

Penerapan Random Oversampling dan Algoritma Boosting untuk Memprediksi Kualitas Buah Jeruk

Immanuel Khrisna Ananda ^{1,*}, Ahmad Zainul Fanani ¹, Dicky Setiawan ¹, Duta Firdaus Wicaksono ²

¹ Program Studi Teknik Informatika, Universitas Dian Nuswantoro, Indonesia

² Program Studi Sistem Informasi, Universitas Dian Nuswantoro, Indonesia

* Correspondence: 111202013030@mhs.dinus.ac.id

Copyright: © 2024 by the authors

Received: 30 April 2024 | Revised: 2 Mei 2024 | Accepted: 17 April 2024 | Published: 20 Juni 2024

Abstrak

Data tahun 2019 menunjukkan produksi jeruk dunia meningkat menjadi 79 juta ton. Meskipun berbagai jenis jeruk dijual di pasar Indonesia, namun masih banyak yang menjual jeruk berkualitas rendah. Tujuan dari penelitian ini adalah menerapkan teknik random oversampling dan algoritma boosting untuk memprediksi kualitas buah jeruk. Dengan sumber data dari dataset publik Orange Quality Analysis Dataset penelitian ini merupakan penelitian prediktif atau eksperimental dengan menggunakan teknik random oversampling untuk menangani ketidakseimbangan data dan menggabungkan dengan metode algoritma boosting seperti adaboost, gradient boosting, light GBM, dan catboost dengan objek data yang diteliti ukuran, berat, tingkat kemanisan, tingkat keasaman, dan yang lainnya. Dengan melihat akurasi dari beberapa algoritma boosting yang dipakai menunjukkan bahwa algoritma catboost menghasilkan tingkat akurasi tertinggi sebesar 91,42%. Diharapkan produsen jeruk dapat menghasilkan produk yang berkualitas tinggi dan meminimalkan terjadinya jeruk berkualitas rendah sehingga konsumen mendapatkan jeruk berkualitas tinggi. Hal ini juga dapat membantu produsen memasarkan buah jeruk tidak hanya di dalam negeri tetapi juga ke luar negeri.

Kata kunci: algoritma *boosting*; kualitas jeruk; *random oversampling*

Abstract

According to the 2019 data, global orange production has increased significantly, reaching 79 million tons. However, despite the availability of various types of oranges in Indonesian markets, many vendors still sell low-quality oranges. To address this issue, researchers have applied random oversampling and boosting algorithms to predict orange quality, using the public Orange Quality Analysis Dataset. This study uses random oversampling to address data imbalance and combines it with boosting algorithms like Adaboost, Gradient Boosting, Light GBM, and CatBoost. The data features considered include size, weight, sweetness level, acidity level, and others. The accuracy of the boosting algorithms used varied, with CatBoost showing the highest accuracy rate of 91.42%. The hope is that this research can help orange producers create high-quality products and reduce the occurrence of low-quality oranges, ultimately providing consumers with better oranges. Additionally, this can help producers market their oranges both domestically and internationally.

Keywords: *boosting algorithms*; *orange quality*; *random oversampling*

PENDAHULUAN

Industri pertanian dan agribisnis memegang peranan penting dalam menyediakan makanan bagi populasi dunia. Di dalamnya, prediksi kualitas produk menjadi salah satu aspek yang krusial untuk memastikan kepuasan pelanggan dan efisiensi produksi yang optimal (Sidik



et al., 2023). Khususnya dalam kasus buah jeruk, pemantauan mutu yang tepat dapat memengaruhi proses panen, penyimpanan, distribusi, dan akhirnya, pengalaman konsumen. Jeruk yang dijual umumnya jelek dan kusam tampilannya serta bervariasi dalam rasa dan kualitas. Oleh karena itu, seiring dengan meningkatnya produksi jeruk, sistem pemilihan mutu jeruk masih bersifat tradisional atau hanya berdasarkan penglihatan sehingga tidak memberikan klasifikasi yang cukup baik (Ginting & Ginting, 2021). Memanen buah jeruk yang belum matang berdampak pada kualitas buah jeruk yang dipasarkan karena membuat buah jeruk cepat busuk ketika dipasarkan ke konsumen (Bangun & Sihombing, 2021). Jeruk dengan nama latin *Citrus* mempunyai banyak manfaat untuk kesehatan dan gizi yang tinggi karena kandungan dari vitamin C dan gula yang baik (Manihuruk et al., 2019). Beberapa jenis buah yang sering dikembangkan oleh petani yaitu jeruk manis, jeruk limau dan jeruk siam. Jeruk dapat dibudidayakan dan dibudidayakan oleh petani dataran rendah hingga dataran tinggi dengan menggunakan berbagai varietas/benih komersial dan dapat dikonsumsi oleh masyarakat berpendapatan rendah hingga tinggi (Abdullah & Sandy, 2021). Jeruk mempunyai banyak macam rasa sehingga pembeli hanya memanfaatkan informasi dari penjual buah saja. Terkadang informasi yang diberikan penjual tidak sesuai dengan apa yang diharapkan karena penjual masih melakukan cara yang sangat tradisional sehingga informasi yang diberikan bersifat subyektif dan tidak konsisten (Barkah, 2020).

Sistem yang digunakan penjual masih sangat tradisional maka dari itu untuk memenuhi kebutuhan akan sistem prediksi yang dapat mengklasifikasikan kualitas buah jeruk, faktor-faktor yang mempengaruhi seperti ukuran, berat, tingkat kemanisan, tingkat keasaman, kelembutan buah, waktu panen, kematangan buah, warna, variasi dan cacat dapat dilakukan. dipertimbangkan. digunakan sebagai parameter dengan teknik kecerdasan buatan. Langkah pertama melibatkan pengumpulan data dari berbagai sumber untuk mengidentifikasi keanekaragaman ciri-ciri jeruk. Data tersebut kemudian diproses melalui pembersihan dan pra-pemrosesan yang cermat untuk menghilangkan *outlier* dan mempersiapkannya untuk analisis lebih lanjut. Model pembelajaran mesin kemudian dipilih dan dilatih berdasarkan data yang dihasilkan untuk memahami pola yang mendasari kualitas buah jeruk. Melalui proses validasi dan evaluasi menyeluruh, model dapat diuji dan kinerjanya ditingkatkan. Hasilnya adalah sistem prediksi yang memberikan perkiraan kualitas buah jeruk yang akurat, sehingga membantu memantau dan mengendalikan kualitas buah jeruk secara efektif.

Masalah dengan ketidakseimbangan data pada *dataset* adalah hasil pembelajaran mesin cenderung memprediksi, seringkali mengklasifikasikan kelas minoritas sebagai kelas mayoritas (Maulana et al., 2023). Hal ini mengurangi kemampuan kelompok minoritas untuk memprediksi secara akurat, yang sering kali memiliki pengaruh signifikan dalam pengambilan keputusan. Sesuai dengan penelitian sebelumnya menunjukkan teknik *random oversampling* dapat meningkatkan akurasi dalam ketidakseimbangan data (Aryanti et al., 2023). Untuk mengatasi tantangan ini, teknik *random oversampling* telah muncul sebagai solusi yang populer dan efektif (Wicaksono et al., 2024). *Oversampling* telah banyak digunakan karena mempertahankan informasi asli kumpulan data dan mudah digunakan (Jiang et al., 2021)

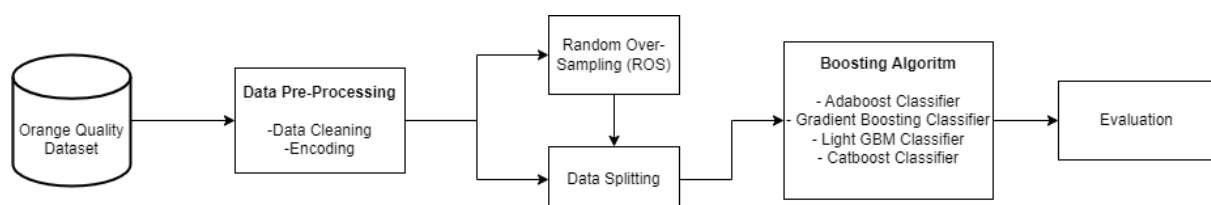
Boosting adalah teknik *assemble* yang menonjol dalam pembelajaran mesin yang menggunakan kombinasi beberapa pembelajar lemah secara berurutan untuk meningkatkan kinerja prediksi atau klasifikasi (Kim et al., 2023). Pembelajaran *ensemble* adalah metode yang menggabungkan berbagai metode pembelajaran mesin tradisional untuk meningkatkan performa model prediktif. Berbagai pendekatan telah diusulkan untuk pembelajaran kelompok. Algoritma augmentasi adalah salah satu pendekatan paling efisien dalam kelompok pembelajaran kelompok (Ganie et al., 2023). Algoritma *boosting* ini dapat disebut meta-algoritma untuk mengurangi bias dan variasi pembelajaran yang diawasi. Kami berpikir untuk membangun model ini untuk menentukan peringkat algoritma akselerasi mana yang tercepat dan paling akurat (Mane et al., 2023).

Pada penelitian yang dilakukan oleh Rahim et al. (2023) menunjukkan bahwa penggunaan metode *synthetic minority oversampling* dan *random forest* menghasilkan nilai akurasi sebesar 92%. Penelitian ini meningkat sebesar 2% dari penelitian sebelumnya. Penelitian selanjutnya Alif & Fahrudin (2024) dengan kombinasi *oversampling* dan *undersampling* dengan metode *naïve bayes*, SVM, dan *random forest* menunjukkan kenaikan nilai yang signifikan. Pada metode *random forest* sebesar 78,53%, *naïve bayes* sebesar 74,83%, dan SVM 73,41%. selain itu, penelitian Diantika (2023) dengan teknik *oversampling* dengan teknik *light GBM* pada klasifikasi *website phishing* dengan nilai akurasi sebesar 96,9%, *recall* sebesar 96,9%, *F1-score* sebesar 96,9%, dan ROC 99,7% yang secara signifikan lebih baik daripada metode yang lain. Dan penelitian Zhu et al. (2021) dengan kombinasi metode *oversampling* dan algoritma *random forest* meningkatkan akurasi dari 44% menjadi 78%.

Beberapa penelitian yang sudah ada menunjukkan bahwa dengan menggunakan teknik *oversampling* dan juga algoritma *boosting* dapat menyeimbangkan data, tetapi penelitian yang khusus membahas tentang kualitas jeruk dengan pengaplikasian teknik *random oversampling* dan beberapa algoritma *boosting* masih jarang ditemukan dan pasti dapat menjadi peluang untuk digali. Penelitian ini bertujuan untuk mengoptimalkan model prediksi kualitas buah jeruk dengan menggabungkan teknik *random oversampling* dan berbagai algoritma *boosting*. Dengan adanya penelitian ini diharapkan dapat memprediksi kualitas buah jeruk dengan lebih akurat dibandingkan tanpa teknik *random oversampling*. Oleh karena itu, penelitian ini dapat memberikan informasi berharga dan membantu petani jeruk untuk memproduksi atau memasarkan buah dengan kualitas lebih baik.

METODE

Penelitian ini bertujuan untuk meningkatkan akurasi perkiraan kualitas buah jeruk dengan menghilangkan ketidakseimbangan data dengan menggabungkan teknik *random oversampling* dan algoritma *boosting*. Kumpulan data yang digunakan tidak seimbang, dengan jumlah sampel jeruk berkualitas tinggi yang jauh lebih sedikit. Oleh karena itu, pengambilan sampel berlebih secara acak digunakan untuk menghasilkan sampel sintesis jeruk berkualitas tinggi. Kemudian beberapa algoritma *boosting* yaitu *adaboost*, *gradient boosting*, *catboost* dan *LightGBM* diterapkan pada dataset *oversample* untuk membuat model prediksi kualitas buah jeruk yang akan dijelaskan pada gambar 1.



Gambar 1. Tahapan penelitian

Pembelajaran mesin adalah pembuatan model prediktif berdasarkan data. Proses pembelajaran mesin yang digambarkan pada gambar mencakup beberapa langkah penting untuk membuat model klasifikasi yang akurat dan efisien. Pertama, data disiapkan melalui pra-pemrosesan, Sebelum memasuki tahap yang lebih mendalam, memastikan bahwa data telah bersih merupakan hal yang penting dalam pemrosesan data (Setiawan et al., 2024), setelah melakukan pembersihan data kemudian *encoding*, *encoding* merupakan salah satu tahap *preprocessing*. Cara ini untuk mengubah value dari fitur yang berupa teks menjadi angka (Kristiawan & Widjaja, 2021) dan dinormalisasi untuk memastikan kualitas data. Data tersebut kemudian dibagi menjadi set pelatihan dan set pengujian untuk pelatihan dan evaluasi model. Ketidakseimbangan kelas dalam data selanjutnya diatasi dengan menggunakan teknik *random*

oversampling, dimana data kelas minoritas direplikasi untuk menyeimbangkan distribusi dengan kelas mayoritas. Algoritma peningkatan kemudian digunakan untuk membangun model klasifikasi secara iteratif. Algoritma ini bekerja dengan membangun model baru pada setiap iterasi, dengan fokus pada kesalahan yang dilakukan pada model sebelumnya. Algoritma peningkatan yang umum digunakan adalah *Adaboost*, *Gradient Boosting*, *Light GBM*, dan *Catboost*. Terakhir, model yang dilatih dievaluasi pada set pengujian menggunakan metrik seperti akurasi, presisi, dan perolehan untuk mengukur performanya.

Pada penelitian ini menggunakan rumus akurasi saja untuk mengetahui nilai dari kualitas buah jeruk untuk mengevaluasi model klasifikasi dengan mendeskripsikan proporsi prediksi yang benar dalam kumpulan data tertentu, Terdapat beberapa singkatan dari rumus persamaan metrik akurasi TP yaitu *true positive*, FN yaitu *false positive*, TN yaitu *true negative*, FN yaitu *false negative* yang dijabarkan pada persamaan (1).

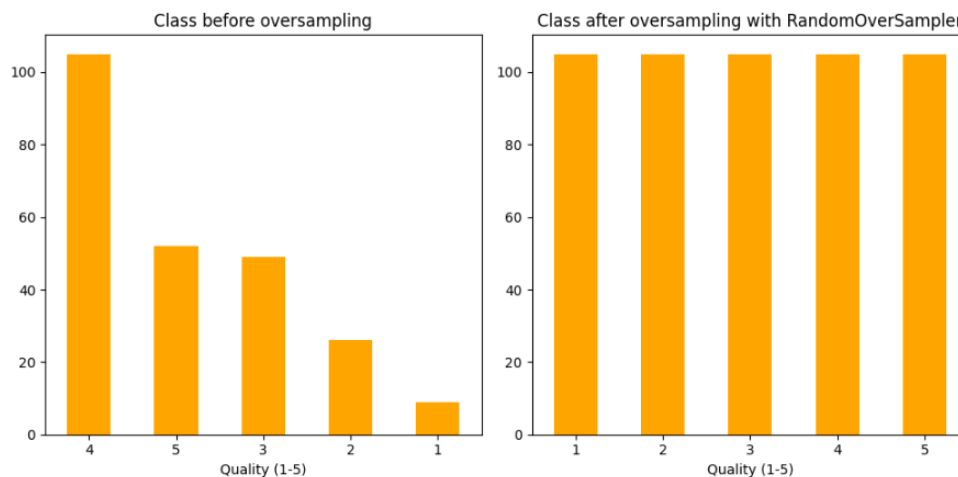
$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \times 100\% \tag{1}$$

HASIL DAN PEMBAHASAN

Hasil

Dataset yang digunakan dalam melakukan prediksi kualitas buah jeruk diambil dari laman Kaggle. *Orange Quality Analysis Dataset* digunakan sebagai bahan utama untuk penelitian ini. Dataset ini memiliki jumlah data sebanyak 241 dengan 11 fitur yang merupakan penunjang untuk mengetahui apakah kualitas jeruk yang diteliti mempunyai kualitas yang baik atau tidak sesuai dengan rentang nilai yang ada. Dataset ini memberikan informasi yang berharga untuk penelitian tentang kualitas jeruk. Dengan fitur-fitur seperti *size* (ukuran dalam cm), *weight* (berat dalam gram), *brix* (tingkat kemanisan), pH (tingkat keasaman), *softness* (tingkat kelembutan dalam rentang 1-5), *harvest time* (hari pasca panen), *ripeness* (tingkat kematangan dalam rentang 1-5), *color* (warna buah), *variety* (jenis varietas jeruk), *blemishes* (cacat/tidak), dan sebagai target adalah *quality* (kualitas buah jeruk dalam rentang 1-5). Dengan memanfaatkan dataset ini, diharapkan penelitian dapat menghasilkan wawasan yang mendalam tentang bagaimana faktor-faktor ini berinteraksi untuk menentukan kualitas jeruk.

Tahapan selanjutnya adalah *data cleaning* (pembersihan data) meliputi pengecekan *missing value* (nilai yang hilang atau kosong) dan *duplicate value* (nilai yang duplikat). Dalam dataset ini, hasil menunjukkan bahwa tidak terdapat adanya data yang hilang atau duplikat. Hal ini memberikan keyakinan bahwa dataset telah siap untuk analisis lanjutan tanpa perlu melakukan tindakan tambahan terkait dengan data yang hilang atau terduplikat.



Gambar 2. Hasil *oversampling*

Tahapan selanjutnya yaitu *oversampling*, karena masih ada dataset yang belum seimbang oleh karena itu dilakukan tahap *random oversampling*. *Random oversampling* merupakan langkah penting dalam penelitian ini untuk mengatasi ketidakseimbangan data dan meningkatkan akurasi prediksi kualitas buah jeruk. Teknik ini membantu model prediktif belajar lebih efisien dan membuat klasifikasi lebih akurat, terutama untuk kategori jeruk premium yang sebelumnya kurang terwakili. Hasil *oversampling* dapat dilihat pada gambar 2.

Terlihat pada gambar 2, sebelum dilakukan *oversampling* data pada buah jeruk masih belum seimbang. data paling banyak berada pada kualitas 4 diikuti kualitas 5, kemudian kualitas 3, kualitas 2, dan kualitas 1. tetapi setelah dilakukannya *oversampling* membuat data kualitas 1 - 5 menjadi seimbang. dalam hal ini menunjukkan bahwa *oversampling* efektif dalam menyelesaikan masalah ketidakseimbangan data, yang pada akhirnya meningkatkan keandalan model prediksi.. Tahap selanjutnya adalah melakukan *splitting data* untuk membagi 2 bagian, yaitu bagian pertama digunakan untuk melatih data dan bagian yang kedua untuk menguji data.

Ada beberapa model yang dipakai dalam penelitian ini. Yang pertama adalah *Adaboost Classifier* yaitu teknik untuk menggabungkan beberapa pembelajaran lemah menjadi pembelajar yang kuat untuk meningkatkan kinerja model prediksi. *Adaboost* memberikan bobot yang berbeda pada setiap instance data. Hal ini memungkinkan model untuk fokus pada instance yang sulit diprediksi oleh model sebelumnya, sehingga meningkatkan kinerja secara bertahap dengan memperbaiki kesalahan prediksi sebelumnya. Dengan pendekatan ini, setiap langkah dalam proses pembelajaran *adaboost* berusaha untuk memperbaiki area yang sulit diprediksi oleh model sebelumnya, sehingga secara keseluruhan, model yang dihasilkan menjadi lebih kuat dan mampu menangani variasi yang lebih kompleks dalam data. Hasil dari beberapa teknik yang telah dicoba dapat dilihat dari tabel 2 dan tabel 3.

Tabel 2. Metrik evaluasi sebelum *oversampling*

Metode	Akurasi	Presisi	Recall	F1-Score
<i>AdaBoost</i>	42,85%	58,50%	42,85%	45,09%
<i>Gradient Boosting</i>	67,34%	69,00%	67,34%	67,09%
<i>Light GBM</i>	65,30%	66,98%	65,30%	65,76%
<i>CatBoost</i>	61,22%	68,20%	61,22%	59,93%

Tabel 3. Metrik evaluasi setelah *oversampling*

Metode	Akurasi	Presisi	Recall	F1-Score
<i>AdaBoost</i>	54,28%	55,12%	54,28%	53,07%
<i>Gradient Boosting</i>	87,61%	87,42%	87,61%	87,36%
<i>Light GBM</i>	88,57%	88,46%	88,57%	88,33%
<i>CatBoost</i>	91,42%	91,53%	91,42%	91,37%

Pada tabel 2 diperlihatkan hasil metriks evaluasi beberapa metode sebelum dilakukannya *oversampling* dengan hasil yang kurang nilai tertinggi pada akurasi hanya diangka 67,34% dengan metode *gradient boosting*. Sedangkanpada tabel 3 diperlihatkan hasil metriks evaluasi dari beberapa metode setelah dilakukan *oversampling* menunjukkan nilai akurasi tertinggi diangka 91,42% dengan metode yang dipakai adalah *Catboost*.

Pembahasan

Penelitian ini menggunakan data dari *Orange Quality Analysis Dataset*, dataset publik sebanyak 241 dataset. Setelah dilakukan seleksi ulang, jumlah data bertambah menjadi 525. Variabel yang diamati adalah ukuran buah jeruk, berat, rasa manis, keasaman, keempukan, waktu panen, kematangan, waktu, varietas, kecacatan dan mutu. Teknik *random oversampling* dan empat algoritma berbeda digunakan dalam percobaan: *Adaboost*, *Gradient Boosting*,

LightGBM dan *Catboost*. Tujuan utamanya adalah memprediksi kualitas buah jeruk dan mengevaluasi tingkat akurasi yang dihasilkan oleh masing-masing algoritma.

Hasil dari metode *adaboost* terlihat sebelum *oversampling* mempunyai akurasi sebesar 42,85%, setelah dilakukannya *oversampling* metode *adaboost* mempunyai akurasi sebesar 54,28%. dalam hal ini terjadi peningkatan nilai akurasi sebesar 11,43%. pada metode *gradient boosting* terlihat sebelum *oversampling* mempunyai nilai akurasi sebesar 67,34%, setelah *oversampling* terjadi peningkatan sebesar 20,27% yaitu menjadi 87,61%. pada metode *light GBM* sebelum *oversampling* mempunyai nilai akurasi sebesar 65,30%. setelah dilakukannya *oversampling* nilai akurasinya menjadi 88,57%, nilai akurasi tersebut meningkat sebesar 23,27%. kemudian metode *catboost* nilai akurasi yang dihasilkan sebelum *oversampling* adalah 61,22%, namun setelah *oversampling* terjadi peningkatan sebesar 30,2% yaitu menjadi 91,42%. Peningkatan ini disebabkan karena data yang tersedia tidak seimbang sebelum dilakukan *oversampling*, yang mana hasilnya hanya mencakup sebagian besar data. Dengan *oversampling* ini, data mayoritas dan minoritas menjadi seimbang, sehingga hasilnya konsisten dan meningkatkan keakuratan data yang diuji.

Hasil pengujian dari keempat algoritma boosting tersebut menandakan bahwa dengan melakukan *oversampling* pada dataset dapat meningkatkan nilai akurasi yang didapat. Hal ini konsisten dengan penelitian sebelumnya yang menggunakan metode *random oversampling*, seperti yang dijelaskan dalam penelitian Lindawati et al. (2023). Sebelum dilakukan *oversampling*, data menunjukkan terdapat 773 *record* yang tepat, sedangkan yang tidak tepat hanya 70. Namun, setelah terpilih kembali, jumlah entri tepat dan tidak tepat sama, masing-masing 773. Selain itu, algoritma boosting khususnya algoritma *random forest* memberikan hasil yang baik dengan presisi 90,04%, *recall* 90,04%, dan presisi 89,05%. Penelitian ini juga merupakan model pengembangan dari penelitian sebelumnya oleh Bahari & Latifa (2023) yang menunjukan hasil temuan yaitu tentang kalsifikasi buah untuk mendeteksi kualitas dan kesegaran buah menggunakan teknik *computer vision*. Berdasarkan temuan mereka masih terdapat kekurangan dikarenakan deteksi tersebut masih general tentang banyak buah. Namun hasil prediksi yang kami dilakukan mendapatkan nilai yang baik sebesar 98,17% untuk klasifikasi kualitas buah - buahan.

SIMPULAN

Hasil temuan kami yang memprediksi kualitas buah jeruk dengan teknik *random oversampling* dan 4 algoritma *boosting* didapatkan hasil terbaik dalam melakukan prediksi yaitu algoritma *boosting catboost* yang mendapatkan nilai akurasi tertinggi yaitu sebesar 91,42%. Dengan adanya sistem prediksi kualitas buah jeruk dengan menggunakan *random oversampling* dan algoritma *boosting* ini diharapkan produsen buah jeruk dapat menghasilkan produk dengan kualitas yang tinggi dan meminimalisir adanya buah jeruk yang kualitasnya buruk, sehingga konsumen dapat mendapatkan buah jeruk dengan kualitas yang baik. Dan juga dapat membantu produsen untuk memasarkan buah jeruk tidak hanya di dalam negeri saja namun dapat dipasarkan ke luar negeri juga.

REFERENSI

- Abdullah, A., & Sandi, K. (2021). Sistem Prediksi Rasa Buah Jeruk Menggunakan Metode K-Nearest Neighbor. *METHODIKA: Jurnal Teknik Informatika dan Sistem Informasi*, 7(2), 7-13. <https://doi.org/10.46880/mtk.v7i2.457>
- Alif, M. D. N., & Fahrudin, N. F. (2024). Performance Analysis of Oversampling and Undersampling on Telco Churn Data Using Naive Bayes, SVM And Random Forest Methods. *E3S Web of Conferences*, 484, 1-13. EDP Sciences. <https://doi.org/10.1051/e3sconf/202448402004>

- Aryanti, R., Misriati, T., & Hidayat, R. (2023). KLIK: Kajian Ilmiah Informatika dan Komputer Klasifikasi Risiko Kesehatan Ibu Hamil Menggunakan Random Oversampling Untuk Mengatasi Ketidakseimbangan Data. *Media Online*, 3(5), 409–416.
- Bahari, S. D. P., & Latifa, U. (2023). Klasifikasi Buah Segar Menggunakan Teknik Computer Vision Untuk Pendeteksian Kualitas Dan Kesegaran Buah. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(3), 1567-1573. <https://doi.org/10.36040/jati.v7i3.6871>
- Bangun, P., & Sihombing, M. (2021). Pengolahan Citra Untuk Identifikasi Kematangan Buah Jeruk Dengan Menggunakan Metode Backpropagation Berdasarkan Nilai HSV. *Jurnal Teknik Informatika Kaputama (JTik)*, 5(1), 85-91.
- Barkah, M. F. (2020). Klasifikasi Rasa Buah Jeruk Pontianak Berdasarkan Warna Kulit Buah Jeruk Menggunakan Metode K-Nearest Neighbor. *Coding Jurnal Komputer dan Aplikasi*, 8(1), 55-66
- Diantika, S. (2023). Penerapan Teknik Random Oversampling Untuk Mengatasi Imbalance Class Dalam Klasifikasi Website Phishing Menggunakan Algoritma Lightgbm. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(1), 19-25. <https://doi.org/10.36040/jati.v7i1.6006>
- Ganie, S. M., Pramanik, P. K. D., Mallik, S., & Zhao, Z. (2023). Chronic kidney disease prediction using boosting techniques based on clinical parameters. *PLoS ONE*, 18(12 December). <https://doi.org/10.1371/journal.pone.0295234>
- Ginting, D., & Ginting, D. G. (2021). Perananan Keagenan Kapal Dalam Melayani Pengisian Air Bersih Untuk Kebutuhan Km. Amrta Vii Pada Pt. Gesuri Lioyd Cabang Kuala Tanjung. *Journal of Maritime and Education (JME)*, 3(2), 245-249. <https://doi.org/10.54196/jme.v3i2.47>
- Jiang, Z., Pan, T., Zhang, C., & Yang, J. (2021). A new oversampling method based on the classification contribution degree. *Symmetry*, 13(2), 1–13. <https://doi.org/10.3390/sym13020194>
- Kim, Y., Lee, T., Hyun, Y., Coatanea, E., Mika, S., Mo, J., & Yoo, Y. (2023). Self-supervised representation learning anomaly detection methodology based on boosting algorithms enhanced by data augmentation using StyleGAN for manufacturing imbalanced data. *Computers in Industry*, 153, 104024. <https://doi.org/10.1016/j.compind.2023.104024>
- Kristiawan, K., & Widjaja, A. (2021). Perbandingan Algoritma Machine Learning dalam Menilai Sebuah Lokasi Toko Ritel. *Jurnal Teknik Informatika Dan Sistem Informasi*, 7(1), 35-46. <https://doi.org/10.28932/jutisi.v7i1.3182>
- Lindawati, L., Fadhli, M., & Wardana, A. S. (2023). Optimasi Gaussian Naïve Bayes dengan Hyperparameter Tuning dan Univariate Feature Selection dalam Prediksi Cuaca. *Edumatic: Jurnal Pendidikan Informatika*, 7(2), 237–246. <https://doi.org/10.29408/edumatic.v7i2.21179>
- Mane, V., Belgaonkar, P., Dabhade, H., & Gandhe, A. (2023). Classification Comparison of Different Boosting Algorithms to Predict and Classify Conditions of Heart Disease. In *Data Science and Intelligent Computing Techniques*, 869–876. Soft Computing Research Society. <https://doi.org/10.56155/978-81-955020-2-8-74>
- Manihuruk, D. S. B. H., Darwin, D., & Munawar, A. A. (2019). Penentuan Kualitas Buah Jeruk (Citrus Sinensis L) Menggunakan Teknologi Laser Photo-Acoustics (LPAS) Dengan Metode Support Vector Machine (SVM). *Jurnal Ilmiah Mahasiswa Pertanian*, 4(2), 377-386. <https://doi.org/10.17969/jimfp.v4i2.10945>
- Maulana, M. D., Hadiana, A. I., & Umbara, F. R. (2023). Algoritma Xgboost Untuk Klasifikasi Kualitas Air Minum. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(5), 3251-3256. <https://doi.org/10.36040/jati.v7i5.7308>

- Rahim, A. M. A., Pratiwi, I. Y. R., & Fikri, M. A. (2023). Klasifikasi Penyakit Jantung Menggunakan Metode Synthetic Minority Over-Sampling Technique Dan Random Forest Clasifier. *Indonesian Journal of Computer Science*, 12(5), 2995-3011. <https://doi.org/10.33022/ijcs.v12i5.3413>
- Setiawan, D., Nugraha, A., & Luthfiarta, A. (2024). Komparasi Teknik Feature Selection Dalam Klasifikasi Serangan IoT Menggunakan Algoritma Decision Tree. *Jurnal Media Informatika Budidarma*, 8(1), 83-93.
- Sidik, A. P., Amin, M., & Wilana, A. (2023). Implementasi Perancangan Klasifikasi Kualitas Buah Jeruk Berdasarkan Warna. *JOURNAL ZETROEM*, 5(1), 72-76.
- Wicaksono, D. F., Basuki, R. S., & Setiawan, D. (2024). Peningkatan Performa Model Machine Learning XGBoost Classifier melalui Teknik Oversampling dalam Prediksi Penyakit AIDS. *Jurnal Media Informatika Budidarma*, 8(2), 736-747.
- Zhu, L., Zhou, X., & Zhang, C. (2021). Rapid identification of high-quality marine shale gas reservoirs based on the oversampling method and random forest algorithm. *Artificial Intelligence in Geosciences*, 2, 76–81. <https://doi.org/10.1016/j.aiig.2021.12.001>