

BAB 6

KESIMPULAN DAN SARAN

6.1 Kesimpulan

Berdasarkan hasil penelitian yang sudah dilakukan pada Tugas Akhir ini, berikut ini adalah beberapa kesimpulan yang dapat diambil:

1. Tahapan Data Science berhasil dilakukan dengan beberapa iterasi untuk mendapatkan iterasi final yang paling baik.
2. Data teks berita yang berkaitan dengan pemilu khususnya kandidat presiden, kandidat wakil presiden, dan partainya sudah berhasil dikumpulkan menggunakan *web scraping* dari laman portal berita Detik dalam rentang waktu antara Januari 2023 hingga November 2023, menggunakan 9 kata kunci yang bersangkutan dengan nama kandidat presiden, kandidat wakil presiden, dan partainya.
3. Melakukan eksplorasi dan penyiapan data berita yang sudah dikumpulkan secara berulang agar mendapatkan hasil yang baik.
4. Mendeteksi sentimen teks berita pemilu yang dikumpulkan dari hasil perbandingan model *machine learning* yang dimiliki oleh *library* MLlib milik Spark dan diambil hasil yang terbaik dari 4 tipe algoritma *machine learning* untuk klasifikasi yaitu *Naive Bayes Classifier*, *Logistic Regression*, *Decision Tree Classifier*, dan *Random Forest Classifier*.
5. Model *machine learning* analisis sentimen yang terbaik dimiliki oleh *Logistic Regression* dengan akurasi 77% pada fitur teks berita, dan akurasi 74% pada fitur judul berita.
6. Hasil analisis popularitas dan sentimen pada data Berita menggunakan model yang dibuat menunjukkan popularitas Kandidat Presiden Nomor Urut 2 Prabowo menjadi teratas, diikuti dengan Kandidat Presiden Nomor Urut 3 Ganjar, dan yang ketiga yaitu Kandidat Presiden Nomor Urut 1 Anies Baswedan.
7. Hasil analisis popularitas dan sentimen pada data Berita menggunakan model yang dibuat menunjukkan popularitas Kandidat Wakil Presiden Nomor Urut 2 Gibran menjadi teratas, diikuti dengan Kandidat Wakil Presiden Nomor Urut 1 Muhaimin, dan yang ketiga yaitu Kandidat Wakil Presiden Nomor Urut 3 Mahfud MD.
8. Hasil analisis popularitas dan sentimen pada data Berita menggunakan model yang dibuat menunjukkan popularitas Partai Pemilu Gerindra menjadi teratas, diikuti dengan Partai Pemilu PDIP, dan yang ketiga yaitu Partai Pemilu Nasdem.
9. Perangkat lunak *Dashboard* telah berhasil dibangun dengan tampilan yang mudah dimengerti oleh pengguna, untuk menampilkan *insights* yang didapat.
10. Penyiapan Data dan Analisis tidak dilakukan secara seluruhnya dengan sistem terdistribusi Spark versi *cluster* dikarenakan terdapat masalah pada *worker* Spark khususnya penggunaan *library* PySpark yang tidak *built-in*.
11. Melakukan beberapa pendekatan metode analisis sentimen yang disesuaikan dengan teks berita pemilu, dan pendekatan yang diambil berupa penggabungan hasil popularitas sentimen dari fitur judul berita dan teks berita.

6.2 Saran

Berdasarkan hasil penelitian dan kesimpulan pada Bagian 6.1, berikut ini adalah beberapa saran yang dapat dilakukan untuk pengembangan selanjutnya:

1. Melakukan tahap *pre-processing* tambahan seperti lematisasi, *Named Entity Recognition* (NER), dan *Part of Speech tagging* (POS).
2. Membuat *cluster* Spark dan Hadoop yang terdistribusi sendiri dengan konfigurasi yang benar untuk proses ETL (Extract Transform Load) data yang lebih terstruktur dan praktis.
3. Melakukan pembuatan model klasifikasi sentimen analisis *machine learning* dengan penggunaan vektor *unigram* dan *bigram* agar mendapatkan model yang lebih baik.
4. Mencoba *tools* Big Data lain untuk mempermudah proses *Extract Transform Load* yang berguna seperti Scoop dan Flume.

DAFTAR REFERENSI

- [1] Pamungkas, S. (2009) *PERIHAL PEMILU*, 1st edition. Laboratorium Jurusan Ilmu Pemerintahan dan Jurusan Ilmu Pemerintahan Fakultas Ilmu Sosial dan Politik Universitas Gadjah Mada, Jl. Socio Justicia 2 Bulaksumur Yogyakarta 55.
- [2] White, T. (2015) *Hadoop: The Definitive Guide*, 4th edition. O'Reilly Media, 1005 Gravenstein Highway North, Sebastopol, CA 95472.
- [3] Chambers, B. dan Zaharia, M. (2018) *Spark The Definitive Guide : Big Data Processing Made Simple*, 1st edition. O'Reilly Media, 1005 Gravenstein Highway North, Sebastopol, CA 95472.
- [4] Version 3.5.0 (2023) *Apache Spark API Reference Manual*. Apache Software Foundation. Massachusetts, USA.
- [5] Narkhede, N., Shapira, G., dan Palino, T. (2017) *Kafka: The Definitive Guide*. O'Reilly Media, Inc., USA.
- [6] Jurafsky, D. dan Martin, J. H. (2020) *Speech and Language Processing An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, 3rd edition. Prentice Hall, Upper Saddle River, NJ, USA.
- [7] Zhai, C. dan Massung, S. (2016) *Text Data Management and Analysis: A Practical Introduction to Information Retrieval and Text Mining*, 1st edition. Association for Computing Machinery and Morgan & Claypool, 822 College Ave 457 Kentfield, CA 94914.
- [8] ISO/IEC 9075-1:2023 (2023) *Information technology Database languages SQL*. ISO/IEC JTC 1/SC 32. Geneva, Switzerland.
- [9] Han, J., Kamber, M., dan Pei, J. (2012) *Data Mining Concepts and Techniques*, 3rd edition. Morgan Kaufmann, Burlington.
- [10] Mitchell, R. (2018) *Web Scraping with Python*, 2nd edition. O'Reilly Media, USA.
- [11] v2.31.0. (2022) *Requests: HTTP for Humans*. Python Software Foundation. 512 Lafayette Boulevard, Suite 2, Fredericksburg, Virginia 22401.
- [12] 4.12.1 (2022) *Beautiful Soup Documentation*. Python Software Foundation. 512 Lafayette Boulevard, Suite 2, Fredericksburg, Virginia 22401.