

BAB VI

PENUTUP

A. Kesimpulan

Berdasarkan dari hasil pengembangan model *Machine Learning* yang ditujukan untuk melakukan tugas Sentimen Analisis. Dalam penelitian ini, untuk melakukan klasifikasi dan prediksi sentimen *Negative* atau *Positive* dalam studi kasus keterbatasan kesempatan kerja dalam bidang yang diminati bagi *Gen Z*. Maka, dihasilkan beberapa kesimpulan yaitu:

1. Langkah-langkah yang diperlukan dalam mengembangkan model *Machine Learning* untuk tugas Sentimen Analisis meliputi, Pengumpulan Data, Analisis Data Eksploratif (*Exploratory Data Analysis*), *Preprocessing Data*, Pelatihan Model, dan Evaluasi Model.
2. Pembangunan model *Machine Learning* untuk tugas Sentimen Analisis berhasil dilakukan menggunakan fungsi MultinomialNB yang didasarkan pada algoritme *Naïve Bayes*, menghasilkan nilai *Accuracy* 72% dari 131 data pengujian.
3. Berdasarkan hasil evaluasi *Classification Report* dengan 131 data pengujian, menghasilkan nilai *F1-Score* 81% untuk kelas 0 (*Negative*) dan 41% untuk kelas 1 (*Positive*), nilai *Precision* 69% untuk kelas 0 dan 100% untuk kelas 1, nilai *Recall* kelas 0 adalah 100% dan kelas 1 adalah 26%.
4. Kemudian, berdasarkan evaluasi model menggunakan *Confusion Matrix* dengan 131 data pengujian menghasilkan nilai *True Negative* 61.83%, *True Positive* 9.92%, *False Negative* 28.24%, dan *False Positive* 0.00%.
5. Nilai *False Negative* 28,24% dari 131 data pengujian mengindikasikan bahwa kemungkinan model dapat melakukan kesalahan prediksi terhadap

data *Positive* diprediksi sebagai *Negative*. Hal ini terjadi karena data kelas sentimen *Positive* dan sentimen *Negative* dalam data pelatihan yang digunakan untuk melatih model jumlahnya tidak seimbang, sehingga model lebih condong menghasilkan prediksi sentimen *Negative*.

6. Berdasarkan hasil analisis sentimen data opini tentang keterbatasan kesempatan kerja dalam bidang yang diminati bagi *Gen Z*, menunjukkan bahwa jumlah data sentimen *Negative* lebih banyak daripada sentimen *Positive*.
7. Pengembangan model *Machine Learning* untuk melakukan tugas Sentimen Analisis dalam Bahasa Indonesia masih memiliki banyak kendala. Karena proses *stemming* menggunakan *library* Sastrawi tidak berhasil dilakukan. Sehingga, program yang dihasilkan masih sangat sederhana dan kurang akurat.
8. Kemudian, jumlah data pelatihan yang kurang banyak dan jumlah kelas sentimen *Negative* dan sentimen *Positive* yang tidak seimbang dalam data pelatihan juga menyebabkan model tidak begitu akurat dalam melakukan klasifikasi dan prediksi pada data baru. Hal ini dapat dibuktikan dengan hasil evaluasi nilai *False Negative* 28.24% dari 131 data pengujian yang digunakan. Dengan demikian, model cenderung memprediksi data sebagai sentimen *Negative*.

B. Saran

Untuk penelitian selanjutnya, pembangunan model *Machine Learning* untuk tugas Sentimen Analisis perlu dilakukan menggunakan algoritme *Machine Learning* yang lain untuk dapat dibandingkan dengan algoritme *Naïve Bayes*. Perlu menggunakan *dataset* berjumlah cukup besar dan jumlah kelas yang seimbang untuk melakukan pelatihan model, agar model dapat melakukan klasifikasi dengan akurat. Melakukan *preprocessing* atau proses pembersihan data adalah hal yang sangat

penting dalam membangun model *Machine Learning*. Melatih model dengan data yang buruk atau “*kotor*” akan menghasilkan model yang buruk dan tidak terlatih dengan benar, membuat model menjadi tidak relevan.



DAFTAR PUSTAKA

- [1] “Natural Language Processing by deeplearning.ai | Coursera.” <https://www.coursera.org/specializations/natural-language-processing> (accessed Feb. 05, 2023).
- [2] “Natural Language Processing (NLP): What Is It & How Does it Work?” <https://monkeylearn.com/natural-language-processing/> (accessed Feb. 02, 2023).
- [3] S. Bhatia, M. Sharma, and K. K. Bhatia, “Sentiment Analysis and Mining of Opinions,” *Studies in Big Data*, vol. 30, no. May, pp. 503–523, 2018, doi: 10.1007/978-3-319-60435-0_20.
- [4] A. D’Andrea, F. Ferri, P. Grifoni, and T. Guzzo, “Approaches, Tools and Applications for Sentiment Analysis Implementation,” *Int J Comput Appl*, vol. 125, no. 3, pp. 26–33, Sep. 2015, doi: 10.5120/IJCA2015905866.
- [5] “About Python™ | Python.org.” <https://www.python.org/about/> (accessed Jun. 21, 2023).
- [6] “8 Steps in the Data Life Cycle | HBS Online.” <https://online.hbs.edu/blog/post/data-life-cycle> (accessed Feb. 03, 2023).
- [7] “What is Exploratory Data Analysis? - GeeksforGeeks.” <https://www.geeksforgeeks.org/what-is-exploratory-data-analysis/> (accessed Jun. 22, 2023).
- [8] “Data Preprocessing in Machine learning - Javatpoint.” <https://www.javatpoint.com/data-preprocessing-machine-learning> (accessed Jun. 22, 2023).

- [9] “Apa itu Visualisasi Data? - Penjelasan tentang Visualisasi Data - AWS.”
<https://aws.amazon.com/id/what-is/data-visualization/> (accessed Jun. 21, 2023).
- [10] “Training ML Models - Amazon Machine Learning.”
<https://docs.aws.amazon.com/machine-learning/latest/dg/training-ml-models.html> (accessed Jun. 21, 2023).
- [11] “Performance Metrics in Machine Learning - Javatpoint.”
<https://www.javatpoint.com/performance-metrics-in-machine-learning>
(accessed Jun. 22, 2023).
- [12] R. S. Jagdale, V. S. Shirsat, and S. N. Deshmukh, *Sentiment analysis on product reviews using machine learning techniques*, vol. 768, no. January. Springer Singapore, 2019. doi: 10.1007/978-981-13-0617-4_61.
- [13] S. Lakshmanan, H. Sharieff, T. Sindhu, and L. Sairamesh, “Comparison of Machine Learning Techniques for Sentimental Analysis on Restaurant Reviews,” *International Journal of Advances in Engineering and Management (IJAEM)*, vol. 2, no. 6, p. 740, 2008, doi: 10.35629/5252-0206740743.
- [14] A. A. Farisi, Y. Sibaroni, and S. al Faraby, “Sentiment analysis on hotel reviews using Multinomial Naïve Bayes classifier”, doi: 10.1088/1742-6596/1192/1/012024.
- [15] A. Prabhat and V. Khullar, “Sentiment classification on big data using Naïve bayes and logistic regression,” *2017 International Conference on Computer Communication and Informatics, ICCCI 2017*, no. March, 2017, doi: 10.1109/ICCCI.2017.8117734.
- [16] “Apa itu Kecerdasan Buatan (AI)? — Amazon Web Services.”
<https://aws.amazon.com/id/machine-learning/what-is-ai/> (accessed Feb. 03, 2023).

- [17] “What is Artificial Intelligence (AI)? | IBM.”
<https://www.ibm.com/topics/artificial-intelligence> (accessed Feb. 03, 2023).
- [18] “Machine learning, explained | MIT Sloan.” <https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained> (accessed Feb. 03, 2023).
- [19] “What is Natural Language Processing? A Guide to NLP in 2023 - viso.ai.”
<https://viso.ai/deep-learning/natural-language-processing/> (accessed Feb. 03, 2023).
- [20] C. Interaksi, “Teknologi Pemrosesan Bahasa Alami sebagai Teknologi Kunci untuk Meningkatkan,” *Gastrointest Endosc*, vol. 92, no. 4, pp. 807–812, 2004.
- [21] “Your Guide to Natural Language Processing (NLP) | by Diego Lopez Yse | Towards Data Science.” <https://towardsdatascience.com/your-guide-to-natural-language-processing-nlp-48ea2511f6e1> (accessed Feb. 05, 2023).
- [22] “Natural Language Processing (NLP): What it is and why it matters | SAS.”
https://www.sas.com/en_us/insights/analytics/what-is-natural-language-processing-nlp.html (accessed Feb. 05, 2023).
- [23] “What is Tokenization | Methods to Perform Tokenization.”
<https://www.analyticsvidhya.com/blog/2019/07/how-get-started-nlp-6-unique-ways-perform-tokenization/> (accessed Feb. 06, 2023).
- [24] “Stemming vs Lemmatization in NLP: Must-Know Differences.”
<https://www.analyticsvidhya.com/blog/2022/06/stemming-vs-lemmatization-in-nlp-must-know-differences/> (accessed Feb. 06, 2023).
- [25] “Stemming and Lemmatization in Python | DataCamp.”
<https://www.datacamp.com/tutorial/stemming-lemmatization-python> (accessed Feb. 06, 2023).

- [26] “Removing stop words with NLTK library in Python | by Banjoko Judah | Analytics Vidhya | Medium.” <https://medium.com/analytics-vidhya/removing-stop-words-with-nltk-library-in-python-f33f53556cc1> (accessed Feb. 06, 2023).
- [27] “ID-Stopwords/id.stopwords.02.01.2016.txt at master · masdevid/ID-Stopwords · GitHub.” <https://github.com/masdevid/ID-Stopwords/blob/master/id.stopwords.02.01.2016.txt> (accessed Feb. 11, 2023).
- [28] “Understanding TF-IDF: A Simple Introduction.” <https://monkeylearn.com/blog/what-is-tf-idf/> (accessed Feb. 09, 2023).
- [29] “TDM (Term Document Matrix) and DTM (Document Term Matrix) | by Heena Girdher | Analytics Vidhya | Medium.” <https://medium.com/analytics-vidhya/tdm-term-document-matrix-and-dtm-document-term-matrix-8b07c58957e2> (accessed Feb. 09, 2023).
- [30] “Feature Extraction Definition | DeepAI.” <https://deepai.org/machine-learning-glossary-and-terms/feature-extraction> (accessed Feb. 09, 2023).
- [31] “What is Supervised Learning? | IBM.” <https://www.ibm.com/topics/supervised-learning> (accessed Feb. 06, 2023).
- [32] “Supervised and Unsupervised Learning in (Machine Learning).” <https://www.simplilearn.com/tutorials/machine-learning-tutorial/supervised-and-unsupervised-learning> (accessed Feb. 06, 2023).
- [33] “5 Types of Classification Algorithms in Machine Learning.” <https://monkeylearn.com/blog/classification-algorithms/> (accessed Feb. 06, 2023).

- [34] “Naive Bayes Classifier in Machine Learning - Javatpoint.”
<https://www.javatpoint.com/machine-learning-naive-bayes-classifier> (accessed Feb. 06, 2023).
- [35] “What Is Data Analysis and How Can You Get Started?”
<https://monkeylearn.com/data-analysis/> (accessed Feb. 06, 2023).
- [36] “Qualitative or Quantitative Research? | MQHRG - McGill University.”
<https://www.mcgill.ca/mqhrg/resources/what-difference-between-qualitative-and-quantitative-research> (accessed Feb. 06, 2023).
- [37] “Sentiment Analysis – What Is It and Why Does It Matter?”
<https://www.nvidia.com/en-us/glossary/data-science/sentiment-analysis/>
(accessed Feb. 07, 2023).
- [38] “What is Sentiment Analysis? - Sentiment Analysis Explained - AWS.”
<https://aws.amazon.com/what-is/sentiment-analysis/> (accessed Feb. 07, 2023).
- [39] “Sentiment Analysis Guide.” <https://monkeylearn.com/sentiment-analysis/>
(accessed Feb. 07, 2023).
- [40] “Guide to Aspect-Based Sentiment Analysis.”
<https://monkeylearn.com/blog/aspect-based-sentiment-analysis/> (accessed Feb. 07, 2023).
- [41] “Aspect-Based Sentiment Analysis | Paper With Code.”
<https://paperswithcode.com/task/aspect-based-sentiment-analysis> (accessed Feb. 07, 2023).
- [42] “What is Model Evaluation? | Domino Data Science Dictionary.”
<https://www.dominodatalab.com/data-science-dictionary/model-evaluation>
(accessed Feb. 07, 2023).

- [43] “Evaluating a Machine Learning Model | by Skyl.ai | Medium.”
<https://medium.com/@skyl/evaluating-a-machine-learning-model-7cab1f597046> (accessed Feb. 07, 2023).
- [44] “Model Evaluation Techniques|Machine Learning Model Evaluation.”
<https://www.analyticsvidhya.com/blog/2021/05/machine-learning-model-evaluation/> (accessed Feb. 07, 2023).
- [45] “What is a confusion matrix?. Everything you Should Know about... | by Anuganti Suresh | Analytics Vidhya | Medium.” <https://medium.com/analytics-vidhya/what-is-a-confusion-matrix-d1c0f8feda5> (accessed Feb. 07, 2023).
- [46] “gen-z noun - Definition, pictures, pronunciation and usage notes | Oxford Advanced Learner’s Dictionary at OxfordLearnersDictionaries.com.”
<https://www.oxfordlearnersdictionaries.com/definition/english/gen-z?q=gen+z> (accessed Feb. 09, 2023).
- [47] “Where Millennials end and Generation Z begins | Pew Research Center.”
<https://www.pewresearch.org/fact-tank/2019/01/17/where-millennials-end-and-generation-z-begins/> (accessed Feb. 09, 2023).
- [48] B. S. Sladek and A. Grabinger, “The first generation of the 21st Century has arrived!,” *XYZ University*, p. 16, 2016, [Online]. Available: www.xyzuniversity.com
- [49] Central Bureau of Statistics, “Hasil Sensus Penduduk 2020,” *Bps.Go.Id*, no. 7, pp. 1–52, 2021, [Online]. Available: <https://papua.bps.go.id/pressrelease/2018/05/07/336/indeks-pembangunan-manusia-provinsi-papua-tahun-2017.html>
- [50] “Karakteristik Negara Maju dan Berkembang, Disertai Contoh dan Permasalahannya - Gramedia Literasi.”

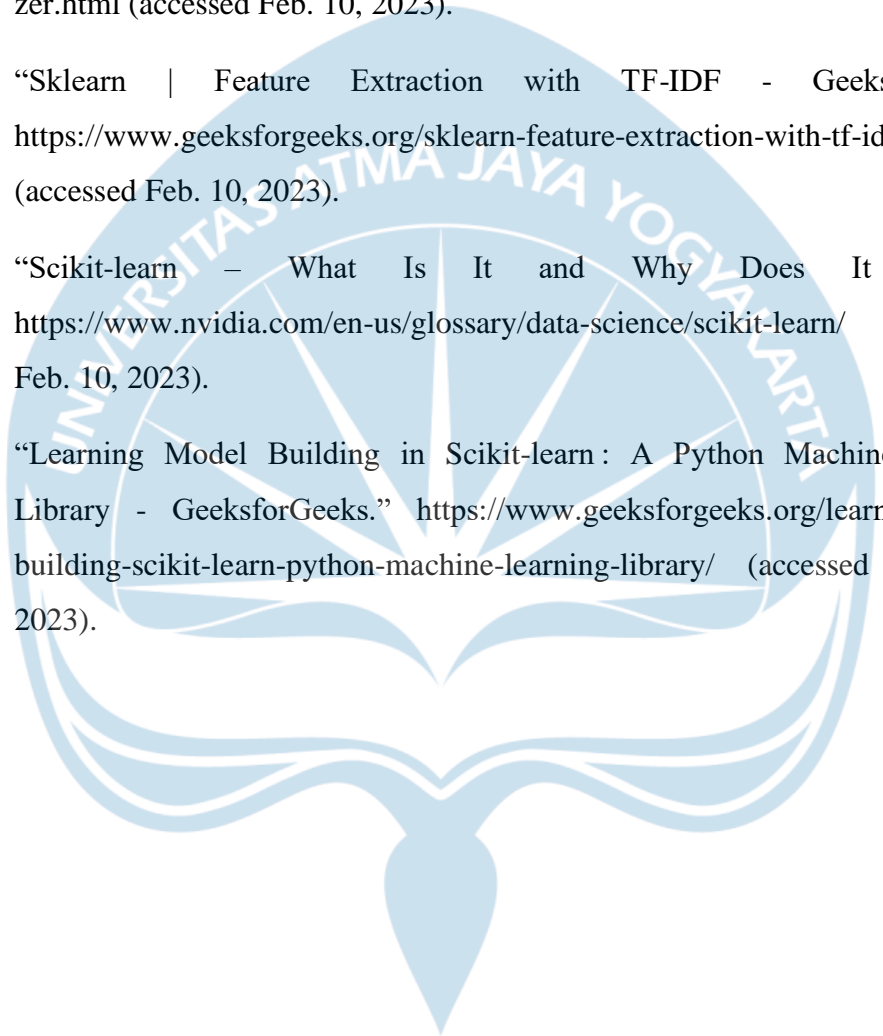
<https://www.gramedia.com/literasi/karakteristik-negara-maju-dan-berkembang/> (accessed Jul. 13, 2023).

- [51] “Top 25 Developed and Developing Countries.” <https://www.investopedia.com/updates/top-developing-countries/> (accessed Jul. 13, 2023).
- [52] T. P. Terbuka, “Badan Pusat Statistik tentang Keadaan Ketenagakerjaan Indonesia Februari 2023,” no. 35, 2023.
- [53] “Developing Economy - Definition, Characteristics, Examples.” <https://www.wallstreetmojo.com/developing-economy/> (accessed Jul. 13, 2023).
- [54] A. Sarlan, S. Basri, and C. Nadam, “Twitter sentiment analysis”, doi: 10.1109/ICIMU.2014.7066632.
- [55] “Introduction to Social Media | University Communications and Marketing.” <https://www.usf.edu/ucm/marketing/intro-social-media.aspx> (accessed Feb. 07, 2023).
- [56] “Apa Itu Blog? Pengertian Blog dan Contohnya Lengkap!” <https://www.hostinger.co.id/tutorial/apa-itu-blog> (accessed Feb. 07, 2023).
- [57] “Blog Definition & Meaning | Dictionary.com.” <https://www.dictionary.com/browse/blog> (accessed Feb. 07, 2023).
- [58] “Train and Test datasets in Machine Learning - Javatpoint.” <https://www.javatpoint.com/train-and-test-datasets-in-machine-learning> (accessed Jul. 18, 2023).
- [59] “What Is Data Preprocessing & What Are The Steps Involved?” <https://monkeylearn.com/blog/data-preprocessing/> (accessed Jun. 18, 2023).

- [60] “What is Exploratory Data Analysis? | IBM.” <https://www.ibm.com/topics/exploratory-data-analysis> (accessed Jun. 18, 2023).
- [61] “What is Data Visualization? | IBM.” <https://www.ibm.com/topics/data-visualization> (accessed Jun. 18, 2023).
- [62] “What is Data Visualization: Tutorial, History, Importance - javatpoint.” <https://www.javatpoint.com/what-is-data-visualization> (accessed Jun. 18, 2023).
- [63] “What is a Trained Model?. Or...what does ‘training an ML model’... | by Federico Trotta | Towards Data Science.” <https://towardsdatascience.com/what-is-a-trained-model-5c872cfa8448> (accessed Jul. 19, 2023).
- [64] “What Is Python Used For? A Beginner’s Guide | Coursera.” <https://www.coursera.org/articles/what-is-python-used-for-a-beginners-guide-to-using-python> (accessed Jun. 21, 2023).
- [65] “Google Forms: Online Form Creator | Google Workspace.” <https://www.google.com/forms/about/> (accessed Feb. 07, 2023).
- [66] “MS Excel Definition - javatpoint.” <https://www.javatpoint.com/ms-excel-definition> (accessed Feb. 11, 2023).
- [67] “Jupyter Notebook for Machine Learning - A Gentle Introduction - viso.ai.” <https://viso.ai/deep-learning/jupyter-notebook-for-machine-learning/> (accessed Feb. 07, 2023).
- [68] “Project Jupyter | Home.” <https://jupyter.org/> (accessed Feb. 07, 2023).

- [69] “Python for Data Analysis, 3E - 4 NumPy Basics: Arrays and Vectorized Computation.” <https://wesmckinney.com/book/numpy-basics.html> (accessed Feb. 07, 2023).
- [70] “Chapter 5: Introduction to NumPy — Python Programming for Data Science.” <https://www.tomasbeuzen.com/python-programming-for-data-science/chapters/chapter5-numpy.html> (accessed Feb. 08, 2023).
- [71] “Top 15 Python Libraries for Data Science and Machine Learning.” <https://www.projectpro.io/article/top-5-libraries-for-data-science-in-python/196> (accessed Feb. 11, 2023).
- [72] “Top 10 Python Libraries for Data Science for 2023.” <https://www.simplilearn.com/top-python-libraries-for-data-science-article> (accessed Feb. 11, 2023).
- [73] “Plot line graph from NumPy array - GeeksforGeeks.” <https://www.geeksforgeeks.org/plot-line-graph-from-numpy-array/> (accessed Feb. 11, 2023).
- [74] “pandas · PyPI.” <https://pypi.org/project/pandas/> (accessed Feb. 08, 2023).
- [75] “Matplotlib Library in Python.” <https://www.enjoyalgorithms.com/blog/introduction-to-matplotlib> (accessed Feb. 08, 2023).
- [76] “seaborn: statistical data visualization — seaborn 0.12.2 documentation.” <https://seaborn.pydata.org/> (accessed Feb. 10, 2023).
- [77] “Python seaborn Library - Javatpoint.” <https://www.javatpoint.com/python-seaborn-library> (accessed Feb. 10, 2023).

- [78] “What is Word Cloud | Guide To Create A Word Cloud in Python.”
<https://www.analyticsvidhya.com/blog/2020/10/word-cloud-or-tag-cloud-in-python/> (accessed Feb. 08, 2023).
- [79] “Generating Word Cloud in Python - GeeksforGeeks.”
<https://www.geeksforgeeks.org/generating-word-cloud-python/> (accessed Feb. 08, 2023).
- [80] “regular expressions Definition.”
https://www.intel.com/content/www/us/en/programmable/quartushelp/17.0/reference/glossary/def_reg_express.htm (accessed Feb. 08, 2023).
- [81] “Regex in Python — A-Z. As a data scientist/engineer, having a... | by Shritam Kumar Mund | Analytics Vidhya | Medium.”
<https://medium.com/analytics-vidhya/regex-in-python-a-z-88ebf1c8fed4>
(accessed Feb. 08, 2023).
- [82] “Python String Methods.”
https://www.w3schools.com/python/python_ref_string.asp (accessed Feb. 13, 2023).
- [83] “Dasar Text Preprocessing dengan Python | by Kuncahyo Setyo Nugroho | Medium.”
<https://ksnugroho.medium.com/dasar-text-preprocessing-dengan-python-a4fa52608ffe> (accessed Feb. 09, 2023).
- [84] “Word Vector Encoding in NLP.”
<https://www.enjoyalgorithms.com/blog/word-vector-encoding-in-nlp> (accessed Feb. 09, 2023).
- [85] “Vectorization Techniques in NLP [Guide] - neptune.ai.”
<https://neptune.ai/blog/vectorization-techniques-in-nlp-guide> (accessed Feb. 06, 2023).

- 
- [86] “sklearn.feature_extraction.text.TfidfVectorizer — scikit-learn 1.2.1 documentation.” https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html (accessed Feb. 10, 2023).
- [87] “Sklearn | Feature Extraction with TF-IDF - GeeksforGeeks.” <https://www.geeksforgeeks.org/sklearn-feature-extraction-with-tf-idf/> (accessed Feb. 10, 2023).
- [88] “Scikit-learn – What Is It and Why Does It Matter?” <https://www.nvidia.com/en-us/glossary/data-science/scikit-learn/> (accessed Feb. 10, 2023).
- [89] “Learning Model Building in Scikit-learn: A Python Machine Learning Library - GeeksforGeeks.” <https://www.geeksforgeeks.org/learning-model-building-scikit-learn-python-machine-learning-library/> (accessed Feb. 10, 2023).