

Sentiment Analysis Terkait Pemindahan Ibu Kota Indonesia Menggunakan Metode *Random Forest* Berdasarkan *Tweet* Warga Negara Indonesia

Sentiment Analysis Related to Transportation of Indonesian Capital City Using Random Forest Method Based On Tweet Of Indonesian Citizens

Tanti Cahya Herdiyani¹, Achmad Udin Zailani²

^{1,2}Program Studi Teknik Informatika, Fakultas Teknik, Universitas Pamulang

E-mail: ¹ htanticahya@gmail.com, ²dosen00270@unpam.ac.id

Abstrak

Pemindahan ibu kota Indonesia ke Provinsi Kalimantan Timur menuai beberapa pendapat warga negara di *Twitter*. Penelitian *sentiment analysis* dilakukan untuk mengetahui respon *tweet* terkait pemindahan ibu kota Indonesia setelah disahkannya Undang-Undang IKN serta perbandingan jumlah *tweet* positif, negatif dan netral. Selain itu juga dapat mengetahui tingkat kinerja yang dihasilkan oleh metode *Random Forest* dengan berbagai skenario yang diuji. Pada penelitian sebelumnya yang serupa dilakukan menggunakan metode *naïve bayes*, *lexicon*, *SVM*, *NBC*, *neighbor weighted K-NN*, *random forest* dan *logistic regression*. Pada penelitian tersebut menggunakan *feature extraction* yang dikombinasikan pada setiap model. Dari penelitian tersebut menghasilkan tingkat kinerja yang berbeda-beda dari model yang digunakan. *Sentiment analysis* termasuk *text mining* yang merupakan implementasi dari NLP. Metode TF-IDF digunakan untuk mengubah kata menjadi numerik dengan melakukan pembobotan. *Random forest* merupakan metode pembelajaran melalui pembangunan pohon (*tree*) keputusan dengan melalui serangkaian pelatihan. Metodologi penelitian yang dilakukan yaitu studi pustaka dan analisis dokumen dari 1.639 data *tweet* warga negara Indonesia. Hasil dari penelitian ini respon dari *tweet* diklasifikasikan menjadi positif, negatif dan netral. Perbandingan jumlah *tweet* positif didapatkan lebih tinggi dibandingkan dengan *tweet* negatif dan netral. Dan tingkat kinerja yang dihasilkan yaitu pada *accuracy* mencapai 76%, *recall* sebesar 70%, *precision* sebesar 69% dan *f1-score* sebesar 69%.

Kata kunci: *Sentiment Analysis, Random Forest, TF-IDF, Text Mining, NLP*

Abstract

The move of Indonesia's capital city to the province of East Kalimantan has drawn several opinions from citizens on *Twitter*. *Sentiment analysis* research was conducted to determine the response of tweets related to the relocation of the Indonesian capital after the enactment of the IKN Law and compare the number of positive, negative, and neutral tweets. In addition, it can also determine the level of performance generated by the *Random Forest* method with various tested scenarios. In a similar previous study, using the *nave Bayes* method, *lexicon*, *SVM*, *NBC*, *neighbor weighted K-NN*, *random forest*, and *logistic regression*. In this study, *feature extraction* is combined in each model. These studies resulted in different levels of performance from the model used. *Sentiment analysis* includes *text mining* which is an implementation of NLP. The TF-IDF method is used to convert words to numeric by weighting them. *Random forest* is a learning method through constructing a decision tree through a series of training. The research methodology is a literature study and document analysis from 1,639 tweets of Indonesian citizens. This study's results were that the tweets' responses were classified as positive, negative, and neutral. The comparison of the number of positive tweets is higher than that of negative and neutral tweets. And the level of performance produced is 76% accuracy, 70% recall, 69% precision, and 69% f1-score.

Keywords: *Sentiment Analysis, Random Forest, TF-IDF, Text Mining, NLP*

1. PENDAHULUAN

Jakarta merupakan ibu kota Indonesia yang ditetapkan berdasarkan hukum melalui Undang-Undang Nomor 10 Tahun 1964 yang disahkan pada tanggal 31 Agustus 1964 oleh Presiden Republik Indonesia saat itu Ir. Soekarno [1]. Sebagai ibu kota negara, Jakarta merupakan daerah yang memiliki peranan sebagai pusat pemerintahan negara, pusat bisnis, pusat keuangan, pusat pendidikan, pusat industri pariwisata, pusat perdagangan dan jasa, bandar udara dan pelabuhan laut terbesar di Indonesia [2]. Jakarta mengalami pertumbuhan yang sangat pesat, pembangunan yang tiada henti dan penambahan penduduk di Jakarta tiap tahun kian meningkat menjadikan Jakarta sebagai kota yang sangat maju di Indonesia. Dibalik kemajuan kota yang dialami oleh Jakarta, menimbulkan beragam masalah seperti kepadatan penduduk, kemacetan, kesenjangan sosial, banjir, polusi udara, krisis air bersih dan masalah lainnya.

Dalam pidato yang disampaikan Bapak Presiden Ir. H. Joko Widodo melalui kanal *Youtube* resmi Sekretariat Presiden pada tanggal 26 Agustus 2019 [2]. Beliau menyampaikan lokasi ibu kota yang baru yaitu di Provinsi Kalimantan Timur yang letaknya di sebagian besar Kabupaten Penajam Paser Utara dan Kabupaten Kutai Kartanegara. Alasan pemindahan ibu kota dikarenakan beban kota Jakarta sebagai ibu kota sudah terlalu berat, bukan hanya dari sisi fungsi tetapi dari pertumbuhan penduduk yang juga kian meningkat, jumlah penduduk di pulau Jawa berkisar 56% dari total seluruh warga di Indonesia. Menurut data dari Badan Pencatatan Statistik jumlah penduduk di Daerah Khusus Ibu Kota Jakarta pada tahun 2021 berjumlah 10.609.681 jiwa [3]. Atas dasar kepadatan penduduk ini yang menimbulkan permasalahan seperti kemacetan yang susah ditangani, kesenjangan sosial, ekonomi tidak merata, polusi udara, pertumbuhan urbanisasi, krisis air bersih, ancaman bahaya bencana alam banjir, gempa bumi-tsunami [2]. Alasan lainnya dipindahkan ke lokasi tersebut karena letak daerahnya yang strategis berada di tengah-tengah Indonesia yang diharapkan ekonomi akan merata setiap daerah, lokasinya berdampingan dengan kota-kota yang sudah berkembang yaitu Balikpapan dan Samarinda, infrastruktur yang relatif lengkap, resiko bencana minimal, ditambah pada lokasi tersebut terdapat lahan yang dikuasai pemerintah seluas 180.000 hektar.

Pemindahan ibu kota merupakan rencana pembangunan jangka menengah nasional yang dilaksanakan secara bertahap sampai dengan tahun 2045 [2]. Tahap awal pembangunan dilakukan dalam kurun waktu 2022-2024 dengan menghabiskan anggaran kurang lebih sejumlah 466T. Pemindahan ibu kota ini diperkuat dengan disahkannya Undang-Undang Republik Indonesia Nomor 3 Tahun 2022 tentang Ibu Kota Negara, pada tanggal 15 Februari 2022 [4]. Semenjak beredarnya berita rencana pemindahan ibu kota Indonesia, beragam pendapat positif, negatif dan netral disampaikan Warga Negara Indonesia. Mereka memberikan pendapatnya baik secara langsung maupun di media sosial seperti *Facebook*, *Instagram*, *Line*, *TikTok* dan salah satunya yaitu *Twitter*. *Twitter* merupakan sosial media yang memungkinkan pengguna dapat mengakses informasi yang sedang terjadi saat ini dengan cepat. *Twitter* digunakan seseorang untuk mengekspresikan pendapatnya dengan pesan yang biasa disebut *tweet*. *Tweet* yang dikirimkan dapat berupa teks, gambar maupun video.

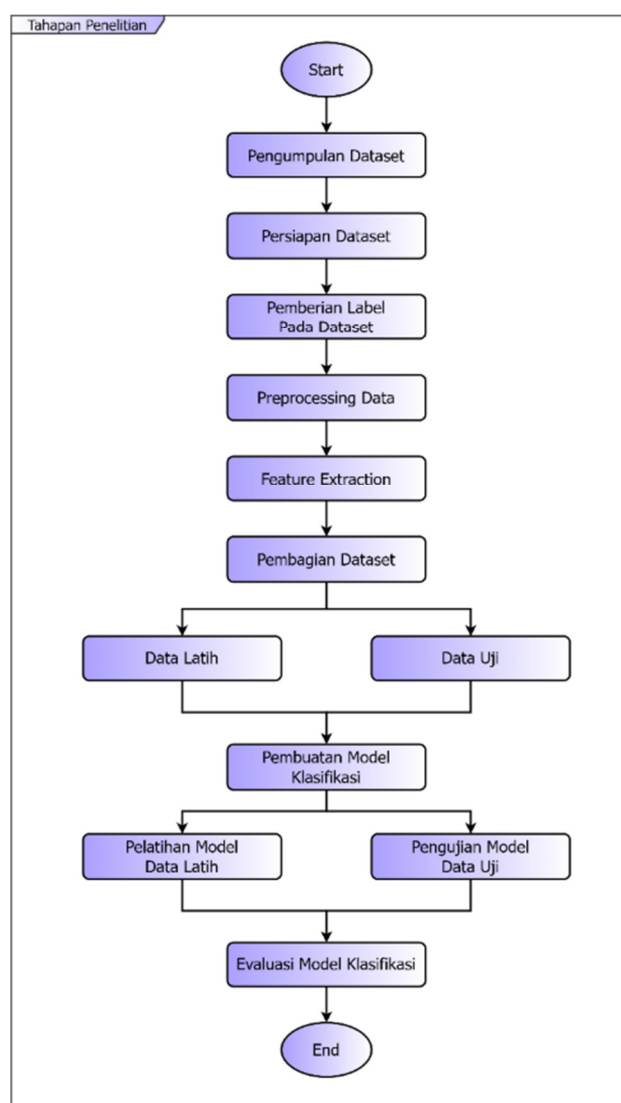
Walaupun saat ini Undang-Undang IKN telah disahkan tetapi masih menuai beberapa pendapat warga negara di *Twitter*. Pendapat tersebut berupa pendapat positif, negatif, ataupun netral akan pemindahan ibu kota Indonesia. Untuk mengetahui respon *tweet* warga negara terkait pemindahan ibu kota Indonesia setelah disahkan Undang-Undang IKN maka dilakukan penelitian *sentiment analysis*. *Sentiment analysis* merupakan suatu proses memahami, mengekstrak dan mengolah data tekstual dengan otomatis guna mendapatkan informasi sentimen yang berada pada suatu kalimat dalam opini seseorang [5]. Terdapat beberapa metode yang digunakan untuk *sentiment analysis* diantaranya *Naïve Bayes*, *Support Vector Machine*, *Lexicon*, *Neighbor Weighted K-Nearest Neighbor*, *K-Nearest Neighbor*, *Decision Tree*, *Random Forest* [6,7,8,9,10]. *Random Forest* merupakan metode pengklasifikasian terdiri dari beberapa pohon keputusan yang dibangun dengan menggunakan vektor acak [11]. Metode *Random Forest* dipilih karena proses *training* dengan metode ini jauh lebih cepat dibanding dengan

decision tree. Selain itu, pada metode *Random Forest* memiliki tingkat akurasi yang tinggi dibandingkan dengan model *Logistic Regression* pada kasus yang sama.

Selain mengetahui respon *tweet* warga negara terkait pemindahan ibu kota Indonesia, *sentiment analysis* juga dapat mengetahui perbandingan jumlah *tweet* positif, negatif dan netral terkait pemindahan ibu kota Indonesia setelah disahkannya Undang-Undang IKN. Dan juga dapat mengetahui tingkat kinerja yang dihasilkan oleh metode *Random Forest* dengan berbagai skenario yang diuji dalam *sentiment analysis* terkait pemindahan ibu kota Indonesia.

2. METODE PENELITIAN

Pada penelitian ini penulis melakukan penelitian kuantitatif yang terdiri dari beberapa tahapan yang akan dilakukan. Berikut adalah perancangan penelitian yang digambarkan dalam bentuk *flowchart* pada penelitian *sentiment analysis* terkait pemindahan ibu kota Indonesia.



Gambar 1. Tahapan Penelitian

2.1 Pengumpulan Dataset

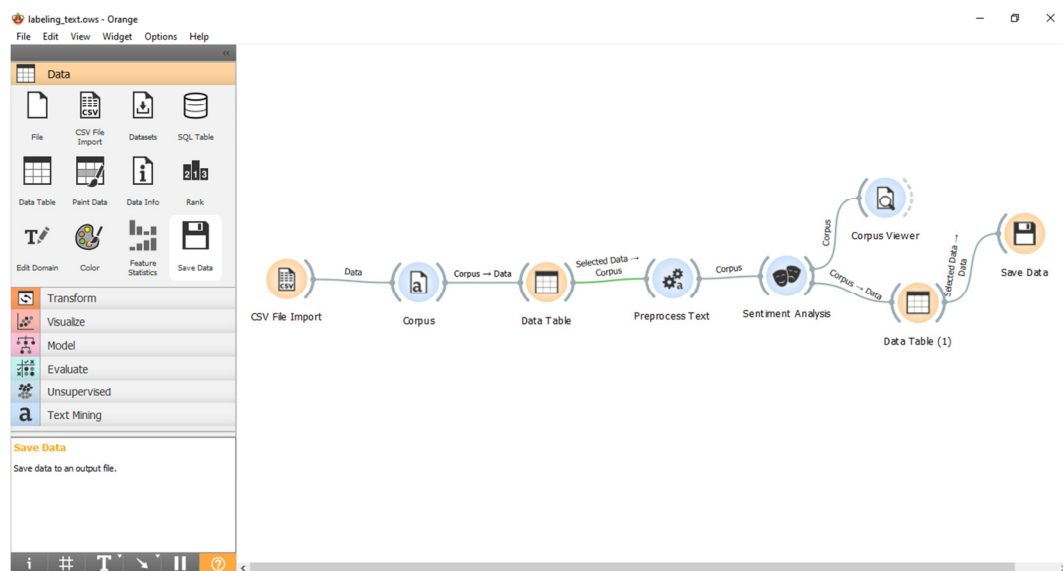
Pengumpulan dataset dilakukan dengan teknik *crawling* menggunakan program pada *python* dengan bantuan *Twitter API* untuk memperoleh *keys* dan token yang digunakan dalam proses penarikan data *tweet*. Data *tweet* yang diambil dengan teknik *crawling* dan disimpan dalam bentuk json, kemudian dikonversi ke format csv.

2.2 Persiapan Dataset

Persiapan dataset dilakukan setelah pengumpulan data. Pada penelitian ini didapati data yang telah ditarik dan dikumpulkan berjumlah 6.925 data *tweet*, dengan jangka waktu penarikan data yaitu dimulai dari tanggal 10 April 2022 sampai 07 Juni 2022. Setelah itu dilakukan proses *cleanings* yaitu pembersihan data dari karakter nama *tweet*, angka, *hashtag*, *retweet*, *link* dan penghapusan data-data kosong dan data-data yang terdeteksi sama atau duplikat. Hasil dari tahapan ini yaitu didapatkan 1.639 data *tweet* yang bisa digunakan digunakan untuk tahapan penelitian berikutnya.

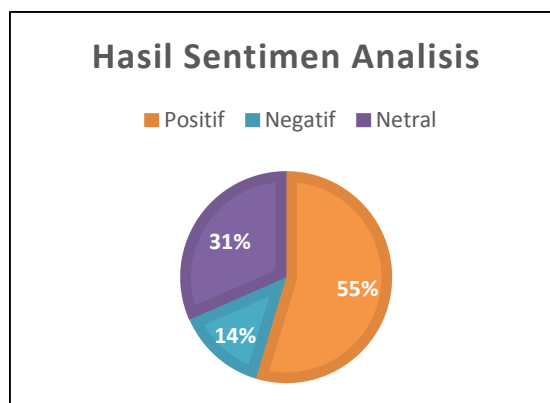
2.3 Pemberian Label Pada Dataset

Setelah dataset sudah disiapkan maka langkah berikutnya adalah pemberian label pada dataset. Pemberian label ini dilakukan menggunakan bantuan aplikasi *Orange Data Mining*. Proses pemberian label pada dataset dapat dilihat seperti pada gambar di bawah ini.



Gambar 2. Proses Pemberian Label Dataset

Pada tahapan ini dihasilkan sentimen yang bernilai positif sebanyak 896 data, untuk sentimen negatif sebanyak 518 data, dan untuk sentimen netral sebanyak 225 data. Persentase hasil *sentiment analysis* pada penelitian ini dapat dilihat pada gambar berikut.



Gambar 3. Persentase Hasil Sentimen Analisis

Berdasarkan gambar di atas dapat diketahui bahwa sentimen positif memiliki persentase sebesar 55% dari data total, untuk sentiment negatif memiliki persentase sebesar 14% dari data total, dan sentimen netral memiliki persentase sebesar 31% dari data total. Dari ketiga kelas sentiment tersebut dapat disimpulkan bahwa sentimen positif lebih banyak 41% daripada sentiment negatif dan 24% lebih banyak dari sentimen netral.

2.4 Preprocessing Data

Preprocessing data dilakukan dengan mengubah data teks yang masih sembarang atau belum terstruktur menjadi data yang berstruktur, dengan tujuan untuk memudahkan dalam pemrosesan oleh sistem. Terdapat beberapa tahapan dari *preprocessing* data yang dilakukan pada penelitian ini diantaranya sebagai berikut.

a. Case Folding

Case folding merupakan tahapan *preprocessing* data dimana semua karakter yang ada pada data *tweet* akan diubah atau dikonversi menjadi huruf kecil (*lowercase*). Sebagai contoh dari *case folding* dapat dilihat seperti tabel di bawah ini.

Tabel 1. Case Folding

Data Sebelum Dilakukan Case Folding	Data Setelah Dilakukan Case Folding
'Kemudian, pemindahan ibu kota akan menurunkan kesenjangan antar kelompok pendapatan dan indikasi ketimpangan akan menyempit.'	'kemudian, pemindahan ibu kota akan menurunkan kesenjangan antar kelompok pendapatan dan indikasi ketimpangan akan menyempit.'

b. Remove Punctuation

Remove punctuation merupakan tahapan *preprocessing* data yang dilakukan untuk menghilangkan tanda baca yang terdapat pada data *tweet*, seperti titik, koma, tanda seru, dan tanda baca lainnya. Sebagai contoh dari *remove punctuation* dapat dilihat seperti tabel di bawah ini.

Tabel 2. Remove Punctuation

Data Sebelum Dilakukan Remove Punctuation	Data Setelah Dilakukan Remove Punctuation
'Kemudian, pemindahan ibu kota akan menurunkan kesenjangan antar kelompok pendapatan dan indikasi ketimpangan akan menyempit.'	'kemudian pemindahan ibu kota akan menurunkan kesenjangan antar kelompok pendapatan dan indikasi ketimpangan akan menyempit'

c. *Tokenize*

Tokenize merupakan proses pemecahan kata dari bentuk kalimat menjadi satuan kata tunggal dari data *tweet*. Sebagai contoh dari *tokenize* dapat dilihat seperti tabel di bawah ini.

Tabel 3. *Tokenize*

Data Sebelum Dilakukan <i>Tokenize</i>	Data Setelah Dilakukan <i>Tokenize</i>
'Kemudian, pemindahan ibu kota akan menurunkan kesenjangan antar kelompok pendapatan dan indikasi ketimpangan akan menyempit.'	['kemudian', 'pemindahan', 'ibu', 'kota', 'akan', 'menurunkan', 'kesenjangan', 'antar', 'kelompok', 'pendapatan', 'dan', 'indikasi', 'ketimpangan', 'akan', 'menyempit']

d. *Stopword Removal*

Stopword Removal merupakan proses yang dilakukan untuk menghapus kata yang tidak memiliki makna apapun dari data *tweet*. Sebagai contoh dari *stopword removal* dapat dilihat seperti tabel di bawah ini.

Tabel 4. *Stopword Removal*

Data Sebelum Dilakukan <i>Stopword Removal</i>	Data Setelah Dilakukan <i>Stopword Removal</i>
'Kemudian, pemindahan ibu kota akan menurunkan kesenjangan antar kelompok pendapatan dan indikasi ketimpangan akan menyempit.'	['pemindahan', 'kota', 'menurunkan', 'kesenjangan', 'kelompok', 'pendapatan', 'indikasi', 'ketimpangan', 'menyempit']

e. *Stemming*

Stemming merupakan proses penguraian suatu kata menjadi kata dasarnya tanpa imbuhan di awal ataupun di akhir kata. Sebagai contoh dari *stemming* dapat dilihat seperti tabel di bawah ini.

Tabel 5. *Stemming*

Data Sebelum Dilakukan <i>Stopword Removal</i>	Data Setelah Dilakukan <i>Stopword Removal</i>
'Kemudian, pemindahan ibu kota akan menurunkan kesenjangan antar kelompok pendapatan dan indikasi ketimpangan akan menyempit.'	['pindah', 'kota', 'turun', 'senjang', 'kelompok', 'dapat', 'indikasi', 'timbang', 'sempit']

2.5 *Feature Extraction*

Feature Extraction yang digunakan peneliti dalam penelitian ini adalah metode TF-IDF untuk menentukan nilai frekuensi dari sebuah kata pada dataset dan frekuensi di dalam dataset. Perhitungan dari TF-IDF ini yang digunakan untuk menentukan seberapa relevan sebuah kata pada dataset yang digunakan. Dalam penggunaan metode TF-IDF peneliti menggunakan *library* yang telah disediakan oleh *scikit learn* yaitu dengan menggunakan *TfidfVectorizer*. Pada prosesnya *TfidfVectorizer* mengubah kumpulan dokumen mentah menjadi matriks fitur TF-IDF.

2.6 *Pembagian Dataset*

Dataset kemudian akan dibagi menjadi dua, yaitu data latih dan data uji. Pembagian dataset dilakukan dengan perbandingan data latih 80% yaitu sebanyak 1.311 data dan data uji 20% yaitu sebanyak 328 data. Berikut adalah potongan *code* program yang digunakan untuk pembagian dataset.

2.7 Pembuatan Model Klasifikasi

Pada tahap ini data yang telah dibagi akan menjalankan tugasnya masing-masing. Data latih digunakan untuk memberi pelatihan pada model yang digunakan, setelah itu model akan diuji dengan menggunakan data uji. Dalam penelitian ini peneliti menggunakan metode *random forest* untuk pembuatan model klasifikasi. Pada penelitian ini peneliti menggunakan *library* dari *scikit learn* yaitu *RandomForestClassifier*. Pada metode *Random Forest* digunakan untuk mencari klasifikasi fitur yang terbaik di antara subset fitur yang acak, dengan begitu dapat menghasilkan model yang lebih baik. Berikut adalah potongan *code* dalam pembuatan model klasifikasi *Random Forest*.

2.8 Evaluasi Model Klasifikasi

Tahap terakhir dari penelitian ini yaitu evaluasi model klasifikasi. Dalam melakukan evaluasi model klasifikasi peneliti akan menganalisa dan meng-evaluasi kinerja dari model klasifikasi *random forest*. Pada tahap ini menggunakan *confusion matrix* yang isinya menggambarkan kinerja dari model klasifikasi yang telah digunakan berdasarkan dari hasil prediksi dibandingkan dengan data yang benar yang disajikan dalam bentuk matriks. Evaluasi dari kinerja model yakni dari segi *accuracy*, *recall*, *precision*, *f1-score* dan *macro avg*.

3. HASIL DAN PEMBAHASAN

Hasil pengujian yaitu nilai kinerja dari metode *Random Forest* terdapat dua kali pengujian dengan menggunakan lima skenario yang berbeda dan memiliki hasil kinerja yang berbeda seperti yang dijabarkan sebagai berikut.

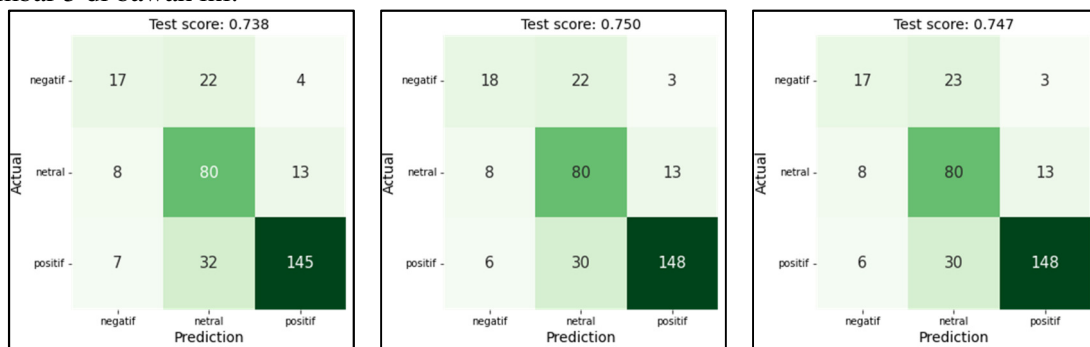
a. Pengujian Pertama

Pada tahap pengujian ini data yang digunakan dibagi menjadi dua yaitu data latih dengan presentase 80% dan data uji dengan presentase 20% dari total seluruh data. Pada pengujian pertama memiliki lima skenario yang berbeda yang bertujuan untuk mengetahui nilai *accuracy*, *recall*, *precision*, *f1-score* dan *macro avg* dari setiap pengujian. Skenario pengujian yang akan dilakukan dapat diketahui pada tabel berikut:

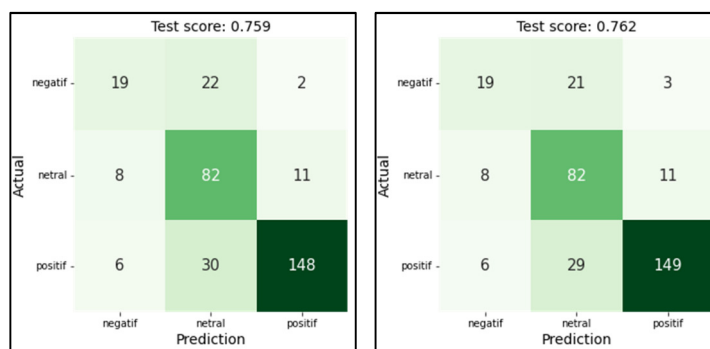
Tabel 6. Skenario Pengujian Pertama

random_state	criterion	max_features	n_estimators
0	gini	None	100
			200
			300
			400
			500

Berdasarkan hasil pengujian pertama diperoleh *confusion matrix* seperti gambar 4 dan gambar 5 di bawah ini.



Gambar 4. *Confusion Matrix* *n_estimators* 100 (kiri), 200 (tengah) dan 300 (kanan)



Gambar 5. *Confusion Matrix* n_estimators 400 (kiri) dan 500 (kanan)

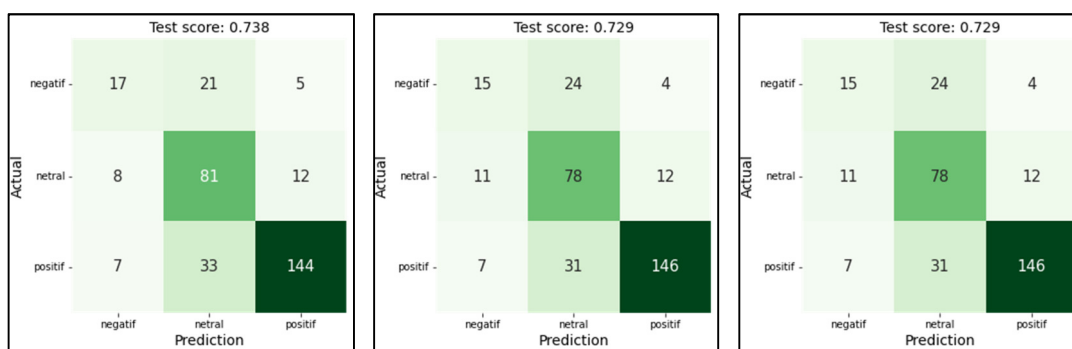
b. Pengujian Kedua

Pada pengujian kedua data yang digunakan juga dibagi menjadi dua yaitu data latih dengan presentase 80% dan data uji dengan presentase 20% dari total seluruh data. Pada pengujian kedua memiliki lima skenario yang berbeda yang bertujuan untuk mengetahui nilai *accuracy*, *recall*, *precision*, *f1-score* dan *macro avg* dari setiap pengujian. Skenario pengujian yang akan dilakukan dapat diketahui pada tabel berikut:

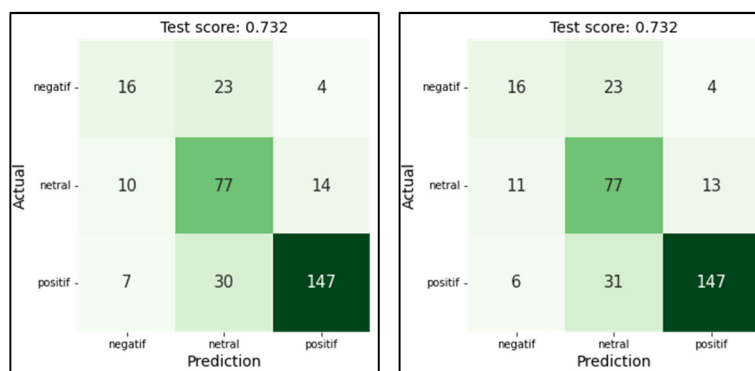
Tabel 7. Skenario Pengujian Kedua

random_state	criterion	max_features	n_estimators
0	entropy	None	100
			200
			300
			400
			500

Berdasarkan hasil pengujian pertama diperoleh *confusion matrix* seperti gambar 6 dan gambar 7 di bawah ini.

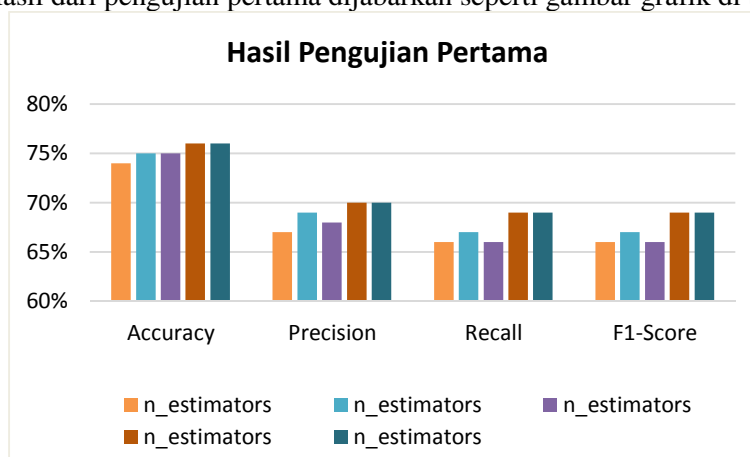


Gambar 6. *Confusion Matrix* n_estimators 100 (kiri), 200 (tengah) dan 300 (kanan)



Gambar 7. Confusion Matrix n_estimators 400 (kiri) dan 500 (kanan)

Evaluasi kinerja model klasifikasi yaitu tahapan untuk peneliti menganalisa dan melakukan evaluasi kinerja dari model klasifikasi *random forest*. Pada penelitian yang dilakukan ini Hasil dari pengujian pertama dijabarkan seperti gambar grafik di bawah ini.



Gambar 8. Grafik Hasil Pengujian Pertama

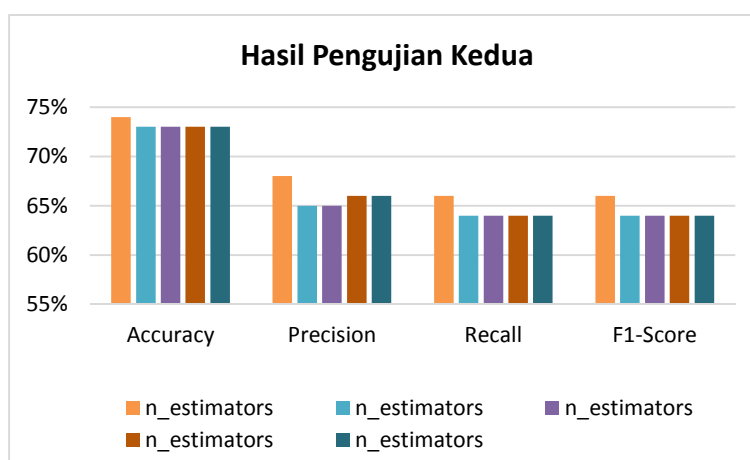
Berdasarkan grafik di atas maka didapatkan hasil analisis pada pengujian pertama yaitu sebagai berikut:

- Pada skenario pertama dengan parameter $n_estimators = 100$ menghasilkan *accuracy* sebesar 74%, *precision* sebesar 67%, *recall* sebesar 66% dan *f1-score* sebesar 66%.
- Pada skenario kedua dengan parameter $n_estimators = 200$ menghasilkan *accuracy* sebesar 75%, *precision* sebesar 69%, *recall* sebesar 67% dan *f1-score* sebesar 67%.
- Pada skenario ketiga dengan parameter $n_estimators = 300$ menghasilkan tingkat *accuracy* sebesar 75%, *precision* sebesar 68%, *recall* sebesar 66% dan *f1-score* sebesar 66%.
- Pada skenario keempat dengan parameter $n_estimators = 400$ menghasilkan *accuracy* sebesar 76%, *precision* sebesar 70%, *recall* sebesar 69% dan *f1-score* sebesar 69%.
- Pada skenario kelima dengan parameter $n_estimators = 500$ menghasilkan *accuracy* sebesar 76%, *precision* sebesar 70%, *recall* sebesar 69% dan *f1-score* sebesar 69%.

Dari kelima skenario yang dilakukan oleh peneliti pada pengujian pertama, maka didapatkan tingkat akurasi tertinggi yaitu sebesar 76% pada skenario keempat dan kelima. Sementara tingkat akurasi terendah yaitu sebesar 74% pada skenario pertama. Perhitungan *precision* tertinggi yaitu sebesar 70% pada skenario keempat dan kelima. Sementara perhitungan *precision* terendah yaitu sebesar 67% pada skenario pertama. Perhitungan *recall* tertinggi yaitu sebesar 69% pada skenario keempat dan kelima. Sementara perhitungan *recall* terendah yaitu sebesar 66% pada skenario pertama dan ketiga. Perhitungan *f1-score* tertinggi yaitu sebesar 69% pada skenario keempat dan kelima. Sementara perhitungan *f1-score* terendah

yaitu sebesar 66% pada skenario pertama dan ketiga.

Sedangkan untuk hasil dari pengujian kedua dijabarkan seperti gambar grafik di bawah ini.



Gambar 4. 1 Grafik Hasil Pengujian Kedua

Berdasarkan grafik di atas maka didapatkan hasil analisis pada pengujian kedua yaitu sebagai berikut:

- Pada skenario pertama dengan parameter $n_estimators = 100$ menghasilkan *accuracy* sebesar 74%, *precision* sebesar 68%, *recall* sebesar 66% dan *f1-score* sebesar 66%.
- Pada skenario kedua dengan parameter $n_estimators = 200$ menghasilkan *accuracy* sebesar 73%, *precision* sebesar 65%, *recall* sebesar 64% dan *f1-score* sebesar 64%.
- Pada skenario ketiga dengan parameter $n_estimators = 300$ menghasilkan *accuracy* sebesar 73%, *precision* sebesar 65%, *recall* sebesar 64% dan *f1-score* sebesar 64%.
- Pada skenario keempat dengan parameter $n_estimators = 400$ menghasilkan *accuracy* sebesar 73%, *precision* sebesar 66%, *recall* sebesar 64% dan *f1-score* sebesar 64%.
- Pada skenario kelima dengan parameter $n_estimators = 500$ menghasilkan *accuracy* sebesar 73%, *precision* sebesar 66%, *recall* sebesar 64% dan *f1-score* sebesar 64%.

Dari kelima skenario yang dilakukan oleh peneliti pada pengujian kedua, maka didapatkan tingkat akurasi tertinggi yaitu sebesar 74% pada skenario pertama. Sementara tingkat akurasi terendah yaitu sebesar 73% pada skenario kedua sampai kelima. Perhitungan *precision* tertinggi yaitu sebesar 68% pada skenario pertama. Sementara perhitungan *precision* terendah yaitu sebesar 65% pada skenario kedua dan ketiga. Perhitungan *recall* tertinggi yaitu sebesar 66% pada skenario pertama. Sementara perhitungan *recall* terendah yaitu sebesar 64% pada skenario kedua sampai kelima. Perhitungan *f1-score* tertinggi yaitu sebesar 66% pada skenario pertama. Sementara perhitungan *f1-score* terendah yaitu sebesar 64% pada skenario kedua sampai kelima.

Berdasarkan hasil analisa dan evaluasi yang dilakukan oleh penulis pada pengujian pertama dan kedua maka dapat disimpulkan bahwa tingkat akurasi tertinggi yaitu mencapai 76%, dengan nilai *recall* sebesar 70%, *precision* sebesar 69% dan *f1-score* sebesar 69% terdapat pada pengujian pertama dengan skenario keempat dan kelima. Sedangkan pada pengujian kedua tingkat akurasi tertinggi hanya mencapai 74%, dengan nilai *recall* sebesar 68%, *precision* sebesar 66% dan *f1-score* sebesar 66% pada skenario pertama.

4. KESIMPULAN DAN SARAN

4.1 Kesimpulan

Berdasarkan penelitian yang telah dilakukan oleh peneliti, maka dapat ditarik kesimpulan yaitu sebagai berikut:

- a. Respon dari warga negara Indonesia melalui *tweet* terkait pemindahan ibu kota Indonesia setelah disahkannya Undang-Undang IKN diklasifikasikan menjadi respon positif yaitu opini yang mendukung pemindahan ibu kota, respon negatif yaitu opini yang tidak mendukung pemindahan ibu kota dan netral yaitu opini yang tidak mendukung atau menolak pemindahan ibu kota.
- b. Jumlah *tweet* positif, negatif dan netral terkait pemindahan ibu kota Indonesia setelah disahkannya Undang-Undang IKN didapati untuk sentimen positif memiliki persentase sebesar 55% dari data total, untuk sentiment negatif memiliki persentase sebesar 14% dari data total, dan sentimen netral memiliki persentase sebesar 32% dari data total. Sehingga perbandingan jumlah *tweet* positif, negatif dan netral yaitu didapati pada sentimen positif lebih tinggi 41% dibandingkan dengan sentiment negatif dan 23% lebih tinggi daripada sentimen netral. Karena banyaknya persentase dari *sentiment* positif artinya warga negara Indonesia lebih banyak yang mendukung pemindahan Ibu Kota Indonesia.
- c. Tingkat kinerja paling optimal yang dihasilkan oleh metode *Random Forest* dengan berbagai skenario dalam analisis sentimen terkait pemindahan ibu kota Indonesia setelah disahkannya Undang-Undang IKN terdapat pada pengujian dengan parameter *random_state* yaitu 0, *criterion* yaitu *gini*, *max_features* yaitu *None*, *n_estimators* bernilai 400 dan 500 dengan hasil tingkat akurasi tertinggi yaitu mencapai 76%, dengan nilai *recall* sebesar 70%, *precision* sebesar 69% dan *f1-score* sebesar 69%.

4.2 Saran

Penulis menyadari bahwa terdapat beberapa hal yang harus dipertimbangkan untuk mengembangkan penelitian ini agar menjadi lebih baik lagi diantaranya sebagai berikut:

- a. Dataset yang digunakan sebaiknya ditambah lebih banyak agar dapat memaksimalkan *training* dalam pembelajaran oleh model klasifikasi.
- b. Pada *preprocessing* data sebaiknya menambahkan proses *lemmatisasi* berbahasa Indonesia agar kata-kata dapat dikelompokkan dengan kata yang serupa maknanya dan dalam memberikan *labeling* sebaiknya dilakukan oleh ahli linguistik sehingga diharapkan dapat meningkatkan kinerja dari model klasifikasi.
- c. Penggunaan fitur ekstraksi dapat ditambahkan atau menggunakan metode yang berbeda sehingga menjadi perbandingan dari hasil kinerja model yang dihasilkan pada penelitian ini.

DAFTAR PUSTAKA

- [1] S. J. D. RI, “Undang-Undang Republik Indonesia Nomor 10 Tahun 1964 Tentang Pernyataan Daerah Khusus Ibukota Jakarta Raya Tetap Sebagai Ibu Kota Negera Republik Indonesia Dengan Nama Jakarta,” Dewan Perwakilan Rakyat Republik Indonesia, 2016. [Daring]. Tersedia pada: <https://www.dpr.go.id/dokjdi/document/uu/1430.pdf>.
- [2] S. Presiden, “Presiden Jokowi Mengumumkan Lokasi Ibu Kota Baru, Istana Negara, 26 Agustus 2019,” Sekretariat Presiden, 2019. [Daring]. Tersedia pada: <https://youtu.be/9ZSAWmN9PFg>.
- [3] B. P. S. P. D. Jakarta, “Jumlah Penduduk Provinsi DKI Jakarta Menurut Kelompok Umur dan Jenis Kelamin 2019-2021,” Badan Pusat Statistik, 2022. [Daring]. Tersedia pada: <https://jakarta.bps.go.id/indicator/12/111/1/jumlah-penduduk-provinsi-dki-jakarta->

menurut-kelompok-umur-dan-jenis-kelamin.html.

- [4] S. J. D. RI, “Undang-Undang Republik Indonesia Nomor 3 Tahun 2022 Tentang Ibu Kota Negara,” *Dewan Perwakilan Rakyat Republik Indonesia*, 2016. [Daring]. Tersedia pada: <https://www.dpr.go.id/dokjdih/document/uu/1791.pdf>.
- [5] S. Akhmad, P. P. Adikara, dan R. C. Wihandika, “Analisis Sentimen Kebijakan Pemindahan Ibukota Republik Indonesia dengan Menggunakan Algoritme Term-Based Random Sampling dan Metode Klasifikasi Naïve Bayes,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, Vol. 3, No. 10, hal. 10086–10094, 2019.
- [6] N. Tri Romadloni, I. Santoso, dan S. Budilaksono, “Perbandingan Metode Naive Bayes, KNN, dan Decision Tree Terhadap Analisis Sentimen Transportasi KRL Commuter Line,” *J. IKRA-ITH Inform.*, Vol. 3, No. 2, hal. 1–9, 2019.
- [7] P. Yuniasari dan F. Maspiyanti, “Analisis Sentimen Data Tweet menggunakan Metode Support Vector Machine (Studi Kasus: Pemindahan Ibukota Baru Republik Indonesia),” *J. Informatics Adv. Comput.*, Vol. 2, No. 1, hal. 2021, 2021.
- [8] A. H. Dyo fatra, N. H. Hayatin, dan C. S. K. Aditya, “Analisa Sentimen Tweet Berbahasa Indonesia Dengan Menggunakan Metode Lexicon Pada Topik Perpindahan Ibu Kota Indonesia,” *J. Repos.*, Vol. 2, No. 7, hal. 977, 2020, doi: 10.22219/repositor.v2i7.937.
- [9] M. I. D. Sakariana, Idriati, dan C. Dewi, “Analisis Sentimen Pemindahan Ibu Kota Indonesia Dengan Pembobotan Term BM25 dan Klasifikasi Neighbor Weighted K-Nearest Neighbor,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, Vol. 4, No. 3, hal. 748–755, 2020.
- [10] M. P. Agustina dan H. Hendry, “Sentimen Masyarakat Terkait Perpindahan Ibukota Via Model Random Forest dan Logistic Regression,” *Aiti*, Vol. 18, No. 2, hal. 111–124, 2021.
- [11] G. A. Sandag, “Prediksi Rating Aplikasi App Store Menggunakan Algoritma Random Forest,” *CogITo Smart J.*, Vol. 6, No. 2, hal. 167, 2020, doi: 10.31154/cogito.v6i2.270.167-178.