

Analisis Sentimen Opini Publik Menggunakan Metode BiLSTM Pada Media Sosial Twitter**Ardian Nur Romadhan¹, Ema Utami², Anggit Dwi Hartanto³**¹ PJJ MTI Universitas Amikom Yogyakarta. Email: ardian.1251@students.amikom.ac.id² PJJ MTI Universitas Amikom Yogyakarta. Email: ema.u@amikom.ac.id³ PJJ MTI Universitas Amikom Yogyakarta. Email: anggit@amikom.ac.id

Abstract

Twitter is one of the social media to express public opinion regarding the issues currently being discussed. The pros and cons of the Constitutional Court's decision to grant all judicial review requests related to Article 29 letter e and Article 34 of the Corruption Eradication Commission Law became the subject of discussion on social media, in this case Twitter. The purpose of this study was to compare the level of accuracy of TF-IDF and Word2Vec term weighting using the BiLSTM model in sentiment preference analysis of public opinion regarding the extension of the term of office of the KPK leadership and the KPK Law. The research conducted is useful for analyzing the analysis of public opinion on Twitter into positive or negative classes. By using the BiLSTM method, opinion classification is carried out by researchers. The weighting method used in this study is TF-IDF and Word2Vec which are useful for measuring the weight of each word in a tweet. With the BiLSTM deep learning method and a combination of TF-IDF term weighting and Word2Vec term weighting, it can provide the best performance evaluation results at a ratio of 80:20 with a precision value of 85.00%, 86.85% recall, and an F1-score of 85.92%. The test results show that the TF-IDF has an accuracy of 84.84% and the WIDF weighting has an accuracy of 85.72%. From this study it was found that from the comparison results the BiLSTM model with Word2Vec was superior to the BiLSTM model with TF-IDF.

Keywords: sentiment analysis, BiLSM, TF-IDF, Word2Vec, Twitter.

Abstrak

Twitter merupakan salah satu media sosial untuk mengekspresikan opini publik mengenai isu yang sedang dalam perbincangan. Pro-kontra terhadap keputusan MK atas pengabulan seluruh permohonan judicial review terkait Pasal 29 huruf e dan Pasal 34 UU KPK menjadi perbincangan di sosial media, dalam kasus ini Twitter. Tujuan penelitian ini adalah untuk mengetahui perbandingan tingkat akurasi term weighting TF-IDF dan Word2Vec menggunakan model BiLSTM dalam preferensi sentiment analisis opini publik terkait perpanjangan masa jabatan pimpinan KPK dan UU KPK. Penelitian yang dilakukan berguna untuk menganalisa analisis opini publik di Twitter kedalam kelas positif atau negatif. Dengan menggunakan metode BiLSTM, klasifikasi opini dilakukan oleh peneliti. Metode pembobotan yang digunakan dalam penelitian ini adalah TF-IDF dan Word2Vec yang berguna untuk mengukur bobot setiap kata pada sebuah tweet. Dengan metode *deep learning* BiLSTM dan kombinasi antara *term weighting* TF-IDF dan *term weighting* Word2Vec, dapat memberikan hasil evaluasi performansi terbaik pada rasio 80:20 dengan nilai *precision* 85.00%, *recall* 86.85%, dan *F1-score* 85.92%. Hasil pengujian menunjukkan bahwa TF-IDF memiliki akurasi sebesar 84.84%. dan pembobotan Word2Vec memiliki akurasi sebesar 85.72%. Dari penelitian ini didapati bahwa dari hasil Perbandingan model BiLSTM dengan Word2Vec lebih unggul dibanding dengan model BiLSTM dengan TF-IDF.

Kata Kunci: analisis sentimen, BiLSM, TF-IDF, Word2Vec, Twitter.

1. Pendahuluan

Sosial media banyak menghasilkan berbagai macam informasi seperti topik politik atau keputusan pemerintah atas pemberlakuan undang-undang baik UU / RUU / RKUHP / KUHP. Berbagai platform media sosial seperti Twitter sering kali dimanfaatkan pengguna untuk beropini atau menyampaikan berbagai pendapatnya mengenai isu yang sedang menjadi perbincangan dengan mengunggah teks. Keputusan Mahkamah Konstitusi yang dibacakan oleh Ketua MK Anwar Usman pada Kamis 25 Mei 2023 yang mengabulkan seluruh permohonan pengujian UU No. 19 Tahun 2019 tentang perubahan kedua atas UU No. 30 Tahun 2002 (Argawati, 2021, 2023; Putri, 2023), menjadi fenomena atas opini publik. Fenomena tersebut menjadi isu yang diperbincangkan oleh pengguna twitter salah satunya tentang masa jabatan pimpinan KPK. Opini merupakan pernyataan atau pendapat terhadap suatu isu yang kontroversial akibat dari diskusi tentang fenomena yang menimbulkan perbedaan opini (Sudrajat & Rohida, 2022). Pada penelitian terdahulu telah memanfaatkan *machine learning* dalam mendapatkan tingkat akurasi terbaik dari tiap pembobotan kata. Urgensi pada penelitian ini yaitu menggunakan metode selain *machine learning* sebagai alternatif dalam pengkalsifikasi data teks yaitu menggunakan *deep learning*.

Pada penelitian terdahulu telah dilakukan kajian terkait topik analisis sentimen opini publik dengan memanfaatkan media sosial Twitter. Salah satunya dilakukan oleh (Amada dkk., 2023) mengklasifikasikan sentimen opini tentang Permenkominfo No.5 Tahun 2020 menggunakan algoritma SVM, NB dan K-NN dan hasil penelitiannya menunjukkan bahwa akurasi SVM 83%, NB 78%, dan K-NN 83%. Lebih lanjut pada penelitian yang dilakukan oleh (Reiki dkk., 2022) mengklasifikasikan sentimen opini tentang UU No. 3 Tahun 2022 tentang pemindahan Ibu Kota Negara (IKN) menggunakan algoritma SVM dengan pembobotan kata TF-IDF, TF-RF, dan TF-BIN.ICF dan hasil penelitiannya menunjukkan bahwa pembobotan TF-IDF masih yang terbaik dibandingkan bobot term TF-BIN.ICF dan TF-RF. Penelitian serupa menggunakan metode SVM dilakukan oleh (Sukma dkk., 2020) mengklasifikasikan sentimen opini tentang *omnibus law* dan hasil penelitian menunjukkan bahwa algoritma terbaik dengan tingkat akurasi 91,80% yaitu algoritma SVM. Sedangkan pada penelitian yang dilakukan oleh (Sandryan dkk., 2021) mengklasifikasikan sentimen opini tentang Undang-Undang Cipta Kerja menggunakan algoritma *backpropagation* dengan pembobotan kata TF-IDF dan hasil penelitian menunjukkan bahwa akurasi sebesar 95%, precision sebesar 98%, recall sebesar 92.4%, dan *f-measure* sebesar 95.1%. Selain itu pada penelitian yang dilakukan oleh (Khurniawan & Ruldeviyani, 2020) pada penelitiannya mengklasifikasikan sentimen opini tentang Revisi UU Komisi Pemberantasan Korupsi (KPK) 2019 menggunakan algoritma SVM, *Decision Tree*, dan NB dan hasil penelitiannya menunjukkan bahwa algoritma SVM memiliki performa dan akurasi tertinggi sebesar 81,70%, diikuti oleh NB sebesar 80,90%, dan *Decision Tree* sebesar 74,55%.

Pada penelitian sebelumnya telah dilakukan klasifikasi menggunakan *machine learning*, sehingga pada penelitian ini akan dilakukan dengan metode *deep learning* menggunakan *Bidirectional Long Short-Term Memory* (BiLSTM). BiLSTM dipilih sebagai metode pada penelitian ini karena dapat mengklasifikasikan pendapat ke dalam klasifikasi yang telah ditentukan yaitu sentimen positif ataupun negatif. BiLSTM menjadi pilihan sebagai metode klasifikasi karena metode ini cocok untuk klasifikasi teks dan analisis metode ini memberikan hasil yang baik karena beroperasi di kedua arah yaitu maju (*forward*) dan mundur (*backward*) (Lin dkk., 2023). Penelitian ini menggunakan pembobotan kata (*term weighting*) *Term Frequency-Inverse Document Frequency* (TF-IDF) dan Word2Vec dalam melakukan klasifikasi teks berbahasa Indonesia dan untuk menunjukkan pengaruh suatu kata dalam isi dokumen.

Tujuan penelitian ini adalah untuk mengetahui perbandingan tingkat akurasi *term weighting* TF-IDF dan Word2Vec menggunakan model BiLSTM dalam preferensi sentiment analisis opini publik perpanjangan masa jabatan pimpinan KPK dan UU KPK. Beberapa langkah yang dilakukan dalam penelitian yang diteliti kali ini yaitu pengumpulan data, pelabelan, preprocessing, pembobotan kata, pemodelan, pengujian, dan evaluasi performa. Penelitian ini diharapkan dapat menjadi rekomendasi bagi pemerintah untuk mendapatkan pandangan publik, serta sebagai acuan dalam proses pengambilan keputusan sesuai kebutuhan.

2. Metode

Pengumpulan data dilakukan menggunakan teknik *scraping* datatweet pada Twitter menggunakan *snsrape*. Data yang dikumpulkan merupakan data *tweet* yang mengandung kata “Perpanjangan Masa Jabatan Pimpinan KPK” atau hastag “#UUKPK dan “#KPKRakusJabatan”. Data yang diambil adalah sebanyak 15.000 *tweet*.

Data selanjutnya akan dilakukan *preprocessing*, meliputi : *case folding* adalah, tahapan yang mengubah kata-kata pada teks menjadi huruf kecil atau menyeragamkan teks, dengan mengubah huruf *uppercase* menjadi *lowercase*. Normalisasi 1 adalah, tahapan untuk memperbaiki kata menjadi kata normal. Pada normalisasi 1 menggunakan kamus alay dari indo-bercamart. *Tokenizing* adalah, proses pemecahan suatu teks, berdasarkan setiap penyusun katanya. *Filtering (stopword removal)* adalah, proses menghapus kata yang tidak, mengandung makna (*stopword*) dan mengambil kata-kata, penting, hasil, dari, normalisasi. *Stemming* adalah, proses pemetaan kata menjadi bentuk, kata dasarnya. Normalisasi 2 adalah, tahapan untuk memperbaiki kata menjadi kata normal. Pada normalisasi 2 menggunakan kamus indonesian-word.

Setelah dilakukan *preprocessing* selanjutnya diseleksi untuk mendapatkan data yang sesuai berdasarkan kebutuhan. Proses seleksi dilakukan dengan menyaring data yang tidak mengandung opini (Wibowo dkk., 2020). Data yang diperoleh dari hasil seleksi adalah sebanyak 11.448 data dari 15.000 *tweet* data. Pemberian label pada data dilakukan dengan bantuan IndoBERT ke dalam klasifikasi kelas yang telah diputuskan, yakni negatif dan juga positif. Metode IndoBERT diimplementasikan untuk mendeteksi penggunaan kalimat *abusive* pada teks bahasa Indonesia, dimana merupakan model berbasis BERT Indonesia yang dilatih pada *dataset* Indo4B (Wilie dkk., 2020). Hasil dari pelabelan diperoleh *tweet* dengan kelas positif dan kelas negatif dapat dilihat pada Tabel 1.

Tabel .1 Polaritas Sentimen pada Data

<i>label</i>	<i>t</i>	<i>it</i>
<i>ive</i>	;	%
<i>ve</i>	;	%

Sumber: Data diolah 2023

Polaritas sentiment yang tersaji pada Tabel 1. dapat dilihat bahwa untuk label *negative* sebanyak 5.823 data dan label *positive* sebanyak 5.625 data, dengan demikian sentiment dapat diketahui sentiment opini publik berlabel *negative* lebih banyak yaitu yaitu 50,86% dibanding dengan label *positive* yaitu 49,14%.

Pembobotan kata adalah proses perhitungan bobot tiap kata pada dokumen sehingga dapat diketahui kemiripan suatu kata pada dokumen tersebut (Parasati dkk., 2020). Pembobotan kata pada penelitian yang dilakukan yakni dengan penggunaan *Term Frequency-Inverse Document Frequency (TF-IDF)* dan *Word2Vec* sebagai metode penelitiannya. TF-IDF merupakan metode untuk menentukan frekuensi relatif dari setiap kata-kata (Rahayu dkk., 2022). Sementara itu *Word2Vec* merupakan model yang mampu merepresentasikan hubungan semantik antar kata dengan baik (Hendrawan dkk., 2022).

Bidirectional Long Short-Term Memory (BiLSTM) digunakan sebagai metode guna mengklasifikasi data yang ada dalam penelitian ini. BiLSTM adalah jenis *neural network* yang menggunakan arsitektur LSTM pada kedua arah input, yaitu maju dan mundur untuk menghasilkan *output*. Pendekatan ini memungkinkan model untuk mempelajari konteks baik dari masa lalu maupun masa depan dari input, yang berguna dalam banyak tugas pemrosesan bahasa alami (NLP), seperti pemrosesan teks dan analisis sentimen. BiLSTM menggunakan dua unit LSTM yang terhubung secara terbalik: satu LSTM yang memproses input secara maju, dan satu LSTM lagi yang memproses input secara mundur (Feldman & Sanger, 2006; Lin dkk., 2023). Dalam proses pelatihan, BiLSTM belajar untuk memprediksi *output* berdasarkan konteks yang terdapat pada kedua arah input. Model ini mampu menangkap informasi kontekstual pada level yang lebih tinggi dan memiliki kemampuan untuk menangani masalah kesulitan jangka panjang (*long-term dependencies*) pada data yang kompleks.

Pengujian dilakukan untuk menguji performa model dalam melakukan klasifikasi. Pada proses pengujian, dilakukan *split validation* yakni data dibagi menjadi data training dan data testing (Nurhuda dkk., 2016). Data training merupakan data yang telah didapatkan dan digunakan untuk pembentukan model yang dapat memprediksi apakah *tweet* tersebut masuk ke kategori positif atau negatif.

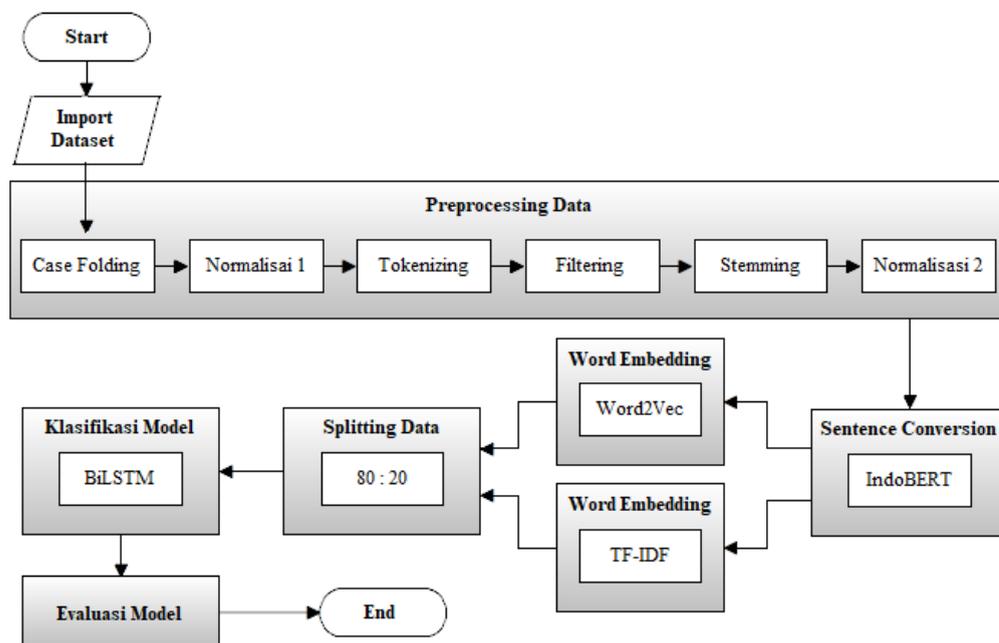
Sedangkan data testing merupakan data yang dipakai untuk menguji keakuratan model klasifikasi. Kedua data akan dilakukan pembagian rasio sesuai yang telah ditentukan yaitu 80:20.

Performa dari prediksi yang dilakukan oleh algoritma BiLSTM terhadap TF-IDF dan Word2Vec dievaluasi menggunakan teknik *Confusion Matrix*. *Confusion Matrix* memberikan informasi perbandingan nilai aktual dengan nilai prediksi dan terdapat empat nilai yang dihasilkan. Empat nilai tersebut dapat dijelaskan sebagai berikut : TP - *True Positive* adalah data aktual yang bernilai positif diprediksi benar, TN - *True Negative* adalah data aktual yang bernilai negatif diprediksi benar, FP - *False Positive* adalah data aktual yang bernilai negatif diprediksi positif, FN - *False Negative* adalah data aktual yang bernilai positif diprediksi negatif (Munawar & Riadi, 2019; Rangkuti dkk., 2018; Wijaya dkk., 2021).

Sedangkan matriks evaluasi yang digunakan pada penelitian ini, yaitu: *Accuracy* adalah, rasio prediksi benar (positif dan negatif) dari jumlah seluruh data, *Precision* adalah, rasio prediksi benar yang positif dengan perbandingannya terhadap jumlah seluruh hasil prediksi positif, *Recall* adalah, rasio prediksi benar positif dengan perbandingannya terhadap seluruh data yang sudah benar positif, *F1 Score* adalah, perhitungan rerata antara *precision* dan *recall* (Fauziah dkk., 2018; Lamasigi, 2021).

3. Hasil dan Pembahasan

Pada penelitian ini, analisis sentimen opini publik dilakukan dengan menggunakan metode *deep learning* BiLSTM dengan ekstraksi fitur TF-IDF dan Word2Vec. Model penelitian yang akan dibangun dalam penelitian ini dapat dilihat pada Gambar 1.



Gambar 1: Model Penelitian

Sumber: Gambar diolah 2023

Model penelitian pada Gambar 1. dimulai dengan import *dataset*, lalu akan dilakukan *preprocessing* data seperti *case folding*, menghapus angka, link, hingga pembobotan kata Word2Vec dan TF-IDF. Proses pemisahan label dilakukan untuk memperkecil kelas pada label, sehingga menjadi 2 label (positif dan negatif). Proses pemisahan data 80%:20 proporsi itu dipilih karena merupakan angka yang ideal. Terakhir, proses pelatihan dan pengujian model BiLSTM untuk mendapatkan hasil.

Membaca dataset merupakan langkah awal yang dilakukan dalam penelitian ini. Dataset yang digunakan merupakan kumpulan data teks (*text mining*) bahasa Indonesia yang terdapat di Twitter terkait Perpanjangan Masa Jabatan Pimpinan KPK dan UU KPK dengan jumlah data yang digunakan

sebanyak 11.448 data, terdiri dari 5.823 data label negatif dan 5.625 data label positif. Berikut adalah contoh dari dataset yang dapat dilihat pada Tabel 2

Tabel 2. Dataset Penelitian

Ulasan	Label
putus panjang jabat kpk rakyat buat apa keluar putus rugi hak demokrasi rakyat cegah lebih dulu informasi bagi	negative
kpk memang revisi harus dasar semangat orientasi kuat peran lembaga kpk serta instrumen kerja bukan justru balik	positive

Sumber: Data diolah 2023

Dataset pada Tabel 2. merupakan contoh data berlabel positif dan negative dari hasil pengolahan yang menggunakan metode IndoBERT. Setelah mendapat hasil dari IndoBERT perlu dilakukan pengecekan secara manual untuk menghindari hasil yang bias dari pelabelan dataset.

Preprocessing merupakan langkah penting untuk meningkatkan efisiensi data (Hendrawan dkk., 2022; Sihombing dkk., 2021). Sebelum melakukan *preprocessing*, *drop* satu data ulasan negatif agar total dataset menjadi seimbang antara data ulasan positif dengan data ulasan negatif. Setelah melakukan hal tersebut, melanjutkan ke tahapan *preprocessing* yang dapat dilihat pada Gambar 2.

Gambar 2: *Preprocessing*

Sumber: Gambar diolah 2023

Pada tahapan *case folding*, semua huruf akan diubah kedalam bentuk huruf kecil. Berikut adalah contoh dari penerapan tahapan *case folding* yang dapat dilihat pada Tabel 3.

Tabel 3. Penerapan *Case Folding*

Teks	Hasil <i>Case Folding</i>
@papajenolica @watch_asra @dennyindrayana Jikapapajenolica watchasra dennyindrayana jika sudah diputuskan oleh MK seperti perpanjangan sudah diputuskan oleh mk seperti perpanjangan jabatan KPK, rakyat bisa berbuat apa? Untuk itu jabatan kpk rakyat bisa berbuat apa untuk itu sebelum MK mengeluarkan putusan yg merugikan sebelum mk mengeluarkan putusan yg hak demokrasi rakyat, dicegah lebih dulu dengan merugikan hak demokrasi rakyat dicegah lebih informasi yg dibagikan Denny.	watchasra dennyindrayana jika sudah diputuskan oleh mk seperti perpanjangan jabatan kpk rakyat bisa berbuat apa untuk itu sebelum mk mengeluarkan putusan yg merugikan hak demokrasi rakyat, dicegah lebih dulu dengan informasi yg dibagikan denny

Sumber: Data diolah 2023

Case Folding merupakan tahapan yang mengubah kata-kata pada teks menjadi *lowercase*. Number, Link, dan punctuation removal merupakan tahapan untuk menghapus angka, URL atau link serta simbol-simbol yang tidak.

Tabel 4. Penerapan Normalisasi 1

Hasil Normalisasi 1
papajenolica watchasra dennyindrayana jika sudah diputuskan oleh mk seperti perpanjangan jabatan kpk rakyat bisa berbuat apa untuk itu sebelum mk mengeluarkan putusan yg merugikan hak demokrasi rakyat dicegah lebih dulu dengan informasi yg dibagikan denny

Sumber: Data diolah 2023

Selanjutnya tahap Normalisasi 1 merupakan tahapan untuk memperbaiki kata menjadi kata normal. Normalisasi pada penelitian ini menggunakan kamus alay dari indo-bercamart dimana kamus alay ini akan mengubah kata-kata atau bahasa slang menjadi kata baku. Berikut adalah contoh penerapan tahapan tokenisasi yang menggunakan data hasil dari Tabel 3.

Tabel 5. Penerapan *Tokenizing*

Hasil *Tokenizing*

['papajenolica', 'watchasra', 'dennyindrayana', 'jika', 'sudah', 'diputuskan', 'oleh', 'mk', 'seperti', 'perpanjangan', 'jabatan', 'kpk', 'rakyat', 'bisa', 'berbuat', 'apa', 'untuk', 'itu', 'sebelum', 'mk', 'mengeluarkan', 'putusan', 'yg', 'merugikan', 'hak', 'demokrasi', 'rakyat', 'dicegah', 'lebih', 'dulu', 'dengan', 'informasi', 'yg', 'dibagikan', 'denny']

Sumber: Data diolah 2023

Setelah melakukan Normalisasi 1, tahap selanjutnya adalah melakukan *Tokenizing* yang bertujuan untuk memisahkan setiap kata yang menyusun suatu dokumen. Umumnya setiap kata terpisahkan dengan kata yang lain oleh karakter spasi, sehingga proses tokenisasi mengandalkan karakter spasi pada dokumen untuk melakukan pemisahan kata (Kane dkk., 2016). Berikut adalah hasil dari penerapan *Tokenizing* yang menggunakan data hasil dari Tabel 4.

Tabel 6. Penerapan Filtering (*Stopword Removal*)

Hasil Filtering (*Stopword Removal*)

['papajenolica', 'watchasra', 'dennyindrayana', 'diputuskan', 'mk', 'perpanjangan', 'jabatan', 'kpk', 'rakyat', 'berbuat', 'apa', 'mk', 'mengeluarkan', 'putusan', 'yg', 'merugikan', 'hak', 'demokrasi', 'rakyat', 'dicegah', 'lebih', 'dulu', 'informasi', 'yg', 'dibagikan', 'denny']

Sumber: Data diolah 2023

Setelah melalui proses *Tokenizing* maka kalimat tersebut menjadi sekumpulan *array* yang setiap selnya berisi kata-kata yang ada pada kalimat tersebut. Selanjutnya dilakukan proses Filtering (*Stopword Removal*) pada tahap ini dilakukan dengan menghapus karakter khusus pada dataset atau kata yang tidak memiliki arti, adapun karakter khusus yang dihapus adalah tanda baca (poin (.), koma (,), tanda tanya, tanda seru, angka numerik dan karater lainnya. Berikut adalah hasil dari penerapan *stopword removal* yang menggunakan data hasil dari Tabel 5.

Tabel 7. Penerapan *Stemming*

Hasil *Stemming*

['papajenolica', 'watchasra', 'dennyindrayana', 'putus', 'mk', 'panjang', 'jabat', 'kpk', 'rakyat', 'buat', 'apa', 'mk', 'keluar', 'putus', 'yg', 'rugi', 'hak', 'demokrasi', 'rakyat', 'cegah', 'lebih', 'dulu', 'informasi', 'yg', 'bagi', 'denny']

Sumber: Data diolah 2023

Setelah melakukan *stopword removal*, tahap selanjutnya adalah melakukan *Stemming* yang bertujuan untuk melakukan konversi kata menjadi kata dasar berdasarkan imbuhan (“komentar” menjadi “komentar”). Berikut adalah hasil dari penerapan *Stemming* yang menggunakan data hasil dari Tabel 6.

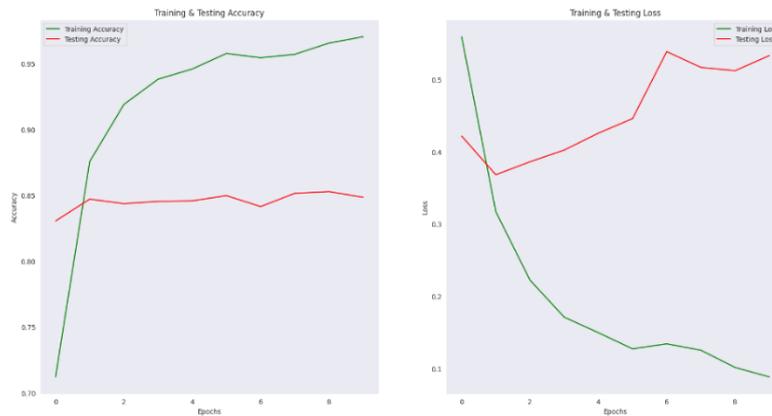
Tabel 8. Penerapan Normalisasi 2

Hasil Normalisasi 2

['putus', 'panjang', 'jabat', 'kpk', 'rakyat', 'buat', 'apa', 'keluar', 'putus', 'rugi', 'hak', 'demokrasi', 'rakyat', 'cegah', 'lebih', 'dulu', 'informasi', 'bagi']

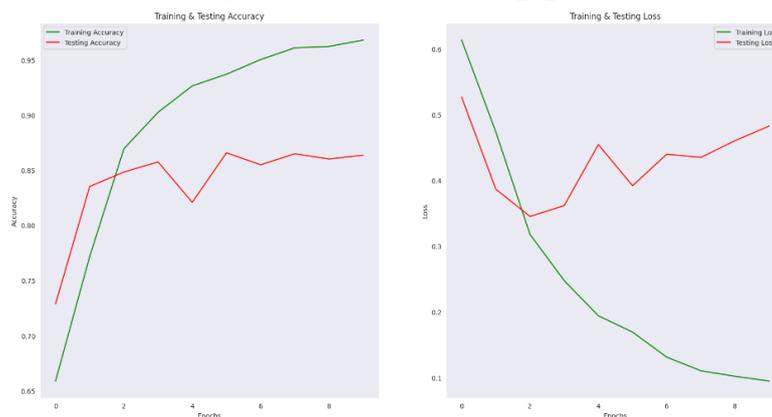
Sumber: Data diolah 2023

Terakhir dilakukan proses Normalisasi 2 pada penelitian ini normalisasi dilakukan sebanyak dua kali dengan tujuan yang sama dengan normalisasi 1 yaitu menyesuaikan kata yang tidak sesuai dengan KBBI. Untuk Normalisasi 2 ini menggunakan kamus indonesian-word. Berikut adalah hasil dari penerapan *Stemming* yang menggunakan data hasil dari Tabel 7.



Gambar 4: Model BiLSTM dengan TF-IDF

Sumber: Gambar diolah 2023

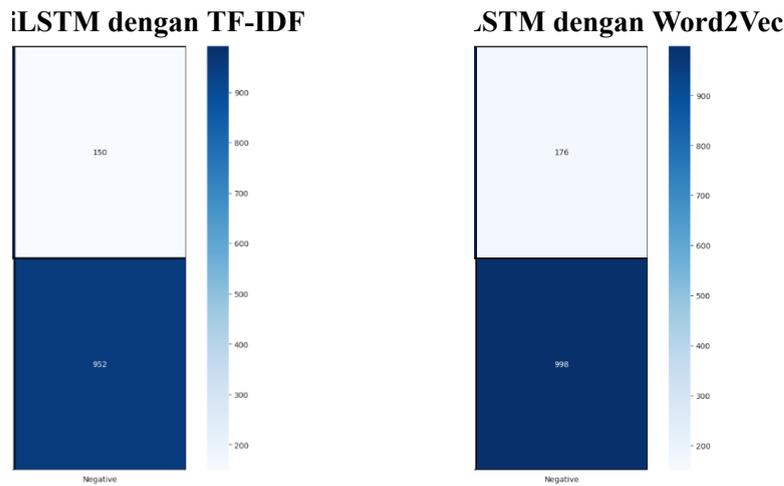


Gambar 5: Model BiLSTM dengan Word2Vec

Sumber: Gambar diolah 2023

Berdasarkan Gambar 4. dan Gambar 5. diatas hasil yang paling baik ditunjukkan pada Gambar 5. Model BiLSTM dengan Word2Vec menghasilkan nilai akurasi terendah pada data validasi diepoch pertama dengan nilai sebesar 0.6588 dan nilai *loss* tertinggi pada data training menghasilkan 0.5266. Pada *epoch* ke 10 model BiLSTM dengan Word2Vec menghasilkan akurasi yang baik dibanding dengan model BiLSTM dengan TF-IDF. Perbandingan terlihat bahwa model sudah cukup baik dengan menunjukkan akurasi data training dengan validasi tidak berbeda jauh sehingga tidak mengalami *over fitting*.

Selanjutnya model akan diuji kepercayaannya pada tahap pengujian dengan data testing Gambar 4 dan Gambar 5. Setelah model menghasilkan prediksinya di masing-masing kelas dilakukan perhitungan tingkat kepercayaan model dengan melihat akurasi, presisi, *recall*, dan *f1-score*. Tujuannya untuk mengetahui seberapa dapat dipercaya model dalam memprediksi kelas. Hasil pengujian data testing ditunjukkan dengan *confusion matrix* seperti Tabel 9 dan Tabel 11 dibawah ini menunjukkan hasil *Confusion Matrix* masing – masing model.



Gambar 6: Visualisasi *Confusion Matrix*
 Sumber: Gambar diolah 2023

Tabel 9. *Confusion Matrix* Model BiLSTM (TF-IDF)

aktual	Prediksi	
	tif	itif
Negatif	150	0
Positif	952	8

Sumber: Data diolah 2023

Tabel 10. *Confusion Matrix* Model BiLSTM (Word2Vec)

aktual	Prediksi	
	tif	itif
Negatif	176	6
Positif	998	2

Sumber: Data diolah 2023

Bedasarkan model yang telah dibuat dimana data test yang berlabel negatif benar diprediksi oleh model BiLSTM dengan TF-IDF sebesar 991, sedangkan pada model BiLSTM dengan Word2Vec sebesar 965 dan salah memprediksi pada model BiLSTM dengan TF-IDF sebesar 197, sedangkan pada model BiLSTM dengan Word2Vec sebesar 151. Begitupula dengan ulasan yang berlabel positif dapat benar diprediksi oleh model BiLSTM dengan TF-IDF sebesar 952, sedangkan pada model BiLSTM dengan Word2Vec sebesar 998. Selanjutnya setelah model menghasilkan prediksinya di masing-masing kelas dilakukan perhitungan tingkat kepercayaan model dengan melihat akurasi, presisi, *recall* dan *f1-score*. Tujuannya untuk mengetahui seberapa besar persentase akurasi untuk dapat dipercaya sebagai model dalam memprediksi kelas.

Tabel 11. Metode *Confusion Matrix*

		Observed	
		true	alse
Predicted Class	True	Positive (TP)	Positive (FP)
	False	Negative (FN)	Negative (TN)

Sumber: (Hamel, 2008)

dengan:

TP adalah *True Positif*, yaitu jumlah data positif yang terklasifikasi dengan benar oleh sistem.

TN adalah *True Negatif*, yaitu jumlah data negatif yang terklasifikasi dengan benar oleh system.

FN adalah *False Negatif*, yaitu jumlah data negatif namun terklasifikasi salah oleh sistem.

FP adalah *False Positif*, yaitu jumlah data positif namun terklasifikasi salah oleh sistem.

Nilai akurasi merupakan perbandingan antara data yang terklasifikasi benar dengan keseluruhan data. Nilai akurasi dapat diperoleh dengan persamaan :

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

Nilai presisi menggambarkan jumlah data kategori positif yang diklasifikasi secara benar dibagi dengan total data yang diklasifikasi positif, Presisi dapat diperoleh dengan persamaan :

$$Presisi = \frac{TP}{TP+FP} \times 100$$

Nilai *recall* menunjukkan beberapa persen data kategori positif yang terklasifikasi dengan benar oleh sistem, *recall* dapat diperoleh dengan persamaan :

$$Recall = \frac{TP}{TP+FN} \times 100\%$$

Sementara itu, *f1-score* merupakan suatu nilai yang digunakan untuk mengetahui nilai kinerja dari algoritma. *f1-score* dapat diperoleh dengan persamaan:

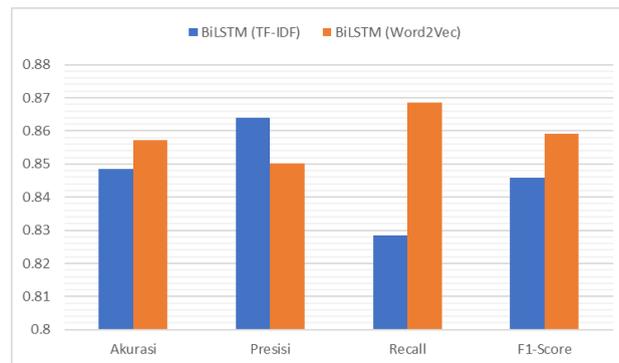
$$f1 - score = 2 \times \frac{presisi \times recall}{presisi+recall}$$

Hasil Perbandingan model dapat dilihat pada Gambar 7 dan Tabel 12. Dari penelitian ini didapati bahwa model BiLSTM dengan Word2Vec lebih unggul dalam perbandingan evaluasi performansi.

Tabel 12. Perbandingan Evaluasi Performansi

Evaluasi Performansi	BiLSTM	
	DF	2Vec
Akurasi	472	4205
Presisi	384	4085
Recall	547	4581
F1-Score	346	4234

Sumber: Data diolah 2023



Gambar 7: Perbandingan Evaluasi Performansi

Sumber: Gambar diolah 2023

Pada Gambar 7. Menunjukkan visualisasi perbandingan performansi terbaik dari Tabel 12. Pada hasil visualisasi tersebut dapat diketahui bahwa penggunaan *term weighting* Word2Vec memiliki akurasi *recall* 82.85%, dan F1-score dengan nilai yang lebih baik disbanding dengan penggunaan *term weighting* TF-IDF. Sementara itu *precision* dari hasil pengujian menggunakan *term weighting* TF-IDF memiliki nilai lebih baik dibanding *term weighting* Word2Vec. Secara keseluruhan hasil visualisasi menunjukkan bahwa penggunaan Word2Vec pada model BiLSTM lebih baik.

4. Kesimpulan

Pada penelitian ini tentang Analisis Sentimen Opini Publik Pada Twitter Menggunakan Metode BiLSTM dengan Pembobotan TF-IDF dan Word2Vec yang bertujuan untuk mengetahui perbandingan tingkat akurasi *term weighting* TF-IDF dan Word2Vec menggunakan model BiLSTM dalam preferensi sentiment analisis opini publik terkait perpanjangan masa jabatan pimpinan KPK dan UU KPK. Dari hasil penelitian dapat diambil kesimpulan bahwa penggunaan Word2Vec lebih baik dibandingkan dengan TF-IDF pada model BiLSTM. Hasil pengujian menunjukkan bahwa TF-IDF memiliki akurasi sebesar 84.84%, *precision* 86.38%, *recall* 82.85%, dan F1-score 84.58%. dan pembobotan Word2Vec memiliki akurasi sebesar 85.72%, *precision* 85.00%, *recall* 86.85%, dan F1-score 85.92%. Dari penelitian ini didapati bahwa dari hasil Perbandingan model BiLSTM dengan Word2Vec lebih unggul dibanding dengan model BiLSTM dengan TF-IDF.

Saran untuk penelitian selanjutnya adalah untuk mencari algoritma yang bisa menangani masalah ketimpangan data agar hasil keputusan model yang dibuat tidak cenderung ke kelas yang paling dominan. Selain itu pada penelitian selanjutnya dapat menambahkan jumlah dataset yang digunakan dan melakukan perbandingan akurasi berdasarkan jumlah dataset yang lebih banyak

1. Ucapan Terimakasih

Ucapan terima kasih disampaikan kepada PJJ Magister Teknik Informatika, Universitas Amikom Yogyakarta.

2. Referensi

- Amada, M., Munawar, M., & Pilliang, M. (2023). Sentiment Analysis of Content Permenkominfo No. 5 of 2020 Using a Classification Algorithm. *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer)*, 8(2), 125–132.
- Argawati, U. (2021). *Nurul Ghufron Uji Aturan Batas Usia Pimpinan KPK*. <https://www.mkri.id/index.php?page=web.Berita&id=18747&menu=2>, Diakses pada 04 Juni 2023
- Argawati, U. (2023). *Masa Jabatan Pimpinan KPK Lima Tahun*. <https://www.mkri.id/index.php?page=web.Berita&id=19196>, Diakses pada 04 Juni 2023
- Fauziah, D. A., Maududie, A., & Nuritha, I. (2018). Klasifikasi Berita Politik Menggunakan Algoritma K-nearest Neighbor. *Berkala Sainstek*, 6(2), 106–114.
- Feldman, R., & Sanger, J. (2006). *The Text Mining Handbook*. Cambridge University Press.

- Hamel, L. (2008). *Model Assessment with ROC Curves. The Encyclopedia of Data Warehousing and Mining* (2nd Edition). Idea Group Publisher.
- Hendrawan, I. R., Utami, E., & Hartanto, A. D. (2022). Comparison of Naïve Bayes Algorithm and XGBoost on Local Product Review Text Classification. *Edumatic: Jurnal Pendidikan Informatika*, 6(1), 143–149.
- Kane, S. N., Mishra, A., & Dutta, A. K. (2016). *Preface: International conference on recent trends in physics (ICRTP 2016)*. 755(1), 0–5.
- Khurniawan, F. S., & Ruldeviyani, Y. (2020). *Twitter Sentiment Analysis: Case Study On The Revision of The Indonesia's Corruption Eradication Commission (KPK) Law 2019*. 1–6.
- Lamasigi, Z. Y. (2021). DCT Untuk Ekstraksi Fitur Berbasis GLCM Pada Identifikasi Batik Menggunakan K-NN. *Jambura Journal of Electrical and Electronics Engineering*, 3(1), 1–6.
- Lin, H., Zhang, S., Li, Q., Li, Y., Li, J., & Yang, Y. (2023). A new method for heart rate prediction based on LSTM-BiLSTM-Att. *Measurement*, 207, 112384.
- Munawar, M. U., & Riadi, Y. U. (2019). Sistem Pendeteksi Berita Hoax Di Media Sosial Dengan Teknik Data Mining Scikit Learn. *Jurnal Ilmu Komputer*, 4(2), 173–179.
- Nurhuda, F., Sihwi, S. W., & Doewes, A. (2016). Analisis sentimen masyarakat terhadap calon Presiden Indonesia 2014 berdasarkan opini dari Twitter menggunakan metode Naive Bayes Classifier. *ITSmart: Jurnal Teknologi dan Informasi*, 2(2), 35–42.
- Parasati, W., Bachtiar, F. A., & Setiawan, N. Y. (2020). Analisis Sentimen Berbasis Aspek pada Ulasan Pelanggan Restoran Bakso President Malang dengan Metode Naïve Bayes Classifier. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN*, 2548, 964.
- Putri, Z. (2023). *Pasal Ini yang Digugat Nurul Ghufron demi Jabatan 5 Tahun Pimpin KPK*. <https://news.detik.com/berita/d-6724584/pasal-ini-yang-digugat-nurul-ghufron-demi-jabatan-5-tahun-pimpin-kpk>. Diakses pada 04 Juni 2023
- Rahayu, S., Yumarlin, M., Bororing, J. E., & Hadiyat, R. (2022). Implementasi Metode K-Nearest Neighbor (K-NN) untuk Analisis Sentimen Kepuasan Pengguna Aplikasi Teknologi Finansial FLIP. *Edumatic: Jurnal Pendidikan Informatika*, 6(1), 98–106.
- Rangkuti, F. R. S., Fauzi, M. A., Sari, Y. A., & Sari, E. D. L. (2018). Analisis Sentimen Opini Film Menggunakan Metode Naïve Bayes dengan Ensemble Feature dan Seleksi Fitur Pearson Correlation Coefficient. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(12), 6354–6361.
- Reiki, M. K. A., Sibaroni, Y., & Setiawan, E. B. (2022). Comparison of Term Weighting Methods in Sentiment Analysis of the New State Capital of Indonesia with the SVM Method. *International Journal on Information and Communication Technology (IJoICT)*, 8(2), 53–65.
- Sandryan, M. K., Rahayudi, B., & Ratnawati, D. E. (2021). Analisis Sentimen Pada Media Sosial Twitter Terhadap Undang-Undang Cipta Kerja Menggunakan Algoritma Backpropagation dan Term Frequency-Inverse Document Frequency. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN*, 2548, 964X.
- Sihombing, L. O., Hannie, H., & Dermawan, B. A. (2021). Sentimen Analisis Customer Review Produk Shopee Indonesia Menggunakan Algoritma Naïve Bayes Classifier. *Edumatic: Jurnal Pendidikan Informatika*, 5(2), 233–242.
- Sudrajat, T., & Rohida, L. (2022). *Efek Media Massa dalam Pembentukan Opini Publik di Masa Pandemi Covid-19*. 4, 519–524.
- Sukma, E. A., Hidayanto, A. N., Pandesenda, A. I., Yahya, A. N., Widharto, P., & Rahardja, U. (2020). *Sentiment Analysis of the New Indonesian Government Policy (Omnibus Law) on Social Media Twitter*. 153–158.
- Wibowo, A. R., Nidya, N., & Rahma, A. F. (2020). Analisis Sentimen Hashtag 'Dirumahaja' Saat Pandemi Covid-19 Di Indonesia Menggunakan NLP. *J. Inform. dan Sist. Inf*, 1(2), 343–353.
- Wijaya, T. N., Indriati, R., & Muzaki, M. N. (2021). Analisis Sentimen Opini Publik Tentang Undang-Undang Cipta Kerja Pada Twitter. *Jambura Journal of Electrical and Electronics Engineering*, 3(2), 78–83.
- Wilie, B., Vincentio, K., Winata, G. I., Cahyawijaya, S., Li, X., Lim, Z. Y., Soleman, S., Mahendra, R., Fung, P., & Bahar, S. (2020). IndoNLU: Benchmark and resources for evaluating Indonesian natural language understanding. *arXiv preprint arXiv:2009.05387*.