Implementasi K-Fold Dalam Prediksi Hasil Produksi Agrikultur Pada Algoritma K-Nearest Neighbor (KNN)

Victor Immanuel Sunarko[1], Denis Lizard Sambawo Dimara[2], Pangestu Sandya Etniko Siagian[3] Daniel Manalu[4] , Fetty Tri Anggraeny [5]
[1, 2, 3, 4, 5] Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional "Veteran" Jawa Timur
Email: [1]21081010145@student.upnjatim.ac.id, [2]21081010159@student.upnjatim.ac.id, [3]21081010180@student.upnjatim.ac.id, [4]21081010189@student.upnjatim.ac.id, [5]fettyanggraeny.if@upnjatim.ac.id

Abstract. *The agricultural sector, especially agriculture in Indonesia, is the backbone of the economy, with the agricultural workforce reaching 38.14 million people in February 2023, or 27.52% of the total national workforce. Despite its great potential, the sector faces significant challenges, including limited land, climate change, and air scarcity, which necessitate the adoption of sustainable agriculture. This research aims to increase the efficiency of agricultural production through the application of artificial intelligence (AI) and data analysis. The methodology used includes data sharing to predict agricultural production results with the k-nearest neighbor (KNN) algorithm. The scenario test was carried out using a k-fold cross-validation and hold-out data-sharing approach. The research results show the highest accuracy of 98.36% using k-fold cross-validation and 97.42% using the hold-out method.*
Keywords: *KNN, K-Fold, Hold-Out, Prediction, Agriculture*

Abstrak. *Sektor agrikultur khususnya pertanian di Indonesia merupakan tulang punggung perekonomian, dengan tenaga kerja pertanian mencapai 38,14 juta orang pada Februari 2023, atau 27,52% dari total tenaga kerja nasional. Meskipun memiliki potensi besar, sektor ini menghadapi tantangan signifikan, termasuk lahan terbatas, perubahan iklim, dan kelangkaan air, yang mengharuskan penerapan pertanian berkelanjutan. Penelitian ini bertujuan untuk meningkatkan efisiensi produksi pertanian melalui penerapan kecerdasan buatan (AI) dan analisis data. Metodologi yang digunakan meliputi pembagian data untuk memprediksi hasil produksi pertanian dengan algoritma k-nearest neighbour (KNN). Uji skenario dilakukan dengan pendekatan k-fold cross-validation dan hold-out data sharing. Hasil penelitian menunjukkan akurasi tertinggi sebesar 98,36% menggunakan k-fold cross-validation dan 97,42% dengan metode hold-out.*
Kata Kunci: *KNN, K-Fold, Hold-Out, Prediksi, Agrikultur*

1. Pendahuluan

Berdasarkan Pusat Data dan Sitem Informasi Pertanian tercatat tenaga kerja pertanian (dalam arti sempit) merupakan tenaga kerja terbesar dengan jumlahnya mencapai 38,14 juta orang pada Februari tahun 2023. Jumlah ini merupakan 27,52% dari jumlah tenaga kerja Indonesia seluruhnya yang berjumlah 138.63 juta orang (*https://satudata.pertanian.go.id/*). Bisnis pertanian merupakan kegiatan pertanian yang bertujuan menghasilkan uang dengan memproduksi produk pertanian. Agribisnis sangat penting bagi perekonomian Indonesia. Ini bertindak seperti tulang punggung, membantu menjaga ekonomi tetap kuat dan bergerak. Sektor ini mencakup berbagai bidang seperti menanam tanaman pangan, perkebunan, kehutanan, peternakan, dan perikanan. Masing-masing bidang ini berkontribusi pada perekonomian dengan menyediakan pekerjaan dan produk. (A. Ravi, dkk 2020)
Komoditas agrikultur, mempunyai potensi yang cukup besar di Indonesia. Komoditas dapat secara signifikan meningkatkan pembangunan ekonomi di berbagai sektor, khususnya sektor industri, dengan meningkatkan nilai komoditas pertanian menjadi produk yang layak secara ekonomi. Permasalahan yang sering dihadapi adalah variabilitas produksi komoditas pertanian yang dipengaruhi oleh berbagai faktor penghambat, baik unsur alam maupun ketersediaan sumber daya (P. Erwin, dkk 2023).

Indonesia menghadapi tantangan yang signifikan di bidang pertanian, termasuk lahan pertanian yang terbatas karena pertumbuhan penduduk, perubahan iklim, dan kelangkaan air. Isu-isu ini menyoroti pentingnya pertanian berkelanjutan, yang dapat mengatasi tantangan ini dengan memastikan produksi pertanian yang stabil, mencegah kerusakan lingkungan, mengurangi biaya, menghasilkan produk pertanian yang lebih baik. (V. Ratri, dkk 2021). Sektor pertanian menghadapi tantangan baik dari faktor lingkungan maupun ketersediaan dan antusiasme sumber daya manusia. Tantangan-tantangan ini dapat menghambat produksi dan pertumbuhan yang konsisten, mempengaruhi keseluruhan output komoditas pertanian. Produksi yang tidak pasti tidak hanya mempengaruhi sektor pertanian tetapi juga memengaruhi ekonomi yang lebih luas. Karena komoditas pertanian strategis untuk pertumbuhan ekonomi, fluktuasi produksinya dapat berdampak pada sektor-sektor terkait, seperti industri dan perdagangan (P. Erwin, dkk 2023). Perencanaan pembangunan di sektor pertanian, pengembangan mesin untuk memprediksi produksi hasil pertanian menjadi semakin diperlukan. (P. Hasdi, dkk 2020)

Pertanian modern diperlukan untuk meningkatkan efisiensi produksi yang tinggi yang dipadukan dengan kualitas produk yang tinggi. Kecerdasan buatan adalah jenis teknologi yang membantu mesin berpikir dan belajar seperti manusia. Untuk meningkatkan efesiensi dan kulitas, pertanian modern menggunakan metode analisis data dengan menggunakan kecerdasan buatan. Di bidang pertanian, kecerdasan buatan dapat membantu menganalisis data secara lebih efektif. Kecerdasan buatan dapat membantu petani membuat keputusan yang lebih baik dan meningkatkan produksi panen (K Sebastian, dkk 2021).

Beberapa model *machine learning* pernah diterapkan untuk memprediksi berhasil meningkatkan performansi model yang dihasilkan dalam berbagai bidang khususnya dibidang pertanian. Penelitian Adin pada tahun 2023 menerapkan algoritma Regresi Linear untuk memprediksi produksi beras di Kabupaten Grobogan. Kinerja model Regresi Linear dievaluasi menggunakan beberapa metrik, seperti *Mean Squared Error (MSE)* dengan hasil 6550,241810 *dan Mean Absolute Error (MAE)* dengan hasil 121247657,9756. Selanjutnya penelitian yang dilakukan oleh P. Fitri, dan kawan-kawan pada tahun 2023 menerapkan algoritma C4.5 untuk memprediksi hasil panen padi di Kabupaten Simalungun. Kinerja model C4.5 menghasilkan tingkat akurasi prediksi sebesar 85%.

Sebagaimana hal tersebut, algoritma *k-Nearest Neighbour (k-NN)* juga menjadi salah satu teknik yang paling sederhana dalam *machine learning*. *K-NN* bekerja dengan mengklasifikasikan data berdasarkan kelas dari tetangga terdekat data tersebut. Algoritma ini sering disebut sebagai teknik *Lazy Learning*, karena proses pembelajarannya dilakukan pada saat runtime, dan juga dikenal sebagai *Memory-based Classification* atau *Case-based Classification* (Cunningham & Delany, 2021).

Dengan demikian, berdasarkan penelitian sebelumnya, metodologi pembagian data digunakan untuk memprediksi hasil produksi pertanian dengan menggunakan algoritma *k-nearest neighbour*. Penelitian ini akan melakukan uji skenario dengan menggunakan pendekatan *k-fold cross-validation* dan *hold-out* data sharing pada algoritma *k-nearest neighbour*, yang bertujuan untuk menghasilkan kinerja kedua metode pembagian data tersebut.

2. Tinjauan Pustaka

2.1. *Machine Learning*

Machine Learning / pembelajaran mesin adalah cabang kecerdasan buatan yang berfokus pada membangun sistem yang dapat belajar dari data dan meningkatkan kinerjanya dari waktu ke waktu tanpa diprogram secara eksplisit. Saat model machine learning memproses lebih banyak data, maka model akan menjadi lebih baik dalam menjalankan tugasnya. Peningkatan ini adalah bagian penting dari *Machine Learning*. Tujuannya adalah agar model membuat prediksi atau keputusan yang lebih akurat karena memperoleh lebih banyak pengalaman (R Susmita 2019).

2.2. *K-Nearest Neighbor* (KNN)

Algoritma KNN bersifat supervised dimana hasil dari *query instance* yang baru diklasifikasikan berdasarkan mayoritas pada kategori. *K-Nearest Neighbor* digunakan pada tahap klasifikasi dimana fungsi telah melewati tahap sebelumnya. Kemudian karakteristik yang sama dihitung sebagai data uji (testing). Dari hasil perhitungan diperoleh jarak dan jarak dari vektor baru ini

dihitung relatif terhadap semua vektor sampel yang ada dan diambil K terdekat (Kurniadi dkk., 2021). Tujuan dari algoritma ini yaitu untuk mengklasifikasikan data baru berdasarkan atribut dan sampel training. Algoritma KNN menggunakan klasifikasi fitur yang berdekatan sebagai nilai prediktif untuk sampel uji baru. (Yahya & Hidayanti, 2020).

Sebagai metode yang sederhana namun efektif, K-Nearest Neighbor sering digunakan pada berbagai permasalahan klasifikasi. Algoritma ini bekerja dengan menemukan pola baru pada data dengan cara menghubungkan pola data yang telah ada. K-Nearest Neighbor dapat diterapkan dengan baik pada dataset berdimensi tinggi, menjadikannya salah satu algoritma yang fleksibel dan sering digunakan di bidang analisis data. (Ramadhan M. A. dkk, 2024)

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^{n}(a_r(x_i) - a_r(x_j))^2} \quad (1)$$

Dimana hasil dari $d(x_i, x_j)$ yaitu pengurangan pada setiap atribut yang dikuadratkan dan dijumlahkan pada nilai terkecil dengan data uji

2.4. *K-Fold Cross-Validation*

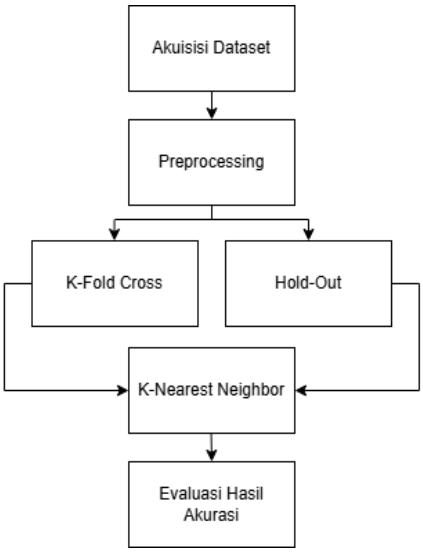
Cross validation adalah sebuah teknik yang membagi kumpulan data menjadi K himpunan bagian atau lipatan. Model ini dilatih pada lipatan K-1 dan diuji pada lipatan yang tersisa. Proses ini diulang sebanyak K kali, dengan setiap lipatan berfungsi sebagai set tes sekali. Tujuan utamanya adalah untuk menilai bagaimana hasil analisis statistik akan digeneralisasi ke kumpulan data independen (Isaac Kofi Nti dkk, 2021).

2.5. *Hold-Out*

Dalam penelitian yang terbaru saat ini, metode *hold-out validation* masih menjadi salah satu teknik utama yang digunakan untuk membagi dataset menjadi dua bagian: *data training* dan *data testing*. Menurut penelitian terkini, metode ini sering digunakan dalam pembelajaran mesin untuk pemilihan model dan *tuning hyperparameter*, terutama pada dataset yang besar dan kompleks (Rizaldi S & Mustakim, 2020).

3. Metode Penelitian

Diperlukan dasar yang mengatur jalannya penelitian ini agar penelitian dapat berjalan dengan baik dan fokus terhadap penyelesaian masalah yang ada. Adapun blok diagram untuk merepresentasikan alur dari penelitian yang akan dijalankan nantinya.



Gambar 1. Metodologi Penelitian

 ISSN: 2579-566X (Online)
ISSN: 2477-5274 (Print)

3.1 Akuisisi Data

Data yang digunakan pada penelitian ini yaitu menggunakan Dataset Hasil Produksi Agrikultur dimana tahap pengambilan data melalui platform pada situs web Kaggle *https://www.kaggle.com/datasets/chitrakumari25/smart-agricultural-production-optimizing-engine* dengan jumlah 2200 record data. Cuplikan data yang telah didapatkan dapat dilihat pada Tabel 1 di bawah ini.

Tabel 1. Akuisisi Data

N	*P*	*K*	*temperature*	*humidity*	*ph*	*rainfall*	*label*
90	42	43	20,87	82	6,5	202	rice
85	58	41	21,77	80	7,0	226	rice
60	60	44	23,00	82	7,8	263	rice
74	74	40	26,49	80	6,9	242	rice
78	78	42	20,12	81	7,62	262	rice
…	…	…	…	…	…	…	…

3.2 *Pre-processing Data*

Sebelum memulai analisis *machine learning*, menjadi langkah wajib untuk memeriksa setiap kolom guna mengidentifikasi variabel apa pun yang menunjukkan sejumlah besar data yang hilang atau error. *Pre-processing Data* adalah langkah pertama dalam *machine learning*, data diubah/diolah sehingga mesin dapat dengan cepat menelusuri atau mengurai data tersebut. Dengan kata lain dapat juga diartikan sebagai model algoritma dapat segera menganalisis fitur data (M. Kiran dkk, 2022). Langkah pertama dalam data *pre-processing* dalam penelitian ini dengan melakukan mengubah nilai data agar berada dalam rentang 0 hingga 1. Lalu mengambil nilai unik dari setiap kolom sehingga tidak ada data yang kembar.

3.3 *Splitting Data*

Selanjutnya dilakukannya proses pra-pengolahan data yaitu *splitting data* dan *data formatting* sehingga data hasil *splitting* tiap skenarionya dapat digunakan untuk proses pelatihan. Adapun juga skenario pelatihan yang dirancangkan untuk mengevaluasi implementasi *K-Fold* dalam algoritma *K-Nearest Neighbor*.

Tabel 2. Skenario Pelatihan

Skenario	*Split*
Hold-out	70:30
	80:20
	90:10
K-Fold	2
	…
	15

Pada proses klasifikasi penelitian ini menggunakan *K-Nearest Neighbor*. Proses analisis dan hasil yang dilakukan menggunakan perhitungan untuk perbandingan performa algoritma *K-Nearest Neighbor* dengan hasil akurasi pada klasifikasi analisis akhir.

3.4 Pembuatan *K-Nearest Neighbor*

Algoritma *machine learning K-Nearest Neighbor* memiliki algoritma dasar berupa *euclidean distance. Euclidean distance* bertanggung jawab untuk menghitung jarak *Euclidean* antara dua titik data untuk menentukan kedekatan antara data. Beranjak ke *K-Nearest Neighbor*, secara garis besar nantinya akan menerima data latih (train), data validasi tunggal (val), dan nilai k sebagai parameter. Untuk setiap titik dalam data latih, fungsi ini menghitung jaraknya ke data validasi menggunakan algoritma *euclidean distance* dan menyimpannya bersama label target. Kemudian, jarak-jarak tersebut

diurutkan secara ascending, dan label dari k tetangga terdekat diambil. Prediksi akhir ditentukan berdasarkan label yang paling sering muncul di antara tetangga-tetangga tersebut, menggunakan pendekatan mode. Kode ini mendemonstrasikan proses prediksi sederhana dalam *K-Nearest Neighbor* dengan fokus pada perhitungan jarak dan klasifikasi berdasarkan tetangga terdekat.

3.5 Evaluasi *K-Nearest Neighbor*

Performa algoritma *K-Nearest Neighbor* nantinya akan dievaluasi dengan membandingkan prediksi model terhadap label sebenarnya dari data validasi. Jika prediksi sesuai dengan label asli, maka penghitung *true answer* akan bertambah satu. Proses evaluasi akan berjalan secara berulang dengan skenario pembagian data, seperti *k-fold cross-validation* atau pembagian train-test split dengan *holdout*, untuk memastikan hasil evaluasi mencerminkan kinerja model secara menyeluruh. Pada akhirnya, akan dilakukan perhitungan total prediksi yang benar dari keseluruhan sampel validasi. Akurasi dihitung sebagai rasio antara jumlah prediksi benar (*true answer*) dengan jumlah total sampel validasi.

4. Hasil dan Pembahasan

Hasil dari penelitian didapatan dari pelatihan berdasarkan skenario uji coba yang telah direncanakan. Pelatihan melibatkan data sekunder yang diambil melalui website Kaggle. Melalui serangkaian uji coba, didapatkan hasil sebagai berikut

Tabel 3. Akurasi K-Fold

K-Fold	*Total Akurasi*
2	97,72%
3	97,86%
4	97,95%
5	98,13%
6	98,22%
7	98,04%
8	98,04%
9	97,99%
10	98,18%
11	98,04%
12	98,04%
13	98,18%
14	98,27%
15	98,36%

Berdasarkan analisis pada Tabel 3, dapat diamati bahwa peningkatan jumlah lipatan (K) dalam teknik validasi *K-Fold* memiliki pengaruh signifikan terhadap kinerja algoritma *K-Nearest Neighbor* (KNN) dalam mengklasifikasikan data. Semakin tinggi nilai K pada pembagian dataset, semakin akurat pula algoritma KNN dalam mengklasifikasikan data. Fenomena ini sejalan dengan prinsip dasar *machine learning*, di mana jumlah data latih yang lebih banyak memberikan representasi yang lebih baik terhadap pola dalam data, sehingga meningkatkan akurasi klasifikasi.

Namun, hasil penelitian ini juga menunjukkan adanya titik jenuh pada nilai K ke-7. Pada titik ini, algoritma KNN mulai mengalami penurunan akurasi, dan peningkatan nilai K setelahnya tidak menghasilkan kenaikan akurasi yang signifikan. Fenomena ini dapat dikaitkan dengan masalah overfitting, di mana model cenderung "menghafal" data latih daripada melakukan generalisasi terhadap pola yang lebih luas. Dalam konteks KNN, hal ini dapat terjadi ketika proses pembentukan neighbor terlalu berdekatan, sehingga data dengan noise atau anomali memiliki kemungkinan lebih besar untuk salah diklasifikasikan.

Selain itu, kelebihan data latih juga dapat menjadi penyebab utama munculnya permasalahan ini. *Neighbor* yang terlalu padat berpotensi membuat algoritma KNN kehilangan kemampuan untuk mengenali data dengan karakteristik unik, khususnya yang memiliki sedikit noise. Oleh karena itu,

meskipun penambahan jumlah lipatan *K-Fold* umumnya meningkatkan akurasi, perlu dilakukan evaluasi terhadap batas optimal untuk menghindari overfitting. Langkah mitigasi seperti pemilihan parameter K yang lebih bijak, penggunaan metode pembobotan neighbor, atau pengurangan noise melalui preprosesing data dapat menjadi solusi potensial untuk memperbaiki kinerja model KNN pada skenario serupa.

Tabel 4. Akurasi Hold-Out

Hold-Out	*Total Akurasi*
70:30	97,42%
80:20	96,81%
90:10	97,27%

Berdasarkan Tabel 4, penggunaan metode pembagian data sebelumnya, yaitu *hold-out*, menunjukkan bahwa algoritma KNN telah mampu mencapai performa yang cukup baik. Namun, metode *hold-out* memiliki keterbatasan, di mana data hanya dipisahkan berdasarkan persentase tertentu dan pembagian dataset dilakukan secara tunggal. Hal ini mengakibatkan akurasi yang diperoleh cenderung bergantung pada distribusi data latih dan data uji pada pembagian tunggal tersebut, sehingga kurang mencerminkan kinerja model secara keseluruhan.

Sebaliknya, metode *K-Fold* menawarkan pendekatan yang lebih komprehensif, dengan membagi dataset menjadi beberapa lipatan (*folds*). Dalam hal ini, *K-Fold* dapat dianalogikan sebagai bentuk *multi-hold-out*, di mana setiap lipatan berfungsi sebagai data uji secara bergantian, sementara lipatan lainnya digunakan sebagai data latih. Akurasi yang diperoleh dari metode ini merupakan rata-rata dari hasil akurasi di setiap lipatan, sehingga memberikan estimasi performa yang lebih general dan *robust* terhadap distribusi data yang beragam.

Perbandingan hasil menunjukkan bahwa metode *K-Fold* menghasilkan akurasi maksimum yang sedikit lebih tinggi dibandingkan metode *hold-out*, dengan selisih sekitar 1%. Perbedaan ini, meskipun kecil, mencerminkan keunggulan *K-Fold* dalam memanfaatkan seluruh dataset secara lebih optimal untuk pelatihan dan pengujian, sekaligus mengurangi kemungkinan bias akibat pembagian dataset tunggal. Dengan demikian, metode *K-Fold* dapat dianggap lebih unggul dalam memberikan gambaran kinerja algoritma KNN yang lebih generalisasi terhadap data baru.

5. Kesimpulan

Penelitian ini mengeksplorasi penggunaan algoritma *K-Nearest Neighbor (KNN)* dalam memprediksi hasil produksi agrikultur, dengan fokus pada perbandingan dua metode pembagian data: *K-Fold Cross-Validation* dan *Hold-Out*. Melalui analisis komprehensif terhadap dataset pertanian dari Kaggle, penelitian menemukan bahwa *K-Fold Cross-Validation* secara konsisten menghasilkan akurasi yang sangat tinggi, mencapai puncak 98,36% pada pembagian 15 lipatan, yang menunjukkan potensi kuat metode ini dalam meningkatkan prediksi di sektor pertanian.

Menariknya, melalui penelitian ini mengungkap dinamika kompleks dalam proses *machine learning*, seperti fenomena titik jenuh akurasi dan risiko *overfitting* yang terjadi setelah tujuh lipatan. Dibandingkan dengan metode *Hold-Out*, *K-Fold* terbukti memberikan estimasi kinerja model yang lebih *robust* dan komprehensif, dengan keunggulan akurasi sekitar 1%. Temuan ini tidak hanya menegaskan efektivitas algoritma KNN dalam memprediksi produksi pertanian, tetapi juga memberikan wawasan penting tentang pentingnya pemilihan metode validasi yang tepat dalam mengembangkan model prediksi yang handal di era pertanian modern.

Referensi

A. Ravi Choirul, & F. Amrie (2020). Implementasi Akuntansi Agrikultur Pada Perusahaan Sektor Pertanian Di Indonesia. Jurnal Ilmiah Akuntansi Universitas Pamulang, 8(2), 85. https://doi.org/10.32493/jiaup.v8i2.4676

C. Pádraig, D. Sarah Jane (2021). K-Nearest Neighbour Classifiers-A Tutorial. ACM Computing Surveys, 54(6). https://doi.org/10.1145/3459665

K. Dedy, S. Andre, W. Linggar Alfithna (2021). Pattern Recognition of Human Face With Photos Using KNN Algorithm. Jurnal Transformatika, 19(1), 17. https://doi.org/10.26623/transformatika.v19i1.3581

K. Sebastian, N. Gniewko (2021). Kujawa, S., and Niedbała, G. Artificial Neural Networks in Agriculture. Agriculture 2021, 11, 497.pdf. Agriculture, 1–6.

M. Adin M. (2024). Implementasi Algoritma Linear Regression untuk Prediksi Produksi Tanaman Padi di Kabupaten Grobogan. Data Sciences Indonesia (DSI), 3(2), 68–78. https://doi.org/10.47709/dsi.v3i2.3118

N. Isaac Kofi, A. Justice, B. Owusu (2021). Performance of Machine Learning Algorithms with Different K Values in K-fold CrossValidation. International Journal of Information Technology and Computer Science, 13(6), 61–71. https://doi.org/10.5815/ijitcs.2021.06.05

P. Erwin, I. Imron Rosyadi NR, R. Sholeh. (2023). Penerapan K-Means Algorithm Untuk Mengidentifikasi Supplier Bahan Baku Pada Pamekasan Application of the K-Means Algorithm To Identify Suppliers of Raw Materials for Agricultural Commodities in Pamekasan. Journal Simantec, 11(2), 147–156.

P. Fitri Annisa, and I. Ali (2024). Sistem Informasi Geografis Menggunakan Algoritma C4.5 untuk Prediksi Hasil Panen Padi di Kabupaten Simalungun. Indonesian Journal of Applied Informatics, 9(1), 155-167.

P. Hasdi, W. Nabilah Ulfa (2020). Jurnal Nasional Teknologi dan Sistem Informasi Penerapan Prediksi Produksi Padi Menggunakan Artificial Neural Network Algoritma Backpropagation. Nasional Teknologi Dan Sistem Informasi, 02, 100–107.

P. Tamilarasi , R. Rani Uma. (2020). Diagnosis of Crime Rate against Women using k-fold Cross Validation through Machine Learning. Proceedings of the 4th International Conference on Computing Methodologies and Communication, ICCMC 2020, Iccmc, 1034–1038. https://doi.org/10.1109/ICCMC48092.2020.ICCMC-000193

R. Said Thaufik, Mustakim (2020). Perbandingan Teknik Pembagian Data untuk Klasifikasi Sarana Akses Air pada Algoritma K- Nearest Neighbor dan Naïve Bayes Classifier. Seminar Nasional Teknologi Informasi, Komunikasi Dan Industri (SNTIKI) 12, 130–137.

R. Susmita (2019). Introduction to Machine Learning and Different types of Machine Learning Algorithms. Proceedings of the International Conference on Machine Learning, Big Data, Cloud and Parallel Computing: Trends, Prespectives and Prospects, COMITCon 2019, 35–39.

R. Muhammad Alviriza, A. Fetty Tri, P. Chrystia Aji (2024). Klasifikasi Curah Hujan Harian Menggunakan Metode K-Nearest Neighbor. JATI (Jurnal Mahasiswa Teknik Informatika), 8(3), 3863–3869. https://doi.org/10.36040/jati.v8i3.9817

V. Ratri, S. Tatie, A. Siti, & F. Anna (2019). Farmers' Perception to Government Support in Implementing Sustainable Agriculture System. Jurnal Ilmu Pertanian Indonesia, 24(2), 168–177. https://doi.org/10.18343/jipi.24.2.168

Yahya, & H. Winda Puspita (2020). Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Efektivitas Penjualan Vape (Rokok Elektrik) pada " Lombok Vape On ". Jurnal Informatika Dan Teknologi, 3(2), 104–114. https://e-journal.hamzanwadi.ac.id/index.php/infotek/article/view/2279/pdf_23

M. Kiran, M. Surajit, & N. Bhushankumar (2022). A review: Data pre-processing and data augmentation techniques. *Global Transitions Proceedings*, *3*(1), 91–99. https://doi.org/10.1016/j.gltp.2022.04.020

ACM Computing Surveys, 54(6). https://doi.org/10.1145/3459665earest Neighbour Classifiers https://doi.org/10.1145/3459665