AGROMIX Terakreditasi Nomor: 23/E/KPT/2019
Website: https://jurnal.yudharta.ac.id/v2/index.php/AGROMIX Volume 10 No 2 (2019), Halaman: 130-137
DOI: https://doi.org/10.35891/agx.v10i2.1636 p-ISSN: 2085-241X ; e-ISSN: 2599-3003

Implementasi metode *K-nearest neighbor* dan bagging untuk klasifikasi mutu produksi jagung

Implementation of the K-nearest neighbor method and bagging for the classification of the quality of corn production

Moch. Lutfi[1]
[1]Program Studi Teknik Informatika, Universitas Yudharta Pasuruan, Indonesia
email: moch.lutfi@yudharta.ac.id

Informasi artikel:
Dikirim: 01/09/2019
ditinjau: 18/09/2019
disetujui: 30/09/2019

ABSTRACT: *Corn is an agricultural crop in the Indonesian community in addition to rice and soybeans, because almost all of the fertile areas for agricultural crop seeds, the quality of corn that must be met as food raw material, is very necessary for farmers producing crops. The k-nearest neighbor algorithm is a method used for the classification process of objects based on training data with the distance closest to the object or often called euclidean distance. In this study, replace imputation is used for the preprocessing stage of missing value data and bagging is used to handle large-scale datasets while k-nearest neighbor is used as a classification of the quality of corn-based on the attributes Varietas, Length, Shape, Color Taste, Season Technique, Pests PH. Based on testing the best accuracy value is 79.30%, precision is 83.04% while recall with a value of 80.93% results is obtained from the results of the performance test of bagging and replace imputation methods on the k-nearest neighbor algorithm by handling missing values.*

Keywords*: corn; classification; imputation; bagging; K-nearest neighbor*

ABSTRAK: Jagung merupakan tanaman pertanian di masyarakat Indonesia selain padi dan kedelai, dikarenakan hampir dari keseluruhan daerahnya yang subur untuk bibit tanaman pertanian, kualitas mutu jagung yang harus dipenuhi sebagai bahan baku pangan, sangat diperlukan untuk petani penghasil panen. Algoritma *k-nearest neighbor* adalah metode yang digunakan untuk proses klasifikasi terhadap objek berdasarkan data *training* dengan jaraknya yang paling dekat dengan objek atau sering disebut *euclidean distance*. Dalam penelitian ini digunakan *replace imputation* untuk tahap *preprocessing* data *missing value* dan *bagging* digunakan untuk menangani dataset dalam sekala besar sedangkan *k-nearest neighbor* digunakan sebagai klasifikasi kualitas mutu jagung berdasar *attribut Varietas*, Panjang, Bentuk, Warna Rasa, Teknik Musim, Hama PH. Berdasarkan pengujian data nilai akurasi terbaik yaitu 79.30%, *precision* yaitu 83.04% sedangkan *recall* dengan nilai 80.93% hasil tersebut di peroleh dari hasil uji kinerja metode *bagging* dan *replace imputation* pada algoritma *k-nearest neighbor* dengan penanganan *missing value*.

Kata Kunci: jagung; klasifikasi; *imputation; bagging; K-nearest neighbor*

Sitasi: Lutfi, M. (2019). Implementasi metode *K-nearest neighbor* dan bagging untuk klasifikasi mutu produksi jagung. *AGROMIX*, *10*(2), 130-137. https://doi.org/10.35891/agx.v10i2.1636

PENDAHULUAN

Jagung merupakan tanaman pertanian di masyarakat Indonesia selain padi dan kedelai, dikarenakan hampir dari keseluruhan daerahnya yang subur untuk bibit tanaman pertanian. Jagung dapat diolah sebagai bahan konsumsi manusia maupun hewan. Tanaman jagung yang banyak di tanam oleh masyarakat di Indonesia adalah tipe

mutiara (Munarto *et al.,* 2014), serta terdapat juga jagung yang bertipe brondong, jagung tipe tepung, jagung gigi kuda, dan jagung manis.

Kualitas mutu jagung yang harus dipenuhi sebagai bahan baku pangan, sangat diperlukan untuk petani penghasil panen. Sebagai bahan acuan dalam penyusunan standar mutu jagung adalah SNI (Indonesia, 1998), dan dapat pula memperhatikan dari semua data serta masukan dari berbagai pihak.

Bojonegoro merupakan salah satu kabupaten Jawa Timur mayoritas masyarakatnya petani juga membudidayakan tanaman jagung. selain itu jagung untuk saat ini dikonsumsi oleh masyarakat Indonesia karena kaya akan gizi, oleh sebab itu kualitas mutu jagung harus terjaga sedemikian rupa karena produksi jagung semakin tahun produktivitasnya menurun oleh berkurangnya lahan tanam. Proses klasifikasi data mining merupakan peran utama dalam proses penggalian informasi (Witten, et al., 2016) dengan mengidentifikasi pola dan hubungan dari sejumlah informasi data yang ada, dalam jumlah kecil maupun dalam sekala besar sehingga diharapkan mampu menemukan pola baru yang bermakna.

Secara umum menentukan kualitas mutu jagung sangatlah sulit karena dari masing-masing daerah lahan tanam memiliki unsur yang berbeda-beda. Oleh karena itu dibutuhkan suatu metode klasifikasi untuk menentukan kualitas mutu jagung yang akan dihasilkan.

Metode klasifikasi data mining ada 6 yaitu *decision tree*, *svm, naive bayes* dan *k-nearest neighbor* (Fayed & Atiya, 2009) dari metode klasifikasi tersebut yang sering digunakan dalam penelitian adalah metode *k-nearest neighbor*. Metode *k-nearest neighbor* digunakan untuk prediksi maupun klasifikasi terhadap objek berdasarkan nilai k tetangga terdekat. Tujuan dari *k-nearest neighbor* adalah untuk prediksi atau klasifikasi objek baru berdasarkan *data training* (Han & Kamber, 2012), nilai parameter k yang digunakan menyatakan jumlah jarak tetangga terdekat yang melibatkan dalam menentukan klasifikasi label kelas sebagai target pada data uji. Dari nilai parameter nilai k jarak tetangga terdekat yang dipilih setelah itu dilakukan voting kelas sebagai target dari nilai parameter nilai k jarak tetangga terdekat.

Algoritma *k-nearest neighbor* memiliki kelebihan karena sederhana, efektif dan sudah banyak diterapkan di berbagai kasus penelitian pada klasifikasi maupun prediksi (Xindong Wu, 2009). Namun kelemahan dari algoritma *k-nearest neighbor* jika diterapkan pada dataset skala besar (Wan, *et al.*, 2012) karena tingkat akurasi dan waktu komputasi kurang baik (Fayed & Atiya, 2009). Adapun dataset yang besar berupa volume yang banyak, atribut atau label data yang banyak, dan informasi yang melebihi sehingga perlu adanya penanganan

DOI: https://doi.org/10.35891/agx.v10i2.1636

khusus dalam pengolahan data salah satunya dengan metode *Computing*. Penelitian yang dilakukan (Neo & Ventura, 2012) menggunakan metode *Direct Boosting* untuk meningkatkan *performance* akurasi *k-nearest neighbor* dengan modifikasi pembobotan jarak tetangga terhadap data latih.

Oleh karena itu dalam penelitian ini perlu adanya metode untuk mengatasi dataset skala besar yang akan diproses sehingga mampu meningkatkan kinerja klasifikasi (Wahono et al., 2014) kualitas mutu produksi jagung. *Bagging* merupakan metode penggabungan atau sering disebut teknik *ensemble* pada teknik klasifikasi ini memisahkan *data training* ke beberapa *data training* yang baru dengan pengambilan sampling data untuk membangun model basis *data training* baru (Wahono, & Suryana, 2013).

Dalam penelitian ini digunakan metode *k-nearest neighbor* sebagai klasifikasi kualitas mutu jagung berdasar atribut Varietas, Panjang, Bentuk, Warna Rasa, Teknik Musim, Hama PH.

Penelitian tentang klasifikasi hasil pertanian telah lama dilakukan, dan sudah banyak hasil penelitian yang dipublikasikan. Sebelum memulai penelitian, perlu adanya kajian terhadap penelitian yang sudah dilakukan, agar dapat mengetahui metode, data, maupun model yang sudah pernah digunakan. Kajian penelitian sebelumnya sebagai rujukan untuk mengetahui *state of the art* tentang penelitian klasifikasi tentang hasil pertanian yang membahas tentang klasifikasi kualitas mutu jagung.

Penelitian yang dilakukan (Munarto et al., 2014) ekstraksi fitur dengan pengambilan ciri dan fitur dari bentuk data yang nantinya nilai yang didapat akan dianalisa untuk proses klasifikasi dengan menggunakan metode *artificial intelligence* yaitu *fuzzy logic*. Metode *fuzzy logic* diimplementasikan karena dapat merepresentasikan dengan baik sehingga mampu menentukan keputusan klasifikasi kualitas biji jagung pecah, biji jagung yang rusak maupun biji jagung yang sempurna.

Penelitian yang dilakukan (Effendy *et al.*, 2018) menggunakan perhitungan klasifikasi data ordinal hanya ditemukan pada tools weka. Dalam penelitian sebelumnya menunjukkan konsep klasifikasi *class ordinal* belum pernah dilakukan. Namun beberapa tahun terakhir ada kemajuan dalam *learning artificial* dalam konsep ordinal menggunakan *machine learning* seperti algoritma *decision trees, neural network,* dan *support vector machines* dan lain-lain sudah mendukung proses klasifikasi ordinal.

Penelitian yang dilakukan (Munanda, 2010) tentang pengembangan aplikasi sistem pendukung keputusan untuk mendiagnosa jenis penyakit tanaman menggunakan metode *Fuzzy MCDM* berbasis web. Penelitian ini digunakan metode *fuzzy Multi Criteria Desicion Making* untuk mengatasi diagnosis masalah penyakit tanaman jagung. Hasil diagnosis masalah

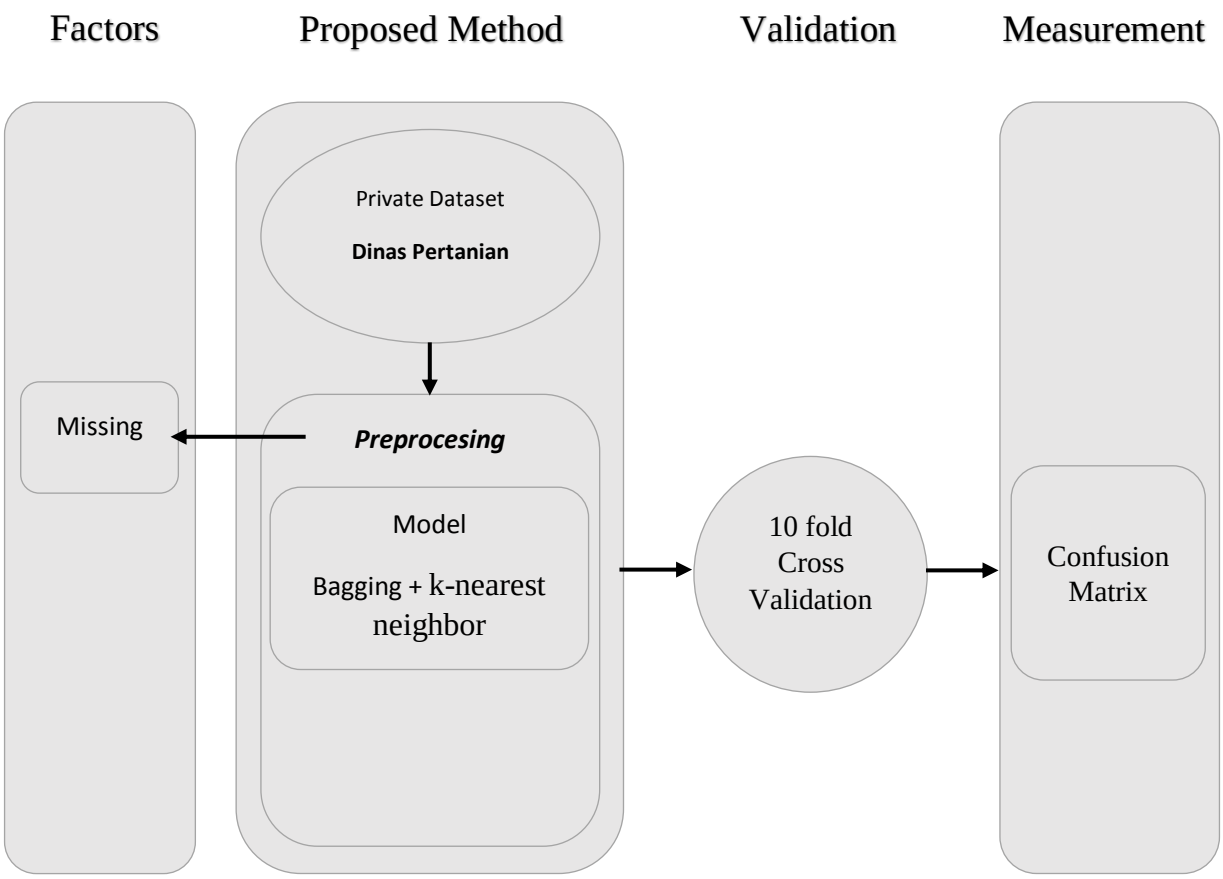
DOI: https://doi.org/10.35891/agx.v10i2.1636

penyakit tanaman jagung diambil dari perhitungan tertinggi dari setiap jenis penyakit yang ada.

Pada penelitian ini akan diterapkan metode klasifikasi *k-nearest neighbor* pada dataset dinas pertanian yang nantinya akan digunakan sebagai penentuan kualitas mutu jagung di kabupaten Bojonegoro.

METODE

Penelitian ini mengusulkan penerapan metode dan aplikasi sistem untuk mengimplementasikan bentuk metode yang diusulkan, selanjutnya menerapkan pada *private dataset* Dinas Pertanian Kabupaten Bojonegoro dan mengukur kinerjanya. Untuk proses klasifikasi kualitas mutu jagung, diusulkan model klasifikasi menggunakan algoritma *k-nearest neighbor*. Validasi pada pengukuran kinerja digunakan 10x-fold *cross validation*. Hasil data uji dianalisa menggunakan *confusion matrix*.



Gambar 1. Model yang diusulkan

Model yang diusulkan dengan algoritma *k-nearest neighbor* pada *dataset private* dinas pertanian kabupaten Bojonegoro akan dibagi menjadi 10 bagian. Secara bertahap data akan dibagi sebagai data latih dan sebagian lainnya digunakan sebagai data uji. Data latih diproses

DOI: https://doi.org/10.35891/agx.v10i2.1636

menggunakan algoritma klasifikasi *k-nearest neighbor*, dan diuji dengan data latih, kemudian diukur kinerjanya menggunakan evaluasi *confusion matrix*.

Tahapan metode dalam penelitian ini adalah:

1. Mengumpulkan data
2. *Preprocessing* data
3. Menerapkan metode *bagging*
4. Menerapkan ke Metode *k-nearest neighbor*
5. Validasi menggunakan 10 *k-fold validation*
6. Evaluasi menggunakan *confusion matrix*

K-nearest neighbor

Algoritma *k-nearest neighbor* merupakan metode klasifikasi terhadap objek berdasarkan *data training* yang jarak tetangga yang paling dekat dengan objek data (Setiawan *et al*., 2015) atau sering disebut *Euclidean distance*. *Data training* diproyeksikan ke dalam ruang berdimensi banyak, dari dimensi masing-masing merepresentasikan bentuk fitur dari data yang ada. Ruang yang dihasilkan akan dibagi berdasarkan *data training*. Dari titik ruang ini sering ditandai dengan kelas c jika kelas c merupakan klasifikasi yang ditemui pada nilai parameter k buah jarak tetangga yang paling dekat dari titik berdasarkan jarak *Euclidean distance* (Han & Kamber, 2012).

Tahapan algoritma *k-nearest neighbor*

Secara umum algoritma *k-nearest neighbor* untuk mencari nilai jarak (k) terdekat atau *Euclidean distance* adalah sebagai berikut:

a. Siapkan dataset produksi jagung
b. Menentukan nilai K.
c. Menghitung kuadrat jarak *euclid distance.*

$$di = \sqrt{\sum_{1=i}^{p} (x1i - x2i)(x_{1i} - x_{2i})^2}$$

d. Mengurutkan objek termasuk pada kelompok jarak tetangga yang paling terdekat atau sering disebut *Euclidean distance.*

Bagging

Metode penggabungan pada *machine learning* yang dikembangkan untuk meningkatkan stabilitas dan akurasi dari algoritma *machine learning* yang digunakan dalam klasifikasi maupun prediksi. Bagging juga dapat mengurangi bentuk varians serta membantu menghindari terjadinya *overfitting* (Wahono, & Suryana, 2013). Meskipun biasanya digunakan pada metode *decision tree* (Abellán, 2013), namun *bagging* juga dapat digunakan pada semua jenis metode *machine learning* dan konsep *bagging* adalah pendekatan model *average*.

Berikut tahapan metode *bagging*:

1. Buat sampel data {(X , $Y1*$) 1 * , (X , $Y2*$), … , 2 * ($Xn*$, $Yn*$)}
2. Output akhir : $C(x) = B^{-1} \sum_{b=1}^{B} C_b(x)$

Confusion matrix

Confusion matrix merupakan tabel informasi klasifikasi yang mengandung hasil dari perhitungan model yang diusulkan secara keseluruhan (Sokolova & Lapalme, 2009). Pengukuran data dengan menggunakan

DOI: https://doi.org/10.35891/agx.v10i2.1636

evaluasi *confusion matrik* melalui akurasi, *presisi,* dan *recall*. Hasil pengukuran dari model yang diusulkan direpresentasikan ke dalam sebuah table informasi hasil klasifikasi untuk memudahkan pembacaan. Adapun rumus perhitungannya(Gorunescu, 2011) yaitu :

$$Akurasi = \frac{TP+TN}{TP+FP+TN+FN} \quad (1)$$

$$Presisi = \frac{TN}{FP+TN} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

		Predicted	
		Negative	Positive
Actual	Negative	TN	TP
	Positive	FN	FP

Keterangan:
TP = True Positif *TN = True Negative*
FP = False Positif *FN = False Negative*

HASIL DAN PEMBAHASAN

Eksperimen dilakukan menggunakan sebuah laptop DELL Latitude E7240 dengan prosesor Pentium®Core i7-4600U@ 2.10 GHz, memori (RAM) 8,00 GB, dan sistem operasi Windows 8 64-bit. Untuk menganalisis hasil pengukuran kinerja digunakan aplikasi Rapid Miner.

Hasil pengukuran kinerja metode *k-nearest neighbor* pada *dataset private* dinas pertanian kabupaten Bojonegoro ditunjukkan pada Tabel 1

Tabel 1 hasil akurasi metode yang diusulkan

accuracy: 79.30%					
	Grade-C	Grade-D	Grade-B	Grade-A	class precision
Pred. Grade-C	2203	42	406	2	83.04%
Pred. Grade-D	144	722	27	3	80.58%
Pred. Grade-B	373	11	967	10	71.05%
Pred. Grade-A	2	1	4	35	83.33%
class recall	80.93%	93.04%	68.87%	70.00%	

Dari tabel di atas bisa disimpulkan bahwa hasil akurasi menggunakan metode *bagging* dan *replace imputation* pada *k-nearest neighbor* adalah sebagai berikut:

a. Nilai TP (*True Positive)* dan nilai selain TP disebut dengan FN (*False Negative)*.
b. Sedangkan *class recall* merupakan kolom yang berisi besar nilai klasifikasi yang tepat. Misalnya *class recall* yang tepat pengklasifikasiannya adalah sebagai berikut:

$$recall = \frac{TP}{TP+FN} = \frac{2203}{2722} = 0.809331374x100 = 80.93\%$$

c. Sedangkan *class precision* merupakan baris yang berisi besar nilai klasifikasi yang tepat. Misalnya *class precision* yang tepat pengklasifikasiannya adalah sebagai berikut:

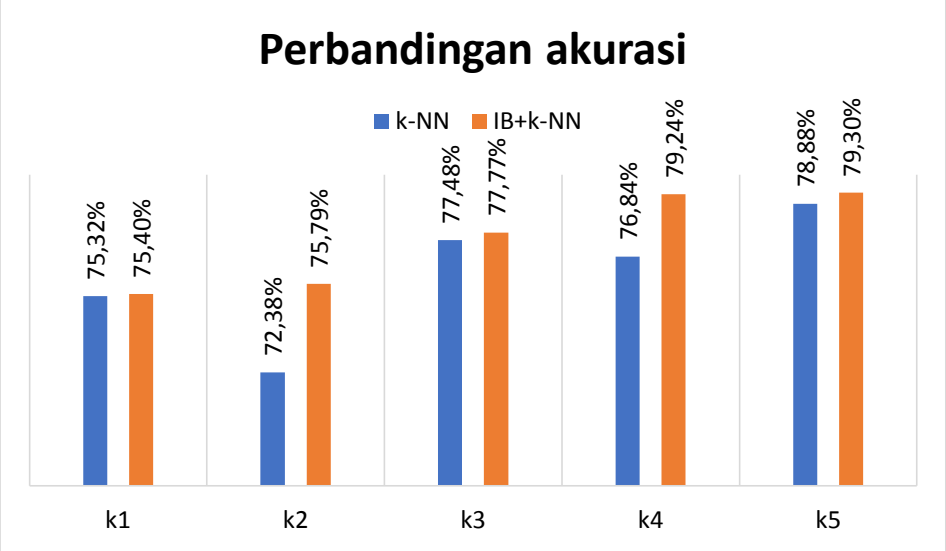
$$precision = \frac{TP}{TP+FN} = \frac{2203}{2650} = 0.831320755x100 = 83.04\%$$

d. Sedangkan *accuracy* merupakan persentase hasil klasifikasi yang benar yang bisa didapatkan dengan cara berikut:

DOI: https://doi.org/10.35891/agx.v10i2.1636

$$Accuracy = \frac{TP + TN}{TP + FN + FP + FN} = \frac{3927}{4952}$$
$$= 0.793012924x100$$
$$= 79.30\%$$

Perbandingan akurasi metode *k-nearest neighbor* dengan metode *Imputation + bagging + k-nearest neighbor.*



Gambar 2. Perbandingan akurasi

KESIMPULAN

Dalam penelitian ini dilakukan perhitungan menggunakan model Ordinal *Class* klasifikasi sebagai target dari *dataset*. Keakuratan data diuji dengan menggunakan 10 *fold validation* yang artinya melatih data sebanyak 10x dan menguji data sebanyak 10x dari jumlah data 4952. Pengujian dilakukan menggunakan parameter k1, k2, k3, k4 dan k5 dengan menghitung nilai *precision, recall,* dan *accuracy*. Berdasarkan pengujian metode yang diusulkan bahwa metode *bagging* dan *replace imputation* pada *k-nearest neighbor* menunjukkan *performance* yang signifikan dibandingkan dengan metode k-NN dan nilai akurasi terbaik yaitu 79.30%, *precision* yaitu 83.04% sedangkan recall dengan nilai 80.93% hasil tersebut di peroleh dari hasil uji kinerja metode bagging dan *replace imputation* pada algoritma *k-nearest neighbor* dengan penanganan *missing value*.

DAFTAR PUSTAKA

Abellán, J. (2013). Ensembles of decision trees based on imprecise probabilities and uncertainty measures. *Information Fusion*, *14*(4), 423–430. https://doi.org/10.1016/j.inffus.2012.03.003

Fayed, H. A., & Atiya, A. F. (2009). A novel template reduction approach for the *K-nearest neighbor* method. *IEEE Transactions on Neural Networks*, *20*(5), 890–896. https://doi.org/10.1109/TNN.2009.2018547

Gorunescu, F. (2011). *Data mining: concepts and techniques*. *Chemistry &* https://doi.org/10.1007/978-3-642-19721-5

Han, J., & Kamber, M. (2012). *Data Mining: Concepts and Techniques (3rd ed)*.

DOI: https://doi.org/10.35891/agx.v10i2.1636

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.

Indonesia, S. N. (1998). SNI 01-4483-1998 Jagung bahan baku pakan Badan Standardisasi Nasional - BSN.

Munanda, E., & Nanang, P. (2010). Perancangan Sistem Pakar untuk Mendiagnosa Penyakit Tanaman jagung menggunakan Fuzzy MCDM berbasis Web. *Litek*, 10, 113-117..

Munarto, R., Permata, E., & Salsabilla, R. (2014), *Klasifikasi Kualitas Biji Jagung Manis Berdasarkan Fitur Warna menggunakan Fuzzy Logic*. Prosiding Simposium Nasional Rekayasa Aplikasi Perancangan dan Industri. Universitas Muhammadiyah Surakarta.

Effendy, F., & Purbandini, P. (2018). Klasifikasi Rumah Tangga Miskin Menggunakan Ordinal Class Classifier. *Jurnal Nasional Teknologi dan Sistem Informasi*, *4*(1), 30-36.

Neo, T. K. C., & Ventura, D. (2012). A direct boosting algorithm for the *k-nearest neighbor* classifier via local warping of the distance metric. *Pattern Recognition Letters*, *33*(1), 92–102. https://doi.org/10.1016/j.patrec.2011.09.028

Setiawan, T. A., Wahono, R. S., & Syukur, A. (2015). Integrasi metode sample bootstrapping dan weighted principal component analysis untuk meningkatkan performa *K-nearest neighbor* pada dataset besar. *Journal of Intelligent Systems*, *1*(2), 76-81.

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing and Management*, *45*(4), 427–437. https://doi.org/10.1016/j.ipm.2009.03.002

Wahono, R S, & Suryana, N. (2013). Combining particle swarm optimization based feature selection and bagging technique for software defect prediction. *International Journal of Software Engineering and Its Applications*, *7*(5), 153–166. https://doi.org/10.14257/ijseia.2013.7.5.16

Wahono, R. S., Suryana, N., & Ahmad, S. (2014). Metaheuristic optimization based feature selection for software defect prediction. *Journal of Software*, *9*(5), 1324–1333. https://doi.org/10.4304/jsw.9.5.1324-1333

Wan, C. H., Lee, L. H., Rajkumar, R., & Isa, D. (2012). A hybrid text classification approach with low dependency on parameter by integrating *K-nearest neighbor* and support vector machine. *Expert Systems with Applications*, *39*(15), 11880–11888. https://doi.org/10.1016/j.eswa.2012.02.068

Xindong Wu, V. K. (2009). *The Top Ten Algorithms in Data Mining*. *Data Mining and Knowledge Discovery*.

DOI: https://doi.org/10.35891/agx.v10i2.1636