

OPTIMASI MODEL SEMI-SUPERVISED LEARNING DENGAN SVM DAN NAÏVE BAYES

Muhammad Siddik Hasibuan^{1*}, Yusuf Ramadhan Nasution²

^{1,2}Imu Komputer, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sumatera Utara
muhammadsiddik@uinsu.ac.id¹, ramadhannst@uinsu.ac.id²

Submitted July 15, 2024; Revised November 29, 2024; Accepted November 30, 2024

Abstrak

Penentuan label dalam text mining menjadi hal terpenting, selain menggunakan *lexicon based* dapat juga menggunakan pemahaman manusia untuk dapat mengartikan kalimat kedalam kategori positif, negatif dan netral. Analisis sentimen untuk mengukur suatu opini yang diambil dari tweet untuk dapat menganalisis *text* tersebut peneliti melakukan proses *learning* dan pengujian data. Fokus penelitian ini adalah untuk dapat menentukan polarisasi *text* kedalam bentuk positif, negatif dan netral dengan model pembelajaran mesin semi *supervised learning*. Pada proses training data sebelumnya dilakukan labelisasi dengan pemahaman manusia untuk mendapatkan label positif, negatif dan netral. selanjutnya pada proses pengujian data tidak dilabeli, peran mesin learning tersebut agar data pengujian mendapatkan label. Teknik *semi-supervised learning* (SSL) digunakan untuk memberi label bagi data yang belum terlabel dengan Algoritma yang digunakan memproses data pelatihan tersebut menggunakan SVM dan NB. Evaluasi statistik yang digunakan adalah *cross-validation* dan untuk mengukur tingkat akurasi dua algoritma tersebut menggunakan *confusion matrix*. SVM mendapat nilai akurasi yang tinggi dalam penelitian ini dibanding dengan NB, SVM mendapat akurasi 88.97% dan NB 83.02%

Kata Kunci : *semi-supervised learning*, SVM, NB

Abstract

Determining labels in text mining is the most important thing, apart from using lexicon based, you can also use human understanding to interpret sentences into positive, negative and neutral categories. Sentiment analysis is to measure an opinion taken from tweets to be able to analyze the text. Researchers carry out a learning process and data testing. The focus of this research is to be able to determine the polarization of text into positive, negative and neutral forms using a semi-supervised machine learning model. In the previous data training process, labeling was carried out using human understanding to obtain positive, negative and neutral labels. Next, in the testing process, the data is not labeled, the role of the learning machine is so that the test data gets a label. The semi-supervised learning (SSL) technique is used to label unlabeled data with the algorithm used to process the training data using SVM and NB. The statistical evaluation used is cross validation and to measure the level of accuracy of the two algorithms using a confusion matrix. SVM received a high accuracy score in this study compared to NB, SVM got a high accuracy score in this study compared to NB, SVM got 88.97% accuracy and NB 83.02%

Keywords : *semi-supervised learning*, SVM, NB

1. PENDAHULUAN

Saat ini, era internet telah mengubah cara pandang masyarakat mengungkapkan pandangan dan pendapatnya. Sekarang sebagian besar dilakukan melalui posting blog, forum online, situs ulasan produk, sosial media, dll [1], [2], [3]. Saat ini, jutaan orang menggunakan media sosial

situs jaringan seperti Facebook, Twitter, Google Plus, dll mengekspresikan emosi, dan pendapat mereka serta berbagi pandangan tentang mereka kehidupan sehari-hari. Media sosial menghasilkan banyak sekali data kaya sentimen dalam bentuk tweet, update status, blog postingan, komentar, review, dll [4], [5], [6], [7], [8]. Apalagi media sosial

memberikan peluang bagi bisnis dengan memberi mereka platform untuk melakukannya terhubung dengan pelanggan mereka untuk beriklan. Kebanyakan orang sangat bergantung pada konten online buatan pengguna luasnya untuk pengambilan keputusan. Selain berbisnis media sosial bisa juga tempat curahan bahkan umpatan seseorang terhadap sesuatu. Media sosial dimasa sekarang ini adalah salah satu aplikasi yang penggunaannya sangat sering dipakai dalam sehari. Menurut Ditjen Informasi dan Komunikasi publik (IKP) situs jejaring sosial yang paling banyak diakses adalah facebook dan twitter. Indonesia sendiri menepati peringkat empat pengguna facebook terbesar di dunia. Jumlah konten yang dihasilkan oleh pengguna terlalu banyak untuk dianalisis oleh pengguna normal. Jadi ada kebutuhan untuk mengotomatisasi ini, berbagai teknik analisis sentimen banyak digunakan[2], [9]. Analisis sentimen (SA) memberi tahu pengguna apakah informasi tersebut tentang produk tersebut memuaskan atau tidak sebelum mereka membelinya[10], [11]. Pemasar dan perusahaan menggunakan data analisis ini untuk memahaminya produk atau layanan mereka sedemikian rupa sehingga dapat ditawarkan sesuai kebutuhan pengguna. Teknik pengambilan informasi tekstual terutama berfokus pada mengolah, mencari atau menganalisis data faktual yang ada. Fakta mempunyai komponen obyektif, namun ada komponen lain yang objektif isi tekstual yang mengungkapkan karakteristik subjektif. Ini isinya terutama opini, sentimen, penilaian, sikap, dan emosi, yang merupakan inti dari Analisis Sentimen (SA)[12]. Ini menawarkan banyak peluang menantang untuk mengembangkan hal baru aplikasi, terutama karena pertumbuhan besar yang tersedia informasi pada sumber online seperti blog dan jejaring sosial[13]. Misalnya rekomendasi item yang diajukan oleh a sistem

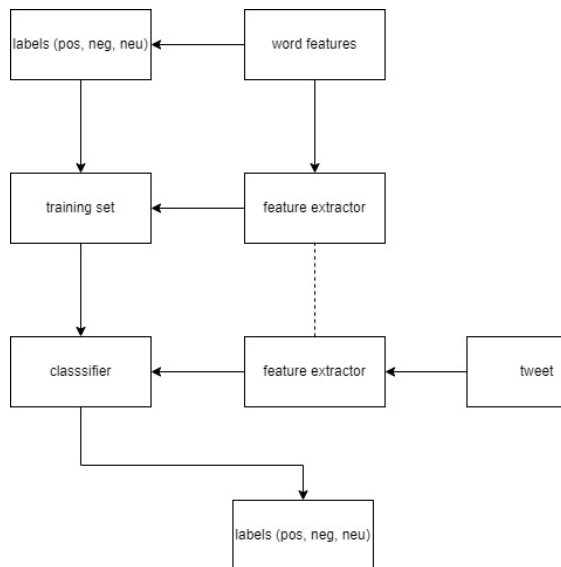
rekomendasi dapat diprediksi dengan mempertimbangkan pertimbangan akun seperti opini positif, negatif dan netral tentang item tersebut dengan memanfaatkan SA[14]. SA dalam penelitian ini digunakan untuk membaca pola-pola opini masyarakat dari media sosial twitter, yaitu mengenai kebijakan pemerintah terhadap penggunaan kendaraan listrik di Indonesia. Pemerintah menerbitkan Perpres Nomor 55 Tahun 2019 tentang percepatan kendaraan listrik berbasis baterai (KBLBB) dan memberikan subsidi insentif pembelian kendaraan listrik pada 2023. Kebijakan ini bertujuan menekan subsidi BBM dan mendorong industri kendaraan listrik, namun menuai kritik karena dinilai kurang tepat sasaran, lebih menguntungkan kalangan mampu, dan seharusnya difokuskan pada kendaraan listrik massal. Meski dianggap dapat mengurangi emisi gas rumah kaca, prioritas utama seharusnya adalah membangun pola pikir masyarakat untuk lebih ramah lingkungan.. Dari banyaknya opini yang berkembang peneliti merasa perlu untuk mempolakan opini-poini tersebut menggunakan polarisasi text untuk dapat mewakili pendapat setuju (positif) dan tidak setuju (negatif) dengan model pembelajaran mesin khususnya pada data yang belum terlabel [15].

Analisis sentimen dapat didefinisikan sebagai proses yang mengotomatisasi penggalan sikap, opini, pandangan dan emosi dari teks,ucapan, tweet, dan sumber basis data melalui *Natural Language Processing* (NLP)[11], [16], [17]. Analisis sentimen melibatkan pengklasifikasian opini dalam teks ke dalam kategori seperti “positif” atau “negatif”[18], [19], [20], [21]. Hal ini juga disebut sebagai analisis subjektivitas, penambangan opini, dan ekstraksi penilaian. Penelitian ini menggunakan teknik semi-supervised learning, semi-supervised learning (SSL)

adalah metode yang efisien untuk menambah data training secara otomatis dari data yang tidak berlabel (unlabeled data). Selain itu, perkembangan dari banyak aplikasi pengolahan bahasa (natural language app) menganggap masalah ini adalah sebuah tantangan dimana data yang tidak berlabel (unlabeled data) relatif dalam jumlah yang berlimpah sedangkan data berlabel (labeled data) jumlahnya agak terbatas [22]. Penelitian lain yang menggunakan SSL juga dilakukan dengan pendekatan multi klasifikasi untuk analisis sentimen [23]. Dari beberapa penelitian tersebut muncul untuk memberikan keterbaruan penggunaan SVM yang sudah teruji baik untuk labelisasi text yang dipasangkan dengan naïve bayes.

2. METODE PENELITIAN

Secara umum analisis sentimen memiliki arsitektur yang dapat dilihat pada gambar 1.

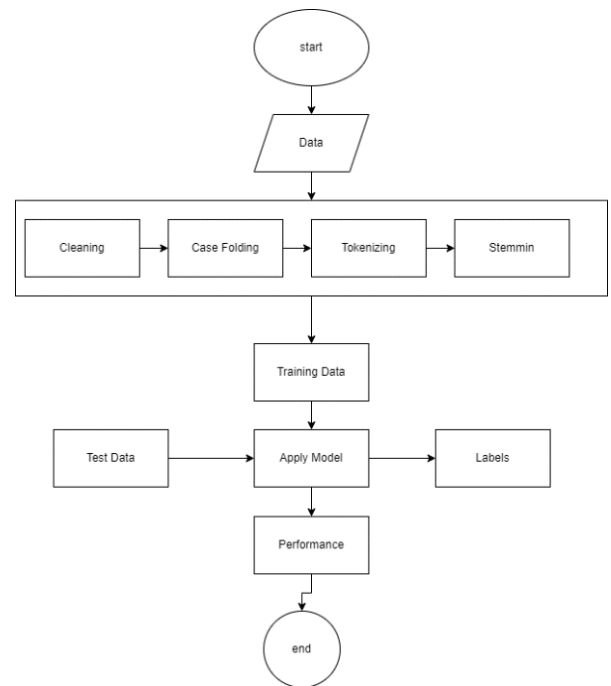


Sumber : Vaibhavi N Patodkar (2016)

Gambar 1. Arsitektur Analisis Sentimen

Secara keseluruhan dalam penelitian ini dataset yang didapat berupa data .csv yang berupa hasil crawling dari twitter berupa tweet masyarakat mengenai kendaraan listrik. Dimensi data yang didapat

sebanyak 1517 data yang belum memiliki label kedalam polaritas positif, negatif dan netral [24], [25]. Data dibagi kedalam komposisi 80:20, bagian untuk pembelajaran model atau *data training* diberi komposisi sebesar 80% sedangkan untuk data uji mendapat komposisi 20%. Data training tersebut akan dilabeling secara mengamati langsung setiap kata kedalam polaritas positif, negatif dan netral. Selanjutnya dari pembelajaran model tersebut akan diuji dengan memasukan data yang belum terpolarisasi kedalam bentuk positif, negatif dan netral. Model ini akan dklasifikasi dengan algoritma *support vector machine* (SVM)[26] dan *Naïve Bayes* (NB) rosesnya dapat diilustrasikan pada gambar 2 berikut.

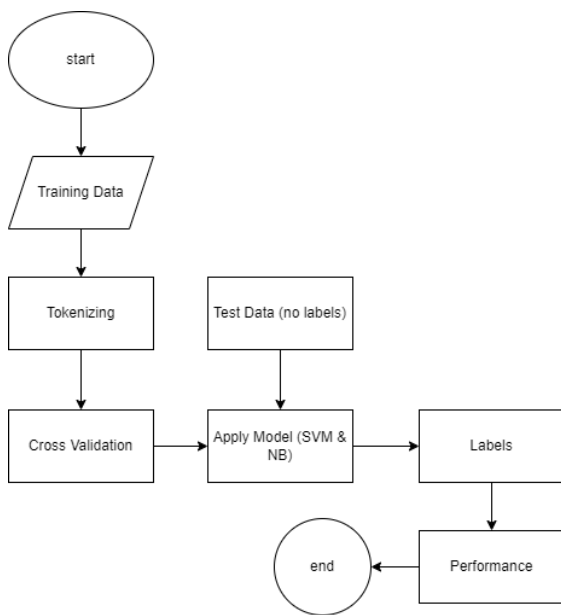


Sumber : penulis

Gambar 2. Kerangka Penelitian

Selanjutnya data uji akan mendapatkan label positif, negatif dan netral dari proses pendekatan pembelajaran mesin. Pembelajaran mesin ini akan mencari pola dari data latih berjumlah 1517 baris data yang sudah dilabel kedalam pola positif, negatif dan netral secara menyeluruh yang dilakukan oleh ahli. Untuk mendapatkan

hasil dari data uji yang sebelumnya tidak memiliki labelisasi kedalam bentuk positif/negatif tersebut, lalu akan menemukan pola baru untuk dapat diberi label tersebut. Menentukan pola ini adalah dengan pendekatan pembelajaran mesin menggunakan algoritma SVM dan NB. Karena menggunakan data yang berbeda dari data latih, peneliti ini menggunakan teknik statistik untuk menguji efektifitas model yang akan diuji. Selanjutnya untuk mendapatkan hasil maksimal agar tidak terjadi data yang overfitting dilakukan cross validation secara K-Fold dengan nilai yang ditentukan adalah 10. Adapun arsitektur untuk mendapatkan model pola positif, negatif dan netral dapat dilihat pada gambar 3.



Sumber : penulis

Gambar 3. Arsitektur model

3. HASIL DAN PEMBAHASAN

Data yang digunakan dalam penelitian ini berasal dari hasil crawling pada platform twitter, data yang diambil adalah hasil tweet sebanyak 1820 baris data, terdiri dari 1517 data latih dan 303 data uji dengan komposisi 80:20. Sebagian data ditampilkan pada tabel 1. Penentuan polarisasi pada data uji dilakukan secara

umum bahasa manusia yang mengartikan kalimat dalam dataset tersebut mengandung arti positif, negatif dan netral. Sedangkan data test atau data uji adalah data yang belum memiliki label yang nantinya akan dilakukan teknik semi supervised learning untuk mencari label terhadap data test tersebut yang berjumlah 303 data.

Tabel 1. Dataset

No	User	Text
1	Sqn Ldr	Saran Sih Bikin Harga Ionic Sama Kayak Brio Insy Alloh Laris Manis
2	lushen ace	PROBLEM SUBSIDI KUALITAS DITURUNIN HARGA DINAIKIN USAHA GITU CARI CUAN SUBSIDI SEBAB INFLASI PALING GEDE
3	Fatih Al-Ayyubi	baik kualitas kembang dulu baik kualitas motor motor pabrikan jepang www.kompas.com
4	yp office	Model jelek kwalitas buruk harga mahal croot
5	Lembur Kuring	syarat @ngaco woy anak muda blom punya rumah blom jd umkm bukan serta kur dapet ngaco sia deui sia deui nu narima subsidi teh ari aing iraha
6	Syarif Airlangga	harga motor mahal masa harga mirip motor beat kualitas bagai langit bumi
7	BajuMax	MOL KEREN YAH BERITA PLUS PADAHAL MOL LOKAL MERK BATRENYA JUTA GARANSI TAHUN ANGGAP BATRE AWET TAHUN RAKYAT INDONESIA PER TAHUN SEKALI KELUAR JUTA BUAT BELI BATRE BARU O SARAN UANG SUBSIDI KELUAR BUAT BATRE PERINTAH PUNYA WEWENANG PERTAMINA KELOLA CONTOH GAS KG

		GAS KG GAS KG PRODUK BARU MISAL BATRE V BATRE V BATRE V CABUT BAYAR PASANG GAS PNP POM SISTEM SWAP BATRE BANYAK TITIK SWAP POINT INDONESIA LUAS
8	Putut Parwoto	proses kenal produk baru butuh waktu ganti kendara bbm jadi kendara butuh waktu tehnologi baru sabar
9	Heru Prasetyo	subsidi tepat sasaran
10	jonan kick ass hole	adil rata terima subsidi jangan jangan pajak PPH msh subsidi pemerintah khianat pancasila sila lima adil seluruh rakyat indonesia mmg mau rata bagus usa subsidi sama sekali untung subsidi rata umum alih kendara bbm lebih cepat efek apa kurang konsumsi subsidi bbm kurang polusi tingkat minat pihak swasta dl ikut partisipasi investasi ekosistem motor
1517 dst		

Selanjutnya data tersebut di *cleaning* untuk menghilangkan duplikasi data serta memberiskan data yang terindikasi *missing value*. Selanjutnya tahap *case folding* adalah untuk mengubah huruf kapital menjadi huruf kecil (*lowercase*). Setelah proses *case folding* selanjutnya tahap *tokenizing* yaitu menghilangkan tanda baca, tag, emotion, dan memotong setiap kalimat kedalam sekumpulan kata untuk menjadi kata dasar. menghilangkan tag, tanda baca, simbul, lambang, link dan imbuhan untuk menjadi kata tunggal. Selanjutnya adalah proses *stemming* yaitu merubah kata kedalam bentuk dasarnya. Proses *praprosesing* telah dilakukan, kemudian melebeli data training kedalam polarisasi positif, negatif dan netral sebanyak 1517 data. Proses polarisasi ini dilakukan langsung oleh peneliti dengan

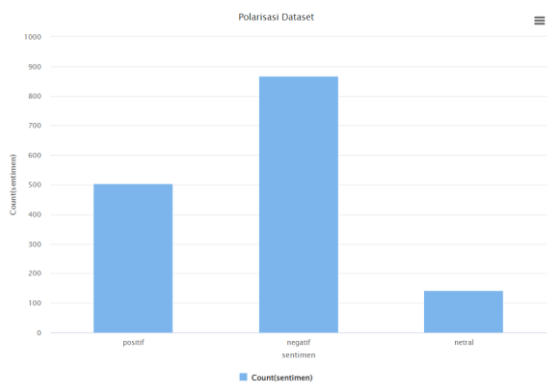
mengamati langsung setiap kata. Data training dapat dilihat pada tabel 2.

Tabel 2. Dataset setelah di praprosesing

No	Text	Labeling
1	kendaraan plg rawan bakar mudah padam	negatif
2	buat hari sih sangat guna jalan jauh sih mikir pake charging station keluar kota saudara kota buat charge banyak stasiun chrargenya jg kaya bensin tinggal isi beberapa menit langsung jalan butuh waktu sebentar kritis banget mungkin nyolokin minimarket or rumah orang	negatif
3	bapak luhut panjaitan yth baik pakai dulu dinas masyarakat lihat hasil jangan dijejeli subsidi laku pakai logika maaf masyarakat beli terimakasih	positif
4	subsidi motor jalan umum tolak system	negatif
5	aneh banget subsidi kendara pribadi pake anggar negara bayar usaha malah dorong program ada ev buat jabat ngasih duit buat kkn proyek ada ev jabat punya saham usaha sembari dapet baru pake anggar negara dapat ras duit pajak masyarakat kecil padahal transportasi publik sendiri pada cuman jakarta kota kota bantu dorong kembang layanan transportasi publik rata nih padahal ahli kota nyata bijak solid transportasi publik turun macet hingga efisiensi sistem gera kota lebih tepat waktu kendara jalan lebih efisien warga kecil nikmat akhir gera ekonomi lebih lancar akibat lain kurang polusi fasilitas dukung transport publik tingkat trotoar lebar lahan pkl rapi toko pinggir jalan hidup dl masyarakat kecil lebih banyak sempat buka usaha jalan banyak lalu transportasi publik sebab semua susah jadi	negatif

	paragraf	
6	musim hujan banjir kek bahaya pake awat ribet ganti baterai mahal	negatif
7	lama bensin laku harga lebih mahal	netral
8	omong kosong berita mantap fakta zero	negatif
..
1517	niat beli ev murah malah ikut dinaikin sama pemerintah terus rakyat jangan kasih sejahtera	negatif

Polarisasi data training terbanyak kedalam label negatif yaitu sebanyak 869 data, selanjutnya label positif mendapat hasil 504 dan netral sebanyak 144 dengan jumlah total data 1517 baris data. Hasil visualisasi data training dapat dilihat pada gambar 4.



Gambar 4. Dataset Learning

Setelah dilakukan proses tokenisasi pada data learning, selanjutnya adalah proses untuk mendapatkan hasil labelisasi pada data uji. Diharapkan proses yang dilakukan tersebut dapat membuat suatu pola yang sesuai dengan yang dilakukan oleh manusia seperti melakukan proses menentukan pola label pada data training. Menentukan pola dalam data uji menggunakan algoritma klasifikasi seperti SVM dan NB. Algoritma tersebut diharapkan dapat bekerja untuk

mengklasifikasi text kedalam pola positif, negatif dan netral . Proses ini juga untuk mengukur kinerja masing masing algoritma SVM atau NB, evaluasi yang digunakan adalah cross validation. Cross validation atau validasi silang Validasi silang mirip dengan metode subsampling acak berulang, tetapi pengambilan sampel dilakukan sedemikian rupa sehingga tidak ada dua pengujian set tumpang tindih. Dalam *k-fold cross-validation*, *learning set* yang tersedia dipartisi menjadi k subset yang terpisah dengan ukuran yang kira-kira sama. Di sini, fold mengacu pada jumlah subset yang dihasilkan. Partisi ini dilakukan secara random sampling kasus dari pembelajaran ditetapkan tanpa penggantian. Model dilatih menggunakan k-1 himpunan bagian, yang bersama-sama mewakili himpunan pelatihan. Maka modelnya adalah diterapkan ke subset yang tersisa, yang dilambangkan sebagai set validasi, dan kinerjanya diukur. Prosedur ini diulangi sampai masing-masing subset k berfungsi sebagai set validasi, dalam penelitian ini K-Fold yang dipakai adalah 10. Adapun hasil pengujian model dengan algoritma SVM dan NB dapat dilihat dibawah ini.

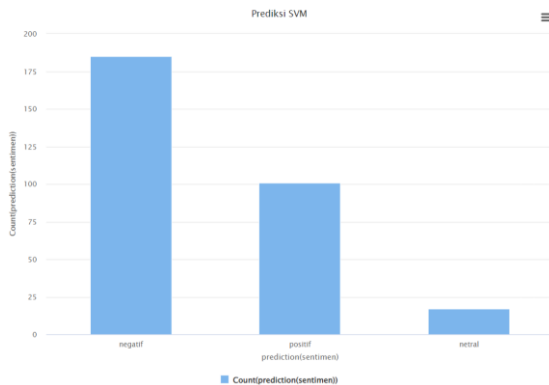
Pengujian SVM

Pengujian dilakukan untuk menentukan label data test dengan algoritma SVM, menggunakan k-fold 10, didapat dilihat pada tabel 3 prediksi dibawah ini.

Tabel. 3 Cofusion Matrix SVM

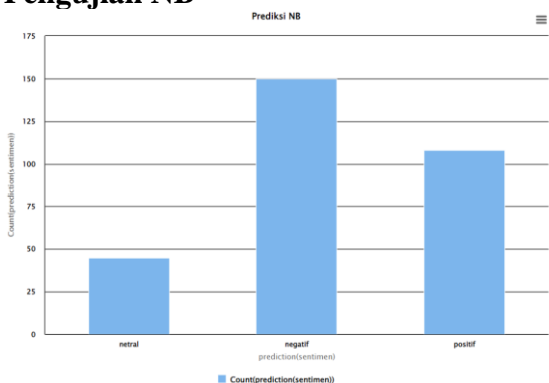
	true positif	true negatif	true netral	class precision
pred. positif	31	3	4	81.58%
pred. negatif	473	863	140	58.47%
pred. netral	0	2	0	0.00%
class recall	6.15%	99.42%	0.00%	

accuracy: 88.97% +/- 1.31% (micro average: 88.97%). Dari pengujian SMV label polarisasi text kedalam label negatif sebesar 158, positif 101 dan netral 17.



Gambar 5. Hasil Pengujian SVM

Pengujian NB



Gambar 6. Hasil Pengujian NB

accuracy: 83.02% +/- 4.07% (micro average: 83.03%)

	true positif	true negatif	true netral	class precision
pred. positif	206	224	42	43.64%
pred. negatif	224	559	63	66.08%
pred. netral	74	85	39	19.70%
class recall	40.87%	64.40%	27.08%	

Dari pengujian SMV label polarisasi text kedalam label negatif sebesar 150, positif 108 dan netral 45.

Hasil dari pengujian dapat dilihat pada tabel 5

Tabel 5. Pengujian Confusion Matrix

Model	Akurasi	Recall	Precision
SVM	88.97%	99.42%	81.58%
NB	83.02%	64.40%	66.08%

4. SIMPULAN

Pengujian yang telah dilihat dapat diketahui SVM memiliki akurasi 88.97% sedangkan NB 83.02% dapat disimpulkan bahwa dalam penelitian ini SVM memiliki tingkat akurasi yang lebih baik dari NB. Model pembelajaran mesin untuk mencari polarisasi label text kedalam bentuk pola positif, negatif dan netral adalah algoritma SVM mendapatkan nilai akurasi lebih baik dari NB. label negatif yaitu sebanyak 869 data, selanjutnya label positif mendapat hasil 504 dan netral sebanyak 144 dengan jumlah total data 1517. Dapat disimpulkan untuk penelitian ini opini masyarakat terhadap kebijakan kendaraan listrik mendapat respon negatif.

UCAPAN TERIMAKASIH

Terima kasih kepada Fakultas Sain dan Teknologi UIN Sumater Utara atas hibah penelitian Program Studi Ilmu Komputer tahun 2023.

DAFTAR PUSTAKA

- [1] K. Arun and A. Srinagesh, "Multi-lingual Twitter sentiment analysis using machine learning," *International Journal of Electrical and Computer Engineering*, vol. 10, no. 6, 2020, doi: 10.11591/ijece.v10i6.pp5992-6000.
- [2] V. A. and S. S. Sonawane, "Sentiment Analysis of Twitter Data: A Survey of Techniques," *Int*

- J Comput Appl*, vol. 139, no. 11, 2016, doi: 10.5120/ijca2016908625.
- [3] M. J. C. Samonte, J. M. R. Garcia, V. J. L. Lucero, and S. C. B. Santos, "Sentiment and opinion analysis on twitter about local airlines," in *ACM International Conference Proceeding Series*, 2017. doi: 10.1145/3162957.3163029.
- [4] R. Syahputra, G. J. Yanris, and D. Irmayani, "SVM and Naïve Bayes Algorithm Comparison for User Sentiment Analysis on Twitter," *Sinkron*, vol. 7, no. 2, 2022, doi: 10.33395/sinkron.v7i2.11430.
- [5] G. K. Shahi, A. Dirkson, and T. A. Majchrzak, "An exploratory study of COVID-19 misinformation on Twitter," *Online Soc Netw Media*, vol. 22, 2021, doi: 10.1016/j.osnem.2020.100104.
- [6] A. Giachanou and F. Crestani, "Like it or not: A survey of Twitter sentiment analysis methods," 2016. doi: 10.1145/2938640.
- [7] A. P. Nardilasari, A. L. Hananto, S. S. Hilabi, T. Tukino, and B. Priyatna, "Analisis Sentimen Calon Presiden 2024 Menggunakan Algoritma SVM Pada Media Sosial Twitter," *JOINTECS (Journal of Information Technology and Computer Science)*, vol. 8, no. 1, 2023, doi: 10.31328/jointecs.v8i1.4265.
- [8] D. S. Utami and A. Erfina, "Analisis Sentimen Pinjaman Online di Twitter Menggunakan Algoritma Support Vector Machine (SVM)," *SISMATIK (Seminar Nasional Sistem Informasi dan Manajemen Informatika)*, vol. 1, no. 1, 2021.
- [9] H. Saif, Y. He, M. Fernandez, and H. Alani, "Contextual semantics for sentiment analysis of Twitter," *Inf Process Manag*, vol. 52, no. 1, 2016, doi: 10.1016/j.ipm.2015.01.005.
- [10] L. Zhang, R. Ghosh, M. Dekhil, M. Hsu, and B. Liu, "Combining lexicon-based and learning-based methods for twitter sentiment analysis," *HP Laboratories Technical Report*, 2011.
- [11] S. Singh and A. Mahmood, "The NLP Cookbook: Modern Recipes for Transformer Based Deep Learning Architectures," *IEEE Access*, vol. 9, 2021, doi: 10.1109/ACCESS.2021.3077350.
- [12] S. H. Sahir, R. S. Ayu Ramadhana, M. F. Romadhon Marpaung, S. R. Munthe, and R. Watrianthos, "Online learning sentiment analysis during the covid-19 Indonesia pandemic using twitter data," *IOP Conf Ser Mater Sci Eng*, vol. 1156, no. 1, 2021, doi: 10.1088/1757-899x/1156/1/012011.
- [13] C. Villavicencio, J. J. Macrohon, X. A. Inbaraj, J. H. Jeng, and J. G. Hsieh, "Twitter sentiment analysis towards covid-19 vaccines in the Philippines using naïve bayes," *Information (Switzerland)*, vol. 12, no. 5, 2021, doi: 10.3390/info12050204.
- [14] S. Boon-Itt and Y. Skunkan, "Public perception of the COVID-19 pandemic on twitter: Sentiment analysis and topic modeling study," *JMIR Public Health Surveill*, vol. 6, no. 4, 2020, doi: 10.2196/21978.
- [15] A. C. Sanders *et al.*, "Unmasking the conversation on masks: Natural language processing for topical sentiment analysis of COVID-19 Twitter discourse," *AMIA Annu Symp Proc*, vol. 2021, 2021.
- [16] M. Zhou, N. Duan, S. Liu, and H. Y. Shum, "Progress in Neural NLP: Modeling, Learning, and Reasoning," 2020. doi: 10.1016/j.eng.2019.12.014.
- [17] L. Deng and Y. Liu, "Epilogue: Frontiers of NLP in the deep

- learning era,” in *Deep Learning in Natural Language Processing*, 2018. doi: 10.1007/978-981-10-5209-5_11.
- [18] L. K. Ramasamy, S. Kadry, Y. Nam, and M. N. Meqdad, “Performance analysis of sentiments in Twitter dataset using SVM models,” *International Journal of Electrical and Computer Engineering*, vol. 11, no. 3, 2021, doi: 10.11591/ijece.v11i3.pp2275-2284.
- [19] N. A. S. Abdullah and N. I. A. Rusli, “Multilingual sentiment analysis: A systematic literature review,” 2021. doi: 10.47836/pjst.29.1.25.
- [20] A. Lighthart, C. Catal, and B. Tekinerdogan, “Systematic reviews in sentiment analysis: a tertiary study,” *Artif Intell Rev*, vol. 54, no. 7, 2021, doi: 10.1007/s10462-021-09973-3.
- [21] K. L. Tan, C. P. Lee, and K. M. Lim, “A Survey of Sentiment Analysis: Approaches, Datasets, and Future Research,” 2023. doi: 10.3390/app13074550.
- [22] Z. Qiu, E. Cho, X. Ma, and W. M. Campbell, “Graph-based semi-supervised learning for natural language understanding,” in *EMNLP-IJCNLP 2019 - Graph-Based Methods for Natural Language Processing - Proceedings of the 13th Workshop*, 2019. doi: 10.18653/v1/d19-5318.
- [23] 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, 2022, doi: 10.26555/ijain.v8i3.929.
- [24] O. Y. Adwan, M. Al-Tawil, A. M. Huneiti, R. A. Shahin, A. A. Abu Zayed, and R. H. Al-Dibsi, “Twitter sentiment analysis approaches: A survey,” *International Journal of Emerging Technologies in Learning*, vol. 15, no. 15, 2020, doi: 10.3991/ijet.v15i15.14467.
- [25] S. E. Saad and J. Yang, “Twitter Sentiment Analysis Based on Ordinal Regression,” *IEEE Access*, vol. 7, 2019, doi: 10.1109/ACCESS.2019.2952127.
- [26] C. P. D. Cyril, J. R. Beulah, N. Subramani, P. Mohan, A. Harshavardhan, and D. Sivabalaselvamani, “An automated learning model for sentiment analysis and data classification of Twitter data using balanced CA-SVM,” *Concurr Eng Res Appl*, vol. 29, no. 4, 2021, doi: 10.1177/1063293X211031485.