



Program Studi Statistika
Fakultas Matematika Dan Ilmu Pengetahuan Alam
Universitas Padjadjaran



AIDA
RESEARCH CENTER FOR
ARTIFICIAL INTELLIGENCE & BIG DATA

ANALISIS SENTIMEN PENGGUNA MEDIA SOSIAL TERHADAP KEBIJAKAN PEMERINTAH DALAM PROGRAM VAKSINASI COVID-19 DI INDONESIA MENGGUNAKAN METODE *BIDIRECTIONAL GATED RECURRENT UNIT (BiGRU)*

Disusun oleh:

Dimas Ananda

NPM. 140610180039

Dosen Pembimbing:

Dr. Anindya Apriliyanti Pravitasari, S.Si., M.Si.

NIP. 19840416 200812 2 004

Dr. Intan Nurma Yulita, M.T.

NIP. 19850704 201504 2 003

Dosen Penguji:

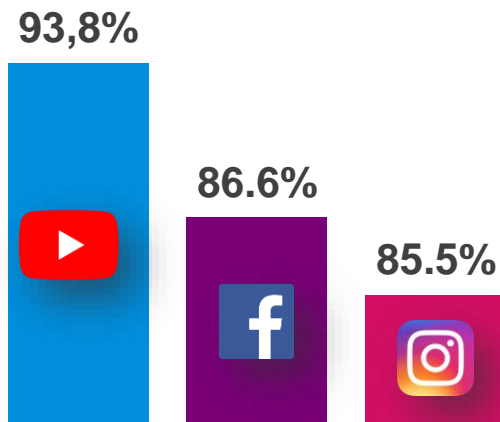
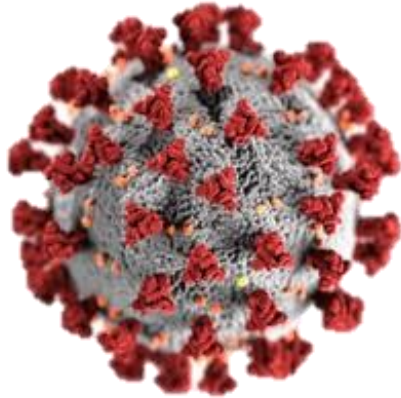
Dr. Irlandia Ginanjar, S.Si., M.Si.

NIP. 19781122 200501 1 003

Sinta Septi Pangastuti, S.Si., M.Stat.

NIP. 19930922 201903 2 023

Latar Belakang



Sumber : We Are Social, Hootsuite, 2021

Melihat pesatnya penyebaran COVID-19 dan bahaya yang akan muncul jika tidak segera ditangani, salah satu cara yang sangat mungkin untuk mencegah penyebaran virus ini adalah dengan mengembangkan **vaksin**.

(Liu C et al. 2020)

Kemunculan kebijakan –kebijakan pemerintah dalam melaksanakan program vaksinasi COVID-19 menimbulkan **respon dan opini** masyarakat di berbagai media, khususnya **Youtube, Facebook, Instagram** sebagai media sosial yang banyak digunakan oleh masyarakat Indonesia.

Analisis sentimen dapat melihat apakah pandangan atau pendapat seseorang tentang suatu isu atau objek cenderung positif, negatif, atau netral. (Rozi et al., 2012).

Kehadiran *big data* menjadi tantangan bagi pakar dalam melakukan analisis sentimen karena melimpahnya data membutuhkan pengolahan data yang tepat untuk digunakan secara komprehensif (Yudistira, 2021)

Machine learning atau pembelajaran mesin dapat mengatasi permasalahan jumlah data yang besar dan pengembangan sistem secara otomatis



Analisis Sentimen

Analisis Sentimen merupakan kajian tentang cara menyelesaikan dan memecahkan masalah **berdasarkan opini masyarakat**, sikap serta emosi suatu entitas, dimana entitas tersebut dapat mewakili individu.
(Wati, 2016)



Negative



Neutral



Positive

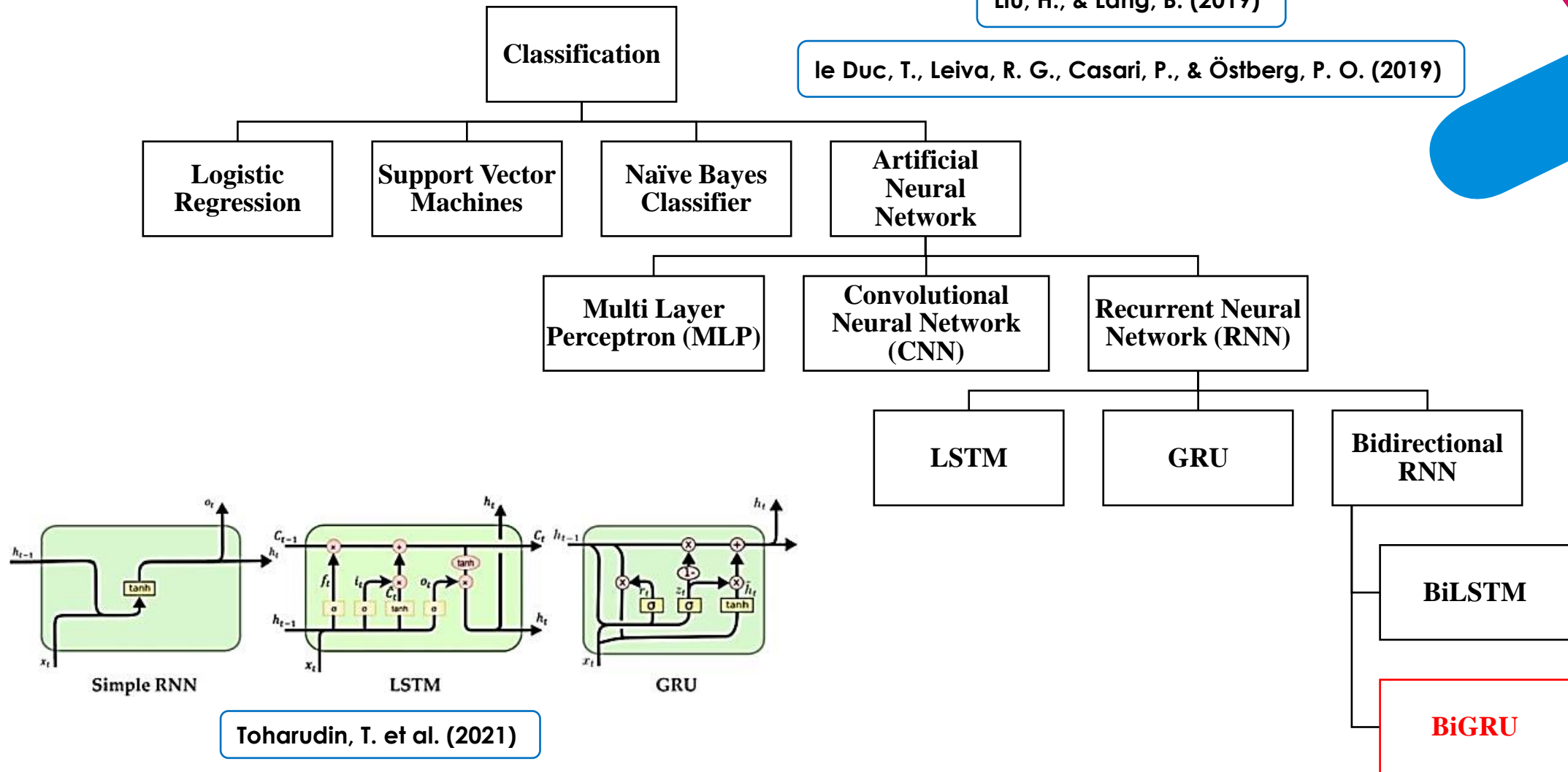
Dilakukannya analisis sentimen ini bertujuan untuk melihat pendapat atau kecenderungan opini terhadap suatu masalah ataupun objek oleh seseorang, apa memiliki kecenderungan **positif, negatif, atau netral**.
(Rozi et al., 2012)

Tinjauan Pustaka



Liu, H., & Lang, B. (2019)

le Duc, T., Leiva, R. G., Casari, P., & Östberg, P. O. (2019)

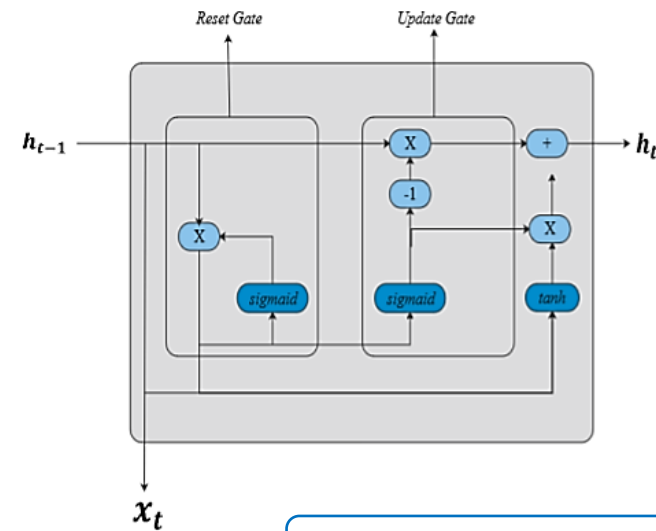


Toharudin, T. et al. (2021)

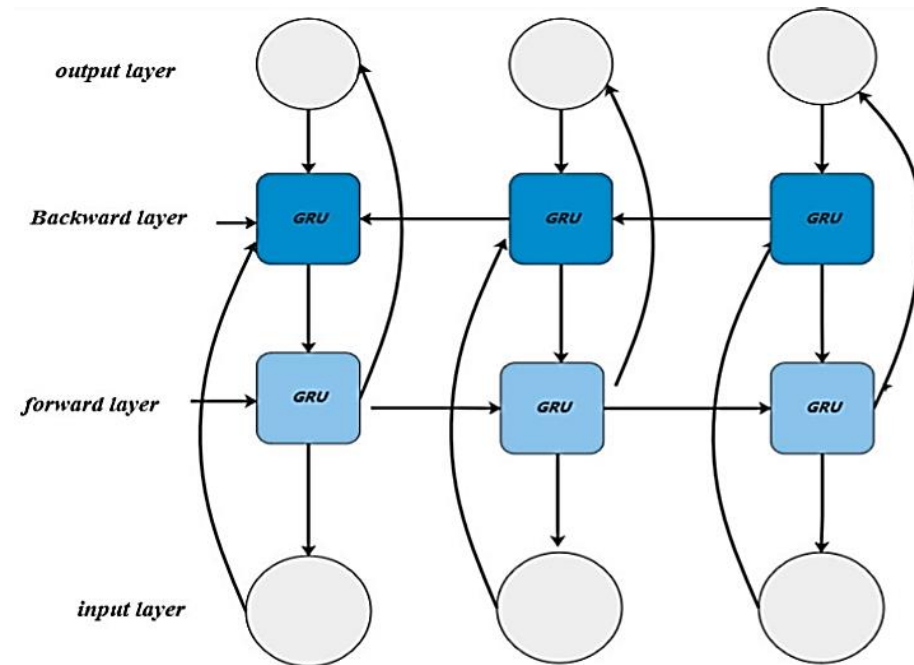


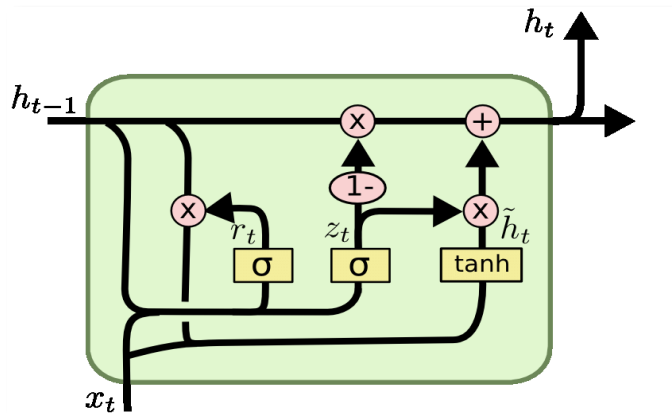
Gated Recurrent Unit (GRU) adalah pengembangan metode RNN yang cukup baru yang diusulkan oleh **K. Cho pada tahun 2014**. GRU efektif dalam tugas sentimen analisis karena kemampuannya untuk mengingat lama dependensi dan sangat berguna untuk teks berukuran besar. **(Biswas et al., 2015)**

Bidirectional Gated Recurrent Unit atau BiGRU adalah model pemrosesan urutan yang terdiri dari **dua GRU**, GRU pertama mengambil input dalam arah maju, dan yang lainnya dalam arah mundur. **(Rana, 2016)**

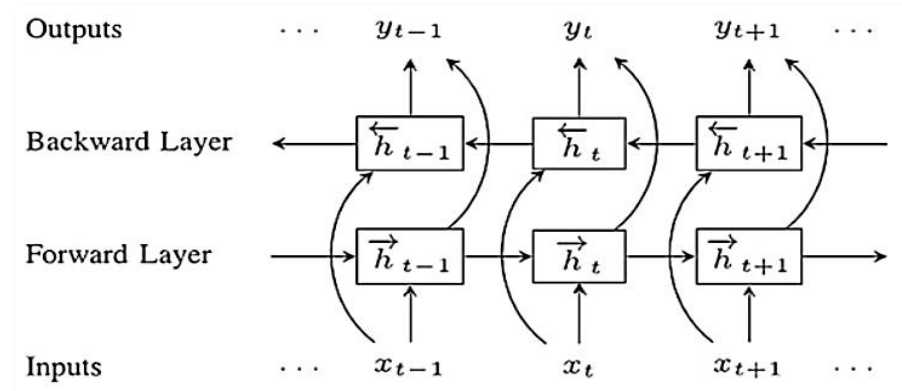


W. Ali, Y. Yang, X. Qiu, Y. Ke & Y. Wang (2021)





$$\begin{aligned} z_t &= \sigma(W_z \cdot x_t + U_z h_{t-1} + b_z) \\ r_t &= \sigma(W_r \cdot x_t + U_r h_{t-1} + b_r) \\ \hat{h}_t &= \tanh(W_h \cdot x_t + r_t \odot U_h h_{t-1} + b_h) \\ h_t &= (1 - z_t) \odot h_{t-1} + z_t \odot \hat{h}_t \end{aligned}$$



$$\begin{aligned} \vec{h}_t &= \sigma(W_{x\vec{h}} x_t + W_{\vec{h}\vec{h}} \vec{h}_{t-1} + b_{\vec{h}}) \\ \overleftarrow{h}_t &= \sigma(W_{x\overleftarrow{h}} x_t + W_{\overleftarrow{h}\overleftarrow{h}} \overleftarrow{h}_{t-1} + b_{\overleftarrow{h}}) \\ h_t &= \vec{h}_t \oplus \overleftarrow{h}_t \end{aligned}$$

Keterangan :

x_t : Input
 z_t : Update gate
 r_t : Reset gate
 \hat{h}_t : Candidate hidden state
 h_{t-1} : Hidden state

h_t : Final hidden state
 $W_z, U_z, W_r, U_r, W_h, U_h$: Parameter Bobot
 b : Parameter bias
 \odot : Perkalian elemen
 σ : Fungsi aktivasi

\vec{h}_t : Final hidden state dari GRU maju
 \overleftarrow{h}_t : Final hidden state dari GRU mundur
 \oplus : Parameter bias

Penulis/ Tanggal	Judul	Metodologi	Hasil dan Pembahasan	Keterangan
Merinda Lestandy, Abdurrahim, & Lailis Syafa'ah/ 2021	Analisis Sentimen Tweet Vaksin COVID-19 Menggunakan <i>Recurrent Neural Network</i> dan <i>Naïve Bayes</i>	RNN dan <i>Naïve Bayes Classifier</i>	Diperoleh hasil akurasi keseluruhan 97,7% untuk RNN, dan untuk <i>Naïve Bayes Classifier</i> 80%	Metode berbasis RNN memperoleh hasil akurasi lebih besar dengan perbedaan 17,7% dengan <i>Naïve Bayes Classifier</i>
Liang Zhou & Xiaoyong Bian/ 2019	<i>Improved text sentiment classification methodbased on BiGRU-Attention</i>	RNN, LSTM, GRU, BiLSTM, BiGRU, BiLSTM + Attention, BiGRU+ Attentiton	BiGRU memperoleh akurasi sebesar 88,56%, RNN 80,74%, LSTM 87,23%, GRU 88,56%, BiLSTM 88,23%, BiLSTM - Attention 89,07%, BiGRU- Attentiton 90,45%.	Metode berbasis Gated Recurrent Unit (GRU) selalu memperoleh akurasi lebih besar dibandingkan <i>simple</i> RNN dan LSTM, baik dengan <i>Bidirectional</i> maupun <i>Attention</i>
Xing Yin, Changhui Liu, & Xiaodong Fang/ 2021	<i>Sentiment analysis based on BiGRU information enhancement</i>	CNN-BiGRU, CNN-BiLSTM- Attention, BERT, BiGRU	BiGRU memperoleh akurasi sebesar 82,6%, CNN-BiGRU 73,6%, CNN-BiLSTM-Attention 74,3%, dan BERT 82,1%	Metode BiGRU dengan hyperparameter terbaik memperoleh hasil akurasi lebih besar dibandingkan BERT, dan metode Hybrid lainnya
Xuanzhen Feng1 & Xiaohong Liu / 2019	<i>Sentiment Classification of Reviews Based on BiGRU Neural Network and Fine-grained Attention</i>	SVM, Mixsupervised+BOW, MBCNN-LSTM, RNN, BiLSTM, BiGRU, Atten-BiGRU, FGAtten- BiGRU, FGANN-BiGRU	Diperoleh hasil akurasi dari SVM sebesar 81,4%, Mixsupervised+BOW 88,9%, MBCNN-LSTM 89,5%, RNN 84,9%, BiLSTM 86.7%, BiGRU 87,9%, Atten-BiGRU 88,7%, FGAtten-BiGRU 90,5%, FGANN-BiGRU 90,9%	Metode dengan unit BiGRU secara kesuluruhah mendapatkan rata-rata akurasi sebesar 89,5% dan paling tinggi mencapai 90,9%. Dimana hasil tersebut lebih baik dibandingkan metode- metode lain yang telah diuji

Confusion Matrix

Confusion Matrix		Prediksi		
		a	b	c
Aktual	a	T_{aa}	F_{ba}	F_{ca}
	b	F_{ab}	T_{bb}	F_{cb}
	c	F_{ac}	F_{bc}	T_{cc}

$$accuracy = \frac{TP}{TP + TN + FP + FN} \times 100\%$$

$$precision = \frac{TP}{TP + FP}$$

$$recall = \frac{TP}{TP + FN}$$

$$True\ Positive = T_{aa}$$

$$True\ Negative = T_{bb} + T_{cc} + F_{bc} + F_{cb}$$

$$False\ Positive = F_{ab} + F_{ac}$$

$$False\ Negative = F_{ba} + F_{ca}$$

$$F_1score = \frac{2}{recall^{-1} + precision^{-1}}$$

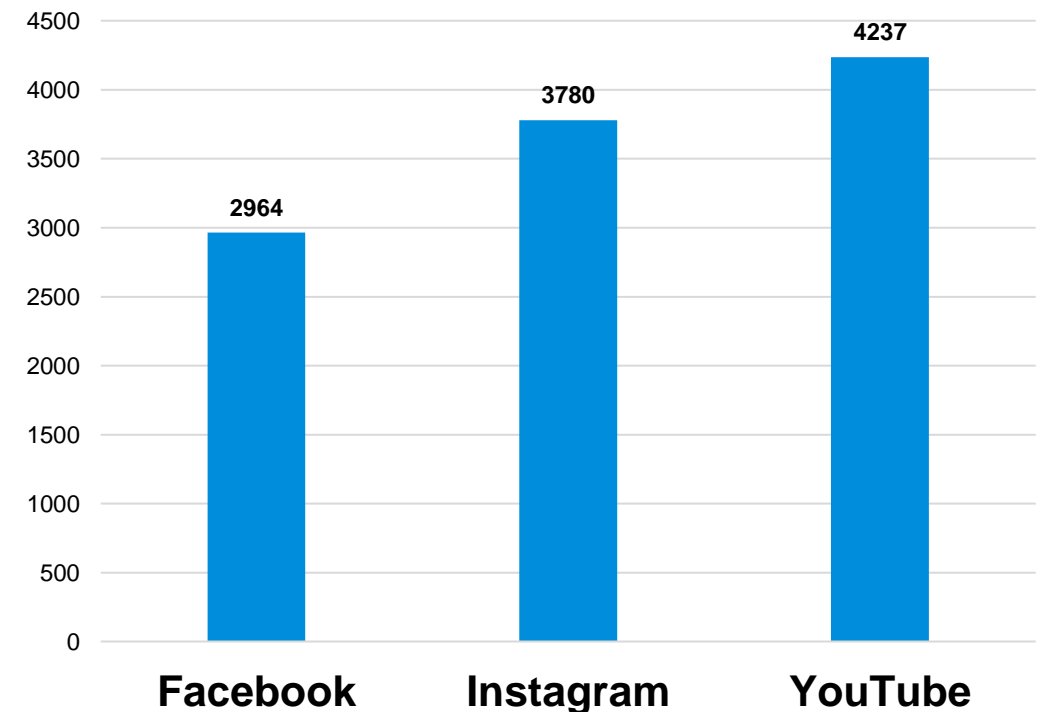
$$= 2 \times \frac{precision \times recall}{precision + recall}$$

$$= \frac{TP}{TP + \frac{1}{2}(FP + FN)}$$



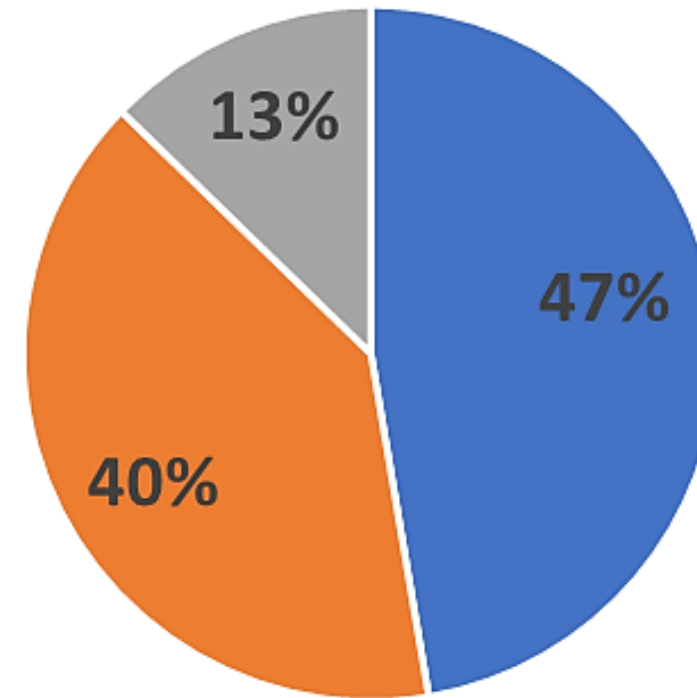
DATA PENELITIAN

Data komentar masyarakat dalam unggahan akun **YouTube, Facebook, dan Instagram** Kementerian Kesehatan Republik Indonesia yang membahas terkait **program vaksinasi COVID-19**. Dataset telah dilabeli ke dalam kelas sentimen **positif, netral, dan negatif** oleh Pusat Riset Kecerdasan Artifisial dan *Big Data* Universitas Padjadjaran.



Contoh Hasil Pelabelan Data

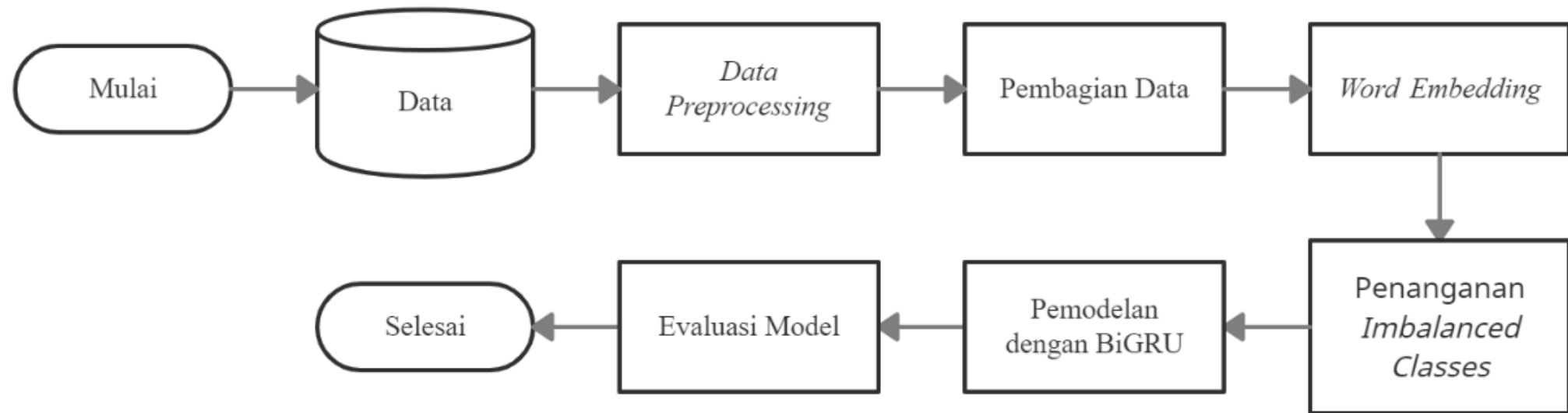
Data	Label
pemerintah sudah berusaha semaksimal mungkin memutus mata rantai virus bahkan sampai geram menangani pandemi ini tinggal kitanya sebagai masyarakat mari kita bantu pemerintah kita jangan saling menyalahkan dimulai dari diri kita sendiri jaga kesehatan	Positif
Vaksin adalah produk biologi yang berisi antigen	Netral
jangan mau divaksin	Negatif



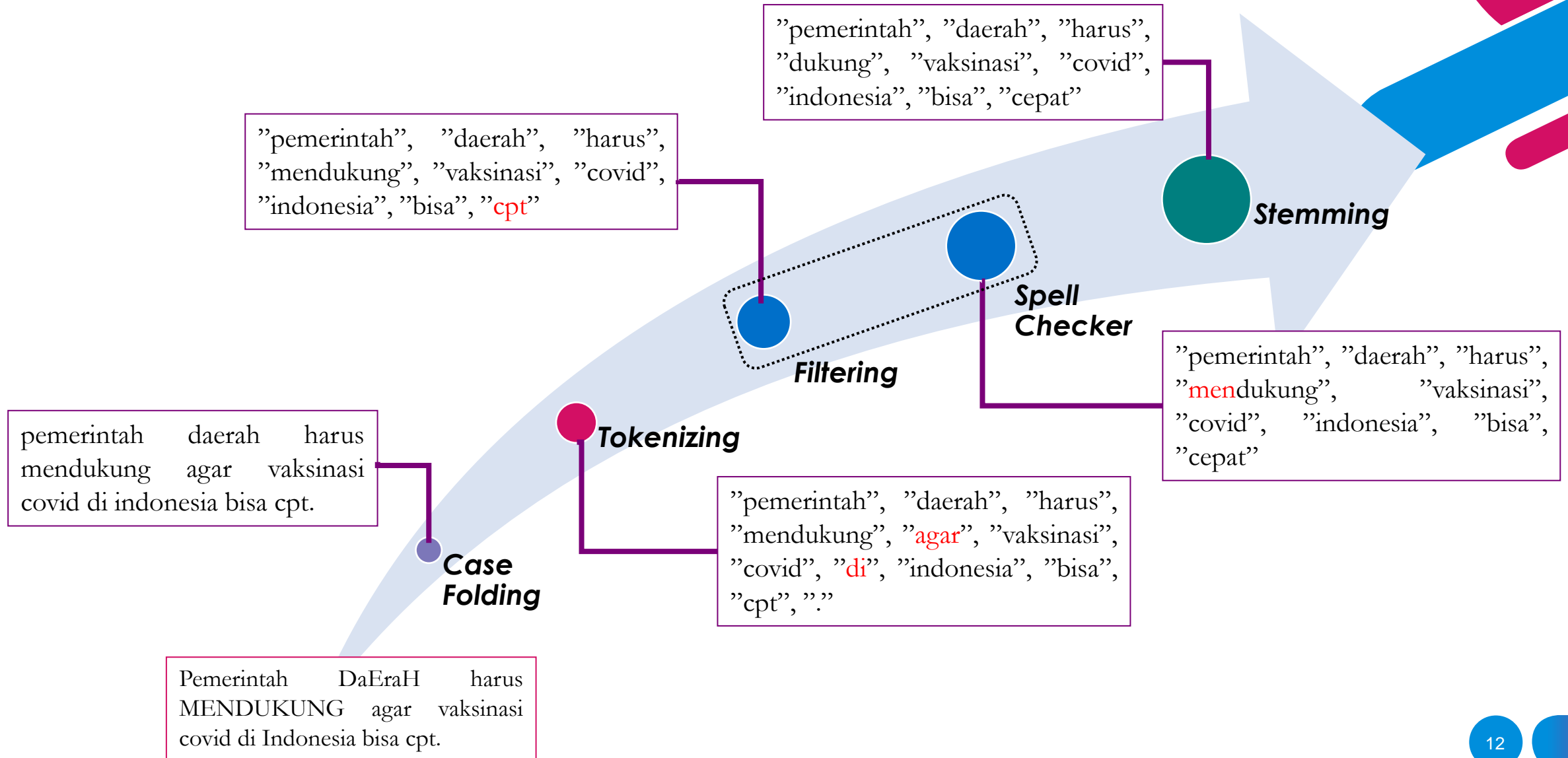
■ Negatif ■ Netral ■ Positif



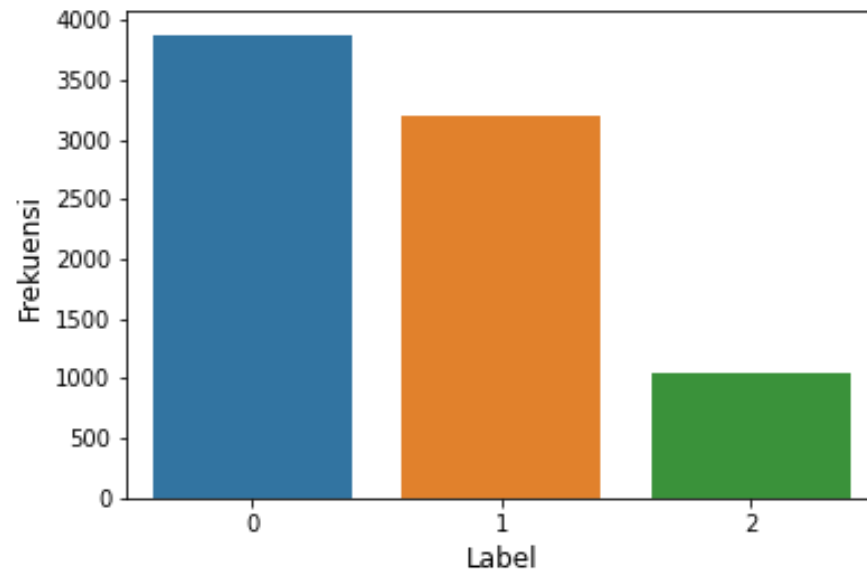
LANGKAH-LANGKAH PENELITIAN



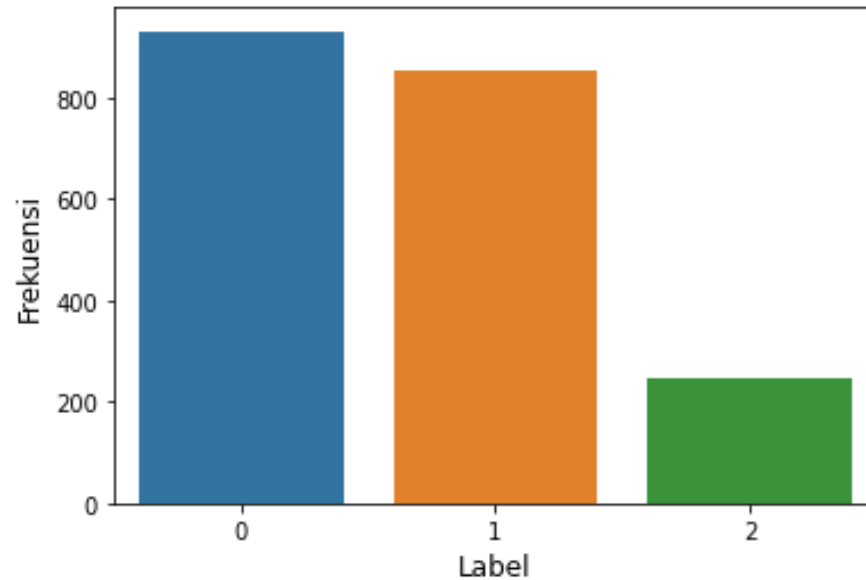
Data Preprocessing



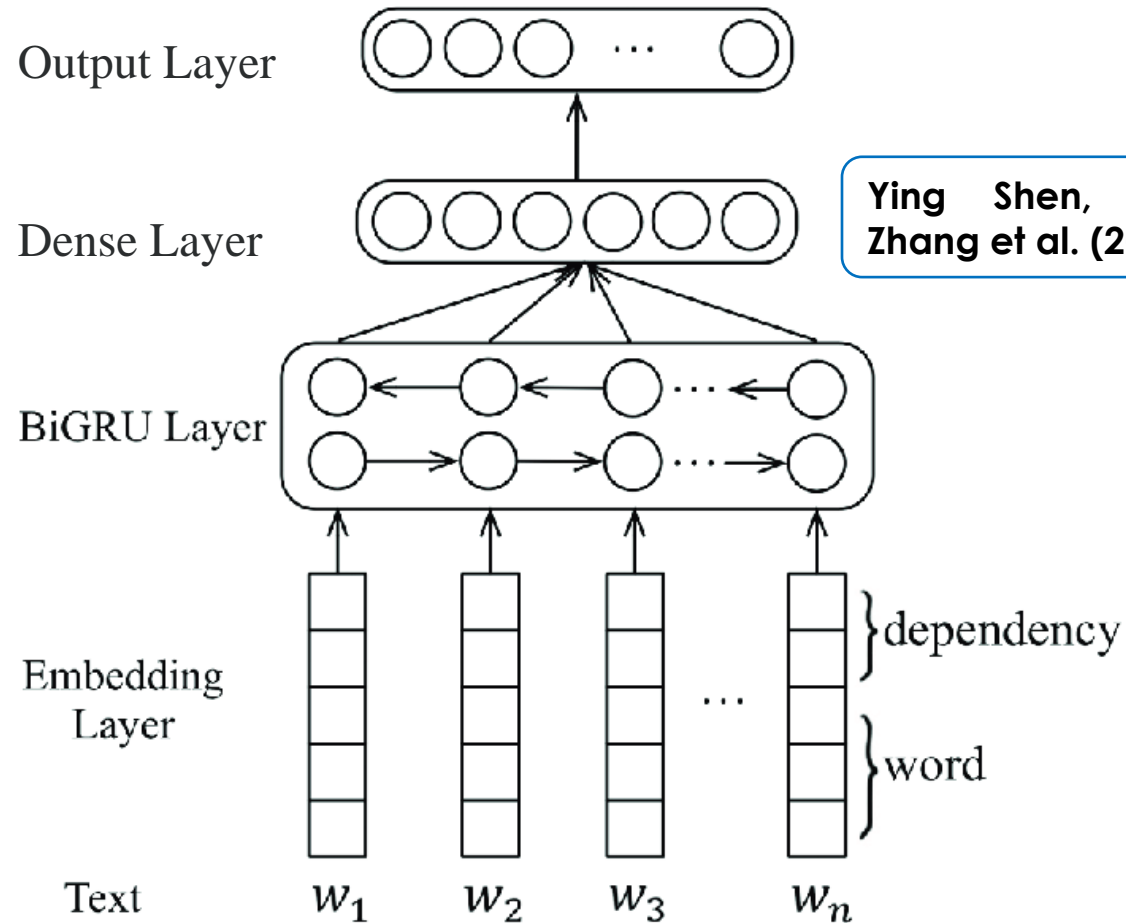
Pada tahap ini membagi data menjadi data pelatihan (*training*) dan data pengujian (*testing*). Pada penelitian ini, proporsi pembagian data yang akan digunakan adalah **80%** untuk data pelatihan dan **20%** untuk data pengujian.



a) Data Pelatihan (*training*)



b) Data Pengujian (*testing*)



Ying Shen, Lizhu Zhang et al. (2018)

Dikarenakan penelitian ini merupakan *multiple classification* yaitu sentimen positif, netral, dan negatif, maka digunakan *loss function categorical cross entropy* dan dengan fungsi optimasi menggunakan 'Adam'.

Untuk menghindari masalah utama dalam proses pelatihan yaitu *overfitting*, pada penelitian ini menggunakan fungsi *Early Stopping* yang menghentikan proses *training* apabila akan terjadi *overfitting* pada model yang dibuat.

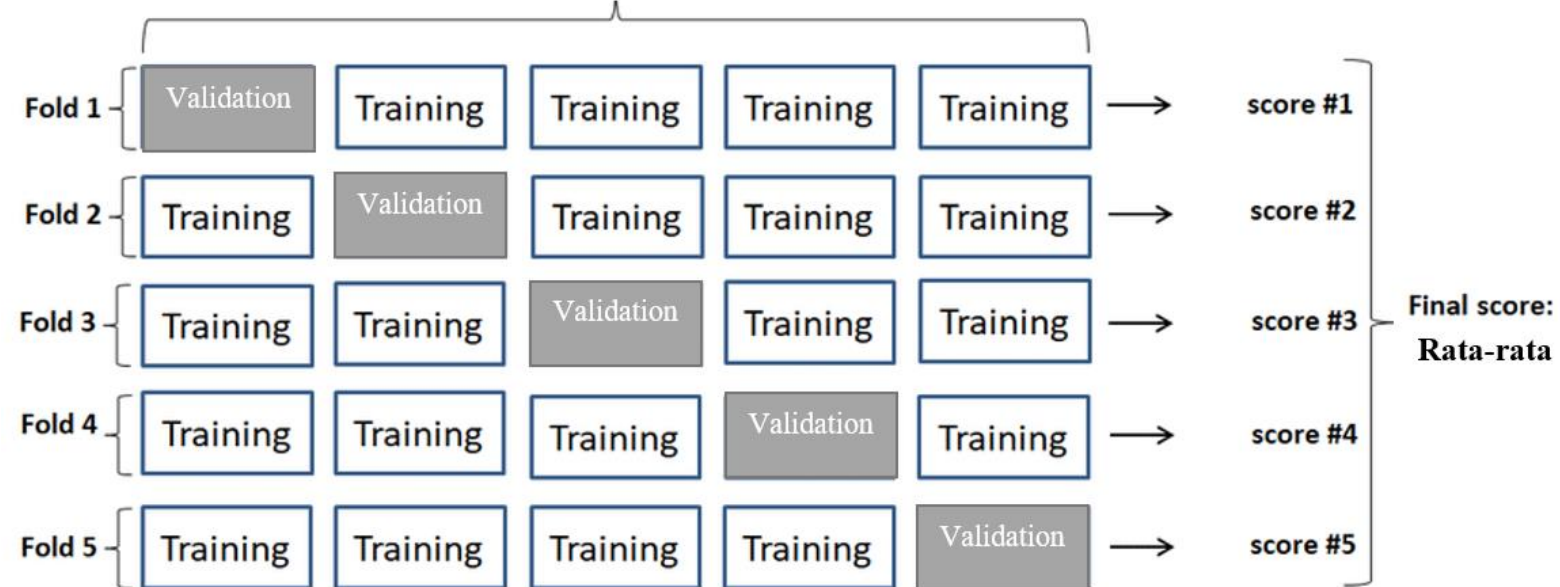
Cross validation adalah metode statistik untuk mengevaluasi dan membandingkan algoritma pembelajaran, metode *cross validation* yang digunakan adalah *5-fold cross validation*. *5-fold cross validation* pada penelitian ini digunakan untuk mencari hyperparameter terbaik untuk pemodelan BiGRU yang optimal

Elgeldawi et al., (2021)

Dalam proses pelatihan model BiGRU digunakan *epoch* maksimal sebesar 100 dan menggunakan beberapa skema pengujian yaitu:

- Membandingkan penggunaan **unit BiGRU**, banyak **neuron** pada dense layer, **dropout**, dan **batch size**.
- Menggunakan **Word2Vec** dan tanpa menggunakan **Word2Vec**.
- Menerapkan penanganan dan tanpa penanganan **Imbalanced Classes**.

Data dibagi kedalam 5 partisi



Pemodelan dengan Bidirectional Gated Recurrent Unit (BiGRU)

• Proses Pelatihan dan Validasi Model

Membandingkan penggunaan **unit BiGRU**, banyak **neuron** pada dense layer, **dropout**, dan **batch size**.

Pengujian Ke-	Unit BiGRU	Loss	Accuracy	F1-score
1	4	0.844	0.606	0.591
2	10	0.849	0.604	0.588
3	20	0.830	0.621	0.610
4	43	0.835	0.616	0.601
5	130	0.835	0.618	0.605
6	391	0.840	0.616	0.602

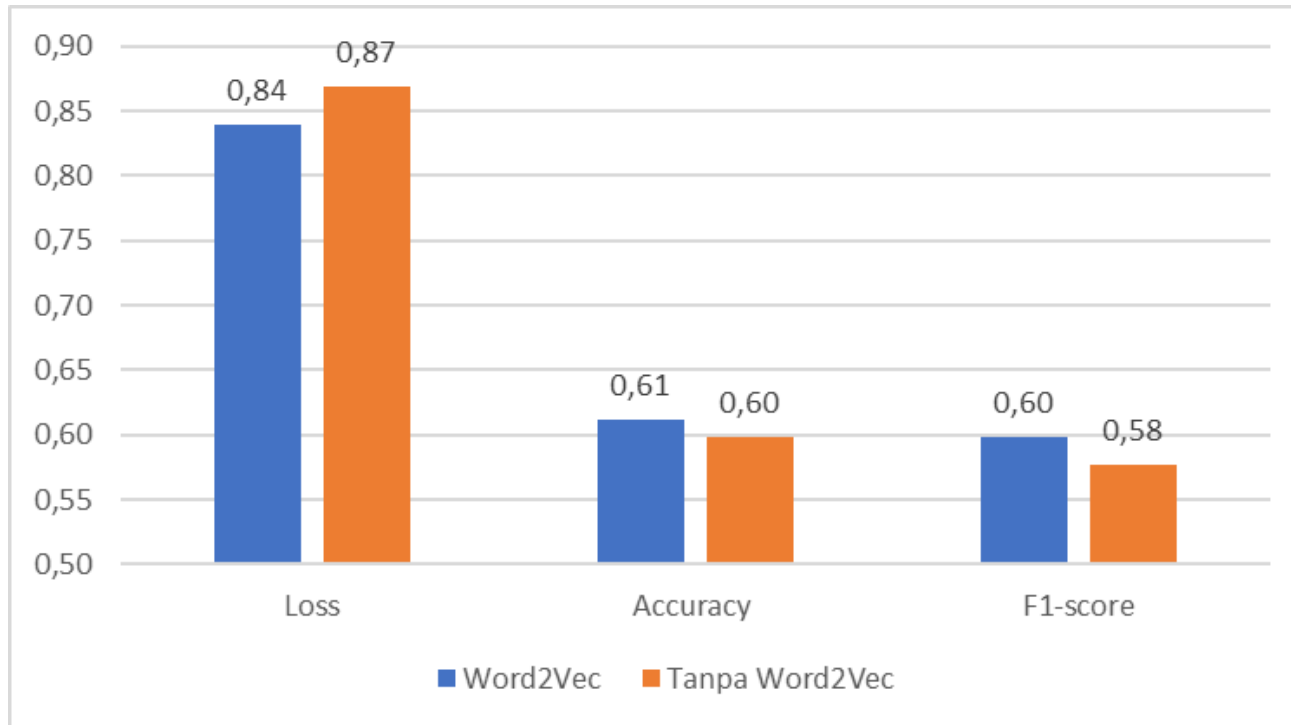
Pengujian Ke-	Dense Layer Neuron	Loss	Accuracy	F1-score
1	5	0.852	0.610	0.585
2	10	0.831	0.615	0.598
3	15	0.826	0.613	0.601
4	20	0.837	0.618	0.605
5	30	0.830	0.621	0.610
6	60	0.832	0.614	0.603

Pengujian Ke-	Dropout	Loss	Accuracy	F1-score
1	0.1	0.866	0.605	0.591
2	0.2	0.840	0.616	0.604
3	0.3	0.830	0.621	0.610
4	0.4	0.826	0.611	0.597
5	0.5	0.834	0.612	0.593
6	0.6	0.845	0.605	0.582

Pengujian Ke-	Batch Size	Loss	Accuracy	F1-score
1	16	0,833	0,617	0,604
2	32	0,842	0,610	0,597
3	64	0,829	0,621	0,609
4	128	0,830	0,618	0,605
5	256	0,838	0,613	0,599
6	512	0,840	0,610	0,592

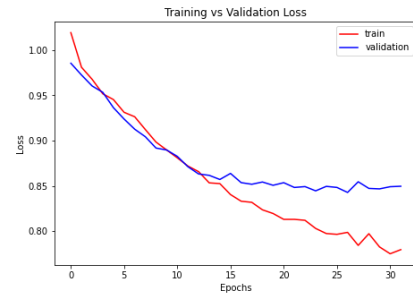
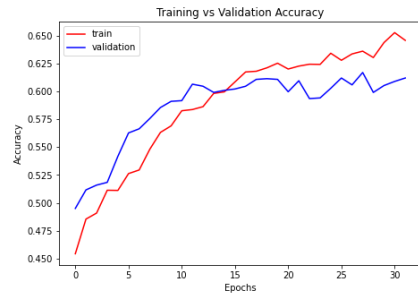


Membandingkan penggunaan Word2Vec dan tanpa menggunakan Word2Vec.

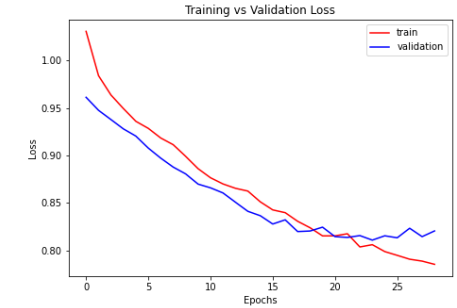
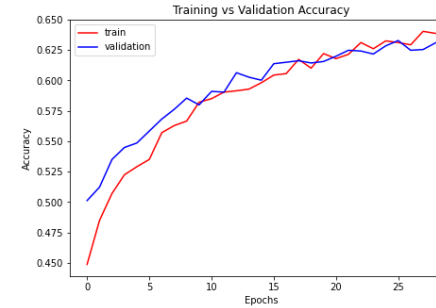


Word Embedding	Loss	Accuracy	F1-score
Word2Vec	0.839	0.611	0.598
Tanpa Word2Vec	0.869	0.598	0.577

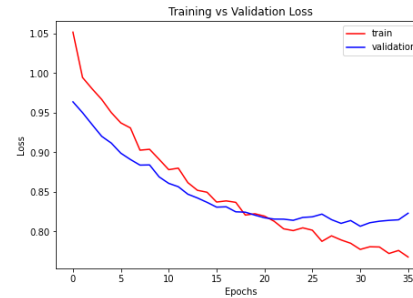
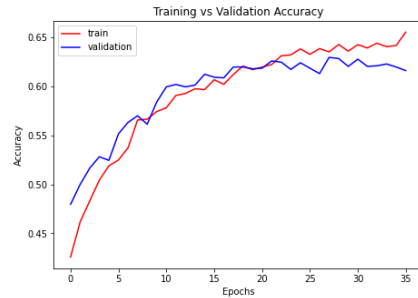
Fold ke-1



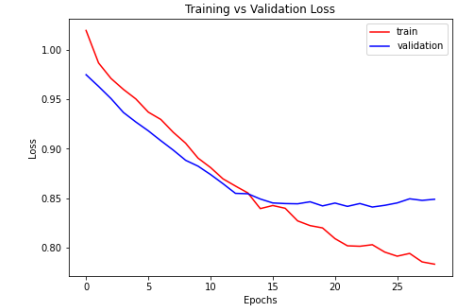
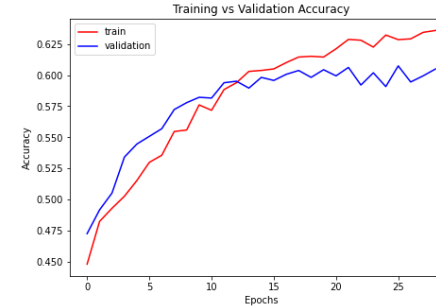
Fold ke-4



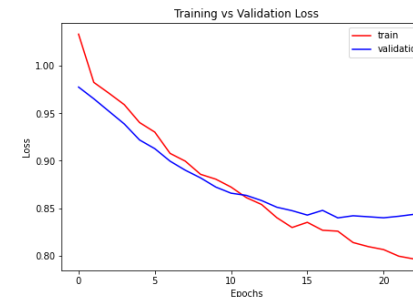
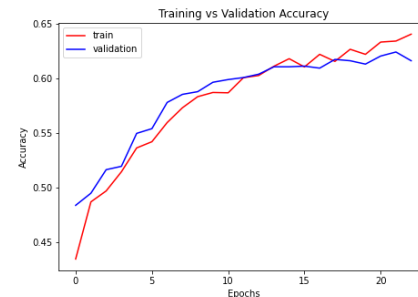
Fold ke-2



Fold ke-5



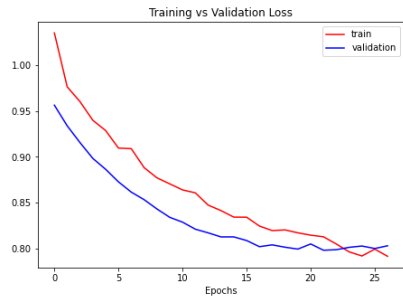
Fold ke-3



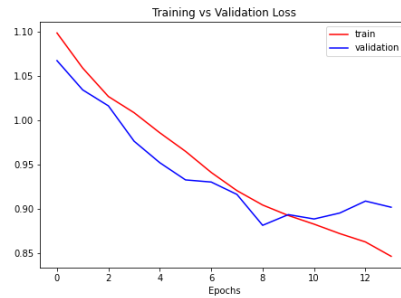
<i>Fold ke-</i>	<i>Epoch</i>	<i>Loss</i>	<i>Accuracy</i>	<i>F1-score</i>
1	32	0.843	0.606	0.597
2	36	0.807	0.628	0.623
3	29	0.811	0.622	0.617
4	29	0.841	0.602	0.593
5	23	0.840	0.617	0.607
Rata-rata (<i>Score</i>)		0,828	0,615	0,607



Grafik Loss



a) Tanpa Penanganan

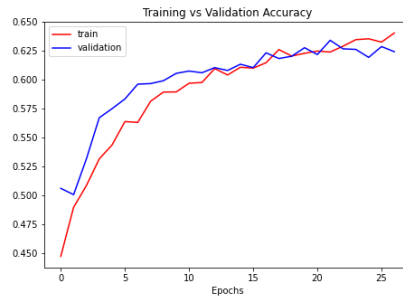


b) Random Over Sampling

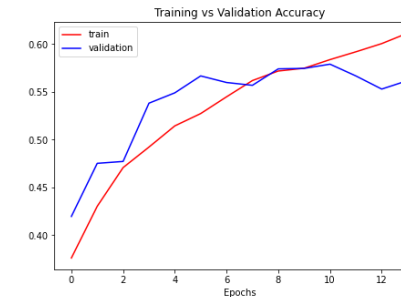


c) Random Under Sampling

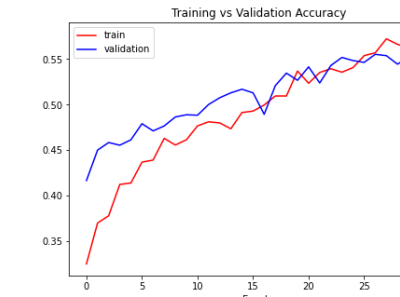
Grafik Accuracy



a) Tanpa Penanganan

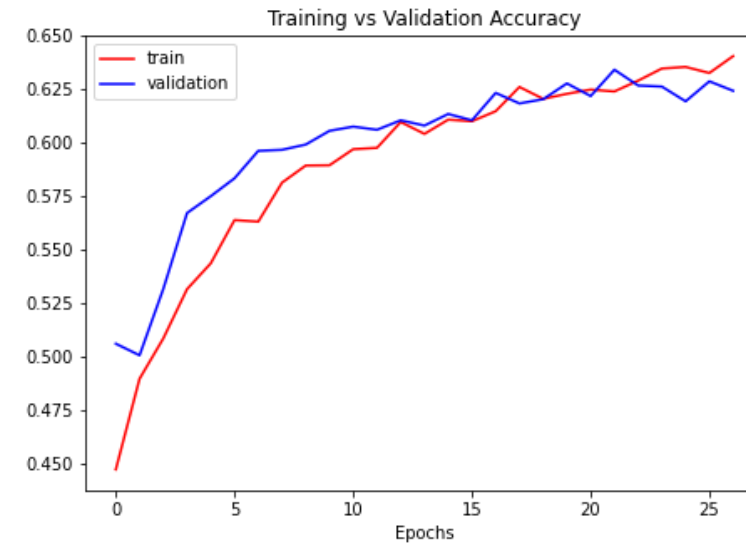
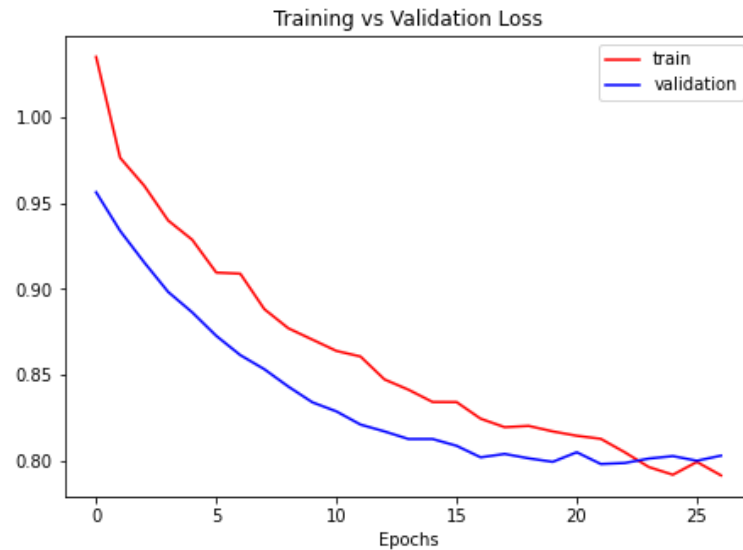


b) Random Over Sampling



c) Random Under Sampling

<i>Imbalanced Classes</i>	<i>Loss</i>	<i>Accuracy</i>	<i>F1-score</i>
Tanpa Penanganan	0.798	0.634	0.623
Random Over Sampling	0.881	0.574	0.599
Random Under Sampling	0.933	0.555	0.487



Confusion Matrix		Prediksi		
		Negatif	Netral	Positif
Aktual	Negatif	716	207	9
	Netral	309	494	50
	Positif	58	111	79

	Negatif	Netral	Positif
<i>Precision</i>	0.66	0.61	0.57
<i>Recall</i>	0.77	0.58	0.32
<i>F1-score</i>	0.71	0.59	0.41
<i>Accuracy</i>	0.63		

Metode	<i>Accuracy</i>	<i>F1-score</i>
<i>Support Vector Machine (SVM)</i>	0.569	0.529
<i>Naïve Bayes Classifier</i>	0.538	0.520
<i>Random Forest Classifier</i>	0.571	0.534
<i>Multi Layer Perceptron (MLP)</i>	0.533	0,500
<i>Simple Recurrent Neural Network (Simple RNN)</i>	0.518	0,498
<i>Long Short-Term Memory (LSTM)</i>	0.589	0,578
<i>Bidirectional Long Short-Term Memory (BiLSTM)</i>	0.591	0,578
<i>Bidirectional Gated Recurrent Unit (BiGRU)</i>	0,615	0,607



Kesimpulan

1. Didapatkan pemodelan arsitektur BiGRU yang paling optimal yaitu 1 *embedding layer*, lapisan BiGRU dengan unit sebanyak 20 unit, 1 lapisan *dense layer* dengan 30 *neuron*, menggunakan *dropout* dengan ukuran 0.3, *batch size* sebesar 64, menggunakan *loss function categorical cross entropy*, fungsi optimasi menggunakan ‘Adam’, *epoch* sebanyak 27 iterasi, *word embedding* dengan Word2Vec, dan tanpa penanganan *imbalanced classes*
2. Secara keseluruhan model BiGRU memperoleh nilai akurasi 63% dari data uji. Selain itu model mendapatkan *weighted average f1-score* 62%, rata-rata *precision* 61 % dan rata-rata *recall* sebesar 56 %.
3. Metode BiGRU mendapatkan nilai tertinggi apabila dibandingkan dengan metode SVM, *Naïve Bayes Classifier*, *Random Forest Classifier*, MLP, *Simple RNN*, LSTM, dan BiLSTM

Saran

1. Pada penelitian selanjutnya diharapkan dapat menambah jumlah *dataset* agar menghasilkan model yang lebih baik. Penambahan jumlah dataset dapat dilakukan dengan memperbanyak sumber media sosial.
2. Dalam penelitian selanjutnya diharapkan menambah dan berkolaborasi lebih banyak dengan ahli atau pakar khususnya spesialis di bidang linguistik, karena kontribusi pakar sangat penting pada penelitian analisis sentimen khususnya pada tahap pelabelan data.
3. Model BiGRU pada penelitian ini diharapkan dapat dilanjutkan ketahap *deployment* atau tahap pengembangan menjadi sebuah aplikasi agar dapat digunakan oleh pemerintah dalam mengevaluasi kebijakan dalam program vaksinasi COVID-19

TERIMA KASIH



- Abdelgwad, M. M., Soliman, H. A., Taloba, A. I., & Farghaly, M. F. (2021). *Arabic aspect based sentiment analysis using bidirectional GRU based models*.
- Akkaya, B. (2019). Comparison of Multi-class Classification Algorithms on Early Diagnosis of Heart Diseases. *Y-BIS 2019 Conference: Recent Advances in Data Science and Business Analytics*. https://www.academia.edu/41940316/Comparison_of_Multi_class_Classification_Algorithms_on_Early_Diagnosis_of_Heart_Diseases
- Biswas, S., Chadda, E., & Ahmad, F. (2015). Sentiment Analysis with Gated Recurrent Units. *Advances in Computer Science and Information Technology (ACSIT)*, 2(11), 59–63. <http://www.krishisanskriti.org/acsit.html>
- Brownlee, J. (2017). *Long Short-Term Memory Networks With Python: Develop Sequence Prediction Model With Deep Learning*. Machine Learning Mastery.
- Burkov, A. (2019). *The Hundred-Page Machine Learning*.
- Choi, D., Shallue, C. J., Nado, Z., Lee, J., Maddison, C. J., & Dahl, G. E. (2019). *On Empirical Comparisons of Optimizers for Deep Learning*. <https://arxiv.org/abs/1910.05446v3>
- Collins, C., Hasan, S., & Ukkusuri, S. (2013). A Novel Transit Rider Satisfaction Metric: Rider Sentiments Measured from Online Social Media Data. *Journal of Public Transportation*, 16(2), 2. <https://doi.org/http://doi.org/10.5038/2375-0901.16.2.2>
- Dahl, G. E., Sainath, T. N., & Hinton, G. E. (2013). Improving deep neural networks for LVCSR using rectified linear units and dropout. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 8609–8613. <https://doi.org/10.1109/ICASSP.2013.6639346>
- Darujati, C., & Gumelar, A. B. (2012). *PEMANFAATAN TEKNIK SUPERVISED UNTUK KLASIFIKASI TEKS BAHASA INDONESIA*.
- DiPietro, R., & Hager, G. D. (2019). Deep learning: RNNs and LSTM. *Handbook of Medical Image Computing and Computer Assisted Intervention*, 503–519. <https://doi.org/10.1016/B978-0-12-816176-0.00026-0>
- Enyinna Nwankpa, C., Ijomah, W., Gachagan, A., & Marshall, S. (2018). Activation Functions: Comparison of Trends in Practice and Research for Deep Learning.
- Ghods, A., & Cook, D. J. (2019). A Survey of Techniques All Classifiers Can Learn from Deep Networks: Models, Optimizations, and Regularization. <https://arxiv.org/abs/1909.04791v2>
- Giasryani, N. (2020). Komparasi Algoritma Machine Learning dan Deep Learning untuk Named Entity Recognition : Studi Kasus Data Kebencanaan. *Indonesian Journal of Applied Informatics*, 4(2), 138–144. <https://doi.org/10.20961/IJAI.V4I2.41317>
- Gruber, N., & Jockisch, A. (2020). Are GRU Cells More Specific and LSTM Cells More Sensitive in Motive Classification of Text? *Frontiers in Artificial Intelligence*, 0, 40. <https://doi.org/10.3389/FRAI.2020.00040>
- Härdle, W. K., Prastyo, D. D., & Hafner, C. M. (2014). Support Vector Machines with Evolutionary Model Selection for Default Prediction. *The Oxford Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics*. <https://doi.org/10.1093/OXFORDHB/9780199857944.013.011>



- Härdle, W. K., Prastyo, D. D., & Hafner, C. M. (2014). Support Vector Machines with Evolutionary Model Selection for Default Prediction. *The Oxford Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics*. <https://doi.org/10.1093/OXFORDHB/9780199857944.013.011>
- Hermawan, L., & Ismiati, M. B. (2020). Pembelajaran Text Preprocessing berbasis Simulator Untuk Mata Kuliah Information Retrieval. *Jurnal Transformatika*, 17(2), 188–199. <https://doi.org/10.26623/TRANSFORMATIKA.V17I2.1705>
- Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/NECO.1997.9.8.1735>
- Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., Cheng, Z., Yu, T., Xia, J., Wei, Y., Wu, W., Xie, X., Yin, W., Li, H., Liu, M., ... Cao, B. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet*, 395(10223), 497–506. [https://doi.org/10.1016/S0140-6736\(20\)30183-5](https://doi.org/10.1016/S0140-6736(20)30183-5)
- Janocha, K., & Czarnecki, W. M. (2017). On Loss Functions for Deep Neural Networks in Classification. *Schedae Informaticae*, 25, 49–59. <https://arxiv.org/abs/1702.05659v1>
- Kingma, D. P., & Ba, J. (2014). Adam: A Method for Stochastic Optimization. 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings. <https://arxiv.org/abs/1412.6980v9>
- le Duc, T., Leiva, R. G., Casari, P., & Östberg, P. O. (2019). Machine learning methods for reliable resource provisioning in edge-cloud computing: A survey. *ACM Computing Surveys*, 52(5). <https://doi.org/10.1145/3341145>
- Liu, B. (2012). Sentiment Analysis and Opinion Mining.
- Liu, C., Zhou, Q., Li, Y., Garner, L. v., Watkins, S. P., Carter, L. J., Smoot, J., Gregg, A. C., Daniels, A. D., Jervey, S., & Albaiu, D. (2020). Research and Development on Therapeutic Agents and Vaccines for COVID-19 and Related Human Coronavirus Diseases. *ACS Central Science*, 6(3), 315. <https://doi.org/10.1021/ACSCENTSCI.0C00272>
- Liu, H., & Lang, B. (2019). Machine learning and deep learning methods for intrusion detection systems: A survey. *Applied Sciences (Switzerland)*, 9(20). <https://doi.org/10.3390/APP9204396>
- Mikolov, T., Grave, E., Bojanowski, P., Puhersch, C., & Joulin, A. (2017). Advances in Pre-Training Distributed Word Representations. *LREC 2018 - 11th International Conference on Language Resources and Evaluation*, 52–55. <https://arxiv.org/abs/1712.09405v1>
- Minaee, S., Cambria, E., & Gao, J. (2020). Deep Learning Based Text Classification: A Comprehensive Review. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>
- Mustakim, Indah, R. N. G., Novita, R., Kharisma, O. B., Vebrianto, R., Sanjaya, S., Hasbullah, Andriani, T., Sari, W. P., Novita, Y., & Rahim, R. (2019). DBSCAN algorithm: twitter text clustering of trend topic pilkada pekanbaru. *Journal of Physics: Conference Series*, 1363(1), 012001. <https://doi.org/10.1088/1742-6596/1363/1/012001>
- Nair, V., & Hinton, G. E. (2010). Rectified Linear Units Improve Restricted Boltzmann Machines.
- Nurdin, A., Aji, B. A. S., Bustamin, A., & Abidin, Z. (2020). PERBANDINGAN KINERJA WORD EMBEDDING WORD2VEC, GLOVE, DAN FASTTEXT PADA KLASIFIKASI TEKS. *Jurnal Tekno Kompak*, 14(2), 74–79. <https://ejurnal.teknokrat.ac.id/index.php/teknokompak/article/view/732>



- Minaee, S., Cambria, E., & Gao, J. (2020). Deep Learning Based Text Classification: A Comprehensive Review. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>
- Mustakim, Indah, R. N. G., Novita, R., Kharisma, O. B., Vebrianto, R., Sanjaya, S., Hasbullah, Andriani, T., Sari, W. P., Novita, Y., & Rahim, R. (2019). DBSCAN algorithm: twitter text clustering of trend topic pilkada pekanbaru. *Journal of Physics: Conference Series*, 1363(1), 012001. <https://doi.org/10.1088/1742-6596/1363/1/012001>
- Nair, V., & Hinton, G. E. (2010). Rectified Linear Units Improve Restricted Boltzmann Machines.
- Nurdin, A., Aji, B. A. S., Bustamin, A., & Abidin, Z. (2020). PERBANDINGAN KINERJA WORD EMBEDDING WORD2VEC, GLOVE, DAN FASTTEXT PADA KLASIFIKASI TEKS. *Jurnal Tekno Kompak*, 14(2), 74–79. <https://ejurnal.teknokrat.ac.id/index.php/teknokompak/article/view/732>
- Pangaribuan, Y., & Sagala, M. (2017). Menerapkan Jaringan Saraf Tiruan untuk Mengenali Pola Huruf Menggunakan Metode Perceptron. 2548–1916.
- Pascanu, R., Mikolov, T., & Bengio, Y. (2013). On the difficulty of training recurrent neural networks.
- Putra, M. R. A., Djamal, E. C., & Ilyas, R. (2018). Brain Computer Interface untuk Menggerakkan Robot Menggunakan Recurrent Neural Network. *Prosiding Seminar Nasional Rekayasa Teknologi Informasi | SNARTISI*, 1. <https://ejurnal.lppmunsera.org/index.php/snartisi/article/view/829>
- Rachman, F. F., & Pramana, S. (2020). Analisis Sentimen Pro dan Kontra Masyarakat Indonesia tentang Vaksin COVID-19 pada Media Sosial Twitter. *Indonesian of Health Information Management Journal (INOHIM)*, 8(2), 100–109. <https://inohim.esaunggul.ac.id/index.php/INO/article/view/223>
- Ramachandran, P., Zoph, B., & le Google Brain, Q. v. (2017). SEARCHING FOR ACTIVATION FUNCTIONS.
- Rana, R. (2016). Gated Recurrent Unit (GRU) for Emotion Classification from Noisy Speech. <https://arxiv.org/abs/1612.07778v1>
- Rozi, I. F., Pramono, S. H., & Dahlan, E. A. (2012). Implementasi Opinion Mining (Analisis Sentimen) untuk Ekstraksi Data Opini Publik pada Perguruan Tinggi. *Jurnal EECCIS*, 6. <https://jurnaleeccis.ub.ac.id/index.php/eccis/article/view/164>
- Sakunthala, S., Kiranmayi, R., & Mandadi, P. N. (2018). A review on artificial intelligence techniques in electrical drives: Neural networks, fuzzy logic, and genetic algorithm. *Proceedings of the 2017 International Conference On Smart Technology for Smart Nation, SmartTechCon 2017*, 11–16. <https://doi.org/10.1109/SMARTTECHCON.2017.8358335>
- Sari, E. D. N., Sari, E. D. N., & Irhamah, I. (2020). Analisis Sentimen Nasabah pada Layanan Perbankan Menggunakan Metode Regresi Logistik Biner, Naïve Bayes Classifier (NBC), dan Support Vector Machine (SVM). *Jurnal Sains Dan Seni ITS*, 8(2), D177–D184. <https://doi.org/10.12962/j23373520.v8i2.44565>
- Sari, I. P., & Sriwidodo, S. (2020). Perkembangan Teknologi Terkini dalam Mempercepat Produksi Vaksin COVID-19. *Majalah Farmasetika*, 5(5), 204–217. <https://jurnal.unpad.ac.id/farmasetika/article/view/28082>



- Setiawan, E. I., Ferry, F., Santoso, J., Sumpeno, S., Fujisawa, K., & Purnomo, M. H. (2020). Bidirectional GRU for Targeted Aspect-Based Sentiment Analysis Based on Character-Enhanced Token-Embedding and Multi-Level Attention. *International Journal of Intelligent Engineering and Systems*, 13(5). <https://doi.org/10.22266/ijies2020.1031.35>
- Shahnawaz, S., & Astya, P. (2017). Sentiment analysis: Approaches and open issues. *Proceeding - IEEE International Conference on Computing, Communication and Automation, ICCCA 2017, 2017-January*, 154–158. <https://doi.org/10.1109/CCAA.2017.8229791>
- Shen, Y., Zhang, L., Zhang, J., Yang, M., Tang, B., Li, Y., & Lei, K. (2018). CBN: Constructing a clinical Bayesian network based on data from the electronic medical record. *Journal of Biomedical Informatics*, 88, 1–10. <https://doi.org/10.1016/J.JBI.2018.10.007>
- Tama, V. O. (2018). Analisis Pelabelan dalam Klasifikasi Sentimen Ulasan Produk dengan Menggunakan Algoritma Multinomial Naïve Bayes. Universitas Telkom. <https://openlibrary.telkomuniversity.ac.id/home/catalog/id/146126/slug/analisis-pelabelan-dalam-klasifikasi-sentimen-ulasan-produk-dengan-menggunakan-algoritma-multinomial-na-ve-bayes.html>
- Tampil, Y., Komaliq, H., & Langi, Y. (2017). Analisis Regresi Logistik Untuk Menentukan Faktor-Faktor Yang Mempengaruhi Indeks Prestasi Kumulatif (IPK) Mahasiswa FMIPA Universitas Sam Ratulangi Manado. *D’CARTESIAN:Jurnal Matematika Dan Aplikasi*, 6(2), 56–62. <https://doi.org/10.35799/DC.6.2.2017.17023>
- Tan, P.-N., Steinbach, M., & Kumar, V. (2006). *Introduction to Data Mining*.
- Ting, K. M. (2017). Confusion Matrix. *Encyclopedia of Machine Learning and Data Mining*, 260–260. https://doi.org/10.1007/978-1-4899-7687-1_50
- Wati, R. (2016). Penerapan Algoritma Genetika Untuk Seleksi Fitur Pada Analisis Sentimen Review Jasa Maskapai Penerbangan Menggunakan Naive Bayes. *EVOLUSI : Jurnal Sains Dan Manajemen*, 4(1). <https://ejournal.bsi.ac.id/ejurnal/index.php/evolusi/article/view/604>
- We Are Social, & Hootsuite. (2021, January). *Digital 2021: The Latest Insights Into The State of Digital*. <https://datareportal.com/reports/digital-2021-indonesia>
- WHO. (2020). Wijaya, A. P., & Santoso, H. A. (2016). Naive Bayes Classification pada Klasifikasi Dokumen Untuk Identifikasi Konten E-Government. *Journal of Applied Intelligent System*, 1(1), 48–55. <http://publikasi.dinus.ac.id/index.php/jais/article/view/1032>
- W. Ali, Y. Yang, X. Qiu, Y. Ke and Y. Wang, (2021). Aspect-Level Sentiment Analysis Based on Bidirectional-GRU in SIoT. *IEEE Access*, vol. 9, pp. 69938-69950, doi: 10.1109/ACCESS.2021.3078114.
- Yu, Q., Zhao, H., & Wang, Z. (2019). Attention-based bidirectional gated recurrent unit neural networks for sentiment analysis. *ACM International Conference Proceeding Series*, 116–119. <https://doi.org/10.1145/3357254.3357262>
- Zhang, X. F., Huang, H. Y., & Zhang, K. L. (2009). KNN text categorization algorithm based on semantic centre. *Proceedings - 2009 International Conference on Information Technology and Computer Science, ITCS 2009*, 1, 249–252. <https://doi.org/10.1109/ITCS.2009.57>
- Zhang, Y., Ren, W., Zhu, T., & Faith, E. (2019). MoSa: A modeling and sentiment analysis system for mobile application big data. *Symmetry*, 11(1). <https://doi.org/10.3390/SYM1101011>



no	komentar
1	pemerintah sudah berusaha semaksimal mungkin memutus mata rantai virus bahkan sampai geram menangani pandemi ini tinggal kitanya sebagai masyarakat mari kita bantu pemerintah kita jangan saling menyalahkan dimulai dari diri kita sendiri jaga kesehatan
2	jaga prokes secara ketat kasihan nakes nakes sudah banyak yg meninggal terlalu banyak nyawa melayang sia sia mari kita bantu pemerintah dalam menghadapi pandemi ini semua harus kompak bekerja sama dlm satu tujuan menghentikan laju covid di indonesia
3	biar cepet selesai sembuh total normal kembali semula biar damai sama merasakan penderitaan nya seperti rakyat kecil miskin yg gak punya bantuan sama sekali pekerjaan hentikan gaji pemerintah hentikan BBM seluruhnya biar orang gak kemana
4	jangan lupa mencuci tangan menjaga jarak
5	semoga komentar nya semua disini jadi pertimbangan buat kementerian lain kali sebelum PPKM bantuan diratakan sy mewakili yg terdampak PPKM tagihan apapun tetap jalan selama PPKM tolong sekali lagi tindak tegas oknum yg mencovidkan demi tunjangan dari pemerintah
6	bukan covid nya yg di hindari bukan matinya juga yg di takuti karena semua itu pasti terjadi siapkan bekal akhirat kelak kasihan kalau besok mati
.....
11376	saat ini kita seolah dibuat bodoh oleh peraturan yg berlaku ke rumah sakit harus wajib swab bagi pasien wajar jika orang menduga kalau pasien sengaja di covid karena saat ini tidak ada yg nama nya penyakit flu batuk dan beberapa lainnya kecuali covid

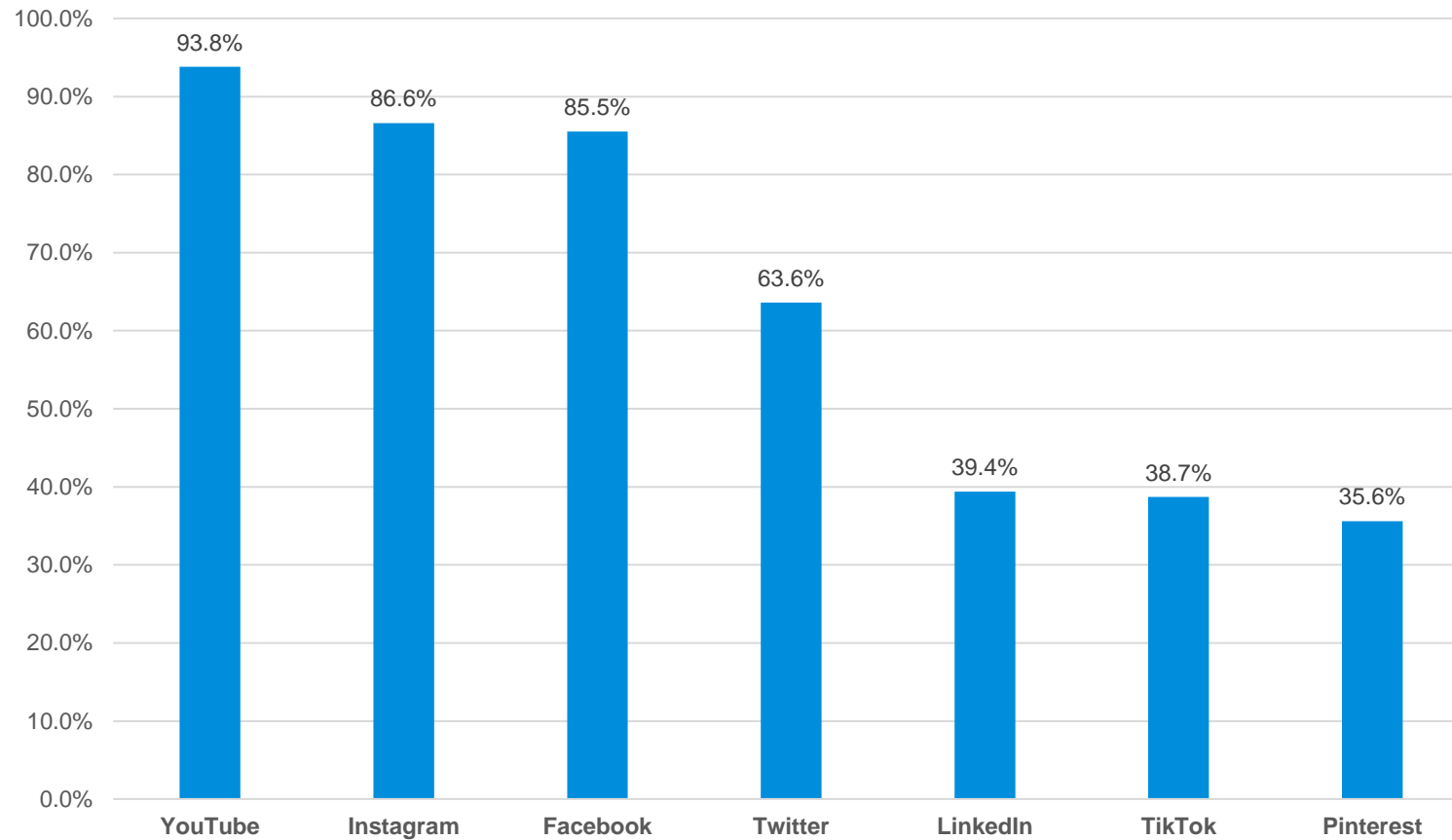




Referensi	Deskripsi Dataset	Metode	Akurasi
A. M. Rahat, A. Kahir, and A. K. M. Masum	Dataset yang digunakan yaitu tweet Airline Reviews yang terdiri dari ulasan positif dan negatif dengan total 10.000 data	SVM	82,48%
		Naïve Bayes	76,56%
M. Wongkar and A. Angdresey	Data dikumpulkan melalui media sosial twitter terkait dengan pasangan calon presiden Republik Indonesia periode 2019 – 2024. Data tersebut berjumlah 443 dengan atribut sentimen yang berisi informasi positif dan negatif	SVM	63,99%
		Naïve Bayes	75,58%
		K-NN	73,34%
L. Kurniasari and A. Setyanto	Dataset yang digunakan adalah dataset yang berisi ulasan dalam bahasa Indonesia dari situs Traveloka. Model yang ada akan digunakan untuk mengklasifikasikan ulasan pengguna menjadi dua kategori, ulasan positif dan negatif.	RNN	91,9%

Ristasari Dwi Septiana, Agung Budi Susanto, Tukiyyat	Data tweets yang mengandung 5 (lima) kata kunci, yaitu “Vaksin Sinovac”, “Vaksin Astrazeneca”, “Vaksin Moderna”, “Vaksin Merah Putih”, dan “Vaksinasi Covid-19”	Naïve Bayes	66,02%
Frizka Fitriana, Ema Utami, Hanif Al Fatta	Data yang digunakan dalam penelitian ini diperoleh dari data postingan Twitter berdasarkan kata kunci vaksin Covid -19. Interval waktu data Twitter diambil dari data tahun 2021 berkisar 1000 data melalui kata vaksin Covid-19	SVM	90,47%
		Naïve Bayes	88,64%
Merinda Lestandy, Abdurrahim Abdurrahim, Lailis Syafa'ah	Data vaksin COVID-19 berjumlah 5000 data. Pengumpulan data dimulai pada tanggal 10 Januari 2021 sampai dengan 21 April 2021. Dari data tersebut kemudian dilakukan pelabelan secara manual dengan 3 pembagian klasifikasi yakni 3800 tweet sentimen positif, 800 tweet sentimen negatif dan 400 tweet sentimen netral	Naïve Bayes	80%
		LSTM	97,7%





Persentase Platform yang Sering Digunakan Pengguna Internet Berusia 16 hingga 64 Tahun di Indonesia pada Januari 2021



Pelabelan secara manual menghasilkan data yang akurat karena manusia dapat membedakan dengan tepat suatu kalimat termasuk ke dalam suatu sentiment **(Tama, 2018)**.

Data komentar yang sudah dikumpulkan pada penelitian ini selanjutnya dilakukan pelabelan data yaitu dengan memberikan tag yang bermakna ke setiap data. Data komentar diklasifikasikan ke dalam kelas sentimen positif, netral, dan negatif. Label yang digunakan ada 3 yaitu “Negatif”, “Netral”, dan “Positif”.

- **Label “Negatif”** adalah komentar yang mengandung kata-kata buruk, ejekan, atau menyatakan kontra.
- **Label “Netral”** merupakan komentar yang memiliki kata yang bermakna informasi, dan tidak memihak pihak pro maupun kontra.
- **Label “Positif”** adalah data komentar mengandung kata-kata baik, pujian, pernyataan setuju, atau dukungan



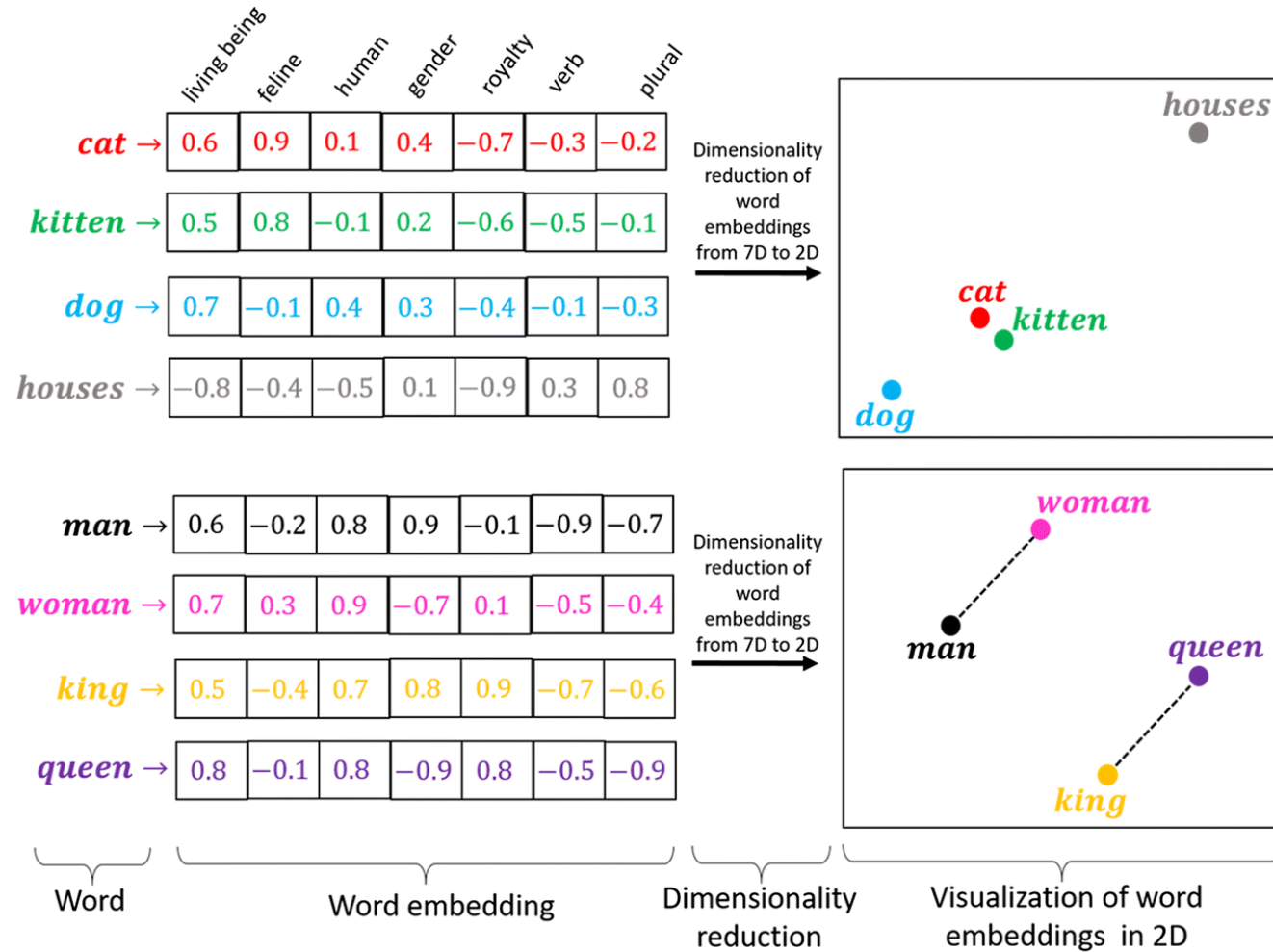
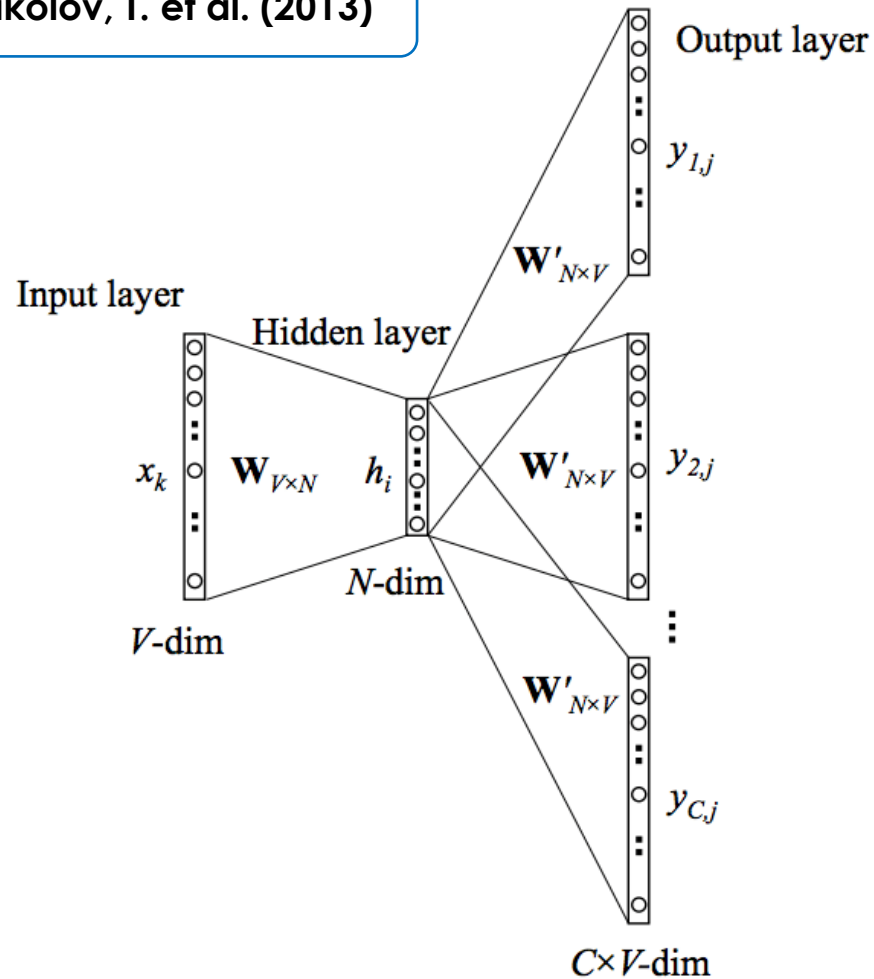
Komentar	Hasil Prediksi	Hasil Label
kita hanya di bodohi bukanya virus yg mati tapi masyarakat yg mati kelaparan anda hanya bisa berbicara tanpa tau keadaan kita pribadi	Negatif: 64% Netral: 27% Positif: 9%	Negatif
Go semangat vaksinasi. Semoga jadi jalan keluar. Tapi abis divaksin hati2 jgn tiba2 sok pede ga pake masker keluyuran ya	Negatif: 12% Netral: 42% Positif: 46%	Positif
vaksin sudah masuk ke daerah selain jabodetabek belum ya	Negatif: 19% Netral: 74% Positif: 7%	Netral



komentar	Label
pemerintah sudah berusaha semaksimal mungkin memutus mata rantai virus bahkan sampai geram menangani pandemi ini tinggal kitanya sebagai masyarakat mari kita bantu pemerintah kita jangan saling menyalahkan dimulai dari diri kita sendiri jaga kesehatan	Positif
Vaksin adalah produk biologi yang berisi antigen	Netral
jangan mau divaksin	Negatif



Mikolov, T. et al. (2013)



Confusion Matrix		Prediksi		
		a	b	c
Aktual	a	T_{aa}	F_{ba}	F_{ca}
	b	F_{ab}	T_{bb}	F_{cb}
	c	F_{ac}	F_{bc}	T_{cc}

$$\begin{aligned} \text{True Positive} &= T_{aa} \\ \text{True Negative} &= T_{bb} + T_{cc} + F_{bc} + F_{cb} \\ \text{False Positive} &= F_{ab} + F_{ac} \\ \text{False Negative} &= F_{ba} + F_{ca} \end{aligned}$$

$$\text{accuracy} = \frac{TP}{TP + TN + FP + FN} \times 100\%$$

$$\text{precision} = \frac{TP}{TP + FP}$$

$$\text{recall} = \frac{TP}{TP + FN}$$

$$\begin{aligned} F_1 &= \frac{2}{\text{recall}^{-1} + \text{precision}^{-1}} \\ &= 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \\ &= \frac{TP}{TP + \frac{1}{2}(FP + FN)} \end{aligned}$$

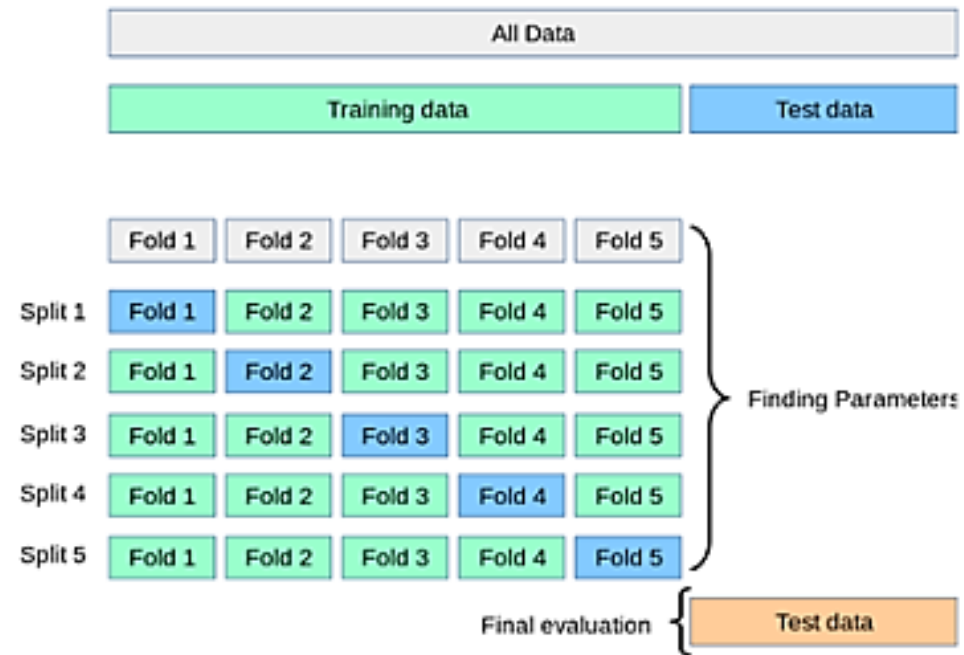


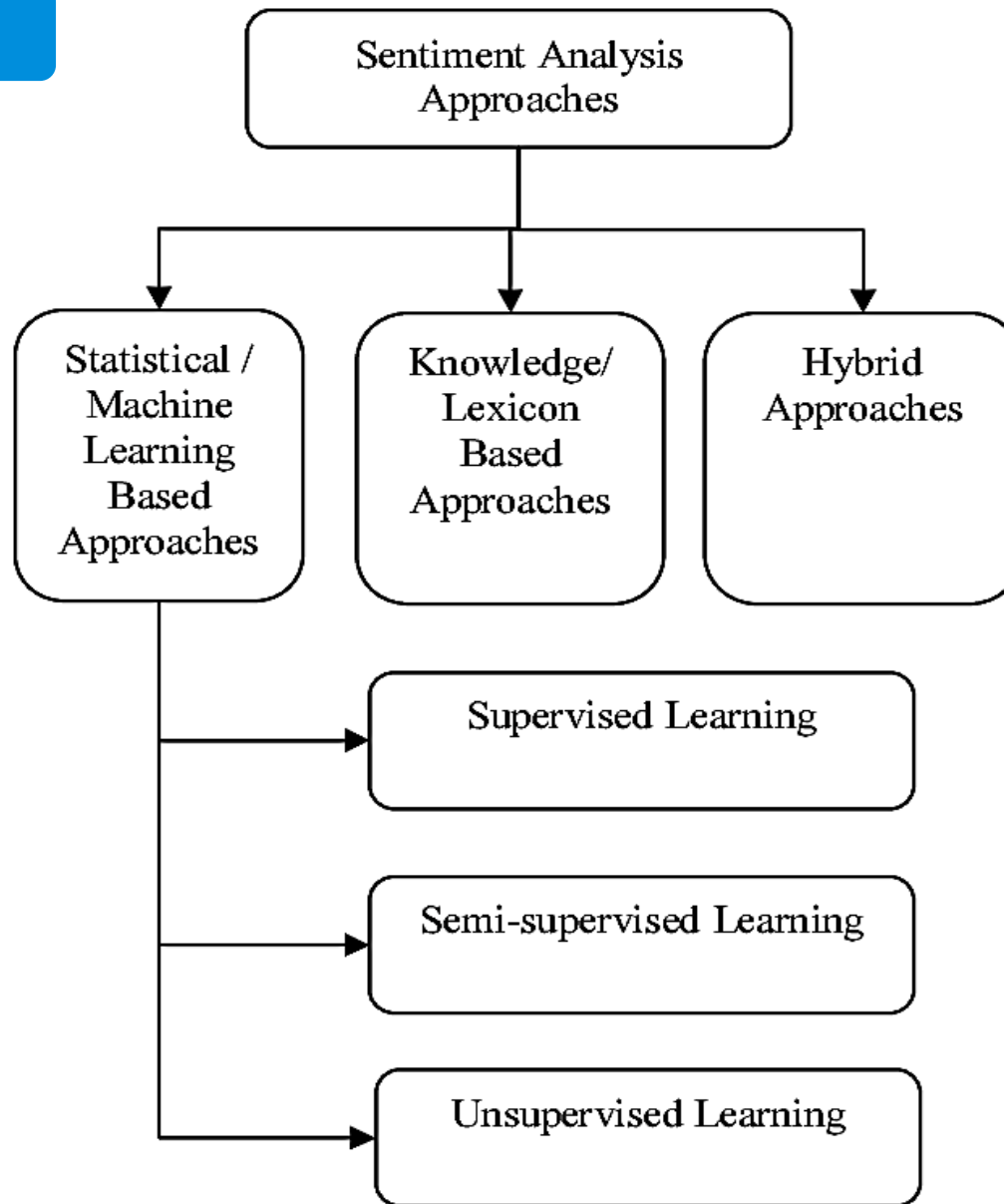
TUJUAN

Mengaplikasikan **metode klasifikasi yang sesuai** pada analisis sentimen untuk mengklasifikasikan opini **positif, netral, dan negatif** dalam dataset sentimen pengguna media sosial terhadap kebijakan pemerintah dalam program vaksinasi COVID-19.



K-Fold Cross Validation





Shahnawaz, & Astya, P. (2017)

Analisis Sentimen Vaksinasi Covid-19 di Jakarta

JSC | 7 months ago

f Facebook

Twitter

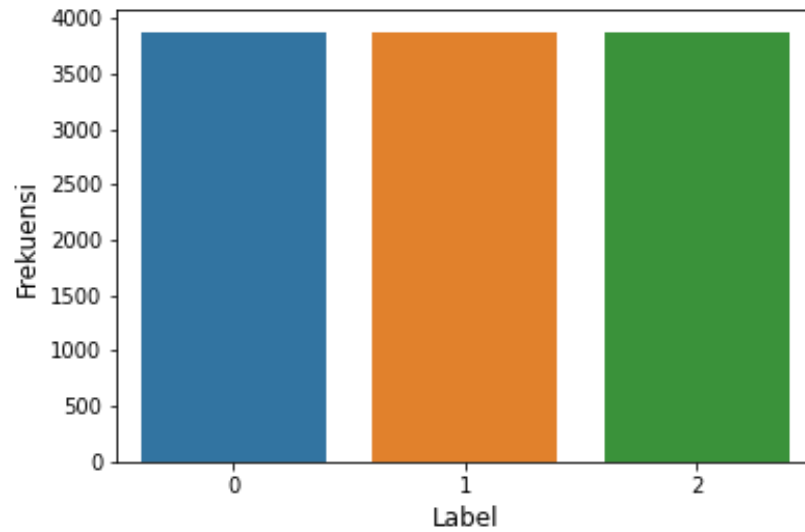
Email



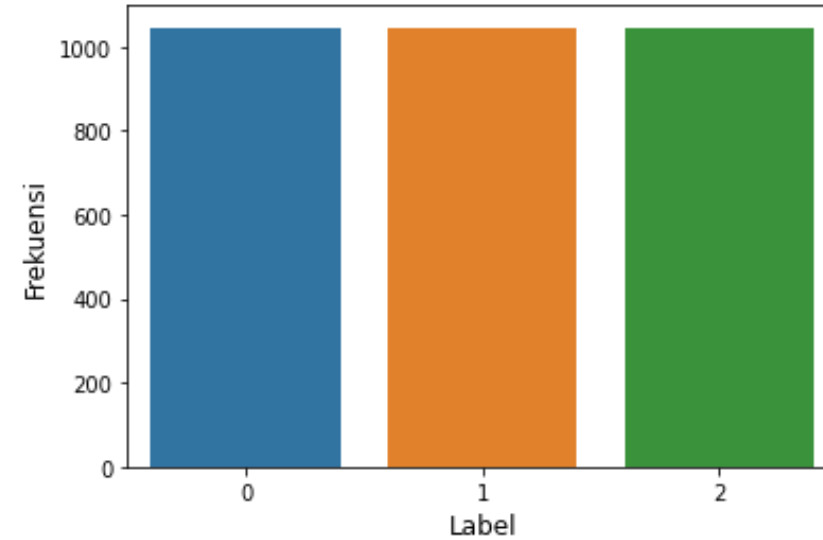
**PENKATEGORIAN
PORNOGRAFI, PERJUDIAN DAN PENIPUAN
PADA EMAIL LAYANAN ADUAN KONTEN
KEMENTERIAN KOMUNIKASI DAN INFORMATIKA**

Okky Robiana Sulaeman

Pranata Komputer Muda Direktorat Pengendalian Aplikasi
Informatika di Kementerian Komunikasi dan Informatika



a) *Random Over Sampling*



b) *Random Under Sampling*

