

Model Semi-supervised Learning Menggunakan Logistic Regression Untuk Anotasi Sentimen

Siti Khomsah^{1*}, Agus Sasmito Aribowo²

¹*Sains Data, Institut Teknologi Telkom Purwokerto, Indonesia*

²*Informatics, Universitas Pembangunan Nasional “Veteran” Yogyakarta, Indonesia*

*Corresponding Author: siti@ittelkom-pwt.ac.id

Abstract

Sentiment analysis uses a labeled dataset to train a classification model. Sentiment datasets are not always available for every experiment. The problem is that preparing sentiment annotated datasets requires experts as annotators and takes a lot of time and cost. Semi-supervised annotation (SSA) is a promising method for overcoming the difficulty of the data labeling process, namely minimizing expert dependency and reducing time and costs. This study aims to build a text annotation model with a semi-supervised annotation model (SSA-Model) for sentiment analysis. The performance of classification model with SSA will be compared to model with baseline. Baseline is a model with datasets labeled manually by experts. SSA-Model is a model with machine learning annotations. The data vectorization stage uses the BoW, and N-gram approaches with $N=1,2,3$. At the same time, the classification model uses Logistic Regression and Naïve Bayes. This study uses the IMDB dataset to test the model. The model is evaluated using a confusion matrix, measuring the accuracy and F1 score. In conclusion, semi-supervised annotations can automatically assign sentiment class labels. However, its performance is affected by the amount of training data and the suitability of the dataset with machine learning algorithms. The more labeled training data used in the SSA model, the better the classification performance. In the IMDB dataset, which consists of two classes, the performance of the classification model with Logistic Regression is better than Naïve Bayes.

Keywords: *sentiment analysis, logistic regression, annotation, semi supervised learning, naïve bayes*

Abstrak

Sentimen analisis menggunakan dataset berlabel untuk melatih model klasifikasi. Dataset berlabel sentimen tidak selalu tersedia untuk setiap percobaan. Permasalahannya, untuk menyiapkan dataset beranotasi sentimen membutuhkan ahli sebagai *anotator*, membutuhkan banyak waktu, dan biaya. *Semi-supervised annotation* (SSA) hadir sebagai metode yang menjanjikan untuk mengatasi sulitnya proses pelabelan data, yaitu meminimalkan ketergantungan pakar, mengurangi waktu dan biaya. Tujuan penelitian ini membangun model anotasi teks dengan pendekatan *semi-supervised annotation* (SSA-Model) untuk sentimen analisis. Model klasifikasi sentimen dengan SSA akan dibandingkan akurasi dengan model *baseline* yaitu model dengan dataset yang diberi label secara manual oleh pakar. Model SSA adalah model dengan anotasi *machine learning*. Tahap vektorisasi data menggunakan pendekatan BoW dan N-gram dengan jumlah $N=1,2,3$. Sedangkan model klasifikasi menggunakan Logistic Regression dan Naïve Bayes. Penelitian ini menggunakan dataset IMDB untuk menguji model. Model dievaluasi menggunakan *confusion matrix*, dilihat dari nilai akurasi dan *F1 score*. Kesimpulannya, *semi-supervised annotation* dapat digunakan untuk memberikan label kelas sentimen secara otomatis. Namun, kinerjanya dipengaruhi oleh jumlah *data training* dan kesesuaian dataset dengan algoritma *machine learning*. Semakin banyak data training berlabel yang digunakan dalam model SSA maka kinerja klasifikasi makin baik. Pada dataset IMDB yang terdiri dari dua kelas, kinerja model klasifikasi dengan Logistic Regression lebih baik dibandingkan Naïve Bayes.

Kata Kunci: *sentiment analysis, logistic regression, annotation, semi supervised learning, naïve bayes*

I. INTRODUCTION

Analisis sentimen bertujuan untuk mendeteksi opini-opini termasuk kategori opini positif, negative, atau netral. Kegunaan dari analisis sentimen di internet adalah untuk mengetahui informasi sentimen publik terkait tokoh politik[1], film[2], hotel dan obyek wisata [3], maskapai penerbangan [4], dan sebagainya. Hasil dari proses klasifikasi sentimen adalah sekumpulan dokumen yang dapat dipakai untuk menyimpulkan kepuasan pelanggan terkait tokoh politik, layanan atau produk. Contohnya, opini "kampanye sang capres tidak menarik...!!" yang diambil dari komentar di internet. Kata "tidak menarik" merefleksikan sentimen negatif. Contoh lainnya, kata "wah bajunya bagus, kainnya berkualitas" adalah komentar terhadap sebuah produk. Kata "bagus" dan "berkualitas" menunjukkan adanya sentimen positif. Pada *supervised learning*, tugas utama model *classifier* analisis sentimen adalah menentukan kelas sebuah teks menjadi positif, netral atau negatif. Pendekatan *supervised learning* hanya bekerja pada dataset *training* berlabel. Pada proses pelabelan atau anotasi kelas secara manual memerlukan pakar untuk menghasilkan dataset berlabel yang berkualitas tinggi. Permasalahan proses anotasi manual antara lain waktu yang lama, biaya tinggi, dan hasilnya bisa bias karena subyektifitas pakar dalam memaknai kalimat atau teks. Berdasarkan kajian literatur, pendekatan *semi-supervised learning* bisa digunakan untuk anotasi [5][6][7]. Pendekatan tersebut menjanjikan menyelesaikan permasalahan anotasi manual oleh pakar.

Penelitian ini bertujuan untuk membangun model anotasi teks dengan pendekatan *semi-supervised annotation* (SSA-Model) untuk sentimen analisis. Penelitian ini merupakan kelanjutan dari penelitian sebelumnya yaitu pada [5] [6] [7]. Sebagian dataset teks yang belum berlabel akan diberikan label oleh pakar. Kemudian, model klasifikasi sentimen akan dibangun berdasarkan dataset tersebut, hasil akurasi model disebut *baseline*. Langkah berikutnya, sisa dataset yang belum berlabel akan dianotasi atau diberikan label oleh model *baseline* tersebut. Kemudian, seluruh dataset yang diberi label baik oleh pakar dan model *baseline* akan digunakan untuk membangun model klasifikasi sentimen. Ketercapaian target dari SSA-Model adalah ketika kinerja model yang menggunakan seluruh dataset akan dibandingkan dengan kinerja model *baseline*. Beberapa strategi untuk menemukan model SSA terbaik yaitu dengan ekstraksi fitur *Bag of Word* (BoW), *n-gram*, dan dua *machine learning* berbeda yaitu Naïve Bayes dan Logistic Regression sebagai model *classifier*. Gram pada N-gram yang dimaksud adalah level word atau kata. Strategi lainnya yaitu *multi-classifiers* untuk meningkatkan akurasi model dan menjaga kualitas anotasi tetap baik. Pada mekanisme *multi-classifier*, setiap model bekerja dalam satu jenis tokenisasi. Model *classifier* pertama bekerja pada token *unigram*, model *classifier* kedua bekerja pada *bigram*, dan model *classifier* ketiga bekerja pada *trigram*. Pertanyaan penelitian ini adalah dapatkah kombinasi BoW, *n-gram*, dan kedua *machine learning* dapat mencapai akurasi model *semi-supervised annotation* tinggi dibandingkan dengan model *baseline*. *Machine learning* yang digunakan ada dua yaitu Logistic Regression dan Naïve Bayes. Pertanyaan berikutnya adalah berapakah jumlah dataset beranotasi yang dibutuhkan untuk training model *semi-supervised annotation* sehingga secara signifikan menaikkan akurasi model. Maka untuk menguji model, penelitian menggunakan dataset yang umum digunakan pada penelitian analisis sentimen, yaitu dataset IMDB.

II. LITERATURE REVIEW

Beberapa penelitian menunjukkan bahwa kesulitan pada anotasi secara manual dapat diatasi dengan *semi-supervised annotation* (SSA). Beraneka metode *semi-supervised annotation* memberikan akurasi lebih tinggi, misalnya pada penelitian [8] dan [9]. Penelitian *semi-supervised annotation* pada data IMDB [8] menerapkan *semi-supervised deep neural networks* dengan beberapa parameter uji coba. Percobaan mereka menghasilkan model dengan akurasi 81%-82%, dan tidak berbeda jauh dengan model *baseline*-nya(82%). Model SSA pada AraSenCorpus [9] dipakai untuk anotasi korpus berbahasa Arab, menggunakan tweets beranotasi berukuran kecil. Hasilnya, pada metode *two-way* AraSenCorpus berhasil meningkatkan kinerja model dari 80.37% menjadi 87.4% pada dataset SemEval 2017 dan dari 79.77% menjadi 85.2% menggunakan ASTD dataset. Pada *three-way* memberikan akurasi 69.4% pada SemEval

2017, sementara sistem sebelumnya memberikan 63.38%. Kemudian hasil pengukuran menggunakan pengujian F1 score, model mereka meningkatkan kinerja klasifikasi dari 64.10% menjadi 68.1% pada ASTD dataset.

Kombinasi *supervised* dan *unsupervised* untuk anotasi data digunakan oleh Balakrishnan [10] dalam membangun *semi-supervised annotation* pada teks berbahasa Inggris yang diperoleh dari tiga aplikasi *digital payment*. Penelitiannya [10] melibatkan unsur *sentiment analysis* dan *emotion analysis*. Algoritma klasifikasi menggunakan *hybrid supervised* dan *unsupervised machine learning*. Penelitian mereka menggunakan algoritma Support Vector Machine, Random Forest, dan Naïve Bayes. Hasil penelitian mereka menunjukkan bahwa Random Forest mendapatkan F1 score = 73.8% dan Cohen's Kappa = 52.2% (untuk sentimen analisis). Untuk emotion analisis model mereka mendapatkan F1 score = 58.8% dan Cohen's Kappa = 44.7% [10]. Selain machine learning, deep learning memberikan kinerja baik pada anotasi otomatis, dibuktikan oleh Alahmary [11] yang mengajukan pendekatan *semi-automatic* untuk anotasi dataset twitter berdialek Saudi. Model mereka menghasilkan akurasi 83% menggunakan Naïve Bayes classifier. Model mereka juga menggunakan tiga *deep learning classifiers* yaitu convolutional neural network (CNN), long short-term memory (LSTM), dan bidirectional long short-term memory (Bi-LSTM). Hasil penelitian mereka menunjukkan bahwa kinerja dari deep learning classifiers, khususnya CNN melebihi SVM. CNN melampaui semua *classifier* yang lain dengan akurasi tertinggi yaitu 87% [11].

III. RESEARCH METHOD

Pada bagian ini membahas tentang metodologi penelitian untuk persiapan data, vektorisasi, ekstraksi fitur, pemodelan menggunakan Logistic Regression, Naïve Bayes, validasi model, dan arsitektur model.

A. Data Collection

Penelitian ini menggunakan dataset IMDB yang berisi dataset revid film sebanyak 50,000 *record* berlabel. IMDB adalah dataset untuk klasifikasi sentiment biner dengan label positif dan negatif.

B. Data Cleaning and Preprocessing

Dataset IMDB terdiri atas komentar-komentar perfilman dalam bahasa Inggris. Dataset tersebut butuh analisis awal, terkait jumlah kalimat, kata, dan keberadaan simbol-simbol khusus. Proses-proses untuk merubah data tidak terstruktur tersebut menjadi lebih terstruktur harus melalui beberapa tahap seperti tokenisasi, (unigram, bigram, and trigram), konversi ke *lowercase*, menghapus nomor, *stopword*, dan semua *non-alphabetic* karakter, dan tanda baca, kemudian dilakukan *stemming*.

C. Vectorization

Bag of Word (BoW) atau sering juga disebut sebagai *Term frequency* adalah algoritma untuk menghitung bobot setiap kata dalam dokumen-dokumen. *Term frequency* juga sering disebut sebagai frekuensi kemunculan kata t pada dokumen D . Bobot *Term frequency* adalah nilai kemunculan kata t pada dokumen D dibagi dengan total kata dalam dokumen D . Setiap kata unik akan menjadi fitur utama dokumen. Maka BoW akan menghasilkan matriks vektor berukuran besar. Jika jumlah fitur adalah n dan jumlah dokumen adalah $sum(D)$, maka ukuran matriks adalah $n \times sum(D)$.

D. Model Classifier

Logistic Regression adalah salah satu jenis algoritma yang digunakan untuk proses-proses klasifikasi data. Logistic regression adalah jenis analisis statistik yang sering digunakan untuk pemodelan prediktif. Dalam pendekatan analitik ini, variabel dependennya bersifat kategoris, bisa berupa positif atau negatif (pada regresi biner) atau berbagai pilihan multi kelas seperti positif, netral, negatif, atau sangat negatif (disebut regresi multinomial). Pada dataset sentiment, algoritma ini dibutuhkan untuk memprediksi apakah opini-opini tersebut masuk pada kelompok sentimen positif atau negatif. Sedangkan Naïve Bayes juga digunakan untuk berbagai penelitian analisis sentimen dalam bahasa Indonesian [12][13] dan dataset komentar video [14]. Naive Bayes sudah dikenal sebagai *machine learning* yang memberikan akurasi tinggi pada *binary sentiment analysis* [12][13][14]. Aturan bayes ada pada rumus (1).

$$P(y|X) = \frac{P(X|y)P(y)}{P(X)} \dots\dots\dots (1)$$

dimana :

- P (y) = Kemungkinan y benar
- P (X) = Kemungkinan X benar
- P (y|X) = Kemungkinan y benar jika X benar
- P (X|y) = Kemungkinan X benar jika y benar

Naive Bayes dapat digunakan pada klasifikasi biner atau multi kelas. Metode ini diterapkan pada teknik-teknik supervised classification.

E. Validation

Kinerja model SSA akan divalidasi menggunakan *confusion matrix*. Confusion matrix berkerja dengan cara membandingkan antara hasil prediksi dengan kelas aktual menggunakan aturan dalam TABEL 1.

TABEL 1
CONFUSION MATRIX

		Actual	
		Positive	Negative
Predicted	Positive	True Positive / TP	False Positive / FP
	Negative	False Negative / FN	True Negative / TN

Penelitian ini menggunakan dua parameter untuk validasi model, yaitu akurasi dan F1score. Akurasi adalah rasio dataset yang diprediksi dengan benar terhadap semua dataset dalam percobaan. Akurasi pada persamaan (2) adalah alat pengukuran akurasi yang baik tetapi hanya dapat diperlakukan pada *symmetric dataset*, yaitu ketika jumlah *false positive* dan *false negative* kurang lebih sama.

$$Accuracy = (TP+TN) / (TP+FP+FN+TN) \dots\dots\dots (2)$$

F1 score (pada persamaan (3)) adalah bobot rata-rata dari *Precision* dan *Recall*. *F1 score* biasanya lebih bermanfaat daripada akurasi, khususnya jika hasil memiliki distribusi kelas yang tidak merata.

$$F1\ score = 2*(Recall * Precision) / (Recall + Precision) \dots\dots\dots(3)$$

Presisi (pada persamaan (4)) adalah rasio dari dataset yang berlabel benar diprediksi benar dibandingkan dengan total dataset prediksi secara benar. Presisi tinggi berkaitan dengan rendahnya angka *low positive*.

$$Presisi = TP/TP+FP \dots\dots\dots(4)$$

Recall (Sensitivity) adalah kebalikan dari presisi, recall diformukan oleh persamaan 5.

$$Recall = TP/TP+FN \dots\dots\dots(5)$$

2.3. Semi-Supervised Learning Architecture

Penelitian ini mengajukan arsitektur model pada Fig 1. Arsitektur model pada Fig.1 menjelaskan langkah-langkah SSA. Proses SSA dimulai dengan membaca dataset beranotasi sebagai data training, data testing, dan unlabelled data (pada box berwarna abu-abu dalam Fig 1). Training data (yang sudah dalam kondisi bersih) akan diubah menjadi vektor menggunakan metode BoW. Hasil vektorisasi BoW ada tiga buah yaitu : *unigram*, *bigram*, dan *trigram*. Gram pada n-gram yang dimaksud adalah kata atau word. Ketiga vektor tersebut digunakan untuk membentuk tiga model menggunakan LogisticRegression (LR) dan Naïve Bayes (NB) sebagai pembanding pada eksperimen berikutnya. Ketiga model bekerja secara terpisah (menggunakan prinsip *stacking*). Ketiga model digunakan untuk menganotasi *unlabeled data*. *Unlabeled data* juga diubah menjadi vektor menggunakan BoW. *Unlabeled data* akan dianotasi oleh setiap model. *Pseudo* label yang dihasilkan berwujud tiga buah dokumen yang terklasifikasi masing-masing oleh tiga model tersebut. Sebuah label dianggap sebagai label dengan nilai *confidence* tinggi jika didukung oleh jumlah bobot dibagi total bobot dari beberapa model. Nilai *confidence* akan dibandingkan dengan *threshold*. *Threshold* digunakan untuk memilih manakah *annotated data* (dengan *pseudo-labels*) yang

bernilai *confidence* tinggi yang layak diubah statusnya menjadi bagian dari training data. Dokumen dengan nilai *confidence* tinggi akan digabungkan dengan *data training*.

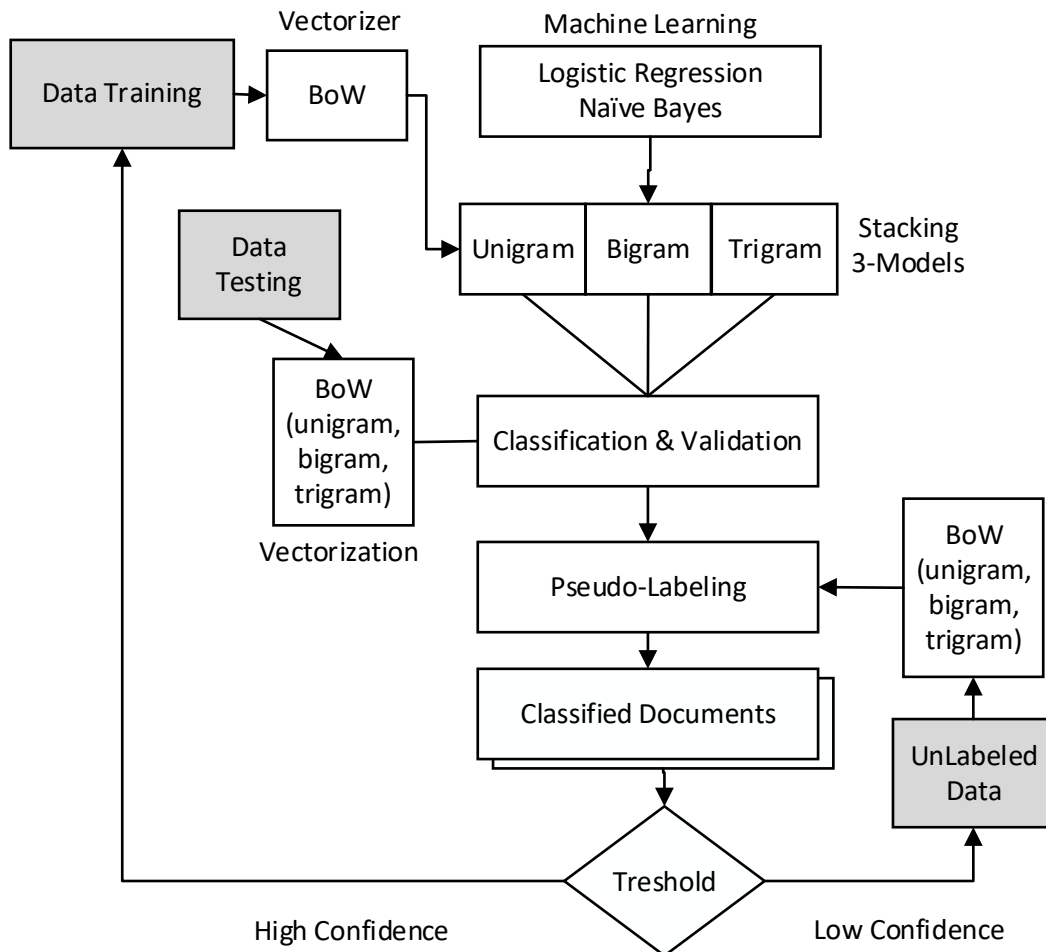


Fig.1. Arsitektur SSA-Model

SSA-Model akan diulang sebanyak 5 kali atau sampai semua *unlabeled-data* dikonversi menjadi *data training*. Hasil klasifikasi dokumen belum berlabel. *Output* dari model adalah *data training* (DT) yang telah diberi anotasi oleh mesin dan pakar.

IV. RESULT AND DISCUSSION

SSA-Model diuji menggunakan dataset IMDB. Dataset ini sering digunakan untuk pengujian model *sentiment analysis* seperti pada penelitian [15], [16], dan [17]. Percobaan pengujian model, empat kondisi IMDB dataset disiapkan dan diberi nama : Dataset1, Dataset2, Dataset3, dan Dataset4. Pada tahap pemodelan, data dibagi menjadi *data training* dan *data testing* dengan perbandingan 8:2. Jumlah *data testing* berlabel pada Dataset1, Dataset2, Dataset3, and Dataset4 adalah 10000 (20% dari keseluruhan IMDB data). Jumlah *data training* berlabel (*annotated dataset*) pada Dataset1, Dataset2, Dataset3, dan Dataset4 adalah 625, 1250, 2500, dan 5000. Sisanya digunakan sebagai *unlabeled dataset* (atau sebagai *unannotated dataset*). Seperti halnya pada penelitian sebelumnya[5], model *baseline* pada percobaan dengan Dataset1, Dataset2, Dataset3, dan Dataset4 dilatih menggunakan *data training* tanpa *pseudo-label*. Kemudian model *baseline* model juga diuji menggunakan data uji berlabel. Hasil pengukuran akurasi ada pada Tabel 2. Pengujian menggunakan pengukuran F1 score ada pada Tabel 3.

TABEL 2.
AKURASI SSA MODELS PADA DATASET IMDB

Experiment		Accuracy					
		Naïve Bayes			Logistic Regression		
No	Treshold	Baseline	SSA	Diff	Baseline	SSA	Diff
Dataset1	60%	0,87	0,82	-0,05	0,87	0,82	-0,05
	70%	0,87	0,82	-0,05	0,87	0,83	-0,04
	80%	0,85	0,75	-0,1	0,86	0,8	-0,06
	90%	0,85	0,75	-0,1	0,86	0,79	-0,07
Dataset2	60%	0,81	0,8	-0,01	0,83	0,81	-0,02
	70%	0,81	0,75	-0,06	0,82	0,8	-0,02
	80%	0,81	0,77	-0,04	0,82	0,76	-0,06
	90%	0,81	0,77	-0,04	0,82	0,75	-0,07
Dataset3	60%	0,83	0,76	-0,07	0,84	0,79	-0,05
	70%	0,83	0,75	-0,08	0,83	0,79	-0,04
	80%	0,83	0,7	-0,13	0,83	0,73	-0,1
	90%	0,83	0,7	-0,13	0,82	0,72	-0,1
Dataset4	60%	0,82	0,79	-0,03	0,83	0,81	-0,02
	70%	0,82	0,79	-0,03	0,83	0,81	-0,02
	80%	0,82	0,71	-0,11	0,82	0,78	-0,04
	90%	0,82	0,71	-0,11	0,82	0,77	-0,05
Average		0,8300	0,7588		0,8356	0,7850	

Tabel 2 menunjukkan akurasi dari model *baseline* dan *semi-supervised annotation (SSA)* model pada pengujian Dataset1, Dataset2, Dataset3, dan Dataset4, pada beberapa angka threshold yang berbeda. Tabel 2 menampilkan 64 buah SSA-model pada IMDB dataset, dan memberikan hasil yang berbeda dari penelitian sebelumnya [5]. Pertama, pada model klasifikasi *baseline* menunjukkan akurasi dan *F1 score* tidak terkait dengan jumlah *data training*.

TABEL 3.
F1-SCORE SSA MODELS PADA DATASET IMDB

Experiment		F1-Score					
		Naïve Bayes			Logistic Regression		
No	Threshold	Baseline	SSA	Diff	Baseline	SSA	Diff
Dataset1	60%	0,87	0,82	-0,05	0,86	0,83	-0,03
	70%	0,87	0,82	-0,05	0,86	0,83	-0,03
	80%	0,85	0,75	-0,1	0,86	0,76	-0,1
	90%	0,85	0,75	-0,1	0,85	0,76	-0,09
Dataset2	60%	0,81	0,79	-0,02	0,83	0,76	-0,07
	70%	0,81	0,75	-0,06	0,82	0,79	-0,03
	80%	0,81	0,77	-0,04	0,82	0,75	-0,07
	90%	0,81	0,77	-0,04	0,82	0,71	-0,11
Dataset3	60%	0,83	0,75	-0,08	0,84	0,77	-0,07
	70%	0,83	0,75	-0,08	0,83	0,79	-0,04
	80%	0,83	0,69	-0,14	0,83	0,76	-0,07

Experiment		F1-Score					
		Naïve Bayes			Logistic Regression		
No	Threshold	Baseline	SSA	Diff	Baseline	SSA	Diff
	90%	0,83	0,7	-0,13	0,83	0,75	-0,08
Dataset4	60%	0,82	0,79	-0,03	0,84	0,8	-0,04
	70%	0,82	0,79	-0,03	0,83	0,8	-0,03
	80%	0,82	0,71	-0,11	0,83	0,82	-0,01
	90%	0,82	0,7	-0,12	0,83	0,81	-0,02
	Average	0,8319	0,7563		0,8363	0,7806	

Pada percobaan Tabel 3 menggunakan Dataset2, akurasi dan *F1 score* lebih rendah daripada Dataset3 dan Dataset4. Akurasi dan *F1 score* pada kondisi model *baseline* menggunakan Logistic Regression lebih tinggi daripada Naïve Bayes. Kedua, hasil dari *semi-supervised annotation* model menunjukkan akurasi dan *F1score* linier dengan jumlah *data training*. Tapi linieritasnya berkebalikan dengan angka *threshold*. Akurasi SSA dipengaruhi angka *threshold*. Angka *threshold* yang rendah memberikan akurasi dan *F1 score* tinggi karena *threshold* rendah menghasilkan pseudo-labeled dataset lebih banyak daripada *threshold* tinggi. Secara umum akurasi dan *F1 score* dari SSA Logistic Regression lebih tinggi daripada Naïve Bayes model. Ketiga, perbedaan antara rata-rata akurasi dari model *baseline* dan rata-rata akurasi dari SSA-Model Naïve Bayes hampir sama dengan Logistic Regression. Perbedaan antara F1 score kondisi *baseline* dengan rata-rata *F1 score* pada SSA-Model di Logistic Regression lebih tinggi daripada Naïve Bayes pada SSA model. Hal ini menunjukkan bahwa pada IMBD dataset, Logistic Regression lebih baik dalam menjaga akurasi proses SSA daripada Naive Bayes.

V. Conclusion

Penelitian ini adalah tentang pengembangan model *semi-supervised sentimen annotation*(SSA) dengan tujuan untuk memberikan label secara semi otomatis. Penelitian ini menerapkan BoW untuk membangun vektor dataset dan N-Gram (N=1,2,3) serta Logistic Regression dan Naïve Bayes. Kesimpulan pertama, akurasi anotasi semi otomatis (semi supervised annotation) pada klasifikasi sentimen sangat bergantung pada jumlah data latih berlabel yang digunakan. Kesimpulan kedua adalah akurasi SSA sangat tergantung pada kecocokan dataset dengan machine learning yang dipakai. Pada dataset IMDB yang terdiri atas dua kelas, kinerja Logistic Regression lebih baik daripada Naïve Bayes. Saran untuk penelitian berikutnya adalah melakukan percobaan pada dataset yang lain dan beberapa jenis machine learning yang lain atau deep learning.

REFERENCES

- [1] D. F. Budiono, A. S. Nugroho, and A. Doewes, "Twitter sentiment analysis of DKI Jakarta's gubernatorial election 2017 with predictive and descriptive approaches," *Proceedings - 2017 International Conference on Computer, Control, Informatics and its Applications: Emerging Trends In Computational Science and Engineering, IC3INA 2017*, vol. 2018-Janua, pp. 89–94, 2017, doi: 10.1109/IC3INA.2017.8251746.
- [2] A. H. Abdulhafiz, "Novel opinion mining system for movie reviews in Turkish," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 8, no. 2, pp. 94–101, 2020, doi: 10.18201/ijisae.2020261590.
- [3] H. Imaduddin, Widyawan, and S. Fauziati, "Word Embedding Comparison For Indonesian Language Sentiment Analysis," *Proceeding - 2019 International Conference of Artificial Intelligence and Information Technology, ICAIIT 2019*, pp. 426–430, 2019, doi: 10.1109/ICAIIIT.2019.8834536.
- [4] R. Monika, S. Deivalakshmi, and B. Janet, "Sentiment Analysis of US Airlines Tweets Using LSTM/RNN," *Proceedings of the 2019 IEEE 9th International Conference on Advanced Computing, IACC 2019*, pp. 92–95, 2019, doi: 10.1109/IACC48062.2019.8971592.
- [5] A. S. Aribowo, H. Basiron, and N. F. A. Yusof, "Semi-supervised learning for sentiment classification with ensemble multi-classifier approach," *International Journal of Advances in Intelligent Informatics*, vol. 8, no. 3, pp. 349–361, 2022, doi: 10.26555/ijain.v8i3.929.
- [6] N. H. Cahyana, S. Saifullah, Y. Fauziah, A. S. Aribowo, and R. Drezewski, "Semi-supervised Text Annotation for Hate Speech Detection using K-Nearest Neighbors and Term Frequency-Inverse Document Frequency," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 10, pp. 147–151, 2022, doi: 10.14569/ijacsa.2022.0131020.
- [7] W. Wisnalmawati, A. S. Aribowo, and Y. Herawati, "Semi-supervised Learning Models for Sentiment Analysis on Marketplace Dataset," *International Journal of Artificial Intelligence & Robotics (IJAIR)*, vol. 4, no. 2, pp. 78–85, 2022, doi: 10.25139/ijair.v4i2.5267.
- [8] V. L. Shan Lee, K. H. Gan, T. P. Tan, and R. Abdullah, "Semi-supervised Learning for Sentiment Classification using Small Number of Labeled Data," *Procedia Computer Science*, vol. 161, pp. 577–584, 2019, doi: 10.1016/j.procs.2019.11.159.

- [9] A. Al-Laith, M. Shahbaz, H. F. Alaskar, and A. Rehmat, "Arasencorpus: A semi-supervised approach for sentiment annotation of a large arabic text corpus," *Applied Sciences (Switzerland)*, vol. 11, no. 5, 2021, doi: 10.3390/app11052434.
- [10] V. Balakrishnan, P. Y. Lok, and H. Abdul Rahim, "A semi-supervised approach in detecting sentiment and emotion based on digital payment reviews," *Journal of Supercomputing*, vol. 77, no. 4, pp. 3795–3810, 2021, doi: 10.1007/s11227-020-03412-w.
- [11] R. Alahmary and H. Al-Dossari, "A semiautomatic annotation approach for sentiment analysis," *Journal of Information Science*, 2021, doi: 10.1177/01655515211006594.
- [12] R. A. Maisal, A. N. Hidayanto, N. F. Ayuning Budi, Z. Abidin, and A. Purbasari, "Analysis of sentiments on Indonesian YouTube video comments: case study of the Indonesian government's plan to move the capital city," in *1st International Conference on Informatics, Multimedia, Cyber and Information System*, 2019, pp. 121–124, doi: 10.1109/ICIMCIS48181.2019.8985228.
- [13] A. N. Muhammad, S. Bukhori, and P. Pandunata, "Sentiment analysis of positive and negative of YouTube comments using naïve bayes-support vector machine (NBSVM) classifier," in *International Conference on Computer Science, Information Technology, and Electrical Engineering*, 2019, vol. 1, pp. 199–205, doi: 10.1109/ICOMITEE.2019.8920923.
- [14] R. Novendri, A. S. Callista, D. N. Pratama, and C. E. Puspita, "Sentiment analysis of YouTube movie trailer comments using naïve bayes," *Bulletin of Computer Science and Electrical Engineering*, vol. 1, no. 1, pp. 26–32, 2020, doi: 10.25008/bcsee.v1i1.5.
- [15] R. S. Kumar, A. F. Saviour Devaraj, M. Rajeswari, E. G. Julie, Y. H. Robinson, and V. Shanmuganathan, "Exploration of sentiment analysis and legitimate artistry for opinion mining," *Multimedia Tools and Applications*, 2021, doi: 10.1007/s11042-020-10480-w.
- [16] V. Kumar and B. Subba, "A Tf-idf vectorizer and SVM based sentiment analysis framework for text data corpus," *26th National Conference on Communications, NCC 2020*, pp. 1–6, 2020, doi: 10.1109/NCC48643.2020.9056085.
- [17] Y. Pan, Z. Chen, Y. Suzuki, F. Fukumoto, and H. Nishizaki, "Sentiment analysis using semi-supervised learning with few labeled data," *Proceedings - 2020 International Conference on Cyberworlds, CW 2020*, pp. 231–234, 2020, doi: 10.1109/CW49994.2020.00044.