

BAB 6

KESIMPULAN DAN SARAN

Bagian ini berisi kesimpulan dari hasil penelitian dan saran untuk pengembangan perangkat lunak di masa depan.

6.1 Kesimpulan

Kesimpulan yang didapat dari pengujian ini adalah sebagai berikut:

1. Telah berhasil mengumpulkan data-data yang ada di koleksi Perpustakaan UNPAR menggunakan *webscraping* di laman Repositori UNPAR.
2. Karya tulis ilmiah yang ada di Repositori UNPAR telah diolah sebelumnya menggunakan pra-pemrosesan data dari NLP dan pemilihan fitur.
3. Pembuatan sistem otomatis untuk pelabelan jurusan dan topik di setiap karya tulis ilmiah yang ada di Perpustakaan UNPAR berhasil dilakukan.
4. Model klasifikasi jurusan dan model LDA untuk topik untuk karya tulis ilmiah di Perpustakaan UNPAR telah terbentuk dan siap digunakan, model klasifikasi digunakan untuk menebak jurusan sementara model LDA digunakan untuk penentuan topik dan kata yang terkait.
5. Evaluasi dari model klasifikasi dan model LDA untuk melabeli dokumen dan hasilnya terlihat cukup baik pada beberapa kondisi, meskipun dibutuhkan ukuran dokumen yang lebih besar dan waktu pemrosesan yang lebih lama untuk pengujian yang lebih akurat.
6. Perangkat lunak berbasis *desktop* untuk peluncuran model klasifikasi dan otomatisasi label telah dibuat menggunakan Python.
7. Metode pemilihan fitur dengan membuang kata yang paling sering muncul dari seluruh *dataset*, kemudian memecah data berdasarkan jurusan, dan akhirnya mengambil kata-kata yang paling sering muncul dari jurusan tersebut dipilih karena dianggap paling efektif. Alasan mengapa kata-kata yang sering muncul di tiap jurusan menjadi penting dalam pengolahan data adalah karena mereka menunjukkan karakteristik khusus dari jurusan tersebut.
8. Hasil *lemmatization* dapat membuat model klasifikasi menjadi kabur karena banyak jurusan yang menggunakan kata yang sama tetapi berbeda artinya. Pada Hubungan Internasional, kata "usaha" digunakan sebagai persamaan kata dari upaya, sementara kata "usaha" dalam ekonomi dianggap sebagai perusahaan. Hal itulah yang menyebabkan kesalahan-kesalahan dalam penebakan data.
9. Algoritma klasifikasi yang paling cocok untuk kasus penambahan teks adalah *naive bayes* karena memiliki akurasi yang sedikit lebih tinggi dan kompleksitas lebih rendah dibandingkan regresi logistik.

6.2 Saran

Berikut adalah saran untuk penelitian berikutnya:

1. Untuk penelitian lebih lanjut dapat digunakan pemodelan topik yang lain sebagai perbandingan
2. Menentukan topik dan kata kunci dengan lebih akurat dengan memilih fitur yang lebih baik

3. Untuk penelitian lebih lanjut sebaiknya menggunakan teknik-teknik yang lebih baik dari stemming agar bisa melihat konteks dengan akurat.
4. Untuk penelitian lebih lanjut bisa menambahkan data dengan mengambil kata kunci yang tersedia di repository unpar.

DAFTAR REFERENSI

- [1] Yao, M. (2021) 4 different approaches to natural language processing understanding. <https://www.topbots.com/4-different-approaches-natural-language-processing-understanding/#genesis-content>. Accessed: 2023-09-02.
- [2] Glints (2021) Natural language processing adalah: Pengertian, fungsi, dan contoh. <https://tex.stackexchange.com/questions/615190/misc-bibliography>. Accessed: 2023-09-02.
- [3] Jurafsky, D. dan Martin, J. H. (2009) *Speech and language processing : an introduction to natural language processing, computational linguistics, and speech recognition*. Pearson Prentice Hall, Upper Saddle River, N.J.
- [4] Bird, B., Klein, E., dan Loper, E. (2018) Applied text analysis with python: Enabling language-aware data products with machine learning. <https://www.oreilly.com/library/view/applied-textanalysis/9781491963036/ch04.html>. Accessed: 2023-09-10.
- [5] Brilliant Classification. <https://brilliant.org/wiki/classification>. Accessed: 2023-09-10.
- [6] of Computer Science Binus University, S. (2020) Confusion matrix. <https://socs.binus.ac.id/2020/11/01/confusion-matrix>. Accessed: 2023-09-10.
- [7] Blei, D. M., Ng, A. Y., dan Jordan, M. I. (2003) Latent dirichlet allocation. *J. Mach. Learn. Res.*, **3**, 993–1022.
- [8] Ng, A. Y. dan Jordan, M. I. (2001) On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes. *Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic*, Cambridge, MA, USA NIPS'01 841–848. MIT Press.
- [9] Morris, B. (2021) Compare lda topic modeling in sklearn and gensim. <https://www.kaggle.com/code/morrisb/compare-lda-topic-modeling-in-sklearn-and-gensim/notebook>. Accessed: 2022-04-14.
- [10] Siswadi, I. (2016) Mengenal konsep penetapan kata kunci. *Jurnal Pustakawan Indonesia*, **12**.
- [11] Setyowati, E. dan Proboyekti, U. (2021) Kesimpulan, saran dan abstrak. https://www.academia.edu/5102068/KESIMPULAN_SARAN_Dan_ABSTRAK. Accessed: 2023-02-17.