Klasifikasi untuk Prediksi Cuaca Menggunakan Esemble Learning

Amril Mutoi Siregar¹; Tukino²; Sutan Faisal³; Ahmad Fauzi⁴; Ilman kadori⁵

1, 2, 3, 4 Fakultas Teknologi dan Ilmu Komputer, Universitas Buana Perjuangan Karawang;

Sekolah Tinggi Manajemen Informatika dan Komputer MIC Cikarang

1 amrilmutoi@ubpkarawang.ac.id

2 tukino@ubpkarawang.ac.id

3 sutan.faisal@ubpkarawang.ac.id

4 afauzi@ubpkarawang.ac.id

5 ilmankadori@mic.ac.id

ABSTRACT

Knowing the weather condition is very important for human life, which is expressed with parameters such as temperature, wind speed, air pressure, rainfall, those which are considered as the main components for the weather. The purpose of knowing the weather is important, mainly for rural communities engaged in agriculture, plantation and others. Therefore, an application that can ease the people to get the weather information needs to be made so that it can be utilized by the community. The problem is on how to make an accurate and automatic weather prediction, so that everyone can operate it. This study proposes an ensemble learning to calculate the opportunities of each existing group attribute and determine which approach is the most optimal, meaning that grouping can be done based on the category that the users entered in the application. Ensemble learning involves several algorithms to find the average value of accuracy. The expected results can be a referral for building a weather accounting application. The accuracy result is 81.21%, and 18.79% for MSE.

Keywords: weather, ensemble learning, classification, prediction

ABSTRAK

Mengetahui kondisi cuaca adalah hal yang sangat penting untuk kehidupan manusia, yang dinyatakan dengan parameter seperti suhu, kecepatan angin, tekanan udara, curah hujan, yang menjadi komponen utama untuk cuaca. Tujuan mengetahui cuaca adalah penting, terutama bagi masyarakat pedesaan yang bergerak dibidang pertanian, perkebunan dan lain lain. Karena itu, sebuah aplikasi untuk mendapatkan informasi mengenai cuaca perlu dibuat sehingga dapat dimanfaatkan oleh masyarakat. Permasalahan yang muncul adalah bagaimana membuat prediksi cuaca akurat dan otomatis sehingga dapat dioperasikan oleh setiap orang. Penelitian ini mengajukan ensemble learning untuk menghitung peluang dari masing-masing atribut kelompok yang ada dan menentukan pedekatan mana yang paling optimal, artinya pengelompokan dapat dilakukan berdasarkan kategori yang pengguna masukkan pada aplikasi. Ensemble learning melibatkan beberapa algoritma untuk mencari nilai rata rata akurasi. Hasil yang diharapkan bisa menjadi rujukan untuk membangun aplikasi perakiraan cuaca. Hasil akurasi adalah 81.21% dan MSE 18.79%.

Kata kunci: cuaca, ensemble learning, klasifikasi, prediksi

1. PENDAHULUAN

Sekarang ini faktor cuaca mempegaruhi kehidupan manusia terutama disector pertanian, penerbangan dan lainnya. Cuaca sangat sulit diprediksi karena banyak faktor sebagai variable yang menentukan tentang cuaca, sering hanya melihat kondisi udara pada waktu yang relatif singkat, yang dapat diformulakan ke beberapa atribut misalnya tekanan, kecepatan angin, curah hujan, suhu, dan fenomena atmosfer sebagai komponennya. Kondisi cuaca merupakan penting dan tidak akan lepas dari kehidupan manusia. Pendekatan ini didukung dengan menggunakan teknologi yang tersedia adalah penerapan sistem cuaca penentuan saat.

Mengingat pentingnya prediksi cuaca, banyak penelitian mengenai hal tersebut semakin berkembang. Dengan itu dibutuhkan prakiraan cuaca sebagai langkah antisipasi untuk memperkecil dampak yang akan terjadi. Prediksi diharapkan memiliki keakuratan tinggi terhadap cuaca agar aktivitas manusia tidak terganggu misalnya pertanian, perkebunan, penerbangan ini sangat tergantung pada kondisi cuaca, agar kegiatan tersebut lancar. Prakiraan iklim dan cuaca adalah bagian dalam sistem informasi yang dapat digunakan untuk dapat memantau kondisi alam masa akan datang. Pada hakekatnya, sistem informasi iklim dan cuaca adalah suatu cara dapat dilakukan untuk memaksimalkan untuk memantau, hingga menjadi bentuk evaluasi atau klasifikasi dan memprediksi cuaca atau iklim. Prediksi cuaca merupakan suatu usaha dalam perkembangan kondisi udara yang lalu, sekarang, dan yang akan dating khususnya dalam kaitan mengantisipasinya. Beberapa metode yang digunakan untuk prediksi secara matematis dan statistic seperti Regressive (AR), AR-Integrated Moving Average (ARIMA) kedua metode tersebut telah digunakan untuk prakiraan suhu udara. Sekarang ini masih ada yang melakukan metode konvensional tersebut banyak didapatkan beberapa kelemahan. Dalam prakiraan dengan metode statistik, terdapat kelemahan salah satunya seperti presentasi ketidaktepatan ramalan pada saat kondisi cuaca yang sangat sulit untuk diprediksi secara matematis.

Penentuan cuaca untuk mendapatkan informasi perlu dibuat sehingga dapat dimanfaatkan oleh masyarakat. Permasalahan yang muncul bagaimana membuat prediksi cuaca denga otomatis sehingga dapat dilakukan oleh setiap orang. Dalam penelitian ini mengajukan beberapa algoritme *Navie Bayes, Decision Tree, Random forest* untuk menghitung peluang dari satu kelas dari masingmasing atribut kelompok yang ada dan menentukan kelas mana yang paling optimal, artinya pengelompokan dapat dilakukan berdasarkan kategori yang pengguna masukkan pada aplikasi. Sistem prediksi telah dibuat mendapatkan tingkat akurasi Navie Bayes sebesar 77.22% dengan standar deviasi sebesar 29 %, tingkat akurasi Decision Trees[1].

Dalam menentukan cuaca merupakan proses penerapan ilmu pengetahuan untuk mendapatkan keadaan atmosfer bumi. Proses penentuan cuaca menggunakan metode, untuk medapatkan hasil yang baik. Penggunaan metode yang baik, sekarang ini menjadi suatu kegiatan yang sering dilakukan penelitian cuaca atau atmosfer. Untuk mendapatkan hasil yang lebih baik dan lebih akurat, peneliti didorong untuk terus menerus meneliti atmosfer atau alam yang berhubungan dengan parameter cuaca dan mengembangkan banyak metode untuk memenuhi kebutuhan informasi tentang cuaca atau kondisi atmosfer yang ada.

Metode pengklasifikan probabilistic sederhana seperti Naïve bayes classifier, dapat diandalkan, dan mudah digunakan, efesien, efektif, dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset. Naive Bayes dikemukakan oleh ilmuan inggis Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya[2].

Menggunakan *evaluation folds* untuk membandingkan *ensembles learning* terhadap komponen model, dan melawan BRTs. Menggunakan daerah di bawah receiver-Operating Curve

karakteristik (AUC) dan *log-likelihood* untuk menilai kinerja model. Hasil ensemble model dengan baik, tetapi tidak konsisten lebih baik daripada komponen untuned model individu atau BRTs[3].

Menggunakan metode dengan *evolutionary algorithm* dengan memperkenalkan *Salp Swarm algoritma* (SSA) untuk optimasi bobot. Metode yang diusulkan mengungguli semua metode lain dan metode umum Ensemble. Hasilnya dibandingkan dengan pengukuran dengan menggunakan indeks kinerja perwakilan dan menunjukkan akurasi yang memuaskan[4].

Ensemble learning sangat populer dalam dapat meningkatkan akurasi dan diterapkan dengan menggunakan tiga base classifiers. Hasil model terbaik dengan akurasi tertinggi adalah *bagging* dengan *base classifier decision tree algorithm* (95,312%). Percobaan menunjukkan bahwa Ensemble classifier model dilakukan lebih baik daripada Base yang hanya klasifikasi[5].

Paper yang berjudul A Weighted Voting Ensemble Self-Labeled Algorithm for the Detection of Lung Abnormalities from X-Rays " dan itu ditulis oleh Livieris et al.[6].

Paper selanjutnya dengan judul Exploring an Ensemble of Methods that Combines Fuzzy Cognitive Maps and Neural Networks in Solving the Time Series Prediction Problem of Gas Consumption in Greece". Paper ini menyajikan model peramalan waktu seri yang inovatif untuk prediksi permintaan konsumsi gas di Yunani[7].

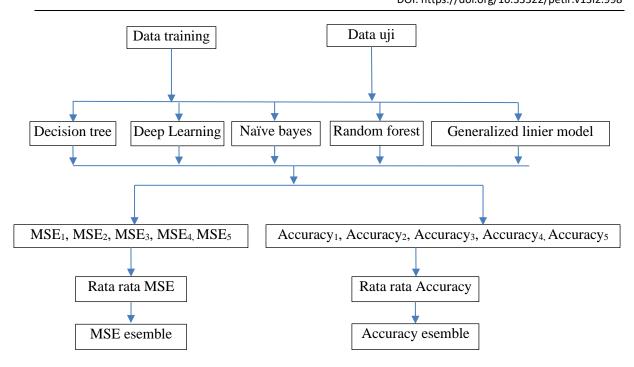
2. METODE/PERANCANGAN PENELITIAN

Metode penelitian ini menggunakan metode esemble learning yang bertujuan untuk meminimalkan kesalahan klasifikasi dengan menggunakan dataset dari Kaggle, data berhubungan dengan klasifikasi cuaca di Australia yang terdiri dari 142.193 record dan 24 attribut, dengan data selama 10 tahun. Pengamatan cuaca dilakukan harian dari berbagai stasiun cuaca Australia. Target *RainTomorrow* berarti, apakah hujan dihari berikutnya? Ya atau tidak. Metode yang diusulkan dalam penelitian ini adalah Ensemble learning dasar yang akan kita gunakan adalah Naïve bayes [1], generalized linier model, deep learning, decision tree, dan random forest [1]. Selain itu, kita juga akan menggunakan metode Bagging. Metode Bagging, yang berasal dari "bootstrap aggregation" adalah jenis lain ansambel learning. Selain itu, metode Bagging menggunakan beberapa versi metode pembelajaran dan berdasarkan metode prediksi dan memperoleh hasil yang berasal dari hasil rata rata semua metode. Dalam hal ini ide utama adalah untuk menggunakan pembelajaran yang terdiri dari Naïve bayes, generalized linier model, deep learning, decision tree dan random forest. metode membuat beberapa salinan set pelatihan dan menghasilkan pohon keputusan untuk set ini. Setiap rangkaian Pelatihan penyalinan diperoleh dengan memilih secara acak N pola data pelatihan dari N dengan penggantian, di mana N menunjukkan seluruh Dataset size.

2.1. Ensemble learning with different base classifiers

Pendekatan ini telah muncul dan sangat populer dengan keragaman anggota ansambel yang berasal dari berbagai jenis pengklasifikasi bukannya mengandalkan subset yang berbeda dari contoh yang dihasilkan oleh teknik sampling. Ketika keragaman dijamin, setiap classifier dibedakan dari satu sama lain pada kinerja prediktif dan biasanya akhir kinerja akhir lebih baik.

Metode Bagging menggunakan beberapa versi metode pembelajaran dan berdasarkan metode prediksi dan memperoleh hasil yang berasal dari hasil rata rata semua metode. Berikut metode yang yang diusulkan dalam penelitian ini:



Gambar 1. Metode yang diusulkan

2.2. Pengumpulan Data training dan data uji

Penelitian ini, memiliki beberapa tahap yang harus termasuk pengumpulan data yang akn digunakan, semua data dan atribut tidak digunakan semua, semua value atribut harus melalui beberapa tahap pengolahan data awal (data persiapan). Mendapatkan data yang berkualitas, beberapa teknik dilakukan sebagai berikut [8], berikut attribute yang gunakan dalam penelitian ini adalah: Date, Tanggal observasi, Location, MinTemp, MaxTemp, Rainfall, Evaporation, Sunshine, WindGustDir, WindGustSpeed, WindDir9am, WindDir3pm, WindSpeed9am, WindSpeed3pm, Humidity9am, Humidity3pm, Pressure9am, Pressure3pm, Cloud9am, Cloud3pm, Temp9am, Temp3pm, RainToday, RISK_MM, RainTomorrow.

2.3. Data cleaning

Kualitas data yang baik adalah kunci dasar untuk data yang menghasilkan kualitas yang baik, noise data yang masih outlier atau noise, data yang tidak lengkap dihilang nilai atribut, dan data yang tidak konsisten dalam atribut pengisian. Tahapan dalam pembersihan data adalah menghilangkan outlier, nilai-nilai yang tidak lengkap atau hilang (missing value), algoritma naif bayes memiliki keunggullan tersendiri yang dapat menangani data yang tidak lengkap atau hilang (missing value), dan data data yang tidak konsisten adalah tetap serta memecah redundansi yang disebabkan oleh interogasi data.

2.4. Data transformasi

Langkah selanjutnya adalah teknik integrasi data yang digunakan untuk menganalisis korelasi data, atribut berlebihan dan data ganda, dan transformasi digunakan untuk meningkatkan akurasi dan efisiensi algoritma. Kelebihan algoritma naif bayes mampu memproses data yang nominal, terus menerus, dan ordinal. Oleh karena itu nilai setiap atribut yang terkandung dalam dataset tidak harus diubah.

2.5. Algoritma Naive Bayes

Ini adalah teknik klasifikasi berdasarkan teorema Bayes dengan asumsi kebebasan di antara prediktor. Dalam istilah sederhana, sebuah classifier Naive Bayes mengasumsikan bahwa kehadiran fitur tertentu dalam kelas tidak terkait dengan kehadiran fitur lainnya [9],[10]. Sebagai contoh, buah dapat dianggap sebagai apel jika merah, bulat, dan sekitar 3 inci diameter. Bahkan jika fitur ini tergantung pada satu sama lain atau pada keberadaan fitur lainnya, semua properti ini secara independen berkontribusi pada probabilitas bahwa buah ini adalah apel dan itulah sebabnya ia dikenal sebagai 'naif'. Naive Bayes model mudah untuk membangun dan sangat berguna untuk set data yang sangat besar. Seiring dengan kesederhanaan, Naive Bayes dikenal mengungguli metode klasifikasi yang sangat canggih sekalipun.

Bayes teorema menyediakan cara untuk menghitung probabilitas posterior $P(c \mid x)$ dari P(c), P(x) dan $P(x \mid c)$. Lihat persamaan di bawah ini:

$$P(C \mid x) = \frac{P(x \mid C) P(C)}{P(x)}$$
 (1)

2.6. Algoritme Decision tree

Algoritma pohon keputusan satu rumpun algoritma pembelajaran diawasi. Tidak seperti algoritma belajar diawasi lainnya, algoritma pohon keputusan dapat digunakan untuk memecahkan masalah regresi dan klasifikasi juga. Tujuan penggunaan pohon keputusan adalah untuk membuat model pelatihan yang dapat digunakan untuk memprediksi kelas atau nilai variabel target dengan mempelajari aturan keputusan sederhana yang disimpulkan dari data sebelumnya data Latihan [11],[12]. Dalam keputusan pohon, untuk memprediksi sebuah label kelas untuk sebuah catatan kita mulai dari akar pohon. Kita membandingkan nilai dari atribut root dengan atribut record. Atas dasar perbandingan, mengikuti cabang yang sesuai dengan nilai itu dan melompat ke simpul berikutnya.

2.7. Algoritma Random forest

Random forest seperti namanya, terdiri dari sejumlah besar pohon keputusan individu yang beroperasi sebagai ansambel. Setiap pohon individual Random forest mengeluarkan prediksi kelas dan kelas dengan suara terbanyak menjadi prediksi model. Visualisasi dari model Random forest membuat prediksi. Konsep fundamental di balik Random forest adalah sederhana tapi kuat satu kebijaksanaan keramaian [13]. Dalam ilmu data berbicara, alasan bahwa model Random forest bekerja dengan baik adalah: Sejumlah besar model yang relatif tidak berkorelasi (pohon) yang beroperasi sebagai forest akan mengungguli salah satu model konstituen individu. Korelasi rendah antara model adalah kuncinya.

2.8. Algoritma Deep learning

Deep learning juga dikenal sebagai pembelajaran terstruktur mendalam, merupakan bagian dari sebuah keluarga yang lebih luas dari metode pembelajaran mesin berdasarkan jaringan syaraf tiruan dengan pembelajaran representasi. Belajar dapat diawasi, semi diawasi atau tanpa pengawasan [14],[15].

Arsitektur pembelajaran yang mendalam seperti jaringan syaraf jaringan kepercayaan yang mendalam, recurrent neural network dan convolutional neural network telah diterapkan pada bidang termasuk visi komputer, visi mesin, pengenalan ucapan, pengolahan bahasa alami, pengenalan audio, penyaringan jaringan sosial, terjemahan mesin, bioinformatika, desain obat, analisis gambar medis,

DOI: https://doi.org/10.33322/petir.v13i2.998

inspeksi bahan dan program papan permainan, di mana mereka telah menghasilkan hasil yang sebanding dengan dan dalam beberapa kasus melebihi kinerja ahli manusia [15][16].

Jaringan syaraf tiruan (ANNs) terinspirasi oleh pemrosesan informasi dan node komunikasi terdistribusi dalam sistem biologis. ANNs memiliki berbagai perbedaan dari otak biologis. Secara khusus, jaringan saraf cenderung statis dan simbolis, sementara otak biologis kebanyakan organisme hidup dinamis dan analog [17],[18].

2.9. Algoritma model linear generalisasi

Dalam statistik, model linear generalisasi (GLM) adalah generalisasi yang fleksibel dari regresi linear biasa yang memungkinkan untuk variabel respon yang memiliki kesalahan distribusi model selain distribusi normal. The GLM generalizes regresi linier dengan memungkinkan model linier yang akan berhubungan dengan variabel respon melalui fungsi link dan dengan memungkinkan besarnya varians dari setiap pengukuran menjadi fungsi dari nilai diprediksi. Model linier Generalized diformulasikan oleh John Nelder dan Robert Wedderburn sebagai cara untuk menyatukan berbagai model Statistik lainnya, termasuk regresi linear, regresi logistik dan regresi Poisson [19]. Mengusulkan metode kuadrat setidaknya reweighted untuk estimasi kemungkinan maksimum dari parameter model. Estimasi kemungkinan maksimum tetap populer dan merupakan metode default pada banyak paket komputasi Statistik. Pendekatan lain, termasuk pendekatan Bayesian dan kuadrat paling cocok untuk varians stabil tanggapan, telah dikembangkan.

2.10. Evaluasi model

Accuracy

Evaluasi klasifikasi biner membandingkan dua metode menetapkan atribut biner, salah satunya biasanya metode standar dan yang lainnya sedang diselidiki. Ada banyak metrik yang dapat digunakan untuk mengukur kinerja classifier atau predictor, bidang yang berbeda memiliki preferensi yang berbeda untuk metrik tertentu karena tujuan yang berbeda [20].

Confusion matriks adalah hasil prediksi pada masalah klasifikasi. Jumlah prediksi yang benar dan salah diringkas dengan menghitung nilai dan dicapai masing-masing kelas. Ini memberi kita wawasan tidak hanya ke dalam kesalahan yang dibuat oleh classifier tetapi yang lebih penting jenis kesalahan yang dibuat.

Tabel 1. Confusion Matrix

	Class 1: Positif	Class 2: Negatif
Class 1: Positif	TP	FN
Class 2: Negatif	FP	TN

Keterangan:

Class 1 = Positif

Class 2 = Negatif

True positive (TP), False negative (FN), True negative (TN), False positive (FP). Formula untuk menghitung accuracy sebagai berikut:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

Means square erorr MSE

Dalam statistik, mean kuadrat error (MSE) atau mean kuadrat deviasi (MSD) dari sebuah pengukur (dari sebuah prosedur untuk memperkirakan sebuah kuantitas yang tidak diamati)

mengukur rerata dari kuadrat dari kesalahan-yaitu, rata kuadrat perbedaan antara perkiraan nilai dan nilai aktual. MSE adalah fungsi risiko, sesuai dengan nilai yang diharapkan dari kesalahan kuadrat kerugian. Fakta bahwa MSE hampir selalu sangat positif (dan bukan nol) adalah karena keacak atau karena pengukur tidak memperhitungkan informasi yang dapat menghasilkan perkiraan yang lebih akurat.

$$MSE = \frac{1}{2} \sum_{i=1}^{N} (fi - yi)^{2}$$
Dimana N adalah jumlah data
$$fi = nilai \ returnd \ oleh \ model$$

$$yi = nilai \ actual \ pada \ data \ i$$
(3)

3. HASIL DAN PEMBAHASAN

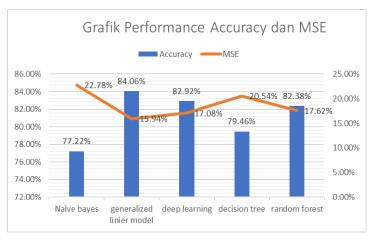
3.1. Hasil Penelitian

Perubahan iklim dan cuaca merupakan permasalah yang dihadapi hampir seluruh dunia yang meng klasifikasi dan memprediksi. Permasalahan banyak variable yang mempegaruhi sehingga cukup sulit dan tak terduga. Perubahan iklim dan cuaca adalah pemanasan global yang dilakukan manusia yang membuat bertambah sulit memecahkan masalah cuaca. Hasil dari penelitian ini dapat jadikan untuk mengukur tingkat akurasi dan MSE terhadap cuaca. Setelah ini dapat dibuat aplikasi prediksi atau klasifikasi cuaca, dengan melihat tingkat akurasi yang paling tinggi dalam penelitian ini. Berikut ini adalah tabel dari hasil akurasi dan MSE dari berbagai algoritma.

Metode	Accuracy	MSE	
Naïve bayes	77.22%	22.78%	
Decision tree	79.46%	20.54%	
Random forest	82.38%	17.62%	
Deep learning	82.92%	17.08%	
Generalized linier model	84.06%	15.94%	

Tabel 2. Performance Accuracy Algoritme Klasifikasi

Tabel 2 memperlihatkan performance semua metode yang diukur berupa akurasi dan MSE, dari tabel diatas akurasi tertinggi algoritme GLM (Generalized linier model) adalah 84.06% dan MSE sebesar 17.08%, disusul Deep learning sebesar 82.92% dan MSE 17.62%. untuk mempermudah membaca performance, dapat dilihat bentuk grafik dibawah ini:



Gambar 2. Grafik Performance Accuracy dan MSE

Tabel 3 adalah yang paling tertinggi diantara beberapa metode metode yang lain. Hasil penelitian ini dapat algoritme mengungguli algoritme lainya, berikut akurasi dan MSE yang paling tinggi dan MSE terendah seperti tabel dibawah ini:

Tabel 3. Hasil penelitian metode esemble learning

Accuracy	81.21%
MSE	18.79%

3.2. Pembahasan Penelitian

Hasil penelitian yang telah dilakukan, menggunakan model evaluasi confusion matrix untuk menghitung accuracy dan MSE untuk menghitung classfication erorr. Accuracy adalah yang paling popular menghitung keberhasilan algoritme dalam menyelesaikan masalah, karena sebelum membuat aplikasi prediksi, sebaiknya harus mengukur kinerja algoritme yang akan digunakan, seperti dalam penelitian ini dengan membandingkan lima algoritme dan digabungkan lagi semua algoritme yang dinamakan esemble metode dengan bagging, hasilnya dilihat dari berbagai bahasan adalah akurasi yang didapat adalah 81.21% dengan MSE 18.79%. berarti hasil esambel pada penelitian ini adalah lebih rendah dibanding dengan performance dari GLM (Generalized linier model), ini bisa terjadi tipe dataset yang digunakan, karena algoritme yang digunakan bermacam macam karakteristik, misalnya kelompok decision tree lebih rendah hasil akurasinya, dan deep learning diatas rata rata esambel learning, naïve bayes dibawah rata rata esambel learning.

4. KESIMPULAN DAN SARAN

4.1. Kesimpulan

Berdasarkan hasil Penelitian ini dengan pendekatan esambel learning, dapat diambil beberapa kesimpulan yaitu hasil akurasi dan MSE dari metode esemble learning adalah 81.21% sebagai akurasi dan 18.79% untuk MSE. Untuk rumput decision tree yang terbaik adalah random forest sebesar 82.38% untuk akurasi dan 17.62% adalah MSE. Algoritme deep learning memiliki performance adalah 82.92% sebagai akurasi dan 17.08% untuk MSE. Performance algoritme tertinggi diantar algoritme yang lainnya yaitu 84.06% sebagai akurasi dan 15.94% untuk MSE.

4.2. Saran

Peneliti dapat memberikan memberikan saran berdasarkan hasil penelitian ini, kedepannya bisa menggunakan esemble learning dengan metode stacking dan boosting karena penelitian sekarang ini adalah menggunakan Bagging. Dan penelitian berikut bisa menggunkan algoritme yang yang lain seperti dengan metode yang sama sama classifier.

UCAPAN TERIMAKASIH

Penulis mengucapkan terima kasih kepada teman teman yang mendukung, yang telah memberikan dukungan moril kepada penulis sehingga penelitian dapat dilakukan sampai selesai.

DAFTAR PUSTAKA

[1] Siregar Amril M; Faisal Sutan; Cahyana Yana; Priyatna Bayu, "Perbandingan Algoritme Klasifikasi Untuk Prediksi Cuaca," AIMS, vol. 3, no. 1, 2020, doi: https://doi.org/10.32627/aims.v3i1.92.

- [2] A. Saleh, "Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga," Citec Journal, vol. 2, no. 3, p. 207, 2015, [Online]. Available: https://citec.amikom.ac.id/main/index.php/citec/article/view/49.
- [3] T. Hao, J. Elith, J. J. Lahoz-Monfort, and G. Guillera-Arroita, "Testing whether ensemble modelling is advantageous for maximising predictive performance of species distribution models," Ecography (Cop.)., vol. 43, no. 4, pp. 549–558, Apr. 2020, doi: 10.1111/ecog.04890.
- [4] S. K. Goudos and G. Athanasiadou, "Application of an Ensemble Method to UAV Power Modeling for Cellular Communications," IEEE Antennas Wirel. Propag. Lett., vol. 18, no. 11, pp. 2340–2344, Nov. 2019, doi: 10.1109/LAWP.2019.2926784.
- [5] N. Nai-Arun and P. Sittidech, "Ensemble Learning Model for Diabetes Classification," Adv. Mater. Res., vol. 931–932, pp. 1427–1431, May 2014, doi: 10.4028/www.scientific.net/AMR.931-932.1427.
- [6] P. Livieris, I.E.; Kanavos, A.; Tampakas, V.; Pintelas, "A weighted voting ensemble self-labeled algorithm for the detection of lung abnormalities from X-rays," Algorithms, vol. 12, no. 64, 2019.
- [7] G. Papageorgiou, K.I.; Poczeta, K.; Papageorgiou, E.; Gerogiannis, V.C.; Stamoulis, "Exploring an Ensemble of Methods that Combines Fuzzy Cognitive Maps and Neural Networks in Solving the Time Series Prediction Problem of Gas Consumption in Greece," Algorithms, vol. 12, no. 235, 2019.
- [8] V. . Tharun, R. Prakash, and S. R. Devi, "Prediction of Rainfall Using Data Mining Techniques," in 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), Apr. 2018, pp. 1507–1512, doi: 10.1109/ICICCT.2018.8473177.
- [9] M. E. Maron, "Automatic Indexing: An Experimental Inquiry," J. ACM, vol. 8, no. 3, pp. 404–417, Jul. 1961, doi: 10.1145/321075.321084.
- [10] D. J. Hand and K. Yu, "Idiot's Bayes: Not So Stupid after All?," Int. Stat. Rev. / Rev. Int. Stat., vol. 69, no. 3, p. 385, Dec. 2001, doi: 10.2307/1403452.
- [11] X. Wu et al., "Top 10 algorithms in data mining," Knowl. Inf. Syst., vol. 14, no. 1, pp. 1–37, Jan. 2008, doi: 10.1007/s10115-007-0114-2.
- [12] S. M. Piryonesi and T. E. El-Diraby, "Data Analytics in Asset Management: Cost-Effective Prediction of the Pavement Condition Index," J. Infrastruct. Syst., vol. 26, no. 1, p. 04019036, Mar. 2020, doi: 10.1061/(ASCE)IS.1943-555X.0000512.
- [13] Tin Kam Ho, "The random subspace method for constructing decision forests," IEEE Trans. Pattern Anal. Mach. Intell., vol. 20, no. 8, pp. 832–844, 1998, doi: 10.1109/34.709601.
- [14] Y. Bengio, A. Courville, and P. Vincent, "Representation Learning: A Review and New Perspectives," IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 8, pp. 1798–1828, Aug. 2013, doi: 10.1109/TPAMI.2013.50.
- [15] D. Ciresan, U. Meier, and J. Schmidhuber, "Multi-column deep neural networks for image classification," in 2012 IEEE Conference on Computer Vision and Pattern Recognition, Jun. 2012, pp. 3642–3649, doi: 10.1109/CVPR.2012.6248110.
- [16] J. Schmidhuber, "Deep learning in neural networks: An overview," Neural Networks, vol. 61, pp. 85–117, Jan. 2015, doi: 10.1016/j.neunet.2014.09.003.
- [17] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," Nature, vol. 381, no. 6583, pp. 607–609, Jun. 1996, doi: 10.1038/381607a0.

- DOI: https://doi.org/10.33322/petir.v13i2.998
- [18] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.
- [19] J. A. Nelder and R. W. M. Wedderburn, "Generalized Linear Models," J. R. Stat. Soc. Ser. A, vol. 135, no. 3, p. 370, 1972, doi: 10.2307/2344614.
- [20] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," BMC Genomics, vol. 21, no. 1, p. 6, Dec. 2020, doi: 10.1186/s12864-019-6413-7.