarakat untuk beropini. Bagi Universitas Negeri Semarang (UNNES) sendiri, opini publik atau masyarakat dapat dimanfaatkan dalam proses evaluasi guna meningkatkan mutu pendidikan.

Opini masyarakat dapat dikelompokkan atau diklasifikasikan dengan menggunakan teknik analisis sentimen. Terdapat banyak metode yang telah dilakukan oleh peneliti sebelumnya terkait analisis *sentiment*, diantaranya adalah *machine learning* dan *deep learning*. Buku ini menjelaskan dengan terperinci tentang rancang bangun sebuah sistem aplikasi di mana data mentah yang diperoleh dari internet diolah dengan melalui beberapa tahap pemrosesan sehingga didapatkanlah data keluaran yang diinginkan.

Penerbit widina
www.penerbitwidina.com

ISBN 978-623-500-205-7