

Analisis Sentimen Terhadap Istana Garuda Di Ibukota Nusantara (IKN) Menggunakan Algoritma Random Forest Dan Support Vektor Machine

*Muhamad Jihansyah ¹, Agung Budi Santoso ², Arya Adyaksa Waskita ³

^{1,2,3)} Program Studi Teknik Informatika S-2, Universitas Pamulang

Email: ¹jihansyamuhamed1@gmail.com, ²dosen02680@unpam.ac.id, ³aawaskita@unpam.ac.id

ABSTRACT

The relocation of Indonesia's capital city (IKN) to East Kalimantan is a national strategic project that has sparked diverse public opinions, particularly regarding the construction of Garuda Palace. This study aims to analyze public sentiment toward the Garuda Palace project using Random Forest and Support Vector Machine (SVM) algorithms and to compare their performance based on accuracy, precision, recall, and F1-score. This research offers three key novelties. First, it focuses on public opinion regarding the Garuda Palace project at IKN, which is underexplored in both local and international literature. Second, the use of Inset and Senti labeling techniques introduces a novel approach to sentiment categorization. Third, the comprehensive evaluation of Random Forest and SVM performance provides new insights into their effectiveness in large-scale infrastructure sentiment analysis in Indonesia. The methodology consists of five stages: (1) Data collection through web scraping from Twitter (July-August 2024) using keywords related to "Garuda Palace" and "IKN"; (2) Data preprocessing, including tokenization, stopword removal, stemming, and TF-IDF transformation; (3) Data labeling using Inset and Senti approaches; (4) Model training with Random Forest and SVM algorithms; (5) Model evaluation using confusion matrices and performance metrics such as accuracy, precision, recall, and F1-score. Results indicate that Random Forest achieved 77% (Inset) and 89% (Senti) accuracy, excelling in detecting negative sentiment with an F1-score of 0.93 on the Senti dataset. SVM achieved 89% (Inset) and 91% (Senti) accuracy, performing better in detecting positive sentiment with a precision of 0.96 on the Senti dataset. This study provides valuable insights into public perceptions of national infrastructure projects, supports data-driven decision-making, and serves as a reference for future sentiment analysis systems.

Keywords: Sentiment Analysis, Garuda Palace, IKN, Random Forest, Support Vector Machine (SVM), Twitter

ABSTRAK

Pemindahan Ibu Kota Negara (IKN) ke Kalimantan Timur merupakan proyek strategis nasional yang memicu beragam opini publik, terutama terkait pembangunan Istana Garuda sebagai simbol baru kedaulatan Indonesia. Penelitian ini bertujuan menganalisis sentimen masyarakat terhadap pembangunan Istana Garuda di IKN menggunakan algoritma Random Forest dan Support Vector Machine (SVM), serta membandingkan kinerja keduanya berdasarkan akurasi, precision, recall, dan F1-score. Penelitian ini memiliki kebaruan dalam tiga aspek utama. Pertama, fokus pada opini publik mengenai pembangunan Istana Garuda di IKN yang belum banyak dibahas di literatur lokal maupun internasional. Kedua, penggunaan teknik labeling Inset dan Senti menawarkan pendekatan baru dalam pengkategorian sentimen. Ketiga, evaluasi komprehensif performa Random Forest dan SVM memberikan wawasan baru tentang efektivitas masing-masing algoritma dalam menganalisis sentimen terkait proyek infrastruktur besar di Indonesia. Metodologi meliputi lima tahap: (1) Pengumpulan data melalui web scraping dari Twitter (Juli-Agustus 2024) menggunakan kata kunci terkait "Istana Garuda" dan "IKN"; (2) Preprocessing data mencakup tokenisasi, stopword removal, stemming, dan transformasi TF-IDF; (3) Labeling menggunakan pendekatan Inset dan Senti; (4) Pelatihan model dengan algoritma Random Forest dan SVM; (5) Evaluasi model menggunakan confusion matrix dan metrik akurasi, precision, recall, dan F1-score. Hasil penelitian menunjukkan Random Forest memiliki akurasi 77% (Inset) dan 89% (Senti), unggul dalam mendeteksi sentimen negatif dengan F1-score 0,93 pada dataset Senti. SVM menunjukkan akurasi 89% (Inset) dan 91% (Senti), lebih baik dalam mengidentifikasi sentimen positif dengan precision 0,96 pada dataset Senti.

Penelitian ini memberikan kontribusi signifikan dalam memahami persepsi publik terhadap pembangunan infrastruktur nasional serta menjadi referensi bagi pengembangan sistem analisis sentimen dan pengambilan keputusan berbasis data.

Kata kunci: Analisis Sentimen, Istana Garuda, IKN, Random Forest, Support Vector Machine (SVM), Twitter.

1. PENDAHULUAN

Pemindahan Ibu Kota Negara (IKN) dari Jakarta ke Kalimantan Timur merupakan proyek strategis nasional yang memiliki dampak signifikan terhadap berbagai aspek kehidupan berbangsa dan bernegara. Wacana pemindahan ibu kota sebenarnya telah berlangsung sejak era kepemimpinan Presiden Soekarno hingga Susilo Bambang Yudhoyono, namun baru pada masa pemerintahan Presiden Joko Widodo rencana ini mulai direalisasikan Pada 29 April 2019, dalam sebuah rapat terbatas, Presiden Joko Widodo mengambil keputusan untuk memindahkan ibu kota dari Pulau Jawa, dan setelah melalui berbagai kajian dan audiensi publik, Kalimantan Timur terpilih sebagai lokasi ibu kota baru Indonesia.

Salah satu elemen kunci dalam rencana pembangunan IKN adalah Istana Garuda, yang diproyeksikan menjadi simbol kedaulatan dan identitas nasional Indonesia yang baru. Sebagai proyek berskala besar dengan implikasi sosial, ekonomi, dan politik yang luas, pembangunan Istana Garuda di IKN telah memunculkan beragam opini dan sentimen di kalangan masyarakat Indonesia. Pemerintah berargumen bahwa perpindahan IKN bertujuan untuk mempercepat pertumbuhan ekonomi, mengurangi ketimpangan di luar Jawa, terutama di Indonesia bagian timur, serta membawa manfaat ekonomi seperti peningkatan PDB negara, pemerataan pembangunan, peningkatan lapangan kerja, pengurangan kemiskinan, dan peningkatan kesejahteraan masyarakat secara umum [1]

Beberapa penelitian terdahulu telah menunjukkan efektivitas kedua algoritma ini dalam konteks analisis sentimen. Misalnya [2] membandingkan kinerja berbagai algoritma *machine learning*, termasuk RF dan SVM, dalam analisis sentimen pada (*Amazon Reviews: Unlocked Mobile Phones / Kaggle*, 2016) dan menemukan bahwa kedua algoritma ini menunjukkan performa yang kompetitif. Sementara itu, [3] menggunakan SVM untuk analisis sentimen terhadap kebijakan lolosnya salah satu public figur dari karantina covid 19 dengan hasil yang cukup baik.

2. METODE

2.1. Data dan Sumber Data

Dataset yang digunakan dalam penelitian ini adalah postingan serta komentar yang bersumber dari media sosial X. Pengambilan data pada penelitian ini dilakukan dengan crawling menggunakan twitter API. Lalu dari hasil crawling tersebut didapatkan 8164 data dan file disimpan dalam bentuk CSV. Sumber data yang digunakan dalam penelitian ini adalah data sekunder. Sumber data dalam penelitian ini yaitu mengambil data dari media sosial X. Populasi yang digunakan dalam penelitian ini adalah jutaan komentar pada media sosial X mengenai Istana IKN. Sedangkan sampel pada penelitian ini adalah komentar media sosial X dengan *keyword* “istana ikn”, “istana+ikn”, “istana garuda” dan “istana+garuda”. Sampel yang digunakan yaitu berjumlah 8.164 data. Pada penelitian ini, terdapat sampel berjumlah 8.164 data tweet selama kurang lebih satu bulan yang dikumpulkan pada rentang pengambilan data tweets yaitu dari periode 01 Juli 2024 sampai 25 Agustus 2024.

2.2. Metode Pengumpulan data

Metode yang dilakukan dalam pengumpulan data adalah sebagai berikut

a. Studi Pustaka

Penelitian dilakukan dengan penelusuran yang bersumber dari data media sosial X dengan *keyword* “istana ikn”, “istana+ikn”, “istana garuda” maupun jurnal lainnya.

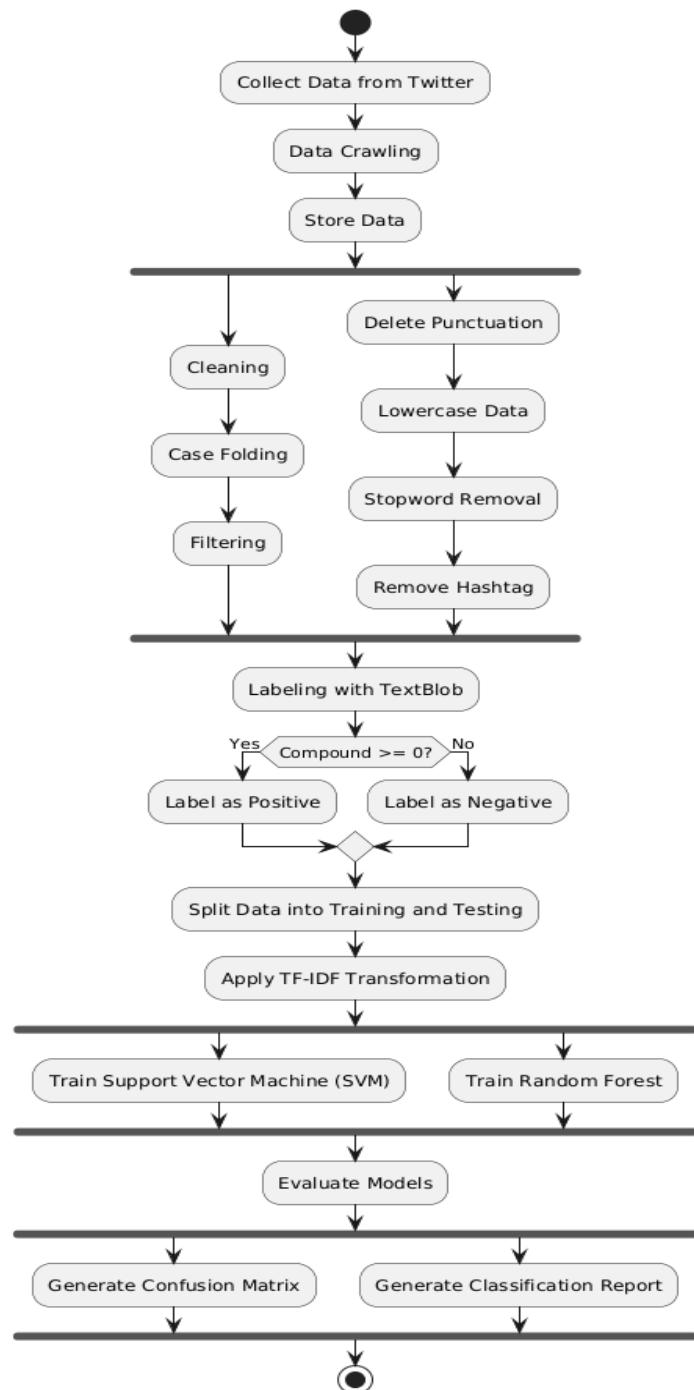
b. Pengumpulan Data Sekunder

Pengumpulan data sekunder merupakan metode pengumpulan data dengan menganalisa objek yang menjadi topik penelitian secara langsung, dalam hal ini pengumpulan data sekunder yang diperoleh dari data media sosial X dengan jumlah data yang digunakan sebanyak 8.164 data *tweet*, yang diambil dari media sosial X dengan *keyword* “istana ikn”, “istana+ikn”, “istana garuda” menggunakan API Twitter.

2.3. Metode Penelitian

Metodologi yang diterapkan dalam penelitian ini mengikuti pendekatan Metode analisis sentimen. Pendekatan ini meliputi tahapan-tahapan berikut: *Crawling Data, Data Labeling, Data Cleaning* dan *Processing, Validation Model*, dan *Evaluation*. Pendekatan ini mengutamakan pengukuran, pengolahan, dan analisis data

secara sistematis dan obyektif untuk mendapatkan kesimpulan akurat yang dapat dipercaya. dalam penelitian sentimen analisis pada istana IKN menggunakan SVM dan Random forest . dengan tahapan sebagai berikut



Gambar 2. Metode Penelitian

2.4. Metode Pengujian

Tahap evaluasi merupakan tahapan yang digunakan untuk menilai apakah model yang dibuat salah satu cara menggunakan *confussion matrix* untuk mendapatkan akurasi ataupun performance dari model [4] ilustrasinya ada di gambar 3 :

Table 1 Confusion Matrix Tabel

Nilai prediksi	Nilai Aktual	
	<i>Positive</i>	<i>Negative</i>
<i>Positive</i>	<i>True Positive (TP)</i>	<i>False Negative (FN)</i>
<i>Negative</i>	<i>False Positive (FP)</i>	<i>True Negative (TN)</i>

Confusion matrix merupakan cara untuk menganalisis performa model klasifikasi, terutama dalam klasifikasi biner, berdasarkan empat metrik utama:

- True Positive (TP)* Prediksi positif yang benar (model memprediksi kelas positif, dan benar-benar positif).
- True Negative (TN)* Prediksi negatif yang benar (model memprediksi kelas negatif, dan benar-benar negatif).
- False Positive (FP)* Prediksi positif yang salah (model memprediksi positif, tetapi sebenarnya negatif).
- False Negative (FN)* Prediksi negatif yang salah (model memprediksi negatif, tetapi sebenarnya positif).

Recall merupakan ukuran yang menunjukkan seberapa baik suatu sistem dalam menemukan kembali data positif yang benar dari seluruh data yang seharusnya teridentifikasi sebagai positif. Nilai *recall* diperoleh dengan membandingkan jumlah prediksi positif yang benar (*true positive*) terhadap seluruh jumlah data yang benar-benar positif (aktual positif) dalam data sebenarnya. Rumus *recall* dapat dituliskan sebagai berikut:

$$Recall = \frac{TP}{TP+FN} \quad (1)$$

Presisi merupakan ukuran ketepatan suatu sistem dalam mengidentifikasi data positif dan negatif secara benar. Nilai presisi diperoleh dengan membandingkan jumlah

prediksi positif yang benar terhadap total prediksi positif yang dihasilkan oleh sistem. Rumus presisi dapat dituliskan sebagai berikut:

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

F1 Score adalah perbandingan rata-rata presisi dan recall yang dibobotkan, dengan rumus sebagai berikut:

$$F1\ Score = \frac{Recall \times Presisi}{Recall + Presisi} \quad (3)$$

Accuracy adalah ukuran yang digunakan untuk menilai seberapa akurat model secara keseluruhan. Dengan kata lain, *accuracy* mengukur tingkat ketepatan model dalam mengklasifikasikan data secara benar, baik untuk kategori positif maupun negatif. Nilai *accuracy* diperoleh dengan membandingkan jumlah prediksi yang benar dengan total seluruh prediksi yang dilakukan oleh model.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

3. HASIL DAN PEMBAHASAN

3.1 Analisis Model Random Forest

Hasil dari penelitian ini untuk algoritma *random forest* untuk strategi labeling menggunakan senti word adalah sebagai berikut:

Table 2 Random Forest Sentiword

	Positiv	Negativ	Total
Positiv	250	218	468
Negative	54	654	708
Total	304	872	1176

Berdasarkan Tabel 2, dari total 1.176 data review yang diuji, sebanyak 904 review berhasil diklasifikasikan dengan benar, terdiri dari 654 review negatif dan 250 review positif. Sementara itu, terdapat 272 review yang diklasifikasikan secara salah; di antaranya adalah 54 review negatif yang salah dikategorikan sebagai positif dan 218 review positif yang salah dikategorikan sebagai negatif. Dengan demikian, model ini mencapai tingkat akurasi sebesar 76,86% dalam mengklasifikasikan sentimen, sementara tingkat kesalahan klasifikasi mencapai 23,14%, menunjukkan performa

model dalam membedakan antara sentimen positif dan negatif, Sedangkan untuk labeling senti hasilnya adalah sebagai berikut

Table 3 Random Forest Inset

	Positiv	Negativ	Total
Positiv	193	118	311
Negative	21	844	865
Total	214	962	1176

Berdasarkan Tabel 4.11, dari 1.176 data review yang diuji, model berhasil mengklasifikasikan 1.037 review dengan benar, terdiri dari 844 review negatif dan 193 review positif. Sebanyak 139 review salah diklasifikasikan, dengan 21 review negatif terdeteksi sebagai positif dan 118 review positif terdeteksi sebagai negatif. Akurasi model mencapai 88,60%, dengan tingkat kesalahan klasifikasi sebesar 13,40%. Model ini menunjukkan kinerja yang lebih baik dalam mengklasifikasikan sentimen negatif dibandingkan sentimen positif, yang terlihat dari jumlah prediksi benar yang lebih tinggi pada kategori negatif, untuk perbandingan hasil Random Forest Classification Report pada Data labeling inset dan labeling senti dapat dilihat dibawah ini:

Table 4 Random Forest Classification Report

Metric	Insert (Dataset Pertama)	Senti (Dataset Kedua)
Precision (Negatif)	0.75	0.88
Recall (Negatif)	0.92	0.98
F1-Score (Negatif)	0.83	0.93
Support (Negatif)	708	865
Precision (Positif)	0.82	0.92
Recall (Positif)	0.53	0.63
F1-Score (Positif)	0.65	0.74
Support (Positif)	468	311
Akurasi (Accuracy)	0.77	0.89

Tabel 4 memberikan gambaran mengenai kinerja model Random Forest yang diterapkan pada dua dataset, yaitu Insert dan Senti. Seluruh metrik dihitung untuk dua kelas, yaitu negatif dan positif, kemudian dirangkum dalam metrik rata-rata (*Macro Average dan Weighted Average*). Berdasarkan hasil analisis, dataset Senti menunjukkan peningkatan yang signifikan dalam semua metrik untuk kelas negatif, dengan nilai *Precision* sebesar 0,88, *Recall* sebesar 0,98, dan *F1-Score* sebesar 0,93, jika dibandingkan dengan dataset Insert. Hal ini menunjukkan bahwa model *Random Forest* pada dataset Senti lebih andal dalam mengklasifikasikan data negatif. Pada kelas positif, model *Random Forest* pada dataset Senti juga memiliki *Precision* yang lebih tinggi (0,92) dibandingkan dengan dataset Insert (0,82). Namun, nilai *Recall* pada dataset Senti (0,63) lebih rendah dibandingkan dengan dataset Insert (0,53). *F1-Score* pada kelas positif juga mengalami peningkatan dari 0,65 pada dataset Insert menjadi 0,74 pada dataset Senti, yang menunjukkan adanya peningkatan keseimbangan antara *Precision* dan *Recall*. Dari segi akurasi keseluruhan, model pada dataset Senti memiliki akurasi yang lebih tinggi, yaitu sebesar 0,89, dibandingkan dengan akurasi pada dataset Insert yang hanya sebesar 0,77. Hal ini mengindikasikan bahwa kinerja keseluruhan model *Random Forest* lebih baik pada dataset Senti. Secara umum, model *Random Forest* menunjukkan kinerja yang lebih baik pada dataset Senti dibandingkan dengan dataset Insert, terutama dalam mengklasifikasikan data negatif dengan lebih akurat dan menghasilkan nilai akurasi yang lebih tinggi. Peningkatan ini terlihat pada semua metrik utama untuk kelas negatif dan sebagian besar metrik untuk kelas positif, sehingga mendukung penggunaan dataset Senti untuk analisis lebih lanjut.

3.2 Analisis Model Support Vector Machine

Hasil dari penelitian ini untuk algoritma *support vector machine* untuk strategi labeling menggunakan senti word adalah sebagai berikut:

Table 5 Support Vector Machine Sentiword

	Positiv	Negativ	Total
Positiv	331	137	468
Negative	43	665	708
Total	374	802	1176

Berdasarkan Tabel 5, dari 1.176 data review yang diuji, model berhasil mengklasifikasikan 996 review dengan benar, terdiri dari 331 review positif dan 665 review negatif, sementara 180 review salah diklasifikasikan. Kesalahan klasifikasi mencakup 137 review positif yang keliru diidentifikasi sebagai sentimen negatif dan 43 review negatif yang salah dikategorikan sebagai sentimen positif. Dengan demikian, akurasi model mencapai 84,71%, sedangkan tingkat kesalahan klasifikasi sebesar 15,29%. Model menunjukkan kinerja yang lebih baik dalam mengklasifikasikan sentimen negatif, namun cenderung lebih sensitif terhadap sentimen negatif dan kurang akurat dalam mengenali review positif, sehingga evaluasi lebih lanjut diperlukan untuk meningkatkan keseimbangan dan akurasi dalam mengenali kedua jenis sentimen secara konsisten. sedangkan untuk labeling senti hasilnya adalah sebagai berikut:

Table 6 Support Vector Machine Inset

	Positiv	Negativ	Total
Positiv	213	98	311
Negative	9	856	865
Total	222	954	1176

Berdasarkan Tabel 6, dari 1.176 data review yang diuji, model berhasil mengklasifikasikan 1.069 review dengan benar (213 positif dan 856 negatif), sementara 107 review salah diklasifikasikan. Kesalahan klasifikasi mencakup 98 review positif yang keliru diidentifikasi sebagai sentimen negatif dan hanya 9 review negatif yang salah diklasifikasikan sebagai sentimen positif. Akurasi model mencapai 91,09%, dengan tingkat kesalahan klasifikasi 8,91%. Model menunjukkan performa sangat baik dalam mengklasifikasikan sentimen negatif (98,96% akurat), namun akurasi dalam mengenali sentimen positif sedikit lebih rendah (68,49%). Kesalahan terbesar terjadi pada review positif yang salah dikategorikan sebagai sentimen negatif, mengindikasikan adanya bias dalam model terhadap sentimen negatif, sehingga diperlukan optimasi lebih lanjut untuk meningkatkan akurasi klasifikasi sentimen positif dan mengurangi bias ini, untuk perbandingan hasil support vector machine report pada data labeling inset dan senti dapat dilihat table dibawah ini:

Table 7 Support Vector Machine Classification Report

Metric	Insert (Dataset Pertama)	Senti (Dataset Kedua)
Precision (Negatif)	0.83	0.9
Recall (Negatif)	0.94	0.99
F1-Score (Negatif)	0.88	0.94
Support (Negatif)	708	865
Precision (Positif)	0.89	0.96
Recall (Positif)	0.71	0.68
F1-Score (Positif)	0.79	0.8
Support (Positif)	468	311
Akurasi (Accuracy)	0.89	0.91

Tabel 7 menggambarkan kinerja model Support Vector Machine (SVM) yang diterapkan pada dua dataset, yaitu Insert dan Senti, dengan evaluasi metrik untuk dua kelas, negatif dan positif. Pada kelas negatif, dataset Senti menunjukkan keunggulan yang jelas dengan nilai Precision, Recall, dan F1-Score yang lebih tinggi dibandingkan dataset Insert, terutama Recall yang mencapai 0,99 pada Senti, lebih baik dari 0,94 pada Insert. Sementara itu, pada kelas positif, dataset Insert memiliki nilai Recall yang sedikit lebih baik, yaitu 0,71 dibandingkan 0,68 pada Senti, yang menunjukkan kemampuan lebih baik dalam mendeteksi data positif. Namun, Precision pada dataset Senti lebih unggul dengan nilai 0,96, dibandingkan 0,89 pada Insert, sehingga menunjukkan keseimbangan yang berbeda antara kedua dataset dalam mengenali kelas positif. Dari segi akurasi keseluruhan, kedua dataset memiliki hasil yang cukup dekat, dengan Insert memperoleh akurasi 0,89 dan Senti sedikit lebih tinggi pada 0,91, menandakan bahwa model SVM bekerja dengan baik pada kedua dataset. Kesimpulannya, dataset Senti lebih unggul dalam mengklasifikasikan kelas negatif dengan metrik yang lebih tinggi, sedangkan dataset Insert lebih baik dalam hal Recall untuk kelas positif meskipun Precision-nya lebih rendah. Dengan demikian, akurasi keseluruhan model SVM pada kedua dataset relatif setara, memperlihatkan performa yang stabil dan efektif dalam analisis sentimen menggunakan kedua dataset tersebut.

3.3. Komparasi Model *SVM* dan *Random Forest*

Berdasarkan analisis perbandingan kinerja antara model Random Forest dan SVM pada dataset Insert dan Senti, dapat disimpulkan beberapa hal sebagai berikut: Pada performa kelas negatif, kedua model menunjukkan Recall yang sangat tinggi, mencapai 0,99 pada dataset Senti untuk Random Forest maupun SVM. Random Forest pada dataset Senti mencatatkan F1-Score kelas negatif terbaik dengan nilai 0,93, yang sedikit lebih rendah dari SVM yang mencapai 0,94. Secara umum, kedua model sangat baik dalam mengidentifikasi kelas negatif, dengan Random Forest sedikit lebih unggul dalam F1-Score pada dataset Senti.

Dalam hal performa pada kelas positif, Precision lebih tinggi pada dataset Senti, di mana Random Forest mencapai 0,92 dan SVM mencapai 0,96. Namun, Recall untuk kelas positif menunjukkan hasil yang berbeda: dataset Insert memperlihatkan performa Recall yang lebih baik untuk kedua model. Model SVM pada dataset Insert memiliki Recall tertinggi yaitu 0,77, lebih tinggi dari nilai 0,68 pada dataset Senti. Sementara itu, Random Forest juga menunjukkan Recall kelas positif yang lebih rendah pada dataset Senti (0,63) dibandingkan dengan dataset Insert (0,75). Ini menunjukkan bahwa meskipun Precision di dataset Senti lebih tinggi, kemampuan model dalam mendekripsi kelas positif (Recall) lebih baik pada dataset Insert.

Mengenai akurasi keseluruhan, model SVM pada dataset Senti mencapai akurasi tertinggi dengan nilai 0,91, sementara Random Forest pada dataset Senti memiliki akurasi 0,89. Meskipun SVM sedikit lebih baik dalam akurasi pada dataset Senti, Random Forest juga menunjukkan akurasi yang sangat baik di kedua dataset, dengan nilai 0,77 pada Insert dan 0,89 pada Senti. Kedua model secara umum menunjukkan akurasi yang baik, dengan sedikit keunggulan SVM pada dataset Senti.

Secara keseluruhan, model Random Forest menunjukkan performa yang lebih stabil dan konsisten, terlihat dari nilai F1-Score yang relatif konsisten untuk kelas negatif dan positif di semua dataset. Di sisi lain, model SVM memiliki keunggulan pada kelas positif di dataset Insert, terutama pada Recall yang mencapai 0,77. Namun, performa SVM menunjukkan sedikit penurunan ketika diterapkan pada dataset Senti. Meskipun SVM unggul dalam Recall untuk kelas positif di dataset Insert, Random Forest tetap menunjukkan konsistensi yang lebih baik di seluruh dataset, khususnya pada kelas negatif. Dengan demikian, dapat disimpulkan bahwa Random Forest menawarkan

performa yang lebih stabil dan konsisten secara keseluruhan, meskipun SVM memiliki kekuatan spesifik pada beberapa aspek.

Table 8 Komparasi Random Forest dan SVM

Metric	RANDOM FOREST		SVM	
	Inset	Senti	Inset	Senti
Precision (Negatif)	0.75	0.88	0.83	0.9
Recall (Negatif)	0.92	0.98	0.94	0.99
F1-Score (Negatif)	0.83	0.93	0.88	0.94
Support (Negatif)	708	865	708	865
Precision (Positif)	0.82	0.92	0.89	0.96
Recall (Positif)	0.53	0.63	0.71	0.68
F1-Score (Positif)	0.65	0.74	0.79	0.8
Support (Positif)	468	311	468	311
Akurasi (Accuracy)	0.77	0.89	0.89	0.91

4. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, proses pengolahan data Twitter untuk analisis sentimen meliputi pengumpulan data sebanyak 8164 tweet terkait "istana ikn" dan "istana garuda" dari 1 Juli hingga 31 Agustus 2024, dilanjutkan dengan pra-pemrosesan data seperti penghapusan karakter khusus, case folding, dan stopword removal. Selanjutnya, data diberi label sentimen positif dan negatif menggunakan dua pendekatan leksikon, yaitu Inset yang berbasis KBBI dan Senti yang menggunakan metode Sentistrength untuk akurasi model yang lebih kuat, kemudian diubah ke representasi numerik menggunakan TF-IDF. Pengolahan data untuk hasil maksimal memanfaatkan kedua pendekatan labeling ini, di mana Inset lebih akurat mendekripsi sentimen negatif, sementara Senti seimbang untuk sentimen positif dan negatif, dan model dioptimasi dengan 1000 trees untuk Random Forest serta kernel linear untuk SVM.

Evaluasi model menunjukkan bahwa SVM mencapai akurasi tertinggi 91% pada dataset Senti, mengungguli Random Forest (89%). Dalam deteksi sentimen negatif, Random Forest pada dataset Senti unggul dengan recall 98%, sedangkan SVM pada dataset Senti menunjukkan precision 90% dan recall 99%, mengindikasikan stabilitas yang lebih baik. Untuk sentimen positif, SVM pada dataset Senti memiliki precision 96%,

lebih baik dalam mengenali sentimen positif, meskipun recall Random Forest lebih rendah (63%) pada dataset yang sama. Secara keseluruhan, Random Forest menunjukkan stabilitas yang lebih baik pada dataset Senti karena distribusi prediksinya merata, sementara SVM lebih sensitif terhadap dataset namun unggul dalam mengenali sentimen positif. Berdasarkan temuan ini, performa Random Forest lebih baik dalam mengidentifikasi sentimen negatif, sedangkan SVM unggul dalam mengenali sentimen positif, khususnya pada dataset Senti. Akurasi tertinggi diperoleh dari SVM pada dataset Senti (91%), menjadikannya model yang direkomendasikan untuk analisis sentimen Istana Garuda di IKN karena performanya yang seimbang dan akurat dalam mendekripsi kedua jenis sentimen.

Penelitian ini memberikan pemahaman mendalam mengenai persepsi masyarakat terhadap Istana Garuda di IKN, yang dapat digunakan sebagai dasar pertimbangan dalam perumusan kebijakan atau strategi komunikasi. Pemahaman atas pola sentimen publik akan membantu pihak terkait dalam mengoptimalkan penerimaan masyarakat terhadap proyek IKN. Ke depan, penelitian ini dapat diperluas dengan melibatkan lebih banyak data dari platform media sosial lain dan mempertimbangkan faktor-faktor temporal yang dapat memengaruhi perubahan sentimen.

5. DAFTAR PUSTAKA

- [1] H. Dhery, A. Assyam, and F. N. Hasan, “Analisis Sentimen Twitter Terhadap Perpindahan Ibu Kota Negara Ke IKN Nusantara Menggunakan Orange Data Mining,” *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 4, no. 1, pp. 341–349, 2023, doi: 10.30865/klik.v4i1.957.
- [2] M. Guia, R. R. Silva, and J. Bernardino, “Comparison of Naive Bayes, support vector machine, decision trees and random forest on sentiment analysis,” *IC3K 2019 - Proc. 11th Int. Jt. Conf. Knowl. Discov. Knowl. Eng. Knowl. Manag.*, vol. 1, no. Ic3k, pp. 525–531, 2019, doi: 10.5220/0008364105250531.
- [3] P. Wahyuningtias, H. Warih Utami, U. Ahda Raihan, H. Nur Hanifah, and Y. Nicholas Adanson, “Comparison of Random Forest and Support Vector Machine Methods on Twitter Sentiment Analysis (Case Study: Internet Selebgram Rachel Venna Escape From Quarantine),” *J. Tek. Inform.*, vol. 3, no. 1, pp. 141–145, 2022, [Online]. Available: <https://doi.org/10.20884/1.jutif.2022.3.1.168>
- [4] M. S. Ummah, “No 主観的健康感を中心とした在宅高齢者における 健康関連指標に関する共分散構造分析Title,” *Sustain.*, vol. 11, no. 1, pp. 1–14, 2019, [Online]. Available: http://scioteca.caf.com/bitstream/handle/123456789/1091/RED2017-Eng-8ene.pdf?sequence=12&isAllowed=y%0Ahttp://dx.doi.org/10.1016/j.regscurboco.2008.06.005%0Ahttps://www.researchgate.net/publication/305320484_SISTEM PEMBETUNGAN TERPUSAT STRATEGI_MESTARI