Jurnal Sistem Komputer dan Informatika (JSON)
Volume 4, Nomor 1, September 2022
e-ISSN 2685-998X
DOI 10.30865/json.v4i1.4813

Seleksi Fitur untuk Prediksi Hasil Produksi Agrikultur pada Algoritma K-Nearest Neighbor (KNN)

Delvi Nur Aini*, Bella Oktavianti, Muhammad Jalal Husain, Dian Ayu Sabillah, Said Thaufik Rizaldi, Mustakim

Program Studi Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia
Email: [1*]12050320493@uin-suska.ac.id, [2]12050326633@uin-suska.ac.id, [3]12050312614@uin-suska.ac.id, [4]12050320317@uin-suska.ac.id, [5]11753101376@uin-suska.ac.id, [6]mustakim@uin-suska.ac.id
Email Penulis Korespondensi: 12050320493@uin-suska.ac.id
Submitted: **08/09/2022**; Accepted: **30/09/2022**; Published: **30/09/2022**

Abstrak− Pertanian merupakan salah satu sektor penggerak ekonomi terbesar di Indonesia. Badan Pusat Statistik (BPS) tahun 2021 mencatat sebesar 37,02% penduduk Indonesia bekerja pada bidang pertanian. Permasalahan yang dihadapi oleh petani saat ini yaitu penurunannya hasil panen baik kuantitas maupun kualitas akibat dari cuaca yang sulit diprediksi membuat petani kesulitan memilih jenis tanaman yang cocok untuk ditanam. Penerapan teknik data mining terdapat permasalahan yang terkait dengan kompleksitas parameter cuaca dan keadaan alam yang menunjang produksi pertanian sehingga sangat penting untuk dilakukan seleksi fitur yakni membentuk fitur yang paling relevan. Penelitian ini melakukan percobaan untuk mengetahui pengaruh penerapan fitur seleksi Principal Component Analysis (PCA) terhadap performa algoritma K-Nearest Neighbor (KNN) yang menghasilkan akurasi paling tinggi sebesar 99,64% pada penelitian ini

Kata Kunci: KNN; Seleksi Fitur; Prediksi; Agrikultur; PCA

Abstract−Agriculture is one of the largest economic driving sectors in Indonesia. The Central Statistics Agency (BPS) in 2021 recorded that 37.02% of Indonesia's population worked in the agricultural sector. The problem faced by farmers today is the decline in yields, both in quantity and quality due to unpredictable weather, making it difficult for farmers to choose the types of plants that are suitable for planting. The application of data mining techniques has problems related to the complexity of weather parameters and natural conditions that support agricultural production, so it is very important to do feature selection, namely to form the most relevant features. This study conducted an experiment to determine the effect of implementing the Principal Component Analysis (PCA) selection feature on the performance of the K-Nearest Neighbor (KNN) algorithm which produces the highest accuracy of 99.64% in this study.

Keywords: KNN; Feature Selection; Prediction; Agriculture; PCA

1. PENDAHULUAN

Badan Pusat Statistik (BPS) pada tahun 2021 mencatat sebesar 37,02% penduduk Indonesia bekerja pada bidang pertanian [1]. Hal ini menunjukkan bahwa pertanian merupakan salah satu sektor penggerak ekonomi terbesar di Indonesia. Badan Pusat Statistik (BPS) menyebutkan bahwa jumlah luas lahan untuk pertanian di Indonesia mencapai 5.2 ribu lebih hektar lahan pertanian sawah dan 9.9 ribu lebih hektar untuk lahan pertanian nonsawah di Indonesia [2]. Namun, Indonesia mengalami kekurangan Sumber Daya Manusia yang berkecimpung pada sektor pertanian. Sehingga, berdampak pada hasil produksi pertanian. Ironisnya, sektor pertanian dihadapkan pada kondisi yang cukup sulit dari berbagai macam permasalahan, khususnya pada konversi lahan, kompetisi dari pemanfaatan yang kurang optimal, degradasi sumber daya lahan serta jumlah tenaga kerja yang setiap tahunnya menurun [3].

Selain perkembangan di Indonesia, Bank Dunia juga menyebutkan terkait hal serupa dimana pada tahun 2017 terbukti hanya 31,5% atau 570 ribu kilometer persegi lahan di Indonesia yang digunakan untuk pertanian. Beberapa peneliti global telah meneliti kebutuhan pangan hingga mencapai 6.600 ton per hari, serta tidak kurang dari 250 juta orang kelaparan di dunia hidup di beberapa daerah. Diperkirakan pada tahun 2035, 65% penduduk akan menghuni perkotaan, khususnya di 16 kota besar di Indonesia akibat permasalahan yang terjadi [4].

Permasalahan para petani yang dihadapi saat ini adalah menurunnya hasil dari panen baik dari segi kuantitas maupun kualitasnya. Ada beberapa faktor yang memberi dampak pada masalah tersebut misalnya perubahan cuaca ekstrim yang sekarang sangat sulit untuk diprediksi, suhu, kelembaban serta pengurangan area lahan yang dialih fungsikan sebagai fasilitas umum. Akibatnya hal ini menyebabkan para petani mengalami kesulitan dalam memilih jenis tanaman yang sesuai untuk ditanam. Sementara itu, seleksi jenis tanaman yang tepat dapat membuat hasil panen meningkat [5]. Salah satu penerapan dibidang pertanian pada pengolahan dan prediksi *output* produksi adalah dengan menggunakan *Smart Agriculture Optimization* dan *automated machine* yang sangat memudahkan terkhusus bagi para petani dalam memprediksi hasil produksi tanaman mereka dan mengurangi resiko gagal panen dengan menggunakan Teknik data mining [6].

Teknik pada data mining merupakan proses dalam mengekstraksi informasi yang dapat diterapkan untuk memprediksi hasil produksi pertanian menggunakan teknik yang efisien dan bermanfaat [7][8]. Dalam penerapan teknik data mining yang efisien dan bermanfaat terdapat permasalahan yang dikaitkan dengan kompleksitas parameter sehingga sangat penting untuk dilakukan seleksi fitur yakni membentuk fitur yang paling relevan pada penelitian yang bertujuan untuk memastikan model pembelajaran mesin memiliki tingkat performansi yang tinggi

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 4, Nomor 1, September 2022
e-ISSN 2685-998X
DOI 10.30865/json.v4i1.4813

dan mengurangi redundasi pada model yang akan diterapkan [9][10]. Beberapa teknik seleksi fitur yang pernah diterapkan untuk memprediksi berhasil meningkatkan performansi model yang dihasilkan dalam berbagai bidang khususnya dibidang pertanian [11][12][13][14][15].

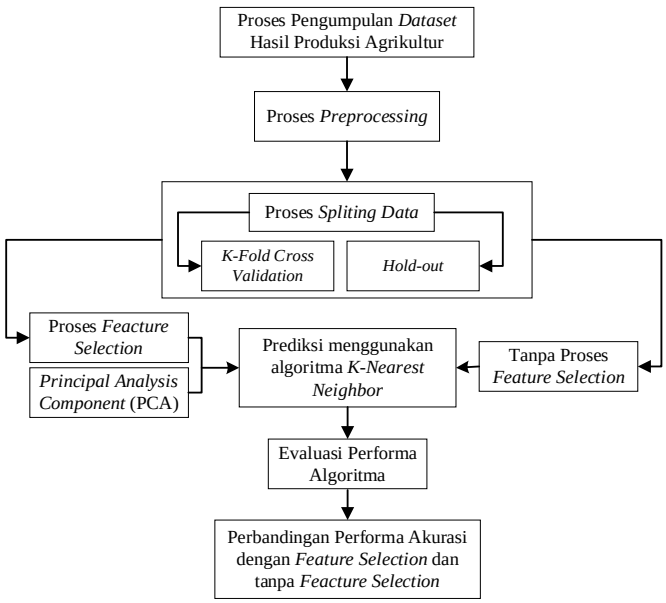
Penelitian Osama dan kawan – kawan pada tahun 2021 melakukan prediksi pada bidang pertanian salah satunya menggunakan algoritma *Random Forest* (RF) untuk mengetahui kadar air kanopi beras dengan menggunakan data hiperspektral. Penelitian ini melakukan dengan pendekatan *feature selection* untuk membentuk fitur yang efisien yang mengusulkan model prediksi yang efisien dan membandingkan tiga metode *feature selection* yang meliputi *Vegetation Indices* (VI), *Model-based Features* (MF), dan *Principal Component Analysis* (PCA). menghasilkan PCA sebagai teknik seleksi fitur yang optimal diterapkan pada permasalahan tersebut yang menghasilkan nilai evaluasi algoritma *Root Mean Square Error* (RMSE) sebesar 0,252 [12]. Penelitian ini tidak menjelaskan evaluasi akurasi yang dihasilkan dari performa algoritma.

Selanjutnya penelitian Bode pada tahun 2017 melakukan perkiraan harga komoditi kopi arabika dengan mengimplementasikan algoritma *K-Nearest Neighbor* (KNN). Penelitian ini melakukan pendekatan teknik *feature selection* yakni *Backward Elimination* untuk memilih variabel yang signifikan dalam melakukan perkiraan nilai jual harga dan komoditi kopi arabika di Indonesia. Adapun penelitian ini menghasilkan akurasi dari evaluasi performansi diatas 95% dengan menggunakan pendekatan teknik feature selection dibandingkan dengan tidak menggunakan teknik *feature selection* sehingga yang dapat diterapkan dalam kasus – kasus tersebut [16].

Berdasarkan pembahasan yang telah dijelaskan pada penelitian sebelumnya dengan melakukan penerapan fitur seleksi untuk Prediksi hasil produksi agrikultur menggunakan algoritma *k-nearest neighbor*. Penelitian ini akan melakukan percobaan terhadap pendekatan teknik *feature selection* untuk melakukan eliminasi atau reduksi pada fitur yang digunakan pada algoritma *k-nearest neighbor* sehingga *output* utama dari penelitian ini adalah melakukan percobaan terhadap perbandingan performa dari penggunaan teknik *feature selection* dan tidak diterapkan.

2. METODOLOGI PENELITIAN

Tahap perencanaan yang digunakan pada penelitian ini yaitu menggunakan Dataset Hasil Produksi Agrikultur dimana tahap pengambilan data melalui platform pada situs web Kaggle (bisa diakses ke: *https://www.kaggle.com/datasets/chitrakumari25/smart-agricultural-production-optimizing-engine*) dengan jumlah 2200 *record* data. Selanjutnya proses pengolahan data yaitu preprocessing dan splitting data. Proses pemilihan fitur seleksi menggunakan *Principal Component Analysisi* (PCA) dimana untuk mengurangi ukuran dataset asli dengan tetap menjaga akurasi kinerja. Pada proses klasifikasi penelitian ini menggunakan *K-Nearest Neighbor* (KNN). Proses analisis dan hasil yang dilakukan menggunakan perhitungan untuk perbandingan performa Algoritma *K-Nearest Neighbor* dengan hasil akurasi pada klasifikasi analisis akhir. Berikut merupakan Metodologi Penelitian yang tertera pada Gambar 1.



Gambar 1. Metodologi Penelitian

2.1 Data Mining

Data mining yaitu suatu rangkaian dalam proses menemukan fakta yang sebelumnya belum diketahui dari data dalam jumlah besar. Informasi yang didapat dari data tersebut menggunakan cara proses pemisahan dan pengenalan yang bisa menjadi informasi bermanfaat [17]. Dalam memproses suatu penelitian tersebut data mining

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 4, Nomor 1, September 2022
e-ISSN 2685-998X
DOI 10.30865/json.v4i1.4813

mamakai teknik matematika, statistika, *machine learning*, dan juga *artificial intelligence* untuk memperoleh informasi yang akurat dari berbagai macam data berukuran besar [18].

2.2 *Feature Selection* dan *Principal Component Analysis*

Feature selection adalah salah satu teknik yang sangat penting serta sering dipakai pada *pre-processing*. Pada teknik ini dilakukan pengurangan jumlah fitur yang terlibat agar dapat menentukan silai kelas target dengan cara mengurangi fitur yang tidak sesuai dan data yang berlebihan. Seleksi Fitur yang dipakai adalah *Principal Component Analysis* (PCA) [19]. PCA merupakan perubahan bentuk linear untuk menentukan sistem koordinat baru yang berasal dari sebuah dataset [20]. Metode PCA digunakan karena cukup efesien dapat mengatasi multikolineritas atau hubungan kuat antara dua variable atau lebih di berbagai kondisi dan dapat menghapuskan korelasi diantara variable bebas sehingga tidak dapat berkolerasi sama sekali dimana PCA dapat ditunjukan pada Persamaan 1

$$(x)' = 1/N \sum_{i=1}^{N} .(xi) \quad (1)$$

Dimana fitur seleksi baru dari dimensi kumpulan data asli lebih rendah dari data dimenasi lama yang dihitung berdasarkan nilai rata – rata *x* pada dimensi [22].

2.3 *K-Fold Cross Validation* dan *Holdout*

Cross validation adalah sebuah teknik untuk memproses sebuah validasi untuk menghasilkan model paling baik. Teknik tersebut selanjutnya memeriksa keberhasilan dari model yang telah dibuat untuk menyusun ulang atau *resampling* terhadap data dan membaginya menjadi dua komponen yakni data uji (*testing)* dan data latih (*training)*. Data *training* ini digunakan saat menyesuaikan model agar bisa mempelajari pola suatu data lalu membuktikan kebenaran pada model yang diterapkan pada data uji yang dijadikan percobaan [23].

Hold-out adalah metode yang menyajikan sejumlah data agar dapat dipakai untuk data testing dan data training. Dataset yang dibagi telah diidentifikasi pada label kelasnya. Saat proses pemrosesan data untuk dibagi sebagai data testing dan data uji, sangat mungkin terjadi overrepresented pada salah satu atau lebih klasifikasi. Pada teknik ini, salah satu komponen yang digunakan untuk menguji pengklasifikasi (*classifier*) dan komponen lainnya untuk uji pengklasifikasi [21]

2.4 *K-Nearest Neighbor*

K-Nearest Neighbor (KNN) merupakan salah satu dari berbagai macam algoritma data mining yang dimana menggunakan seleksi nilai yang cocok pada k (jumlah record data yang paling dekat pada objek), yang dimana klasifikasi pada algoritma ini tergantung pada nilai k tersebut. Saat menentukan nilai k, teknik sederhananya yaitu melakukan algoritma berulang ulang sampai mendapatkan nilai k yang berbeda lalu diambil salah satu nilai k nya dengan kinerja yang paling baik [20]. Berikut Persamaan 2 yang memenuhi pada algoritma *K-Nearest Neighbor*:

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

Dimana hasil dari $d(xi , xj)$ yaitu pengurangan pada setiap atribut yang dikuadratkan dan dijumlahkan pada nilai terkecil dengan data uji [21].

2.5 *Confusion Matrix*

Confusion matrix merupakan metode pada penilaian yang bisa memilih kinerja berdasarkan benar atau salah pada proses klasifikasi. Dalam *confusion matrix* ini terdapat akurasi, *precision*, dan *recall*. Akurasi adalah ukuran yang menentukan antara hasil pengukuran dengan nilai-nilai yang sebenarnya diukur Adapun akurasi pada *confusion matrix* memenuhi Persamaan 3

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

Dimana TP adalah *true positive,* TN adalah *true negative*, FP adalah *false positive* dan FN adalah *false negative* [21].

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Data yang dipakai pada penelitian ini yaitu dataset mengenai Hasil Produksi Agrikultur yang berasal dari *platform* Kaggle yang terdiri dari 2200 *record* data. Atribut yang digunakan adalah Ketersediaan Air di *Nitrogen* (N), *Phosphorous* (P), *Pottasium* (K), suhu (*temperature*), kelembaban (*humadity*), keasaman (pH), dan curah hujan (*rainfall*). Adapun label pada penelitian ini adalah Jenis Tanaman yang terdiri dari *Rice*, *Maize*, *Chickpea*, *Kidney beans*, *pigeonpeas*, *mothbeans*, *mungbean*, *blackgram*, *lentil*, *pomegranate*, *banana*, *mango*, *grape*, *watermelon*,

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 4, Nomor 1, September 2022
e-ISSN 2685-998X
DOI 10.30865/json.v4i1.4813

muskmelon, *apple*, *orange*, *papaya*, *coconut*, *cotton*, *jute*, dan *coffee*. Berikut hasil Data Hasil Produksi Agrikultur pada Tabel 1. Sebagai berikut:

Tabel 1. Data Hasil Produksi Agrikultur

No	N	P	K	*temperature*	*humadity*	pH	*rainfall*	label
1	90,00	42,00	43,00	20,88	82,00	6,50	202,94	rice
2	85,00	58,00	41,00	21,77	80,32	7,04	226,66	rice
3	60,00	55,00	44,00	23,00	82,32	7,84	263,96	rice
4	74,00	35,00	40,00	26,49	80,16	6,98	242,86	rice
5	78,00	42,00	42,00	20,13	81,60	7,63	262,72	rice
…	…	…	…	…	…	…	…	…
2198	118,00	33,00	30,00	24,13	6,72	6,36	173,32	coffee
2199	117,00	32,00	34,00	26,27	5,21	6,76	127,18	coffee
2200	104,00	18,00	30,00	23,60	6,04	6,78	140,94	coffee

3.2 *Feature Selection*

Seleksi fitur pada penelitian ini menggunakan teknik *Principal Component Analysis* (PCA) untuk membentuk fitur – fitur yang baru terhadap penelitian ini. Adapun fitur yang dihasilkan menggunakan teknik *Principal Component Analysis* (PCA) telah dinormaliasikan sehingga terdapat pada Tabel 2. sebagai berikut:

Tabel 2. Normalisasi *Principal Component Analysis* (PCA)

No	pc_1	pc_2	pc_3	pc_4	pc_5	pc_6	label
1	0,583	0,844	1,373	-1,614	-0,308	-0,096	rice
2	0,475	0,785	1,252	-1,792	-1,107	-0,532	rice
3	0,634	0,694	1,179	-1,818	-2,523	-0,538	rice
4	1,048	1,087	1,393	-0,982	-1,448	-0,657	rice
5	0,873	0,659	1,455	-2,334	-1,959	-0,318	rice
…	…	…	…	…	…	…	…
2198	1,158	0,64	1,046	-1,302	0,492	-0,885	coffee
2199	1,219	-0,052	0,181	-0,99	0,601	-1,309	coffee
2200	1,373	-0,056	0,501	-1,219	0,346	-0,573	coffee

3.3 Evaluasi Algoritma *K-Nearest Neighbor*

Berdasarkan penelitian yang dilakukan, penerapan algoritma *K-Nearest Neighbor* menggunakan parameter nilai K = 5 berdasarkan penelitian terdahulu yang terkait. Sehingga menghasilkan nilai akurasi yang dihasilkan algoritma *K-Nearest Neighbor* dengan menggunakan pendekatan teknik *feature selection* dan tanpa menggunakan teknik tersebut dapat dilihat pada Tabel 3. sebagai berikut

Tabel 3. Evaluasi Performa Algoritma *K-Nearest Neighbor*

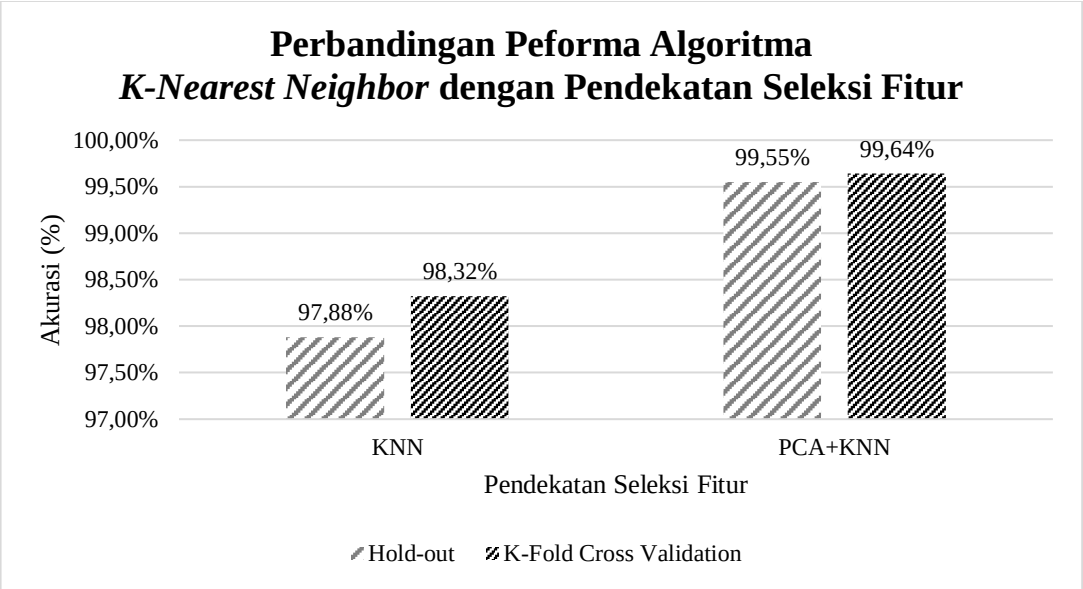
Pembagian Data	Rasio/ Parameter	Tanpa PCA	Dengan PCA
		Evaluasi Akurasi	
Hold-out	70% : 30%	**97.88%**	98.94%
	80% : 20%	97.50%	99.32%
	90% : 10%	97.27%	**99.55%**
K-Fold Cross Validation	K = 5	98.09%	99.41%
	K = 10	98.18%	99.45%
	K = 15	**98.32%**	**99.64%**

Berdasarkan pada Tabel 3. Evaluasi Performa Algoritma *K-Nearest Neighbor* menghasilkan perbedaan terhadap akurasi yang dihasilkan. Evaluasi yang dihasilkan menggunakan pembagian data *Hold-out* tanpa menggunakan pendekatan seleksi fitur dan menggunakan PCA masing – masing menghasilkan akurasi sebesar 97,88% dan 99,55% sehingga memiliki kenaikan akurasi sebesar 1,67% setelah menggunakan penggunaan teknik seleksi fitur PCA. Sedangkan dengan menggunakan pembagian data *K-Fold Cross Validation* menggunakan pendekatan seleksi fitur dan menggunakan PCA masing – masing menghasilkan akurasi sebesar 98,32% dan 99,64% sehingga memiliki kenaikan akurasi sebesar 1,32% setelah menggunakan penggunaan teknik seleksi fitur PCA.

3.4 Perbandingan Performa Algoritma *K-Nearest Neighbor*

Berdasarkan pemaparan pada Tabel 3. Evaluasi Performa Algoritma KNN yang menghasilkan beberapa kombinasi terbaik untuk penerapan Algoritma yang dapat dilihat pada Gambar 1. sebagai berikut

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 4, Nomor 1, September 2022
e-ISSN 2685-998X
DOI 10.30865/json.v4i1.4813



Gambar 2. Perbandingan Performa Algoritma

Pada Gambar 2. Perbandingan Performa Algoritma disajikan Perbandingan Performa Algoritma *K-Nearest Neighbor* dengan Pendekatan Fitur Seleksi. Adapun penggunaan teknik Fitur Seleksi atau *feature selection* memiliki dampak yang cukup signifikan terhadap perubahan performa akurasi pada algoritma. Adapun berdasarkan pada Gambar 1. teknik Fitur Seleksi PCA menghasilkan akurasi paling tinggi sebesar 99,64% dengan kombinasi *K-Fold Cross Validation* sebagai teknik pembagian data pada penelitian ini. Namun, performa algoritma yang tidak menggunakan pendekatan seleksi fitur menghasilkan akurasi paling rendah 97,88% dengan kombinasi *Hold-out* sebagai teknik pembagian data. Sehingga penggunaan PCA dapat meningkatkan akurasi dan performa dari *K-Nearest Neighbor* pada penelitian ini.

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan dapat disimpulkan bahwa penerapan teknik seleksi fitur untuk Prediksi Hasil Produksi Agrikultur menggunakan algoritma *K-Nearest Neighbor* yang menghasilkan teknik seleksi fitur *Principal Component Analysis* (PCA) sebagai peningkatan performa algoritma *K-Nearest Neigbor* (KNN) dengan akurasi paling tinggi sebesar 99,64% dengan kombinasi *K-Fold Cross Validation* sebagai teknik pembagian data pada penelitian ini. Sedangkan, performa algoritma KNN menghasilkan akurasi paling rendah 97,88% dengan kombinasi *Hold-out* sebagai teknik pembagian data pada penelitian ini.

REFERENCES

[1] I. Darmawan, I. Kumara, and D. C. Khrisne, “Smart Garden Sebagai Implementasi Sistem Kontrol dan Monitoring Tanaman Berbasis Teknologi Cerdas,” 2021.
[2] F. Hermawan, H. Junawarko, and T. Informasi, “Rancang Bangun Aplikasi Sistem Informasi Pembelajaran Sektor Pertanian Berbasis Android di Kecamatan Jabung.”
[3] A. Amam and S. Rusdiana, “Pertanian Indonesia dalam menghadapi persaingan pasar bebas,” *J. Agriovet*, vol. 4, no. 1, pp. 37–68, 2021.
[4] R. Nurjasmi, “Review: Potensi Pengembangan Pertanian Perkotaan oleh Lanjut Usia untuk Mendukung Ketahanan Pangan,” 2021. [Online]. Available: http://ejournal.urindo.ac.id/index.php/pertanian
[5] D. Rosian Adhy, “Rancang Bangun Sistem Prediksi Varietas Padi Yang Cocok Dengan Lahan Menggunakan Metode Data Mining Algoritma C4.5.” [Online]. Available: https://jabarprov.go.id/index.php/pages/id/1046
[6] Gunawan, M. Zarlis, P. Sihombing, and Sutarman, “Optimization of the CNN model for smart agriculture,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1088, no. 1, p. 12029, 2021, doi: 10.1088/1757-899x/1088/1/012029.
[7] C. G. Anupama and C. Lakshmi, “A comprehensive review on the crop prediction algorithms,” *Mater. Today Proc.*, no. xxxx, Mar. 2021, doi: 10.1016/j.matpr.2021.01.549.
[8] B. Septia Pranata and D. Putro Utomo, “Bulletin of Information Technology (BIT) Penerapan Data Mining Algoritma FP-Growth Untuk Persediaan Sparepart Pada Bengkel Motor (Study Kasus Bengkel Sinar Service),” *Bull. Inf. Technol.*, vol. 1, no. 2, pp. 83–91, 2020.
[9] S. P. Raja, B. Sawicka, Z. Stamenkovic, and G. Mariammal, “Crop Prediction Based on Characteristics of the Agricultural Environment Using Various Feature Selection Techniques and Classifiers,” *IEEE Access*, vol. 10, pp. 23625–23641, 2022, doi: 10.1109/ACCESS.2022.3154350.
[10] H. Osman, M. Ghafari, and O. Nierstrasz, “The Impact of Feature Selection on Predicting the Number of Bugs,” *arXiv*, Jul. 2018, doi: https://doi.org/10.48550/arXiv.1807.04486.
[11] F. Tasnim and S. U. Habiba, “A Comparative Study on Heart Disease Prediction Using Data Mining Techniques and

Feature Selection," *Int. Conf. Robot. Electr. Signal Process. Tech.*, pp. 338–341, 2021, doi: 10.1109/ICREST51555.2021.9331158.
[12] O. Elsherbiny, Y. Fan, L. Zhou, and Z. Qiu, "Fusion of feature selection methods and regression algorithms for predicting the canopy water content of rice based on hyperspectral data," *Agric.*, vol. 11, no. 1, pp. 1–21, 2021, doi: 10.3390/agriculture11010051.
[13] F. Anowar, S. Sadaoui, and B. Selim, "Conceptual and empirical comparison of dimensionality reduction algorithms (PCA, KPCA, LDA, MDS, SVD, LLE, ISOMAP, LE, ICA, t-SNE)," *Comput. Sci. Rev.*, vol. 40, p. 100378, May 2021, doi: 10.1016/j.cosrev.2021.100378.
[14] M. Gandhi, S. Kothavade, S. Nehete, S. Arlikar, and K. A. Shinde, "Agricultural Production Optimization Engine," *www.irjmets.com @International Res. J. Mod. Eng.*, vol. 5374, [Online]. Available: http://www.irjmets.com
[15] Mustakim, E. Rahmi, M. R. Mundzir, S. T. Rizaldi, Okfalisa, and I. Maita, "Comparison of DBSCAN and PCA-DBSCAN Algorithm for Grouping Earthquake Area," *2021 Int. Congr. Adv. Technol. Eng. ICOTEN 2021*, pp. 0–4, 2021, doi: 10.1109/ICOTEN52080.2021.9493497.
[16] A. Bode, "K-Nearest Neighbor Dengan Feature Selection Menggunakan Backward Elimination Untuk Prediksi Harga Komoditi Kopi Arabika," *Ilk. J. Ilm.*, vol. 9, no. 2, pp. 188–195, Aug. 2017, doi: 10.33096/ilkom.v9i2.139.188-195.
[17] K. C. Pelangi, "Prediksi Produksi Tanaman Pangan di Provinsi Gorontalo Menggunakan Metode K-NN (K-Nearest Neighbor)," vol. 6, no. 2, 2021.
[18] Y. Kurnia, Y. Isharianto, Y. C. Giap, A. Hermawan, and Riki, "Study of application of data mining market basket analysis for knowing sales pattern (association of items) at the O! Fish restaurant using apriori algorithm," in *Journal of Physics: Conference Series*, 2019, vol. 1175, no. 1. doi: 10.1088/1742-6596/1175/1/012047.
[19] E. Prasetyo, "Reduksi Dimensi Set Data dengan DRC pada Metode Klasifikasi SVM dengan Upaya Penambahan Komponen Ketiga," *Pros. SNATIF*, pp. 293–300, 2014.
[20] R. Harun, K. Chandra Pelangi, and Y. Lasena, "Penerapan Data Mining Untuk Menentukan Potensi Hujan Harian Dengan Menggunakan Algoritma K-Nearest Neighbor (KNN)," Online, 2020. [Online]. Available: http://e-journal.stmiklombok.ac.id/index.php/misi
[21] S. T. Rizaldi and M. Mustakim, "Perbandingan Teknik Pembagian Data untuk Klasifikasi Sarana Akses Air pada Algoritma K- Nearest Neighbor dan Naïve Bayes Classifier," in *Seminar Nasional Teknologi Informasi, Komunikasi dan Industri (SNTIKI) 12*, 2020, pp. 130–137.
[22] E. Odhiambo Omuya, G. Onyango Okeyo, and M. Waema Kimwele, "Feature Selection for Classification using Principal Component Analysis and Information Gain," *Expert Syst. Appl.*, vol. 174, no. February, p. 114765, 2021, doi: 10.1016/j.eswa.2021.114765.
[23] Y. Widyaningsih, G. P. Arum, and K. Prawira, "Aplikasi K-Fold Cross Validation Dalam Penentuan Model Regresi Binomial Negatif Terbaik," *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 15, no. 2, pp. 315–322, 2021, doi: 10.30598/barekengvol15iss2pp315-322.