Analisis Sentimen Terhadap SPKLU Menggunakan Metode *Naïve Bayes*

1st Hisyam Mohamad Alam
Program Studi Informatika
Universitas Telkom Surabaya
Surabaya, Indonesia
hisyammoh@student.telkomuniversity.ac.id

2nd Alqis Rausanfita
Program Studi Informatika
Universitas Telkom Surabaya
Surabaya, Indonesia
alqisfita@telkomuniversity.ac.id

3rd Tanzilal Mustaqim
Program Studi Informatika
Universitas Telkom Surabaya
Surabaya, Indonesia
tanzilal@telkomuniversity.ac.id

***Abstrak.* — Tantangan dalam pengembangan SPKLU tidak hanya terbatas pada aspek teknis dan operasional, tetapi juga berkaitan dengan rendahnya Tingkat pengetahuan Masyarakat, Penelitian terdahulu hanya berfokus pada kendaraan Listrik[1], sementara penelitian terhadap infrastruktur pendukung seperti SPKLU ini masih terbatas. permasalahan lain juga muncul pada keterbatasan data opini, untuk menjawab permasalahan ini[2]. penelitian menerapkan Naïve Bayes dan data penelitian di peroleh dari sosial media X, dengan proses crawling[3], setelah itu dilakukan pre-processing dikombinasikan dengan TF-IDF dan metode oversampling rasio 20% - 100% serta metode pembagian data dengan Stratified K-fold dengan nilai K = 5, sehingga mendapatkan hasil evaluasi terbaik, dengan nilai akurasi 87,47%, presisi 64,19%, recall 74,30%, F1-Score 67,01%,metode ini terbukti sangat bagus untuk melakukan penelitian teks dengan data yang terbatas.**

***Kata kunci*— Naïve bayes, Crawling, TF-IDF, Pre-processing, Oversampling.**

I. PENDAHULUAN

Perkembangan kendaraan Listrik di Indonesia terus meningkat sebagai Upaya pengurangan emisi karbon.Pembangunan SPKLU yang telah cukup banyak berperan penting dalam mendukung ekosistem EV. Namun, SPKLU masih menghadapi tantangan operasional, seperti fluktuasi,kebutuhan daya dan efisiensi layanan. Selain factor teknis, persepsi Masyarakat terhadap kualitas dan ketersediaan SPKLU turut mempengaruhi Tingkat kepercayaan adopsi kendaraan Listrik, terutama di era digital Ketika opini public banyak disampaikan melalui media sosial.

Analisis sentimen terhadap media sosial, khususnya *platform X[4]*, dapat dimanfaatkan untuk memahami pandangan masyarakat terhadap layanan SPKLU. Penelitian ini menerapkan algoritma *Naïve Bayes* dengan pembobotan *TF-IDF* untuk mengklasifikasikan sentimen publik secara efisien dan metode *oversampling* untuk menyeimbangkan data, serta pembagian data dengan metode *Stratified k-fold* dengan K = 5. Berbeda dengan penelitian sebelumnya yang lebih berfokus pada kendaraan listrik secara umum, studi ini menitikberatkan pada infrastruktur SPKLU. Hasil penelitian diharapkan memberikan wawasan berbasis data bagi pengelola SPKLU dalam meningkatkan kualitas layanan dan mendukung pengembangan ekosistem kendaraan listrik yang berkelanjutan di Indonesia.[5]

II. KAJIAN TEORI

A. Analisis sentimen

Analisis sentimen digunakan untuk mengidentifikasi kecenderungan opini dalam data teks dengan mengklasifikasikannya ke dalam kategori tertentu.[6] Pada penelitian ini, teks direpresentasikan menggunakan pembobotan *TF-IDF* guna mengekstraksi fitur kata yang relevan, kemudian diklasifikasikan menggunakan algoritma *Naïve Baye*s yang bersifat probabilistik dan efisien untuk data teks berdimensi tinggi. Ketidakseimbangan distribusi kelas diatasi melalui penerapan teknik oversampling *SMOTE* dan *ADASYN[7].* yang membangkitkan data sintetis pada kelas minoritas agar proses pembelajaran model menjadi lebih seimbang. Pendekatan ini diharapkan mampu meningkatkan kinerja klasifikasi serta keandalan hasil analisis sentimen.

B. *Naïve Bayes*

teknik yang digunakan dalam proses klasifikasi. Teknik ini didasarkan pada prinsip-prinsip probabilitas dan statistik yang diperkenalkan oleh seorang ilmuwan asal Inggris bernama Thomas Bayes[8]. Pendekatan ini memprediksi kemungkinan terjadinya sebuah peristiwa di masa mendatang dengan merujuk pada pengalaman dari peristiwa yang telah terjadi di masa lalu. Salah satu karakteristik utama *dari Naïve Bayes Classifier* adalah asumsi kemandirian yang sangat kuat (yang dianggap naif) antara masing-masing variabel atau kondisi, meskipun asumsi ini tidak sepenuhnya benar. Dalam setiap kelas keputusan, *Naïve Bayes* menghitung probabilitas berdasarkan anggapan bahwa kelas keputusan tersebut benar dapat dilihat pada persamaan berikut :

$$p(c_j|d_i) = \frac{p(d_i|c_j)p(c_j)}{p(d_i)} \qquad (1)$$

C. *TF – IDF*

pembobotan kata menggunakan Term Frequency-Inverse Document Frequency (TF-IDF). Pembobotan kata bertujuan untuk mengonversi data teks menjadi bentuk numerik[9] Dalam pembelajaran mesin (machine learning) dan deep learning, data numerik sangat penting untuk memastikan proses pengolahan berjalan optimal,persamaan nya dapat dilihat sebagai berikut :

$$TF - IDF = (t, d) . IDF (t) \quad (2)$$

D. *Oversampling*

Ketidakseimbangan data kelas membuat model sering kali memfokuskan diri pada kelas yang lebih besar dan mengabaikan kelas yang lebih kecil. Salah satu metode untuk mengatasi masalah ini adalah dengan menggunakan teknik oversampling[10], yaitu meningkatkan jumlah data pada kelas yang lebih kecil agar rasio antara kelas-kelas tersebut menjadi lebih seimbang, dua metode yang digunakan Adalah *SMOTE* dan *ADASYN.*

E. *Stratified K-fold*

Stratified K-Fold Cross Validation dengan K = 5 membagi keseluruhan data ke dalam lima bagian dengan menjaga perbandingan jumlah kelas pada setiap bagian. Dalam setiap tahap pengujian, satu fold dimanfaatkan sebagai data pengujian, sedangkan empat fold lainnya digunakan untuk proses pelatihan model[11] Strategi ini diterapkan untuk memperoleh hasil evaluasi yang lebih konsisten dan mencerminkan kondisi data sebenarnya, khususnya ketika dataset memiliki ketidakseimbangan antar kelas.Detail pembagian dapat dilihat pada Tabel berikut.

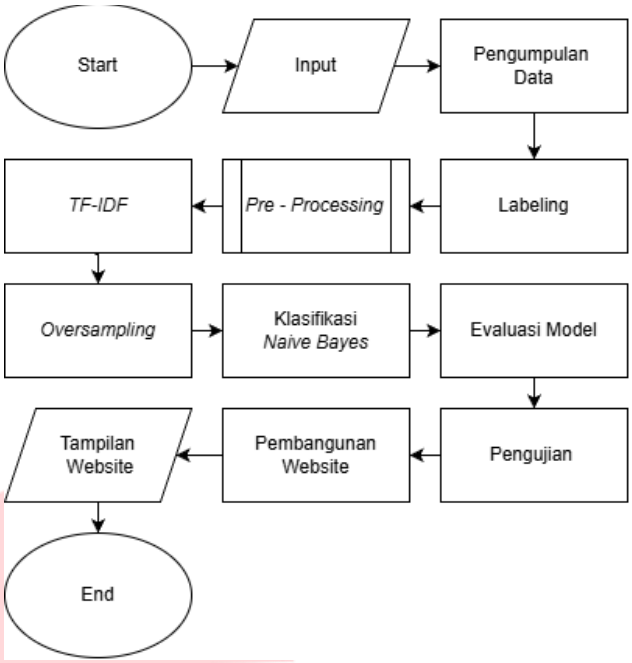
TABEL 1
(DETAIL PEMBAGIAN 5 FOLD)

Fold 1	Fold 1	Fold 1	Fold 1	Fold 1
Fold 2	Fold 2	Fold 2	Fold 2	Fold 2
Fold 3	Fold 3	Fold 3	Fold 3	Fold 3
Fold 4	Fold 4	Fold 4	Fold 4	Fold 4
Fold 5	Fold 5	Fold 5	Fold 5	Fold 5

III. METODE

A. Alur Penelitian

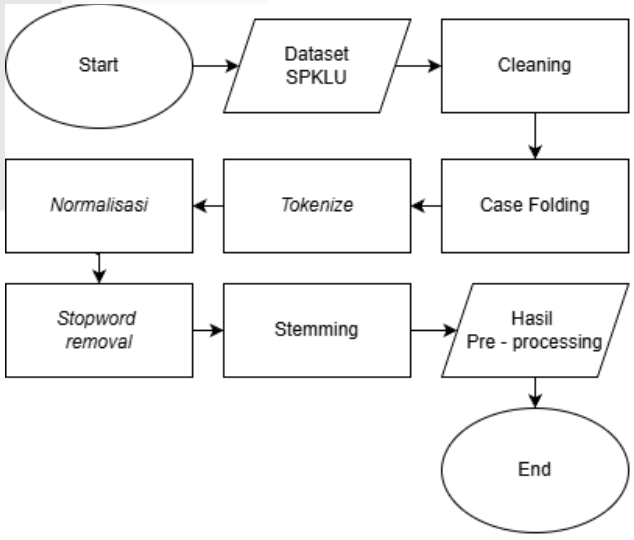
Penelitian ini membahas metodologi analisis sentimen yang disusun secara bertahap dan sistematis. Data teks direpresentasikan menggunakan pembobotan *TF-IDF* dan diklasifikasikan dengan algoritma *Naïve Bayes*. Ketidakseimbangan kelas pada data ditangani melalui penerapan teknik oversampling *SMOTE* dan *ADASYN* sebelum proses pelatihan model. Selanjutnya, kinerja model diuji menggunakan skema *Stratified K-Fold* dan dievaluasi berdasarkan metrik akurasi, presisi, recall, serta F1-score. Model dengan performa terbaik kemudian diimplementasikan ke dalam aplikasi berbasis web sebagai media penyajian hasil analisis sentimen.



GAMBAR 1
(ALUR PENELITIAN)

B. Proses *Pre-Processing*

pre-processing dilakukan melalui serangkaian langkah yang disusun secara sistematis untuk meningkatkan kualitas data teks sebelum dianalisis. Proses diawali dengan *cleansing* untuk menghilangkan elemen yang tidak relevan, seperti tanda baca, URL, angka, dan simbol. Selanjutnya, dilakukan *case folding* dengan mengubah seluruh teks menjadi huruf kecil guna menyeragamkan format data. Tahap berikutnya adalah *tokenization*, yaitu memecah teks menjadi unit kata, yang kemudian dilanjutkan dengan proses normalisasi untuk menyamakan kata tidak baku ke bentuk standar. Setelah itu, *stopword removal* diterapkan untuk menghapus kata-kata umum yang tidak memiliki pengaruh signifikan terhadap makna sentimen. Tahap akhir adalah *stemming* untuk mengembalikan kata ke bentuk dasarnya. Seluruh tahapan *pre-processing* ini bertujuan menghasilkan data yang bersih dan siap digunakan pada proses analisis sentimen selanjutnya.[12]



GAMBAR 2
(ALUR PRE – PROCESSING)

C. Proses *Oversampling*

Dalam penelitian ini, teknik oversampling diterapkan pada data latih untuk mengatasi permasalahan ketidakseimbangan kelas (*class imbalance*) pada dataset. Kondisi ini berpotensi menyebabkan model klasifikasi lebih dominan mengenali kelas mayoritas, sehingga kemampuan prediksi terhadap kelas minoritas menjadi kurang optimal. Oleh karena itu, diperlukan strategi penyeimbangan data sebelum proses pelatihan model dilakukan.

Pendekatan oversampling yang digunakan meliputi *SMOTE* (*Synthetic Minority Over-sampling Technique*) dan *ADASYN* (*Adaptive Synthetic Sampling*) dengan variasi rasio peningkatan kelas minoritas sebesar 20%, 40%, 60%, 80%, hingga 100%[10] . SMOTE menghasilkan sampel sintetis baru dengan menginterpolasikan data kelas minoritas berdasarkan tetangga terdekat, sehingga meningkatkan keragaman data tanpa melakukan duplikasi. Sementara itu, ADASYN secara adaptif menambahkan data sintetis pada area yang memiliki tingkat kesulitan klasifikasi lebih tinggi, sehingga model menjadi lebih sensitif terhadap pola kompleks pada kelas minoritas. Evaluasi terhadap berbagai rasio oversampling dilakukan untuk menentukan konfigurasi yang menghasilkan kinerja klasifikasi paling optimal.

IV. HASIL DAN PEMBAHASAN

A. Hasil Evaluasi

Proses Evaluasi kinerja model pada penelitian ini dilakukan melalui dua pendekatan pengujian yang berbeda. Pendekatan pertama memanfaatkan d*ataset* asli tanpa penerapan *Oversampling*, yang berfungsi untuk menilai performa dasar model klasifikasi. Pendekatan kedua menggunakan *dataset* yang telah melalui proses *Oversampling*. Pada tahap ini, dua metode yang digunakan, yaitu *SMOTE* dan *ADASYN*, masing masing di uji dengan rasio pada data minoritas sebesar 20%,40%,60%,80%, hingga 100%. Seluruh hasil pengujian tersebut di analisis untuk mengetahui pengaruh pemilihan metode dan Tingkat *Oversampling* terhadap kinerja klasifikasi sentimen

B. Evaluasi *SMOTE*

SMOTE diterapkan dengan variasi rasio oversampling sebesar 20% hingga 100% untuk mengevaluasi pengaruh penyeimbangan data terhadap kinerja klasifikasi. *Dataset* awal memiliki distribusi kelas yang tidak seimbang, dengan sekitar 845 data pada kelas mayoritas dan 68 data pada kelas minoritas, sehingga model berpotensi lebih dominan dalam mempelajari karakteristik kelas mayoritas. Pada rasio oversampling 20% dan 40%, penambahan data sintetis pada kelas minoritas masih relatif terbatas, sehingga peningkatan performa model belum menunjukkan perubahan yang signifikan.

Performa yang lebih optimal mulai diperoleh pada rasio 60%, di mana distribusi kelas menjadi lebih seimbang dan representatif. Pada kondisi ini, model *Naïve Bayes* mampu mengenali pola kelas minoritas secara lebih efektif, yang ditunjukkan oleh peningkatan nilai *recall* dan F1-score. Hasil tersebut mengindikasikan bahwa model semakin mampu mengurangi kesalahan klasifikasi pada kelas minoritas tanpa mengorbankan stabilitas kinerja secara keseluruhan.

Sebaliknya, pada rasio *oversampling* yang lebih tinggi, yaitu 80% dan 100%, jumlah data sintetis yang dihasilkan cenderung meningkat secara berlebihan. Kondisi ini berpotensi menyebabkan *overfitting*, di mana model terlalu menyesuaikan diri dengan data latih sehingga kemampuan generalisasi terhadap data baru menurun. Berdasarkan hasil pengujian tersebut, rasio *SMOTE* sebesar 60% dipilih sebagai konfigurasi yang paling optimal karena mampu memberikan keseimbangan terbaik antara peningkatan performa dan kemampuan generalisasi model.

Hasil evaluasi kinerja model terbaik dengan penerapan *SMOTE* pada rasio 60% disajikan pada table berikut.

TABEL 2
(HASIL EVALUASI TERBAIK *SMOTE* 60%)

Fold	Accuracy	Precision	Recall	F1-Score
1	88,21%	64,91%	74,69%	67,99%
2	89,91%	66,79%	72,92%	69,18%
3	85,96%	61,59%	70,78%	64,06%
4	86,40%	63,72%	76,43%	66,91%
5	85,53%	62,89%	75,95%	65,89%

C. Evaluasi *ADASYN*

Penerapan metode *ADASYN* dilakukan untuk menyesuaikan pembangkitan data sintetis berdasarkan tingkat kesulitan klasifikasi pada 68 data kelas minoritas dari 845 data kelas mayoritas. Pendekatan ini bertujuan agar penambahan data tidak dilakukan secara merata, melainkan difokuskan pada area data yang memiliki kompleksitas lebih tinggi. Pada rasio oversampling 20% dan 40%, jumlah data sintetis yang dihasilkan masih relatif terbatas dan hanya mencakup sebagian kecil wilayah data, sehingga peningkatan kinerja model belum menunjukkan perubahan yang signifikan.

Peningkatan performa yang paling optimal diperoleh pada rasio 60%, di mana *ADASYN* mampu menghasilkan sampel sintetis secara adaptif pada area yang sulit diklasifikasikan. Kondisi ini membuat model *Naïve Bayes* menjadi lebih peka terhadap karakteristik kelas minoritas, yang tercermin dari peningkatan nilai *precision*, *recall*, dan F1-score secara seimbang. Hasil ini menunjukkan bahwa model tidak hanya mampu mengenali kelas minoritas dengan lebih baik, tetapi juga tetap mempertahankan kestabilan performa secara keseluruhan.

Sebaliknya, pada rasio *oversampling* yang lebih tinggi, yaitu 80% dan 100%, jumlah data sintetis yang dihasilkan cenderung berlebihan dan berpotensi menimbulkan noise. Dan *overfitting. ADASYN* dengan rasio oversampling 60% dipilih sebagai konfigurasi yang paling optimal karena mampu memberikan keseimbangan terbaik antara peningkatan performa dan stabilitas model,Hasil evaluasi pada *ADASYN 60%* di sajikan pada table berikut.

TABEL 3
(HASIL EVALUASI TERBAIK *ADASYN* 60%)

Fold	Accuracy	Precision	Recall	F1-Score
1	86,90%	61,54%	68,58%	63,75%
2	90,35%	67,61%	73,15%	69,86%
3	85,96%	61,59%	70,78%	64,06%

4	87,28%	66,05%	82,31%	70,02%
5	86,84%	64,16%	76,67%	67,44%

D. Evaluasi Hasil Terbaik dan Terburuk

Berdasarkan hasil evaluasi pada seluruh skenario pengujian, dapat disimpulkan bahwa penerapan teknik *oversampling* memberikan pengaruh yang signifikan terhadap peningkatan kinerja model dibandingkan dengan penggunaan data asli tanpa *oversampling*. Model yang dilatih pada data tidak seimbang cenderung kurang optimal dalam mengenali kelas minoritas, meskipun nilai akurasi yang dihasilkan relatif tinggi.

Pada variasi rasio oversampling antara 20% hingga 40%, peningkatan performa model masih bersifat terbatas karena jumlah data sintetis yang dihasilkan belum cukup untuk merepresentasikan karakteristik kelas minoritas secara menyeluruh. Performa model mulai menunjukkan peningkatan yang lebih stabil pada rasio 60%, di mana distribusi data antar kelas menjadi lebih seimbang dan model mampu mempelajari pola kelas minoritas dengan lebih efektif. Sebaliknya, pada rasio yang lebih tinggi, yaitu 80% hingga 100%, penambahan data sintetis yang berlebihan berpotensi menimbulkan *overfitting [13].*

Di antara seluruh konfigurasi yang diuji, metode *ADASYN* dengan rasio *oversampling* sebesar 60% menghasilkan performa terbaik secara konsisten pada hampir seluruh metrik evaluasi, termasuk *accuracy*, precision, recall, dan F1-score.[14] Hal ini menunjukkan bahwa pendekatan adaptif yang digunakan oleh *ADASYN* lebih efektif dalam menangani ketidakseimbangan data dibandingkan metode *SMOTE*, khususnya pada dataset opini media sosial yang memiliki perbedaan data kelas terutama kelas minoritas yang lebih sedikit daripada kelas mayoritas

Ringkasan hasil evaluasi terbaik dan terburuk dari seluruh skenario pengujian selanjutnya disajikan pada tabel berikut.

TABEL 4
(TABEL HASIL EVALUASI TERBAIK DAN TERBURUK)

Metode	Rasio	Accuracy	Precision	Recall	F1-Score
ADASYN	0.6	0.874676	0.641909	0.742980	0.670254
SMOTE	0.6	0.872033	0.639784	0.741556	0.668050
SMOTE	0.4	0.919356	0.703766	0.637310	0.657921
SMOTE	0.8	0.841362	0.623377	0.762847	0.648529
ADASYN	0.8	0.836984	0.621697	0.771299	0.646779
SMOTE	1.0	0.820329	0.618936	0.794753	0.640941
ADASYN	1.0	0.807171	0.611839	0.787644	0.628228
ADASYN	0.4	0.911472	0.656231	0.606007	0.621112
ADASYN	0.2	0.924623	0.663467	0.510343	0.502603
SMOTE	0.2	0.926377	0.563158	0.505882	0.491985
Tanpa Oversampling	0.0	0.925504	0.462752	0.500000	0.480655

V. KESIMPULAN

Penelitian ini menunjukkan bahwa penerapan algoritma Naïve Bayes dengan pembobotan *TF-IDF* mampu digunakan untuk menganalisis sentimen opini publik terhadap layanan Stasiun Pengisian Kendaraan Listrik Umum (SPKLU)[15]. Hasil pengujian membuktikan bahwa permasalahan ketidakseimbangan kelas pada data media sosial berpengaruh terhadap kinerja model, di mana skenario tanpa oversampling menghasilkan nilai akurasi, *precision*, *recall*, dan *F1-score* yang rendah meskipun akurasinya relatif tinggi. Penerapan teknik *oversampling* terbukti meningkatkan performa klasifikasi, khususnya dalam mengenali kelas minoritas. Konfigurasi terbaik diperoleh pada metode *ADASYN* dengan rasio 60%(0,6) yang menghasilkan keseimbangan optimal antara *accuracy, precision, recall, dan F1-score*. Dengan demikian, kombinasi *Naïve Bayes, TF-IDF*, dan teknik *oversampling* yang tepat dapat menjadi pendekatan yang efektif dalam memahami persepsi masyarakat terhadap layanan SPKLU serta mendukung pengambilan keputusan berbasis data dalam pengembangan infrastruktur kendaraan listrik di Indonesia.

REFERENSI

[1] J. A. Sanguesa, V. Torres-Sanz, P. Garrido, F. J. Martinez, and J. M. Marquez-Barja, “A review on electric vehicles: Technologies and challenges,” Mar. 01, 2021, *MDPI*. doi: 10.3390/smartcities4010022.

[2] M. Y. Nur, K. Nur, and A. Anugrah, “Development of Public Electric Vehicle Charging Stations (SPKLU) in Makassar as an Electric Car Supporting Means,” vol. 13, no. 2, pp. 389–396, 2024, doi: 10.32832/astonjadro.v13i12.

[3] Kharisma Rahayu, M. Khairul Anam, Lusiana Efrizoni, Nurjayadi, and Triyani Arita Fitri, “OPTIMASI TEKNIK VOTING PADA SENTIMEN ANALISIS PEMILIHAN PRESIDEN 2024 MENGGUNAKAN MACHINE LEARNING,” *The Indonesian Journal of Computer Science*, vol. 13, no. 4, Jul. 2024, doi: 10.33022/ijcs.v13i4.4119.

[4] I. Saputra *et al.*, “Analisis Sentimen Pengguna Marketplace Bukalapak dan Tokopedia di Twitter Menggunakan Machine Learning,” *Faktor Exacta*, vol. 13, no. 4, p. 200, Feb. 2021, doi: 10.30998/faktorexacta.v13i4.7074.

[5] T. S. Ramadhan and B. Wibowo, “Risk Analysis of Electricity Demand at Public Electric Vehicle Charging Stations (SPKLU): CVaR Model Approach,” *Quantitative Economics and Management Studies*, vol. 5, no. 3, pp. 528–540, May 2024, doi: 10.35877/454ri.qems2580.

[6] A. Bijaksana, P. Negara, H. Muhardi, and I. M. Putri, “ANALISIS SENTIMEN MASKAPAI PENERBANGAN MENGGUNAKAN METODE NAIVE BAYES DAN SELEKSI FITUR INFORMATION GAIN SENTIMENT ANALYSIS ON AIRLINES USING NAÏVE BAYES METHOD

AND FEATURE SELECTION INFORMATION GAIN," vol. 7, no. 3, pp. 599–606, 2020, doi: 10.25126/jtiik.202071947.

[7] R. A. Nurdian, Mujib Ridwan, and Ahmad Yusuf, "Komparasi Metode SMOTE dan ADASYN dalam Meningkatkan Performa Klasifikasi Herregistrasi Mahasiswa Baru," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 8, no. 1, Apr. 2022, doi: 10.28932/jutisi.v8i1.4004.

[8] Evi Purnamasari, "Prediksi Tingkat Kepuasan Dalam Pembelajaran Daring Menggunakan Algoritma Naive Bayes," 2023.

[9] R. Wati, S. Ernawati, and H. Rachmi, "Pembobotan TF-IDF Menggunakan Naïve Bayes pada Sentimen Masyarakat Mengenai Isu Kenaikan BIPIH," *Jurnal Manajemen Informatika (JAMIKA)*, vol. 13, no. 1, pp. 84–93, Apr. 2023, doi: 10.34010/jamika.v13i1.9424.

[10] M. Anjas Aprihartha *et al.*, "Penyelesaian Masalah Ketidakseimbangan Data Melalui Teknik Oversampling dan Undersampling pada Klasifikasi Siswa Tidak Naik Kelas," *Jurnal Teknik Ibnu Sina)*, vol. 9, no. 01, 2024, doi: 10.36352/jt-ibsi.v9i01.807.

[11] T. Riska Muliani, J. Sumarsono, I. S. Siti Wardatullatifah, P. Studi Teknik Pertanian, and F. Teknologi Pangan dan Agroindustri, "DETEKSI TINGKAT KEMATANGAN BUAH ALPUKAT (Persea americana Mill.) MENGGUNAKAN ALGORITMA KLASIFIKASI DAN METODE STRATIFIED K-FOLD CROSS VALIDATION Detection of Avocado Fruit Ripeness Level Using Classification Algorithm and Stratified K-Fold Cross Validation Method," 2024. [Online]. Available: https://journal.unram.ac.id/index.php/agent

[12] D. Darwis, N. Siskawati, and Z. Abidin, "Penerapan Algoritma Naive Bayes untuk Analisis Sentimen Review Data Twitter BMKG Nasional," vol. 15, no. 1, 2021.

[13] E. Erlin, Y. Desnelita, N. Nasution, L. Suryati, and F. Zoromi, "Dampak SMOTE terhadap Kinerja Random Forest Classifier berdasarkan Data Tidak seimbang," *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 21, no. 3, pp. 677–690, Jul. 2022, doi: 10.30812/matrik.v21i3.1726.

[14] A. Riyani, M. Zidny Naf'an #2, and A. Burhanuddin, "Penerapan Cosine Similarity dan Pembobotan TF-IDF untuk Mendeteksi Kemiripan Dokumen," 2019.

[15] F. Septianingrum and A. S. Y. Irawan, "Metode Seleksi Fitur Untuk Klasifikasi Sentimen Menggunakan Algoritma Naive Bayes: Sebuah Literature Review," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 5, no. 3, p. 799, Jul. 2021, doi: 10.30865/mib.v5i3.2983.