(2025), 3 (6): 342-356 **Scientica** 3021-8209

Jurnal Ilmiah Sain dan Teknologi

PERBANDINGAN METODE *SUPPORT VECTOR MACHINE* (SVM) DAN *RANDOM FOREST* (RF) DALAM MENGKLASIFIKASI EMOSI PADA TWEET BAHASA SUNDA

Yuni [1], Arif Bijaksana Putra Negara [2], Ibnu Arif [3]
Fakultas Teknik
Universitas Tanjungpura Pontianak
yuniwork@outlook.com

Abstract (English)
Language is a crucial element in human interaction, and in Indonesia, the diversity of ethnicities, cultures, and languages often poses challenges in communication. This study aims to identify and classify emotions in tweets written in Sundanese posted on platform X. In this research, a dataset consisting of 2518 Sundanese tweets was collected and labeled with emotions, including angry, fear, joy, and sadness. The emotion classification process was conducted using Machine Learning methods, namely Support Vector Machine (SVM) and Random Forest, to evaluate the effectiveness of each algorithm in identifying emotions. This study also compares the performance of both methods to determine which is more effective in emotion classification. The results indicate that both algorithms achieve high accuracy, with SVM reaching an accuracy of 89.1% and Random Forest 87.3%. Through confusion matrix analysis, SVM proved to be superior in terms of precision and recall, making it more effective in classifying emotions in Sundanese tweets. These findings are expected to contribute to the development of emotion analysis technology in the context of regional languages on social media.

Article History
Submitted: 3 Oktober 2025
Accepted: 6 Oktober 2025
Published: 7 Oktober 2025

Key Words
tweet, X, Sundanese language, Support Vector Machine, Random Forest, emotion classification, SMOTE, TF-IDF

Abstrak (Indonesia)
Bahasa merupakan elemen krusial dalam interaksi manusia, dan di Indonesia, keragaman suku, budaya, dan bahasa seringkali menimbulkan tantangan dalam komunikasi. Penelitian ini bertujuan untuk mengidentifikasi dan mengklasifikasikan emosi dalam tweet berbahasa Sunda yang diposting di platform X. Dalam penelitian ini, dataset yang terdiri dari 2.518 tweet berbahasa Sunda dikumpulkan dan diberi label emosi, termasuk marah, takut, senang, dan sedih. Proses klasifikasi emosi dilakukan menggunakan metode Machine Learning, yaitu Support Vector Machine (SVM) dan Random Forest, untuk mengevaluasi efektivitas masing-masing algoritma dalam mengidentifikasi emosi. Penelitian ini juga membandingkan kinerja kedua metode tersebut untuk menentukan mana yang lebih efektif dalam klasifikasi emosi. Hasil penelitian menunjukkan bahwa kedua algoritma memiliki akurasi yang tinggi, dengan SVM mencapai akurasi 89,1% dan Random Forest 87,3%. Melalui analisis confusion matrix, SVM terbukti lebih unggul dalam hal presisi dan recall, sehingga lebih efektif dalam klasifikasi emosi tweet berbahasa Sunda. Temuan ini diharapkan dapat memberikan kontribusi terhadap pengembangan teknologi analisis emosi dalam konteks bahasa daerah di media sosial.

Sejarah Artikel
Submitted: 3 Oktober 2025
Accepted: 6 Oktober 2025
Published: 7 Oktober 2025

Kata Kunci
tweet, X, bahasa Sunda, Support Vector Machine, Random Forest, klasifikasi emosi, SMOTE, TF-IDF

Pendahuluan

Perkembangan teknologi informasi dan komunikasi telah membawa transformasi besar dalam cara manusia berinteraksi, berkomunikasi, serta mengekspresikan emosi. Salah satu wujud nyata dari transformasi tersebut adalah meningkatnya penggunaan media sosial sebagai wadah ekspresi publik yang luas, cepat, dan spontan. Platform Twitter (kini dikenal sebagai X), sebagai salah satu media sosial dengan jumlah pengguna aktif yang sangat besar, mencatat

lebih dari 500 juta cuitan setiap harinya (Statista, 2023). Tweet atau cuitan yang dibagikan pengguna umumnya bersifat singkat, emosional, dan tidak terstruktur, sehingga menjadi sumber data potensial untuk dianalisis secara komputasional, khususnya melalui pendekatan *Natural Language Processing* (NLP).

Dalam lingkup NLP, salah satu cabang yang berkembang pesat adalah analisis emosi (*emotion classification*), yaitu proses untuk mengidentifikasi serta mengklasifikasikan jenis emosi dalam sebuah teks, seperti joy (senang), anger (marah), fear (takut), dan sadness (sedih). Analisis emosi memiliki berbagai aplikasi penting, antara lain untuk pemantauan opini publik, sistem rekomendasi, layanan pelanggan otomatis, hingga deteksi awal kondisi kesehatan mental (Glenn et al., 2023). Meski demikian, sebagian besar penelitian analisis emosi masih berfokus pada bahasa global seperti Inggris atau bahasa nasional seperti Indonesia. Bahasa daerah, seperti bahasa Sunda, masih relatif belum banyak dieksplorasi, padahal bahasa tersebut aktif digunakan oleh jutaan penutur dalam kehidupan sehari-hari maupun dalam komunikasi digital.

Bahasa Sunda merupakan salah satu bahasa daerah dengan jumlah penutur terbanyak di Indonesia, diperkirakan mencapai lebih dari 42 juta orang (Badan Pengembangan dan Pembinaan Bahasa, 2022). Penutur bahasa Sunda juga aktif menggunakan media sosial, termasuk dalam mengekspresikan opini dan emosi melalui tweet. Untuk mendukung pengembangan NLP berbasis bahasa Sunda, Putra et al. (2020a) menginisiasi pembangunan *Sundanese Twitter Dataset for Emotion Classification*, yaitu dataset yang berisi lebih dari 2.500 tweet berbahasa Sunda yang telah diklasifikasikan ke dalam empat kategori emosi. Dataset ini menjadi landasan penting untuk mengembangkan dan menguji performa berbagai algoritma klasifikasi emosi berbasis bahasa daerah.

Namun, terdapat beberapa tantangan dalam klasifikasi emosi dari data tweet. Pertama, teks tweet cenderung singkat, informal, dan sering kali tidak mengikuti kaidah gramatikal yang baku. Kedua, tweet dalam bahasa daerah mengandung istilah lokal, emotikon, dan campur kode (code-switching), yang menyulitkan pemodelan secara umum. Ketiga, distribusi emosi dalam dataset umumnya tidak seimbang, di mana emosi tertentu (misalnya joy) lebih dominan dibanding emosi lainnya (seperti fear atau sadness). Oleh karena itu, dibutuhkan teknik representasi dan pemodelan yang mampu menangani kompleksitas tersebut secara optimal.

Teknik representasi teks yang umum digunakan dalam NLP adalah *Term Frequency-Inverse Document Frequency* (TF-IDF) dengan konfigurasi *n-gram*, yang mampu menangkap konteks kata dalam frasa pendek (Isnain et al., 2021). Untuk mengatasi ketimpangan kelas emosi dalam dataset, teknik *Synthetic Minority Over-sampling Technique* (SMOTE) menjadi salah satu pendekatan yang terbukti efektif dalam penelitian klasifikasi (Novianto et al., 2024)

Dalam penelitian ini, dua algoritma machine learning yang banyak digunakan untuk klasifikasi teks, yaitu *Support Vector Machine* (SVM) dan *Random Forest* (RF), digunakan untuk mengklasifikasikan emosi pada tweet berbahasa Sunda. SVM dikenal memiliki performa tinggi pada data berdimensi besar dan tidak seimbang (Noroozi et al., 2023), sementara RF unggul dalam menangkap interaksi antar fitur dan memiliki tingkat kestabilan model yang baik (Sidik et al., 2025). Kombinasi teknik TF-IDF, konfigurasi n-gram, dan penggunaan SMOTE diharapkan mampu meningkatkan akurasi klasifikasi dan memberikan gambaran yang lebih jelas mengenai performa masing-masing algoritma dalam konteks bahasa daerah.

Oleh karena itu, penelitian ini difokuskan untuk membandingkan performa SVM dan RF dalam mengklasifikasikan emosi pada tweet berbahasa Sunda, dengan mempertimbangkan konfigurasi n-gram dan kondisi data seimbang serta tidak seimbang. Hasil dari penelitian ini diharapkan dapat berkontribusi dalam pengembangan sistem analisis emosi berbasis bahasa daerah dan memperkaya penelitian NLP di Indonesia yang masih didominasi oleh bahasa-bahasa mayoritas.

Metode Penelitian

Penelitian ini merupakan penelitian kuantitatif komparatif dengan pendekatan eksperimental. Penelitian kuantitatif digunakan karena seluruh proses analisis melibatkan data numerik dan pengukuran yang dapat dihitung secara statistik, seperti akurasi, presisi, recall, dan F1-score.

Pendekatan komparatif dilakukan untuk membandingkan performa dua algoritma klasifikasi teks, yaitu *Support Vector Machine* (SVM) dan *Random Forest* (RF) dalam mengklasifikasikan data emosi yang terdapat pada tweet berbahasa Sunda. Perbandingan dilakukan berdasarkan hasil evaluasi dari masing-masing model setelah diterapkan pada *dataset* yang sama dengan konfigurasi parameter dan teknik ekstraksi fitur yang sebanding.

Sedangkan pendekatan eksperimental diterapkan karena dalam penelitian ini dilakukan percobaan langsung terhadap data melalui bebrapa tahapan pemrosesan seperti *preprocessing* teks, ekstraksi fitur menggunakan TF-IDF, balancing data menggunakan SMOTE, dan penerapan model klasifikasi SVM dan RF. Hasil dari masing-masing percobaan kemudian dianalisis secara sistematis untuk meihat keefektifan dan efisiensi kedua metode dalam klasifikassi emosi.

Dengan menggunakan desain penelitian ini, penulis berharap dapat memperoleh pemahaman yang jelas mengenai kelebihan dan kelemahan masing-masing metode serta memberikan rekomendasi metode yang paling optimal untuk klasifikassi emosi pada data teks pendek berbahasa daerah.

Hasil dan Pembahasan

Hasil penelitian

Penelitian ini mengkaji performa dua metode klasifikasi teks, yaitu *Support Vector Machine* dan *Random Forest*, dalam megklasifikasikan emosi pada tweet berbahasa Sunda. Data di evaluasi menggunakan kombinasi representasi fitur TF-IDF dengan 1-gram hingga 4-gram, dan dilakukan perbandingan antara data yang tidak diseimbangkan (non-SMOTE) dan data yang diseimbangkan (SMOTE). Evaluasi dilakukan menggunakan metrik *accuracy, precision, recall,* dan *F1-score*.

Support Vector Machine

Non-SMOTE

Model dilatih menggunakan representasi fitur TF-IDF dengan konfigurasi n-gram dari 1-gram hingga 4-gram, bertujuan untuk mengetahui performa awal model saat menangani data yang tidak seimbang antar kelas emosi.

Gambar 3.9 menunjukan hasil evaluasi SVM dengan TF-IDF 1-gram, yang memberikan hasil terbaik dengan akurasi mencapai 97.02%. Seluruh metrik *precision, recall,*

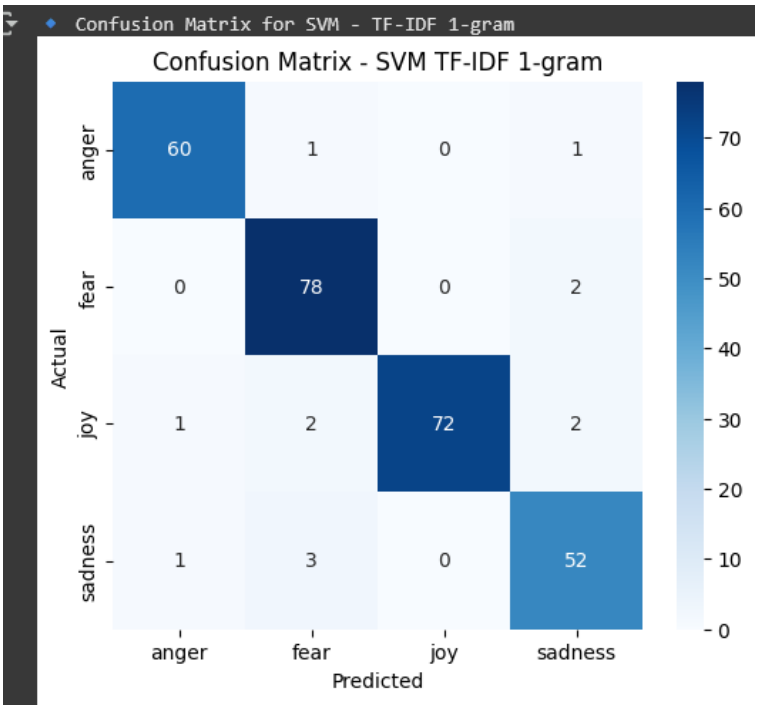
Evaluasi SVM dengan TF-IDF 1-gram

	precision	recall	f1-score	support
anger	0.96	0.96	0.96	141
fear	0.97	0.99	0.98	122
joy	1.00	0.97	0.98	125
sadness	0.95	0.96	0.95	116
accuracy			0.97	504
macro avg	0.97	0.97	0.97	504
weighted avg	0.97	0.97	0.97	504

Akurasi: 0.9702

Gambar 3. 1 Evaluasi SVM 1-gram

F1-score menunjukan nilai tinggi pada semua kelas emosi. Hal ini menunjukan bahwa 1-gram mampu menangkap representasi kata secara efektif.



Gambar 3. 10 Confussion Matrix Untuk SVM 1-gram NON-SMOTE

Gambar 3.10 menunjukan confussion matrix dari model SVM dengan konfigurasi TF-IDF 1-gram tanpa SMOTE. Hasil ini memperlihatkan bahwa model mampu melakukan klasifikasi emosi dengan sangat baik dan konsisten. Kelas *anger* memiliki 60 prediksi benar dari total 62 data, dengan hanya 1 data yang keliru diklasifikasikan sebgai *fear* dan 1 lainnya sebagai *sadness*

Kelas *fear* menunjukkan performa yang hampir sempurna, dengan 78 dari 80 data berhasil diklasifikasikan dengan benar, dan hanya 2 data yang salah sebagai *sadness*. Untuk kelas *joy*, sebanyak 72 data terklasifikasi dengan benar, sementara 5 lainnya keliru diprediksi menjadi *anger*, *fear*, atau *sadness*.

Sementara itu, kelas *sadness* juga menunjukkan hasil yang memuaskan, dengan 52 prediksi benar dari 56 data, dan 4 data lainnya keliru diklasifikasikan sebagai *anger*, *fear*, atau *joy*.

Secara keseluruhan, nilai diagonal pada *confusion matrix* mendominasi, yang mengindikasikan bahwa model SVM 1-gram memiliki kemampuan klasifikasi yang akurat dan seimbang terhadap keempat kategori emosi. Hasil ini juga sejalan dengan skor evaluasi sebelumnya, di mana *precision*, *recall*, dan *f1-score* untuk semua kelas berada di atas 0.95, serta akurasi keseluruhan mencapai 96%

```
🔹 Evaluasi SVM dengan TF-IDF 2-gram
              precision    recall  f1-score   support

       anger       0.90      0.56      0.69        62
        fear       0.49      0.96      0.65        80
         joy       0.95      0.52      0.67        77
     sadness       0.94      0.61      0.74        56

    accuracy                           0.68       275
   macro avg       0.82      0.66      0.69       275
weighted avg       0.80      0.68      0.68       275

Akurasi: 0.6764
```

Gambar 3. 11 Evaluasi SVM dengan TF-IDF 2-gram

Namun, pada konfigurasi TF-IDF 2-gram, akurasi model menurun signifikan menjadi 67%, terutama karena penurunan recall pada kelas *joy* (0.52) dan *anger* (0.56). hal ini mengindikasikan bahwa 2-gram mulai membawa dampak sparsitas fitur yang dimana kata atau frasa yang digunakan menjadi jarang muncul dan kurang general.

```
🔹 Evaluasi SVM dengan TF-IDF 3-gram
              precision    recall  f1-score   support

       anger       0.95      0.14      0.25       141
        fear       0.29      0.99      0.44       122
         joy       1.00      0.26      0.41       125
     sadness       0.93      0.23      0.37       116

    accuracy                           0.40       504
   macro avg       0.79      0.41      0.37       504
weighted avg       0.80      0.40      0.36       504

Akurasi: 0.3968

🔹 Evaluasi SVM dengan TF-IDF 4-gram
              precision    recall  f1-score   support

       anger       1.00      0.06      0.11       141
        fear       0.25      1.00      0.40       122
         joy       1.00      0.09      0.16       125
     sadness       1.00      0.03      0.07       116

    accuracy                           0.29       504
   macro avg       0.81      0.29      0.19       504
weighted avg       0.82      0.29      0.18       504

Akurasi: 0.2877
```

Gambar 3. 2 Evaluasi SVM dengan TF-IDF 3-gram dan 4-gram

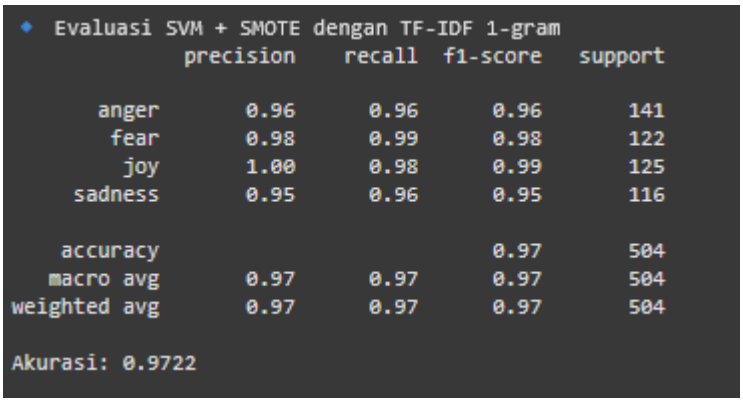
Penurunan performa semakin drastis pada konfigurasi 3-gram dan 4-gram. Pada 3-gram, akurasi turun menjadi 39,68%, dan pada 4-gram bahkan hanya 28.77%. Hasil ini memperlihatkan bahwa n-gram yang lebih besar, model cenderung kesulitan mengenali pola karena kombinasi kata menjadi terlalu spesifik dan jarang muncul di data latih, terutama pada kondisi data yang tidak seimbang.

Performa buruk terlihat jelas pada kelas *anger*, *joy*, dan *sadness*, yang hanya memiliki recall antara 0.03–0.26 pada 3-gram dan 4-gram. Sebaliknya, kelas *fear* justru mengalami peningkatan *recall* signifikan, mencapai 1.00 pada 4-gram, yang menunjukkan kemungkinan bias model terhadap kelas ini akibat distribusi data atau kemunculan frasa tertentu yang dominan.

Secara keseluruhan, hasil ini menunjukkan bahwa meskipun model SVM secara teoritis stabil terhadap data tidak seimbang, pengaruh distribusi data tetap sangat terlihat, terutama saat menggunakan fitur yang kompleks seperti 3-gram atau 4-gram. Oleh karena itu, teknik penyeimbangan seperti SMOTE menjadi penting untuk memastikan semua kelas emosi memiliki representasi yang adil dalam proses pelatihan, terutama ketika menggunakan representasi fitur tingkat lanjut.

SMOTE

Hasil evaluasi model Support Vector Machine (SVM) setelah dilakukan penyeimbangan data menggunakan SMOTE.



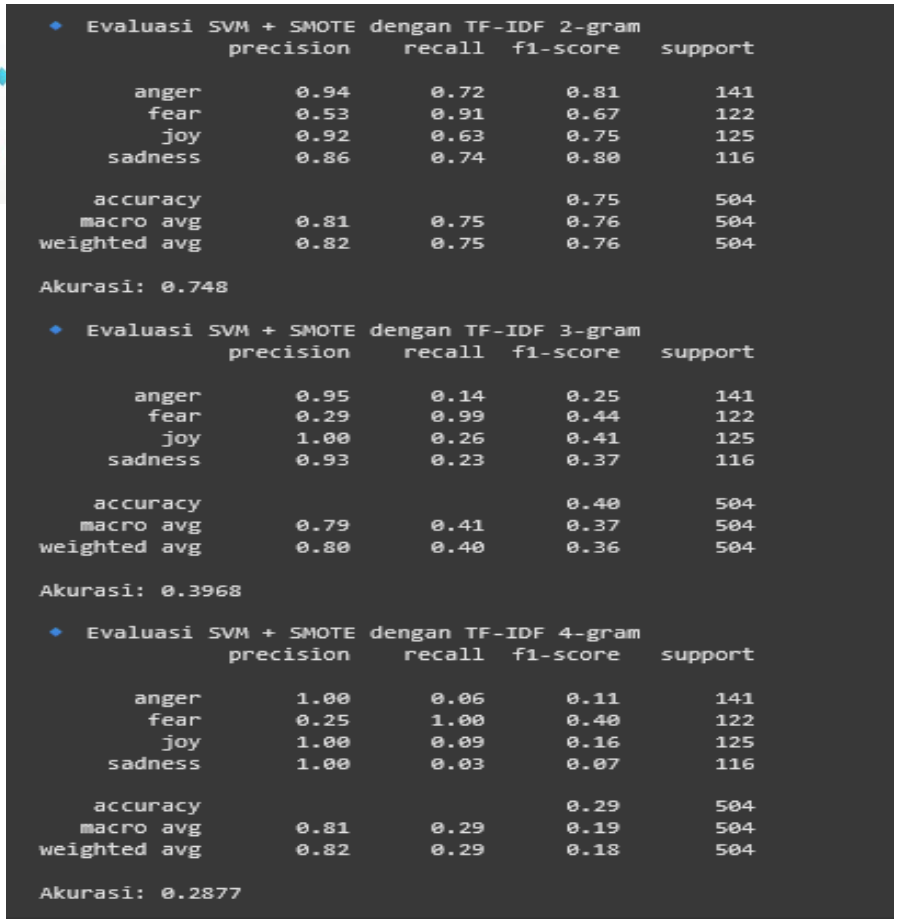
Evaluasi SVM + SMOTE dengan TF-IDF 1-gram

	precision	recall	f1-score	support
anger	0.96	0.96	0.96	141
fear	0.98	0.99	0.98	122
joy	1.00	0.98	0.99	125
sadness	0.95	0.96	0.95	116
accuracy			0.97	504
macro avg	0.97	0.97	0.97	504
weighted avg	0.97	0.97	0.97	504

Akurasi: 0.9722

Gambar 3. 3 Hasil evaluasi SVM + SMOTE 1-gram

Gambar 3.11 memperlihatkan hasil evaluasi model untuk konfigurasi n-gram dari 1-gram hingga 4-gram setelah diterapkannya SMOTE. Hasil tertinggi diperoleh pada konfigurasi 1-gram, dengan akurasi sebesar 97.22% dan f1-score pada seluruh kelas berada pada rentang 0.95 hingga 0.99. Ini menunjukkan bahwa pada kondisi distribusi data yang seimbang, SVM mampu mengenali setiap emosi dengan sangat baik, bahkan tanpa kehilangan performa dibanding model tanpa SMOTE.



🔹 Evaluasi SVM + SMOTE dengan TF-IDF 2-gram

	precision	recall	f1-score	support
anger	0.94	0.72	0.81	141
fear	0.53	0.91	0.67	122
joy	0.92	0.63	0.75	125
sadness	0.86	0.74	0.80	116
accuracy			0.75	504
macro avg	0.81	0.75	0.76	504
weighted avg	0.82	0.75	0.76	504

Akurasi: 0.748

🔹 Evaluasi SVM + SMOTE dengan TF-IDF 3-gram

	precision	recall	f1-score	support
anger	0.95	0.14	0.25	141
fear	0.29	0.99	0.44	122
joy	1.00	0.26	0.41	125
sadness	0.93	0.23	0.37	116
accuracy			0.40	504
macro avg	0.79	0.41	0.37	504
weighted avg	0.80	0.40	0.36	504

Akurasi: 0.3968

🔹 Evaluasi SVM + SMOTE dengan TF-IDF 4-gram

	precision	recall	f1-score	support
anger	1.00	0.06	0.11	141
fear	0.25	1.00	0.40	122
joy	1.00	0.09	0.16	125
sadness	1.00	0.03	0.07	116
accuracy			0.29	504
macro avg	0.81	0.29	0.19	504
weighted avg	0.82	0.29	0.18	504

Akurasi: 0.2877

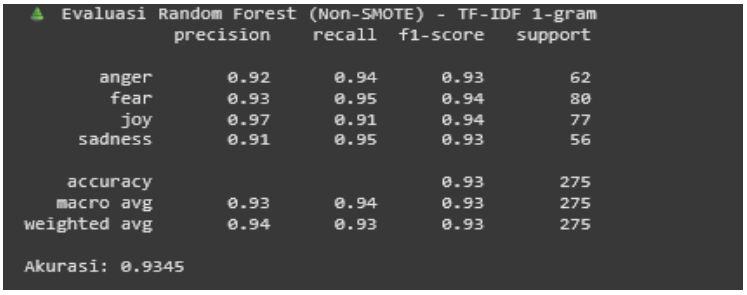
Gambar 3. 4 Evaluasi SVM + SMOTE dengan 2-gram hingga 4-gram

Namun, hasil pada 2-gram hingga 4-gram justru menunjukkan penurunan performa yang signifikan, dan pola ini identik dengan hasil Non-SMOTE pada konfigurasi yang sama. Pada 2-gram, akurasi hanya mencapai 74.8%, dan f1-score untuk kelas *fear* hanya 0.67, meskipun recall cukup tinggi. Pada 3-gram, akurasi turun drastis menjadi 39.68%, diikuti oleh 28.77% pada 4-gram. F1-score untuk kelas *anger*, *joy*, dan *sadness* sangat rendah (berada di bawah 0.41), meskipun precision terlihat tinggi akibat prediksi yang terlalu spesifik ke kelas tertentu (overfitting).

Random Forest

Non-SMOTE

Hasil klasifikasi emosi menggunakan algoritma Random Forest tanpa menggunakan Teknik *balancing data.*



🎄 Evaluasi Random Forest (Non-SMOTE) - TF-IDF 1-gram

	precision	recall	f1-score	support
anger	0.92	0.94	0.93	62
fear	0.93	0.95	0.94	80
joy	0.97	0.91	0.94	77
sadness	0.91	0.95	0.93	56
accuracy			0.93	275
macro avg	0.93	0.94	0.93	275
weighted avg	0.94	0.93	0.93	275

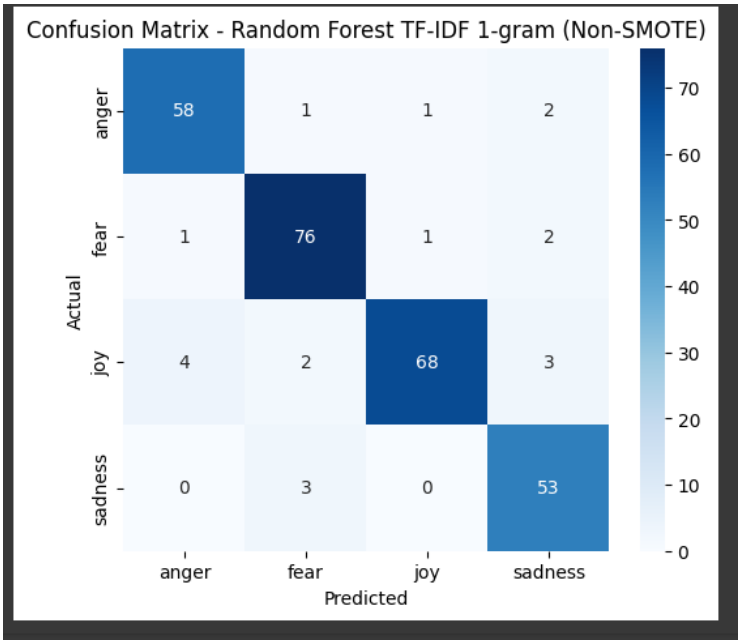
Akurasi: 0.9345

Gambar 3. 13 Evaluasi Random Forest 1-gram

Hasil evaluasi ditampilkan pada Gambar 3.15 Pada konfigurasi TF-IDF 1-gram, model menunjukkan performa yang sangat baik dengan akurasi mencapai 93.45%. Metrik f1-score pada seluruh kelas emosi berada pada rentang 0.93–0.94, menunjukkan bahwa model mampu mengenali semua kelas secara seimbang. Kelas *joy* dan *fear* memiliki performa sangat baik, sedangkan *sadness* dan *anger* juga tetap tinggi baik dari sisi precision maupun recall.

Gambar 3.14 menunjukan *confussion matrix* dari hasil klasifikasi menggunakan algoritma Random Forest 1-gram tanpa SMOTE. Model ini menunjukan kinerja yang cukup baik dalam mengenali empat kelas emosi, dengan prediksi benar yang mendominasi diagonal utama *confussion matrix.*

Pada kelas *anger*, sebanyak 58 dari 62 data berhasil diklasifikasikan dengan benar, sementara 4 lainnya salah diklasifikasikan ke kelas *fear*, *joy*, dan *sadness*. Kelas *fear* memiliki performa terbaik dengan 76 dari 80 tweet diprediksi benar, dan hanya 4 kasus yang diklasifikasikan ke kelas lain.



Gambar 3. 14 Confussion Matrix Random Forest 1-gram (Non-SMOTE)

Kelas *joy* juga menunjukkan performa yang cukup stabil dengan 68 dari 77 tweet diprediksi secara tepat. Meski demikian, masih terdapat beberapa kasus *joy* yang salah diklasifikasikan sebagai *anger* (4 kasus), *fear* (2 kasus), dan *sadness* (3 kasus). Untuk kelas *sadness*, model berhasil memprediksi 53 dari 56 tweet dengan benar, sedangkan 3 tweet lainnya salah diklasifikasikan sebagai *fear*.

Hasil ini menunjukkan bahwa model Random Forest mampu mengenali pola-pola dalam teks berbahasa Sunda dengan cukup baik ketika menggunakan fitur 1-gram. Tingkat akurasi keseluruhan mencapai 93.45%, yang sesuai dengan hasil evaluasi metrik sebelumnya. Namun, dibandingkan dengan model SVM, Random Forest menunjukkan sedikit lebih banyak *misclassification* terutama pada kelas *joy* dan *anger*.

```
🌲 Evaluasi Random Forest (Non-SMOTE) - TF-IDF 2-gram
              precision    recall  f1-score   support

       anger       0.97      0.48      0.65        62
        fear       0.43      0.96      0.60        80
         joy       1.00      0.43      0.60        77
     sadness       0.94      0.55      0.70        56

    accuracy                           0.62       275
   macro avg       0.83      0.61      0.63       275
weighted avg       0.82      0.62      0.63       275

Akurasi: 0.6218
```

Gambar 3. 6 Evaluasi Random Forest 2-gram

```
🌲 Evaluasi Random Forest (Non-SMOTE) - TF-IDF 3-gram
              precision    recall  f1-score   support

       anger       1.00      0.08      0.15        62
        fear       0.32      1.00      0.48        80
         joy       1.00      0.10      0.19        77
     sadness       1.00      0.21      0.35        56

    accuracy                           0.38       275
   macro avg       0.83      0.35      0.29       275
weighted avg       0.80      0.38      0.30       275

Akurasi: 0.3818
```

Gambar 3. 5 Evaluasi Random Forest 3-gram

Namun, ketika n-gram dinaikkan ke 2-gram, terjadi penurunan performa yang cukup signifikan. Akurasi turun menjadi 62.18%, dan terjadi ketimpangan prediksi antar kelas. Kelas fear tetap memiliki *recall* tinggi (0.96), namun *joy* dan *anger* turun drastis, masing-masing hanya memiliki recall 0.43 dan 0.48. F1-score keseluruhan turun di bawah 0.70, mengindikasikan ketidakseimbangan dalam prediksi.

Pada konfigurasi 3-gram, performa model menurun drastis. Akurasi tercatat hanya 38.18%, dengan f1-score kelas *anger*, *joy*, dan *sadness* berada di bawah 0.35, meskipun *precision* masih terlihat tinggi. Hal ini terjadi karena model terlalu sering memprediksi satu kelas tertentu (misalnya *fear*), dan mengabaikan kelas lain — sebuah indikasi bias model terhadap pola yang dominan dalam dataset yang tidak seimbang.

```
🌲 Evaluasi Random Forest (Non-SMOTE) - TF-IDF 4-gram
              precision    recall  f1-score   support

       anger       1.00      0.03      0.06        62
        fear       0.30      1.00      0.46        80
         joy       1.00      0.01      0.03        77
     sadness       1.00      0.04      0.07        56

    accuracy                           0.31       275
   macro avg       0.82      0.27      0.15       275
weighted avg       0.80      0.31      0.17       275

Akurasi: 0.3091
```

Gambar 3. 7 Evaluasi Random Forest 4-gram

Penurunan paling ekstrem terjadi pada 4-gram, di mana akurasi hanya mencapai 30.91%. Meskipun precision untuk semua kelas berada di angka 1.00 (artifisial), namun recall-nya sangat rendah (antara 0.01–0.04) pada *joy*, *sadness*, dan *anger*. Artinya, model hampir tidak bisa mengenali kelas-kelas tersebut dan hanya fokus pada satu atau dua kelas dominan, yang berbahaya dalam sistem klasifikasi emosi.

SMOTE

Evaluasi model *Random Forest* setelah diterapkan SMOTE.

Evaluasi Random Forest + SMOTE - TF-IDF 1-gram

	precision	recall	f1-score	support
anger	0.92	0.95	0.94	62
fear	0.92	0.95	0.93	80
joy	0.99	0.88	0.93	77
sadness	0.90	0.95	0.92	56
accuracy			0.93	275
macro avg	0.93	0.93	0.93	275
weighted avg	0.93	0.93	0.93	275

Akurasi: 0.9309

Gambar 3. 8 Evaluasi Random Forest + SMOTE 1-gram

Hasil evaluasi model disajikan pada Gambar 3.18. Pada konfigurasi TF-IDF 1-gram, performa model sangat baik dengan akurasi sebesar 93.09%, tidak jauh berbeda dari versi Non-SMOTE (93.45%). Nilai precision, recall, dan f1-score pada seluruh kelas juga stabil dan tinggi, menunjukkan bahwa pada konfigurasi ini, SMOTE tidak terlalu berdampak signifikan karena distribusi fitur masih cukup umum dan tidak terlalu spars.

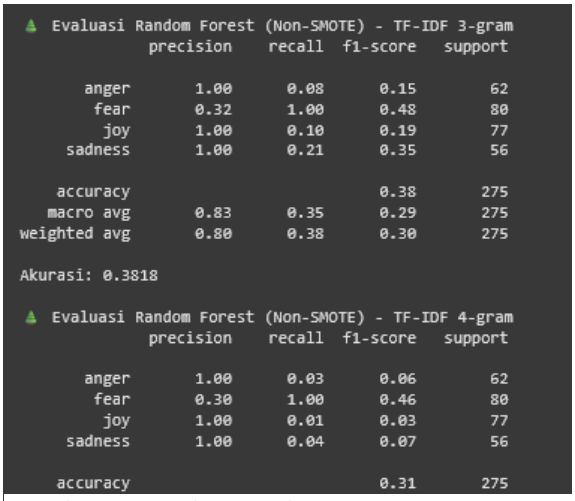
Namun, mulai dari konfigurasi 2-gram, performa model kembali menurun drastis, dengan akurasi hanya 62.18%, yang identik dengan hasil Non-SMOTE. Kelas *fear* memiliki *recall* tinggi (0.96), tetapi *precision*-nya rendah (0.43), dan kelas *joy* hanya dikenali dengan recall 0.43. Ini menandakan bahwa meskipun data sudah diseimbangkan, sparsitas fitur akibat n-gram tinggi tetap menyebabkan model kehilangan konteks yang konsisten.

Evaluasi Random Forest + SMOTE - TF-IDF 2-gram

	precision	recall	f1-score	support
anger	0.97	0.48	0.65	62
fear	0.43	0.96	0.60	80
joy	1.00	0.43	0.60	77
sadness	0.94	0.55	0.70	56
accuracy			0.62	275
macro avg	0.83	0.61	0.63	275
weighted avg	0.82	0.62	0.63	275

Akurasi: 0.6218

Gambar 3. 9 Evaluasi Random Forest + SMOTE 2-gram



```
Evaluasi Random Forest (Non-SMOTE) - TF-IDF 3-gram
              precision    recall  f1-score   support

       anger       1.00      0.08      0.15        62
        fear       0.32      1.00      0.48        80
         joy       1.00      0.10      0.19        77
     sadness       1.00      0.21      0.35        56

    accuracy                           0.38       275
   macro avg       0.83      0.35      0.29       275
weighted avg       0.80      0.38      0.30       275

Akurasi: 0.3818

Evaluasi Random Forest (Non-SMOTE) - TF-IDF 4-gram
              precision    recall  f1-score   support

       anger       1.00      0.03      0.06        62
        fear       0.30      1.00      0.46        80
         joy       1.00      0.01      0.03        77
     sadness       1.00      0.04      0.07        56

    accuracy                           0.31       275
```

Gambar 3. 10 Evaluasi Random Forest + SMOTE 3-gram dan 4-gram

Pada 3-gram, akurasi hanya mencapai 38.18%, dan pada 4-gram turun menjadi 30.91%, sama persis seperti versi Non-SMOTE. F1-score pada kelas *anger*, *joy*, dan *sadness* sangat rendah (antara 0.03 hingga 0.19), meskipun *precision* tampak tinggi akibat prediksi model yang terlalu terpaku pada satu kelas tertentu (bias prediksi).

Hasil Klasifikasi

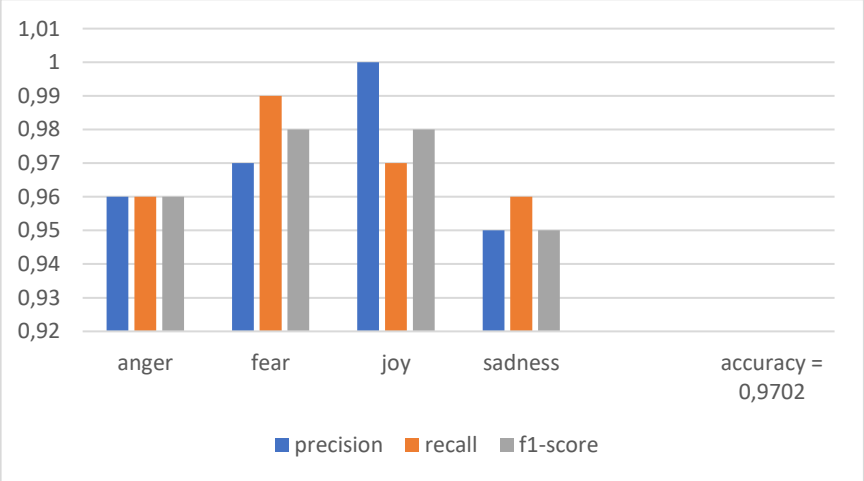
Tabel 3. 1 Tabel ringkasan hasil klasifikasi emosi

Model	**SMOTE**	**N-Gram**	**Accuracy**	**F1-score *Angry***	**F1-score *Fear***	**F1-score *Joy***	**F1-score *Sad***
SVM	No	1-gram	0.9702	0.96	0.98	0.98	0.95
SVM	No	2-gram	0.7500	0.81	0.67	0.75	0.80
SVM	No	3-gram	0.3968	0.25	0.44	0.41	0.37
SVM	No	4-gram	0.2877	0.11	0.40	0.16	0.07
Random Forest	No	1-gram	0.9345	0.93	0.94	0.94	0.93
Random Forest	No	2-gram	0.6218	0.65	0.60	0.60	0.70
Random Forest	No	3-gram	0.3818	0.15	0.48	0.19	0.35
Random Forest	No	4-gram	0.3091	0.06	0.46	0.03	0.07
Random Forest	Yes	1-gram	0.9309	0.94	0.93	0.93	0.92
Random Forest	Yes	2-gram	0.6218	0.65	0.60	0.60	0.70
Random Forest	Yes	3-gram	0.3818	0.15	0.48	0.19	0.35
Random Forest	Yes	4-gram	0.3091	0.06	0.46	0.03	0.07

Tabel 3.2 menyajikan ringkasan hasil klasifikasi emosi yang diperoleh dari penerapan dua algoritma, yaitu *Support Vector Machine* (SVM) dan *Random Forest* (RF), dengan dan

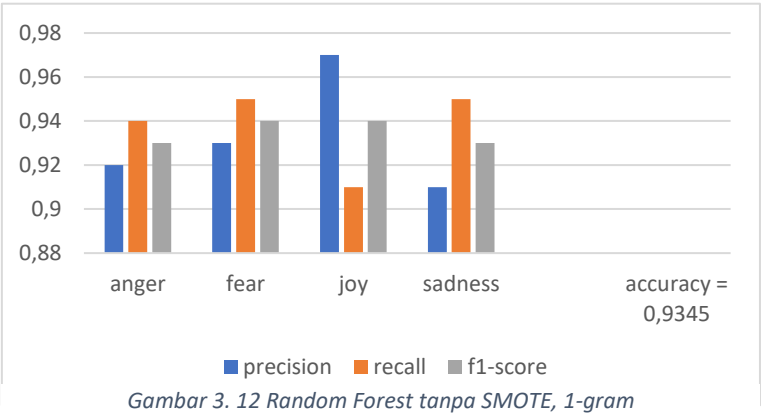
tanpa teknik penyeimbangan data SMOTE. Evaluasi dilakukan terhadap data tweet berbahasa Sunda yang telah melalui preprocessing, dengan representasi fitur menggunakan TF-IDF pada konfigurasi n-gram dari 1-gram hingga 4-gram.

Untuk memberikan gambaran yang lebih detail terkait performa masing-masing model dalam mengklasifikasikan emosi, evaluasi dilakukan tidak hanya berdasarkan nilai akurasi rata-rata, tetapi juga berdasarkan nilai precision, recall, dan F1-score per kelas emosi. Hal ini penting karena dataset yang digunakan bersifat tidak seimbang, di mana jumlah tweet pada tiap kategori emosi berbeda. Tanpa analisis per kelas, model dapat terlihat baik secara keseluruhan tetapi sebenarnya gagal mengenali kelas minoritas. Dengan menampilkan F1-score untuk masing-masing kategori emosi, seperti *joy*, *anger*, *fear*, dan *sadness*, maka dapat diketahui seberapa baik model mengenali tiap jenis emosi, serta seberapa besar dampak penerapan teknik balancing data menggunakan SMOTE terhadap peningkatan performa model, terutama untuk emosi yang jarang muncul.



Gambar 3. 11 SVM tanpa SMOTE, 1-gram

Secara umum, hasil klasifikasi menunjukkan bahwa SVM tanpa SMOTE pada 1-gram menghasilkan akurasi tertinggi sebesar 97.02%, dengan *f1-score* untuk keempat emosi juga sangat baik (berkisar antara 0.95–0.98). Hal ini menunjukkan bahwa fitur 1-gram cukup representatif dalam menangkap ekspresi emosi dalam teks pendek, tanpa perlu balancing data tambahan. Namun, performa SVM mengalami penurunan drastis pada n-gram yang lebih tinggi, khususnya pada 3-gram dan 4-gram, dengan akurasi masing-masing hanya sebesar 39.68% dan 28.77%. F1-score pada emosi seperti *anger* dan *sadness* bahkan berada di bawah



Gambar 3. 12 Random Forest tanpa SMOTE, 1-gram

0,20, mengindikasikan bahwa model tidak mampu mengenali kelas-kelas tersebut dengan baik karena sparsitas fitur.

Sementara itu, *Random Forest* tanpa SMOTE juga menunjukkan performa optimal pada 1-gram, dengan akurasi 93.45% dan nilai f1-score yang merata tinggi di semua kelas. Namun, sama seperti SVM, akurasi Random Forest menurun tajam pada 3-gram (38.18%) dan 4-gram (30.91%). Penambahan SMOTE pada Random Forest tidak memberikan peningkatan yang berarti, dengan nilai akurasi dan f1-score yang hampir identik dengan versi non-SMOTE, terutama pada n-gram tinggi. Hal ini menunjukkan bahwa meskipun SMOTE berhasil menyeimbangkan distribusi data, performa model tetap sangat dipengaruhi oleh kualitas fitur n-gram.

Dari hasil tersebut, dapat disimpulkan bahwa kombinasi terbaik dalam penelitian ini adalah SVM dengan TF-IDF 1-gram tanpa SMOTE, karena memberikan akurasi dan f1-score tertinggi secara konsisten. Sebaliknya, penggunaan n-gram yang lebih tinggi dan model Random Forest menunjukkan sensitivitas yang tinggi terhadap sparsitas dan kompleksitas fitur, yang berdampak pada penurunan performa klasifikasi. Oleh karena itu, strategi representasi fitur dan penyeimbangan data perlu dipertimbangkan secara cermat dalam pengembangan sistem klasifikasi emosi berbasis teks pendek berbahasa daerah.

Kesimpulan dan Saran

Kesimpulan

Penelitian ini bertujuan untuk membandingkan dua algoritma machine learning, yaitu Support Vector Machine (SVM) dan Random Forest (RF), dalam mengklasifikasikan emosi pada tweet berbahasa Sunda. Berdasarkan hasil yang diperoleh, dapat disimpulkan bahwa model SVM dengan representasi fitur TF-IDF 1-gram tanpa penerapan SMOTE memberikan performa terbaik, dengan akurasi mencapai 97,02% dan nilai f1-score yang tinggi dan merata pada keempat kelas emosi*: joy*, *anger*, *fear*, dan *sadness*. Hasil ini menunjukkan bahwa pendekatan sederhana seperti TF-IDF 1-gram cukup efektif dalam menangani teks pendek seperti tweet, serta bahwa SVM memiliki kestabilan performa meskipun data tidak seimbang.Penggunaan SMOTE terbukti memberikan peningkatan performa yang signifikan khususnya pada model *Random Forest*, yang menunjukkan ketergantungan lebih besar terhadap distribusi data seimbang. Di sisi lain, SVM menunjukkan ketahanan terhadap ketidakseimbangan data, sehingga meskipun SMOTE diterapkan, peningkatan performanya tidak terlalu drastis.

Sementara itu, *Random Forest* juga menunjukkan hasil yang kompetitif, terutama pada konfigurasi 1-gram tanpa SMOTE dengan akurasi 93,45%. Namun, model ini cenderung lebih sensitif terhadap distribusi data dan mengalami penurunan performa yang signifikan ketika digunakan pada representasi fitur yang lebih kompleks seperti 3-gram dan 4-gram. Selain itu, penerapan metode *balancing* data menggunakan SMOTE tidak selalu meningkatkan hasil klasifikasi, khususnya pada algoritma SVM yang justru bekerja lebih optimal tanpa balancing.

Penurunan performa juga terlihat pada kedua model ketika digunakan dengan n-gram yang lebih tinggi. Hal ini disebabkan oleh meningkatnya sparsity pada representasi fitur, yang membuat model kesulitan menangkap pola emosi dari teks pendek. Secara keseluruhan, penelitian ini menunjukkan bahwa kombinasi antara algoritma SVM dan representasi fitur sederhana (*unigram*) sangat cocok untuk tugas klasifikasi emosi dalam data berbahasa daerah, serta memberikan kontribusi nyata dalam pengembangan NLP lokal berbasis *machine learning* klasik.

Berdasarkan seluruh hasil analisis, dapat disimpulkan bahwa SVM dengan TF-IDF 1-gram tanpa SMOTE adalah model yang paling direkomendasikan untuk klasifikasi emosi pada tweet berbahasa Sunda, karena memberikan hasil yang konsisten, akurat, dan stabil terhadap berbagai variasi data.

Saran

Kesimpulan dari penelitian ini menunjukan bahwa meskipun beberapa teknik dan model, seperti SVM dengan TF-IDF 1-gram tanpa SMOTE, memberikan hasil yang terbaik dalam pengujian. Untuk meningkatkan relevansi dan kegunaan model dalam praktik industri, beberapa saran yang dapat dipertimbangkan adalah sebagai berikut:

1. Jumlah dan keragaman data perlu ditingkatkan agar model dapat belajar dari lebih banyak variasi emosi dan gaya bahasa dalam bahasa Sunda. *Dataset* yang lebih besar dan seimbang akan memperbaiki performa model, terutama pada kelas-kelas emosi yang secara alami lebih jarang muncul.
2. Penelitian selanjutnya disarankan untuk mengeksplorasi metode representasi fitur berbasis embedding seperti *Word2Vec*, *FastText*, atau *BERT* yang dapat menangkap konteks semantik lebih baik dibandingkan TF-IDF. Pendekatan berbasis deep learning juga dapat dicoba, seperti penggunaan LSTM atau transformer, yang terbukti efektif dalam pemrosesan bahasa alami.
3. Perlu dilakukan analisis linguistik terhadap karakteristik bahasa Sunda, seperti penggunaan dialek atau bentuk informal dalam tweet, agar preprocessing dapat disesuaikan dengan konteks lokal. Hal ini dapat meningkatkan pemahaman model terhadap struktur bahasa.
4. Evaluasi model sebaiknya tidak hanya menggunakan metrik numerik, tetapi juga disertai analisis berbasis *confusion matrix* dan visualisasi, sehingga kesalahan prediksi dapat lebih mudah dipahami dan diperbaiki.
5. Mengingat bahwa precision tinggi tidak selalu berarti performa baik, maka penting untuk menginterpretasikan metrik evaluasi secara holistik, tidak hanya bergantung pada satu nilai tertentu, agar keputusan penggunaan model menjadi lebih tepat dan akurat.
6. Penelitian selanjutnya disarankan untuk menggunakan metode validasi silang (*k-fold cross-validation*) agar hasil klasifikasi dapat lebih terjamin keandalannya dan mampu menghindari *overfitting*

Dengan saran-saran tersebut, diharapkan penelitian lanjutan dapat mengembangkan sistem klasifikasi emosi yang lebih akurat, adaptif terhadap bahasa daerah, serta relevan untuk diterapkan pada aplikasi analisis media sosial dalam konteks lokal.

Daftar Pustaka

Anjani, A. F., Anggraeni, D., & Tirta, I. M. (2023). Implementasi Random Forest Menggunakan SMOTE untuk Analisis Sentimen Ulasan Aplikasi Sister for Students UNEJ. Jurnal Nasional Teknologi Dan Sistem Informasi, 9(2), 163–172. https://doi.org/10.25077/teknosi.v9i2.2023.163-172

Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research, 16, 321–357. https://doi.org/10.1613/jair.953

Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273–297. https://doi.org/10.1007/BF00994018

Glenn, A., LaCasse, P., & Cox, B. (2023). Emotion classification of Indonesian Tweets using Bidirectional LSTM. Neural Computing and Applications, 35(13), 9567–9578. https://doi.org/10.1007/s00521-022-08186-1

Isnain, A. R., Marga, N. S., & Alita, D. (2021). Sentiment Analysis Of Government Policy On Corona Case Using Naive Bayes Algorithm. IJCCS (Indonesian Journal of Computing and Cybernetics Systems), 15(1), 55. https://doi.org/10.22146/ijccs.60718

Joachims, T. (1998). Text categorization with Support Vector Machines: Learning with many relevant features. Proceedings of the 10th European Conference on Machine Learning (ECML-98), 137–142. https://doi.org/10.1007/BFb0026683

Korde, V. ; M. C. N. (2012). Text Classification and Classifiers: A Survey. International Journal of Artificial Intelligence & Applications, 3(2), 85–99. https://doi.org/10.5121/ijaia.2012.3207

Kotsiantis, S. B. (2007). Supervised Machine Learning: A Review of Classification Techniques. In Informatica (Vol. 31, Issue 3). https://www.informatica.si/index.php/informatica/article/view/148/140)

Liu, B. (2012). Sentiment Analysis and Opinion Mining. Synthesis Lectures on Human Language Technologies, 5(1), 1–167. https://doi.org/10.2200/S00416ED1V01Y201204HLT016

Mohammad, S. M., & Turney, P. D. (2013). CROWDSOURCING A WORD–EMOTION ASSOCIATION LEXICON. Computational Intelligence, 29(3), 436–465. https://doi.org/10.1111/j.1467-8640.2012.00460.x

Noroozi, Z., Orooji, A., & Erfannia, L. (2023). Analyzing the impact of feature selection methods on machine learning algorithms for heart disease prediction. Scientific Reports, 13(1). https://doi.org/10.1038/s41598-023-49962-w

Novianto, E., Suhirman, S., & Prasetyo, D. (2024). PERBANDINGAN METODE KLASIFIKASI RANDOM FOREST DAN SUPPORT VECTOR MACHINE DALAM MEMPREDIKSI CAPAIAN STUDI MAHASISWA. JIPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika), 9(4), 1821–1833. https://doi.org/10.29100/jipi.v9i4.5423

Prastyo, P. H., Sumi, A. S., Dian, A. W., & Permanasari, A. E. (2020). Tweets Responding to the Indonesian Government's Handling of COVID-19: Sentiment Analysis Using SVM with Normalized Poly Kernel. Journal of Information Systems Engineering and Business Intelligence, 6(2), 112. https://doi.org/10.20473/jisebi.6.2.112-122

Putra, O. V., Wasmanson, F. M., Harmini, T., & Utama, S. N. (2020b). Sundanese Twitter Dataset for Emotion Classification. CENIM 2020 - Proceeding: International Conference on Computer Engineering, Network, and Intelligent Multimedia 2020, 391–395. https://doi.org/10.1109/CENIM51130.2020.9297929

Russell, S. J. ., Norvig, Peter., & Davis, Ernest. (2010). Artificial intelligence : a modern approach. Prentice Hall.

Sebastiani, F. (2002). Machine learning in automated text categorization. ACM Computing Surveys, 34(1), 1–47. https://doi.org/10.1145/505282.505283

Sidik, P., Made, I., Sunarya, G., Gede, I., & Gunadi, A. (2025). Comparison of Random Forest and Support Vector Machine Methods in Sentiment Analysis of Student Satisfaction Questionnaire Comments at ITB STIKOM Bali. In Journal of Applied Informatics and Computing (JAIC) (Vol. 9, Issue 3). http://jurnal.polibatam.ac.id/index.php/JAIC

Statista. (2023). Number of monetizable daily active Twitter users worldwide from 1st quarter 2017 to 2nd quarter 2023. https://www.statista.com/statistics/970920/worldwide-twitter-daily-active-users/

Wijaya, R. ; K. R. (2022). Perbandingan algoritma dalam klasifikasi sentimen teks. 137–142. https://doi.org/10.1007/BFb0026683