

BAB 6

KESIMPULAN DAN SARAN

6.1 Kesimpulan

Berdasarkan pada penelitian yang telah dilakukan, didapatkan beberapa kesimpulan yaitu :

1. Penarikan data pada *website Rotten Tomatoes* dilakukan dengan metode *webscraping* menggunakan *library Selenium*. Selain selenium dilakukan juga percobaan dengan menggunakan *library BeautifulSoup*, namun saat menggunakan *library BeautifulSoup* gagal untuk menarik lebih dari satu halaman. Data yang telah terkumpul sebanyak 12.350, dengan 3 atribut yaitu nama pengulas, *review* film, dan *rating* atau sentimen. Dengan sentimen positif memiliki 7.756 dan untuk sentimen negatif memiliki 4.594 untuk proporsinya sentimen positif memiliki 62.8% sedangkan untuk sentimen negatif memiliki 37.2%.
2. Ekstraksi fitur dilakukan dengan cara menggunakan nilai TF-IDF. Kata kunci berdasarkan nilai IDF yang sering muncul pada sentimen positif yaitu “*best*”, “*fun*”, dan “*action*”, sedangkan Untuk kata kunci berdasarkan nilai IDF yang sering muncul pada sentimen negatif yaitu “*characters*”, “*man*”, dan “*action*”. Jumlah kemunculan kata pada seluruh dataset kata kunci “*full review*” dan “*review spanish*” memiliki jumlah yang cukup besar yaitu 798 dan kata “*one best*” sebanyak 339. Dilakukan juga menggunakan *postagging* untuk membuang kata benda, *postagging* dilakukan dengan unigram. Nilai TF-IDF pun digunakan untuk pembuatan model.
3. Pembuatan *wordcloud* dilakukan menggunakan dataset *preprocessing* positif unigram terdapat kata kunci “*good*”, “*fun*”, “*best*”, “*performance*”, dan “*entertaining*”. Dataset *preprocessing* Negatif Unigram terdapat kata kunci “*no*”, “*not*”, “*bad*”, “*never*”, dan “*like*”. Kata kunci pada unigram mungkin terdapat yang ambigu seperti pada sentimen positif terdapat “*not*” dan sentimen negatif terdapat “*like*” hal tersebut dikarenakan hanya mengambil unigram bukan bigram. *Wordcloud* postag positif unigram memiliki kata kunci yang cukup baik, namun masih memiliki kata-kata yang dianggap sebagai kurang relevan atau tidak memiliki arti positif maupun negatif. *Wordcloud* postag negatif memiliki kata kunci “*like*” di mana saat dianalisa lebih lanjut kata “*like*” tersebut merujuk kepada kata preposisi bukan kepada kata kerja.
4. Pembuatan model klasifikasi menggunakan tiga model yaitu *decision tree*, *logistic regression*, dan *Random Forest*. Pembuatan model dilakukan dengan menggunakan *document term matrix* nilai tf-idf *preprocessing* unigram serta *postag* unigram, dengan menggunakan data latih 80% dan data uji 20%. Pembuatan dan pengujian model pada data *preprocessing* bigram dan *postag* bigram tidak dapat dilakukan karena dataset yang terlalu besar sehingga terhalang oleh memori RAM peneliti yang tidak memadai.
5. Berdasarkan pengujian menggunakan *k-fold cross validation* model *Logistic Regression* memiliki rata-rata nilai akurasi yang menunjukkan bahwa model dapat mengklasifikasi data dengan benar sebanyak 69% dari 100% data, dan nilai rata-rata presisi yang menunjukkan bahwa sebanyak 58% prediksi kelas positif yang benar-benar positif, dan untuk rata-rata nilai *recall* menunjukkan bahwa model dapat mengindikasikan sebesar 58% data dari total kelas positif yang sebenarnya, sehingga untuk nilai rata-rata *F1-Score* yang menunjukkan bahwa model memiliki keseimbangan yang baik antara recall dan presisi yaitu sebesar 58%. Nilai rata-rata spesifisitas yang cukup tinggi yaitu sebesar 75% menunjukkan bahwa model cukup

efektif dalam mengidentifikasi kasus kelas negatif dengan benar. Dengan menggunakan data latih sebanyak 80% dan data uji sebanyak 20%. Model *logistic regression* dengan data *preprocessing* memberikan hasil yang cukup konsisten dari pengujian menggunakan metrik *accuracy*, *precision*, *recall*, *specifity*, dan *F1-Score*, serta dalam memberikan performa yang paling baik dari semua percobaan metode yang dilakukan. Untuk hasil yang kurang baik memungkinkan dari beberapa nilai *fold* yang lebih rendah dibandingkan nilai *fold* yang lainnya sehingga saat menghitung nilai rata-rata akan menjadi rendah. Dalam model ini pun menunjukkan nilai presisi dan recall hanya 58% yang menunjukkan bahwa model sering menghasilkan prediksi yang tidak tepat untuk kelas positif, sehingga dapat menyebabkan *false positive* dan *false negative*, yang dapat mengindikasikan bahwa model mungkin kurang efektif dalam mengenali pola yang relevan pada data.

6. Hasil analisis ditampilkan pada perangkat lunak berbasis *web* di mana halaman *dashboard* dapat melihat penjelasan dataset, percobaan *wordcloud* dengan menggunakan dataset unigram positif dan negatif *preprocessing* dan unigram positif dan negatif *postag*, selain itu pada halaman *dashboard* juga ditampilkan tabel pengujian model dengan menggunakan data uji. Selanjutnya, pada halaman input *review* pengguna dapat menginputkan *review* film dan perangkat lunak akan mengeluarkan hasil berupa sentimen *fresh* yaitu positif dan *rotten* yaitu negatif. Namun, hasil dari pengujian model masih terdapat kesalahan di mana yang aktualnya positif namun di prediksi negatif. Hal ini dikarenakan model dilatih dan diuji dengan menggunakan dataset unigram sehingga model memprediksi hanya perkata.

6.2 Saran

Berdasarkan penelitian yang telah dilakukan, tujuan dari saran ini ialah untuk memberikan rekomendasi yang bermanfaat bagi pembaca maupun peneliti berikutnya. Berikut merupakan beberapa saran yang dapat dilakukan untuk penelitian selanjutnya, diantaranya :

1. Penelitian selanjutnya dapat mencoba dengan menggunakan model klasifikasi yang lain seperti model klasifikasi *naive bayes*, *support vector machines* dan *Deep Learning*.
2. Penelitian selanjutnya dapat mencoba memfilter kembali kata-kata yang dianggap sebagai *stopword* sehingga hasil *wordcloud* dapat lebih baik.
3. Penelitian selanjutnya diharapkan dapat membuat dan menguji model pada data *preprocessing* bigram dan *postag* bigram.
4. Penelitian selanjutnya diharapkan dapat mencoba melakukan *feature extraction* seperti *Word2Vec* dan menggunakan *N-grams* seperti *trigrams* dan *feature selection* seperti *correlation coefficient*, *chi-squared test*.

DAFTAR REFERENSI

- [1] Nair, V. G. (2014) *Getting Started with Beautiful Soup: Build your own web scraper and learn all about web scraping with Beautiful Soup*. Packt Publishing Ltd, Birmingham.
- [2] Jurafsky, D. dan Martin, J. H. (2023) Regular expressions, text normalization, edit distance. Bagian dari Jurafsky, D. dan Martin, J. H. (ed.), *Speech and Language Processing*. PrenticeHall, New Jersey.
- [3] Feldman, R. dan Sanger, J. (2007) *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press, New York.
- [4] Bird, S., Klein, E., dan Loper, E. (2009) *Natural Language Processing With Python: Analyzing Text with the Natural Language Toolkit*. O'Reilly Media, Sebastopol.
- [5] Jurafsky, D. dan Martin, J. H. (2023) Vector semantics and embeddings. Bagian dari Jurafsky, D. dan Martin, J. H. (ed.), *Speech and Language Processing*. PrenticeHall, New Jersey.
- [6] Krisanto, Y. (2022) Analisis data review ramen. Skripsi. Universitas Katolik Parahyangan, Indonesia.
- [7] Jurafsky, D. dan Martin, J. H. (2023) Sequence labeling for parts of speech and named entities. Bagian dari Jurafsky, D. dan Martin, J. H. (ed.), *Speech and Language Processing*. PrenticeHall, New Jersey.
- [8] Han, J., Kamber, M., dan Pei, J. (2012) *Data Mining: Concepts and Techniques*, 3rd edition. Morgan Kaufmann, Waltham.
- [9] Version 4.0 (2023) *Selenium History*. Selenium. ThoughtWorks, Chicago.
- [10] Version 4.0 (2023) *Finding web elements*. Selenium. ThoughtWorks, Chicago.
- [11] Version 4.0 (2023) *Selenium Overview*. Selenium. ThoughtWorks, Chicago.
- [12] Poletti, N. (2004) The vector space model in information retrieval-term weighting problem. *ResearchGate*, **1**, 01–06.
- [13] Kuhn, M. dan Johnson, K. (2013) *Applied Predictive Modeling*. Springer, New York Heidelberg Dordrecht London.