

BAB V

KESIMPULAN DAN SARAN

Bab ini akan berisikan mengenai kesimpulan yang dapat dihasilkan dari penelitian yang dilakukan terkait model klasifikasi *feature-based* untuk memprediksi *trip type* pada ulasan *online* hotel. Bab ini juga akan berisikan mengenai saran yang dapat digunakan untuk membantu penelitian lanjutan dengan topik dan metode yang sama.

V.1 Kesimpulan

Berdasarkan dari hasil penelitian mengenai model klasifikasi *feature-based* untuk memprediksi *trip type* pada ulasan *online* hotel, dapat ditarik kesimpulan yang menjawab tujuan dari penelitian yang dilakukan. Berikut merupakan kesimpulan dari hasil penelitian:

1. Dalam pembuatan model prediksi dengan metode *feature-based* seperti *tf-idf*, *raw term* dan DFS dapat dimodelkan serta diterapkan untuk memprediksi *trip type* dengan baik. Proses perubahan data mentah hingga sampai pemilihan model terbaik serta parameter terbaik yaitu dengan parameter 300 *feature*, *classifier decision tree* dan metode TF-IDF dan DFS dapat dilakukan dengan baik untuk memprediksi *trip type*.
2. Terdapat 87 *feature* yang berpengaruh dalam metode *tf-idf* serta 59 *feature* yang berpengaruh dalam metode DFS untuk membuat pohon *decision tree* yang digunakan dalam *classifier* terpilih yaitu *Decision Tree*. *Feature* yang paling berpengaruh diantaranya terdapat 'kid', 'business', 'honeymoon', 'nice', 'daughter', 'husband', 'child', 'suite', 'comfortable', 'coffee', 'relaxing', 'airport', 'love', 'dinner', 'modern', 'family', dan 'bar'. *Feature* tersebut berpengaruh dalam menurunkan *impurity* dengan signifikan sehingga model dapat memprediksi *trip type* dengan lebih baik.
3. Performansi akurasi dari model prediksi terbaik didapatkan sebesar 48.31% dan 48.19% untuk metode DFS dan metode TF-IDF. Kedua metode tersebut baik digunakan pada model ini karena tidak terdapat perbedaan rata-rata secara signifikan. Performansi sudah dianggap

cukup baik dari persentase yang dilakukan seorang yang menebak yaitu sebesar 20%.

V.2 Saran

Berdasarkan dari hasil penelitian mengenai model klasifikasi *feature-based* untuk memprediksi *trip type* pada ulasan *online* hotel, terdapat beberapa saran yang dapat diberikan untuk membantu penelitian lanjutan dengan topik serta metode yang sama. Berikut merupakan saran-saran yang dapat diberikan:

1. Dalam memodelkan pada ketiga metode tersebut diharapkan dapat dilakukan otomasi yang lebih baik dalam meng-outputkan data sehingga lebih mudah untuk diproses dalam tahap berikutnya. Penambahan *classifier* lain seperti SVM dan *Linear Regression* ataupun *word vector* dapat digunakan untuk penelitian lebih lanjut.
2. Mengumpulkan lebih banyak data mentah sehingga *feature* dengan karakteristik lain bisa didapatkan sehingga *feature* yang didapatkan penting lebih baik hasilnya. Menggunakan metode *feature selection* lain sehingga dapat dibandingkan hasil fitur terpilih yang dapat menggambarkan kelas.
3. Mencoba lebih banyak parameter dalam memodelkan model sehingga cakupannya dapat menjadi lebih luas.

DAFTAR PUSTAKA

- Badan Pusat Statistik. *Jumlah Kunjungan Wisatawan Mancanegara 2015-2019*
Indonesia: Badan Pusat Statistik. Diunduh dari <https://www.bps.go.id/>
- Badan Pusat Statistik. *Jumlah Kunjungan Wisatawan Mancanegara Menurut Pintu Masuk 2019*. Indonesia: Badan Pusat Statistik. Diunduh dari <https://www.bps.go.id/>
- Badan Pusat Statistik. *Tingkat Penghunian Kamar TPK Hotel Bintang Menurut Kelas di Provinsi Bali*. Bali: Badan Pusat Statistik. Diunduh dari <https://www.bps.go.id/>
- Beel, J., Gipp, B., Langer, S. & Breitinger, C. (2016). Research-paper recommender systems: a literature survey. Erschienen in: international Journal on Digital Libraries; 17,4. -S. 305-338.
- Chawla, N.V., Hall, L.O., Kegelmeyer.P.W., & Bowyer.W.K.(2002). SMOTE: *Synthetic Minority Over-Sampling*. Journal of Artificial Intelligence Research 16 (2002) 321–357.
- Flach, P. (2012). *Machine Learning: The Art and Science of Algorithms that Make Sense of Data: Cambridge*.
- Geckoroutes.com. (2018, 24 Oktober). *Most Popular Travel Booking Websites & Apps*. Diakses pada 28 Maret 2021, dari <https://www.geckoroutes.com>
- Jurafsky, D. & Martin, J.H. (2019). *Speech and Language Processing, Third Edition Draft*.
- Kadhim, A. (2018). *An Evaluation of Preprocessing Techinques for Text Classification*. International Journal of Computer Science and Information Security, 16(6):22-32.
- Kim, S.W., & Gil, J.M. (2019). *Research paper classification systems based on tf-idf and LDA schemes. Human-centric Computing and Information Sciences* 9(1).
- Kobayashi, V.B., Mol, S.T., Berkers, H.A., Kismihók, G., & Den Hartog, D.N. (2017). *Text classification for organizational researchers: A tutorial. Organizational Research Methods*.

- Korfiatis, N. (2012). Evaluating content quality and helpfulness of online product reviews: The interplay of review helpfulness vs. review content. *Electronic Commerce Research and Applications*, 11(93). 205-217.
- Kotler, P. & Keller, K.L. (2009). *Marketing Management 13th Edition*. 208-209.
- Müller, A.C. & Guido, S. (2017). *Introduction to Machine Learning with Python*. Sebastopol: O'Reilly
- Montgomery, D.C. & Runger, G.C. (2011). *Applied Statistics and Probability for Engineers (5th Edition)*: John Wiley & Sons, Inc.
- Statista.com. (2021, 9 Maret). *Trust in online customer reviews 2014-2018*. Diakses pada 28 Maret 2021, dari <https://www.statista.com>
- Statista.com. (2021, 9 Maret). *TripAdvisor number of reviews*. Diakses pada 28 Maret 2021, dari <https://www.statista.com>
- Sproat, R. dan Bedrick, S. (2011). "CS506/606: Text Normalization.
- Trianto, R.B., Triyono, A., dan Arum, D.M.P. *Klasifikasi Rating Otomatis pada Dokumen Teks Ulasan Produk Elektronik Menggunakan Metode N-gram dan Naïve Bayes*. (2020). Jurnal Informatika Universitas Pamulang: Program Studi Teknik Informatika Universitas Pamulang Vol 5, No.3, 295-301.
- Trstenjak, B., Mikac, S., & Donko, D. (2014). *KNN with TF-IDF based framework for text categorization*. *Procedia Engineering*, 69, 1356–1364
- Uysal, A.K. & Gunal, S. (2012). *A novel probabilistic feature selection method for text classification*. *Knowledge-Based Systems*, 36, 226-235.
- Uysal, A.K. & Gunal, S. (2014). *Text classification using genetic algorithm oriented latent semantic features*. *Expert Systems with Applications*, 41, 5938-5947.
- Xiang, Z., Qianzhou, D., Yufeng, M. & Weiguo, F. (2017). *Assessing Reliability of Social Media Data: Lessons from Mining TripAdvisor Hotel Reviews*