EVALUASI MODEL PREDIKSI PRODUKTIVITAS JAGUNG DI INDONESIA MENGGUNAKAN ALGORITMA PEMBELAJARAN MESIN

Ferdinand Murni Hamundu[1]*, Gusti Arviana Rahman[2], Andi Tenriawaru[3], Rashid Armin[4],
[1, 3, 4]Program Studi Ilmu Komputer, FMIPA, Universitas Halu Oleo
[2]Program Studi Statistika, FMIPA, Universitas Halu Oleo

*[1]ferdinand@uho.ac.id,[2]arviana.rahman@uho.ac.id, [3]andi.tenriawaru@uho.ac.id,
[4]rashidarmin@gmail.com

Ketahanan pangan merupakan isu global yang mempengaruhi banyak negara berkembang. Jagung adalah salah satu tanaman pangan terpenting di dunia setelah padi dan gandum. Pada penelitian ini telah diterapkan teknik pembelajaran mesin untuk memprediksi data peramalam produktivitas jagung yang dapat mendukung ketahanan pangan. Algoritma yang digunakan adalah Random Forest, Boosting, dan Bagging. Penelitian ini mengevaluasi beberapa model dengan akurasi sampel. Hasilnya adalah Random forest lebih baik daripada metode yang lain berdasarkan tingkat kesalahan terendah. Hal ini ditunjukkan dengan nilai validitasnya yang paling minimum seperti MSE (6.764), MAPE (9.545), SSE (87570.9), dan R-square (0.8327575). Oleh karena itu, Random Forest dapat diandalkan untuk menyelidiki keakuratan data berkaitan dengan prediksi produktivitas jagung.

***Kata Kunci*— Pembelajaran Mesin, Ketahanan Pangan, Produktivitas Jagung, Random Forest, Boosting, Bagging**

I. PENDAHULUAN

Pertanian memainkan peran penting dalam aktivitas manusia. Tantangan signifikan seperti meningkatnya jumlah penduduk berdampak pada sumber daya yang menimbulkan ancaman dan persaingan bagi ketahanan pangan. Untuk mengatasi masalah kompleks dan terus meningkat di bidang pertanian khususnya ketahanan pangan, para peneliti menawarkan solusi pembelajaran mesin untuk menjadikan pertanian yang berkelanjutan [1].

Ketahanan pangan ini menjadi lebih penting karena makanan bukan hanya kebutuhan dasar tetapi juga hak dasar bagi setiap manusia yang harus dipenuhi [2]. Konsep ketahanan pangan telah berkembang selama seperempat abad terakhir. Konsep ketahanan pangan telah dipertimbangkan pada sejumlah tingkatan: global, regional, nasional, negara bagian, rumah tangga dan individu [3]. Perhatian terhadap ketahanan pangan didominasi dalam beberapa tahun terakhir, baik dari akademisi dan non-akademisi [4].

Algoritma pembelajaran mesin dapat diadopsi yang mengambil informasi secara otomatis menggunakan model statistik atau komputasi dan sangat membantu untuk menemukan faktor atau variabel secara akurat dan untuk meningkatkan kinerja. Algoritma ini tergolong baru sehingga masih berkembang dengan kecepatan yang semakin cepat. Pembelajarn mesin gabungan dari ilmu komputer, *theory* statistik, kecerdasan buatan, dan sains data.

Analitik data memegang peranan penting untuk memastikan ketahanan pangan serta keberlanjutan ekologi di masa depan. Analitik data berkaitan erat dengan pembelajaran mesin, *smart farming* dan *big data* yang berfungsi untuk memodelkan pertanian. Algoritma pembelajaran mesin yang menggunakan data-data pertanian, misalnya, untuk memprediksi hasil pertanian [5]. Output pembelajaran mesin penting dalam memberikan nilai tambah pada ketahanan pangan [6].

Beberapa literatur menyatakan bahwa jagung dapat mendukung ketahanan pangan. Jagung adalah salah satu tanaman pangan terpenting di dunia bersama beras dan gandum, menyediakan setidaknya 30% kalori makanan untuk lebih dari 4,5 miliar orang di 94 negara berkembang. Di beberapa bagian Asia, Afrika dan Amerika Latin, jagung menyumbang lebih dari 20% kalori makanan. Jagung juga merupakan bahan utama dalam pakan ternak dan digunakan secara luas dalam produk industri, termasuk produksi biofuel [7].

Algoritma pembelajaran mesin menjadi alat akurat untuk menganalisa data yang bersifat komplek dan besar, serta berhasil membantu para ilmuwan di bidang sains dan teknologi [8]. Pembelajaran mesin menjadi topik penting untuk penelitian yang bergerak dibidang sains data . Pembelajaran mesin merupakan cabang dari kecerdasan buatan (artificial intelligent), teknik pembelajaran mesin bekerja berdasarkan data empiris [9]. Pembelajaran mesin dirancang untuk memperoleh pengetahuan dari data yang ada [10]. Pembelajaran mesin merupakan gabungan dari

lintas disiplin ilmu seperti *probability theory, statistics, pattern recognition, cognitive science, data mining, adaptive control, neuroscience,* and *theoretical computer science* [11].

Berdasarkan teori-teori yang telah disebutkan bahwa pembelajaran mesin menjadi semakin dapat diandalkan dalam memprediksi data-data yang berkaitan dengan pertanian [12]. Penulis berupaya untuk memprediksi data komoditas pertanian jagung, hal ini berupaya untuk mewujudkan ketahanan pangan dan pertanian berkelanjutan. Pembelajarn mesin yang dipakai dalam penelitian ini adalah *Random Forest, Bagging,* dan *Boosting*.

II. METODE PENELITIAN

A. Fungsi Regresi

Fungsi regresi mempunyai himpunan L dengan N observasi, variable bebas dan variabel tidak bebas seperti pada himpunan $(x_1, y_1),\ldots, (x_i, y_i),\ldots,(x_N, y_N)$ dimana $\boldsymbol{x}_i \in \mathbb{X}$ and $\boldsymbol{y}_i \in \mathbb{Y}$. Fungsi $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ik}, \ldots, x_{iK})$ merepresentasikan fungsi vector yang berisi *K* variabel bebas pada i observasi, dengan y_i. Pada framework ini, supervised learning dapat dinyatakan sebagai fungsi theta $\varphi: \mathbb{X} \to \mathbb{Y}$ dari sebuah himpunan fungsi $\mathcal{L} = (\boldsymbol{x}, \boldsymbol{y})$. Tujuannya adalah untuk menemukan model yang prediksinya $\varphi(\boldsymbol{x})$, juga dilambangkan dengan variabel $\hat{Y}$. Jika *Y* adalah variabel numerik (data kontinu), maka fungsi ini disebut fungsi regresi. Fungsi regresi dapat direpresentasikan $\varphi: \mathbb{X} \to \mathbb{Y}$, dimana $\mathbb{Y} = \mathbb{R}$ [8].

B. Algoritma Pembelajaran Mesin

Algoritme pembelajaran mesin yang diaplikasikan pada penelitian ini adalah *Random Forest, Bagging, dan Boosting*.

Random forest merupakan fungsi ensemble. Algoritma ini bertujuan untuk mengurangi kesalahan, variansi, dan overfitting, maka algoritma *random forest* bekerja membuat sub ruang secara acak dalam memilih sampel. *Random forest* mempunyai N observasi, dengan M variabel serta pemilihan sampel acak m variabel. *Random forest* akan bekerja dengan prosedur m << M pada setiap simpul, m dipilih secara acak pada M untuk mendapatkan split terbaik, dimana nilai m selalu konstan [13].

Algoritma:

Untuk $b = 1$ ke n

1. Membuat contoh *bootstrapped* D_b^* dari set data training D.
2. Kembangkan *tree* menggunakan *m* dari contoh bootstrapped D_b^*.

Untuk mode tertentu

i. Pilih m variabel secara acak.
ii. Tentukan variabel dan nilai split terbaik.
iii. Pisahkan node menggunakan variabel dan nilai split terbaik

Ulangi langgkah i-iii hinggal sampai kriteria berhenti terpenuhi [14].

Boosting bertujuan untuk meningkatkan akurasi model. Boosting didasarkan pada ide untuk menemukan rata-rata setiap perhitungan berdasarkan algoritmenya. Pada kajian ini akan menggunakan algoritma *least squares boosting* (LSB(ε)) [15]. Algoritma LSB(ε) diekspresikan sebagai berikut:

Tetapkan nilai $\varepsilon > 0$ dan jumlah iterasi sebanyak M.

Definisikan $\hat{r}^0 = \boldsymbol{y},\ \hat{\beta}^0 = 0,\ k = 0.$

1. Lakukan untuk $0 \le k \le M$
2. Tentukan *covariates* j_k dan $\tilde{u}_{j_k}$ sebagai berikut:

$$\tilde{u}_m = \underset{u \in \mathbb{R}}{arg\min} \left(\sum_{i=1}^{n} \left(\hat{r}_i^k - x_{im} u \right)^2 \right)$$

$$j_k \in \underset{1 \le m \le p}{\operatorname{argmin}} \sum_{i=1}^{n} \left(\hat{r}_i^k - x_{im} \tilde{u}_m \right)^2$$

for $m = 1, \ldots, p,$ (1)

3. Perbaharui nilai error dan koefisien regresi sebagai berikut:

$$\hat{r}^{k+1} \leftarrow \hat{r}^k - \varepsilon \tilde{u}_{j_k}$$

$$\hat{\beta}_{j_k}^{k+1} \leftarrow \hat{\beta}_{j_k}^k + \varepsilon \tilde{u}_{j_k} \quad \text{and}$$

$$\hat{\beta}_j^{k+1} \leftarrow \hat{\beta}_j^k, j \ne j_k \quad (2)$$

Opitz and Maclin [16] mendefinisikan bagging adalah "bootstrap" metode ensemble yang menciptakan individu untuk ensembelnya dengan melatih setiap *classifier* pada redistribusi acak dari training set. Sama seperti boosting, teknik bagging meningkatkan akurasi *classifier* dengan menghasilkan model komposit yang menggabungkan beberapa *classifier* yang semuanya berasal dari inducer yang sama. Berbeda dengan boosting, dalam banging instance dipilih dengan probabilitas yang sama. Bagging sangat bermanfaat untuk big data. Big data merupakan keterbaruan dari pengumpulan data dalam teknologi komputer. Karena tantangan kompleksitas dan skala pada big data, Bagging adalah algoritma intensif komputasi yang relatif baru dan efektif untuk meningkatkan akurasi regresi. Oleh karena itu, $\emptyset_A$ bergantung pada x, dan mempunyai fungsi distribusi *P*, dengan $\mathcal{L}$ yang terpilih, contoh $\emptyset_A(\boldsymbol{x}) = \emptyset_A(\boldsymbol{x}, P)$. Maka fungsi bagging direpresentasikan sebagai $\varphi_B(\boldsymbol{x}) = \varphi_A(\boldsymbol{x}, P_{\mathcal{L}})$ [17].

Algoritma Bagging sebagai berikut.

i. Membuat contoh *bootstrap* $L_i^* = (Y_i^*, X_i^*)\ (i = 1, \ldots, n)$ berdasarkan distribusi empiris pasangan $L_i = (Y_i, X_i)\ (i = 1, \ldots, n)$.
ii. Gunakan prinsip plug-in untuk menentukan prediktor bootstrapped $\hat{\theta}_n^*(x)$; dengan formula, $\hat{\theta}_n^*(x) = h_n(L_1, \ldots, L_n)(x)$.
iii. θ^_(n;B) (x)=E^* [θ^_n^* (x)] adalah *bagged predictor.*

C. Evaluasi Model

Prosedur dalam menentukan model yang terbaik, terdiri dari beberapa fase sebagai berikut:

Fase I – All Possible Model

$$N = \sum_{j=1}^{k} j\left(C_j^k\right) \quad (3)$$

N adalah jumlah possible yang mungkin, k adalah jumlah variable bebas. Jumlah observasi dengan data sekunder berdasarkan Badan Pusat Statistika dari tahun 1993 sampai dengan 2020 sebagaimana Tabel 1.

Table 1 Produksi dan Luas Lahan Jagung

Tahun	Produksi (Ton)	Luas (Ha)
1993	6,355,214	2,881,466
1994	6,355,214	2,881,466
1995	6,752,146	3,047,378
1996	8,142,863	3,595,700
1997	9,200,807	3,685,459
1998	8,671,647	3,301,795
1999	10,110,557	3,815,919
2000	9,204,036	3,456,357
2001	9,676,899	3,500,318
2002	9,347,192	3,285,866
2003	9,654,105	3,126,833
2004	10,886,442	3,358,511
2005	11,225,243	3,356,914
2006	12,523,894	3,625,987
2007	11,609,463	3,345,805
2008	13,287,527	3,630,324
2009	16,317,252	4,001,724
2010	17,629,748	4,160,659
2011	18,327,636	4,131,676
2012	17,643,250	3,864,692
2013	19,387,022	3,957,595
2014	18,511,853	3,821,504
2015	19,008,426	3,837,019
2016	19,612,435	3,787,367
2017	23,578,413	4,444,369
2018	28,924,015	5,533,169
2019	28,608,770	5,734,326
2020	29,927,856	5,160,000

Fase II – Selected Variable

Dari variabel bebas dan observasi akan diseleksi berdasarkan perspektif Random Forest, Boosting, dan Bagging.

Fase III – Goodness Fit

Goodness Fit dilakukan pada model akhir yang dipilih untuk memeriksa efisiensinya. Data residual akan dikumpulkan dengan mempertimbangkan perbedaan nilai nyata dan yang diharapkan untuk model terbaik. Data residual akan mempertimbangkan antara perbedaan nilai nyata dan nilai yang diharapkan. Langkah tersebut bertujuan untuk mendapatkan Sum of Square Error (SSE), R – Squared, Mean Square Error (MSE), Mean Average Percentage Error (MAPE).

Untuk mengevaluasi algoritma seperti Random Forest, Bagging, dan Boosting digunakan R-squared, MSE, SSE, dan MAPE sebagaimana formula di bawah. SSE, MSE, MAPE, dan R-square (R^2) untuk mengukur perbedaan antara data observasi dan data prediksi yang dihasilkan oleh algoritma tersebut [18].

$$SSE = \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)^2 \quad (4)$$

$$MAPE = \frac{100}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (5)$$

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \quad (6)$$

$$R^2 = \frac{SSR}{SST} = \frac{SST - SSE}{SST} = 1 - \frac{SSE}{SST} \quad (7)$$

Pada formula, y_i merepresentasikan observasi yang sesungguhnya, $\hat{y}_i$ merupakan nilai prediksi, dan n adalah jumlah sampel. Metrik validasi berfungsi sebagai verifikasi apakah algoritma tersebut bekerja secara optimal dalam memprediksi variabel target, atau dengan kata lain bahwa SSE, MSE, MAPE, dan R-square untuk mengukur tingkat kesalahan antara data prediksi dan aktual [18]. SSE, MSE, MAPE, dan R-square digunakan untuk menjelaskan seberapa baik model regressed terhadap data model. Semakin rendah nilai MAPE, MSE dan SSE, semakin tinggi akurat prediksinya [19]. Secara umum, model regresi dengan indeks validasi (MAPE, MSE dan SSE) yang lebih rendah dapat menjelaskan data dengan lebih baik, sedangkan indeks validasi yang lebih tinggi menggambarkan data yang diamati dengan buruk. Kesimpulannya, menghasilkan data yang paling relevan dan memberikan data kesalahan terendah dalam model validasi.

Penelitian ini memanfaatkan R-Studio yang didalamnya lengkap dengan tools yang dapat digunakan untuk melakukan komputasi pembelajaran mesin dengan bahasa pemrograman R

III. HASIL DAN PEMBAHASAN

Metrik validasi seperti MAPE, MSE, SSE, dan R-square diimplementasikan untuk Random Forest, Bagging, dan Boosting. Evaluasi dari metrik validasi diperlukan untuk menilai akurasi dari algoritma tersebut. Metrik validasi berperan penting untuk memverifikasi apakah model tersebut memadai, dan untuk memprediksi dengan benar variabel pada target (variable Y) dalam kisaran akurasi yang rasional. Tabel 2 merangkum hasil prediksi pada algoritma pembelajaran mesin.

Tabel 2 Hasil Validasi Algoritma

ML	MAPE	MSE	R-square	SSE
RF	**9.545**	**6.764**	**0.8327575**	**87570.90**
Bagging	10.499	8.708	0.8355307	89118.80
Boosting	10.417	7.103	0.8555816	96564.47

Tabel 2 menunjukkan bahwa *Random Forest* secara signifikan menunjukkan hasil yang lebih baik daripada yang lain. Kinerja algoritma yang diusulkan dievaluasi secara kuantitatif menggunakan MAPE, MSE, R-Squared, dan SSE. Validasi model, termasuk SSE, MAPE, dan MSE diterapkan untuk mengevaluasi kinerja model. SSE, MAPE, dan MSE mengukur perbedaan antara data observasi dan model estimasi. Secara umum, validasi model yang lebih rendah menunjukkan bahwa model regresi dapat menjelaskan data dengan lebih baik, sedangkan validasi model yang lebih tinggi mengungkapkan bahwa model tersebut kurang menjelaskan data yang diamati.

Random forest juga menunjukkan data kesalahan paling sedikit yang menyediakan data paling relevan dan akurat dalam model regresi. Menurut banyak ahli, beberapa penelitian sebelumnya menunjukkan kecenderungan serupa, dengan hasil bahwa Random Forest akurat untuk masalah regresi dalam data besar atau dimensi ultra-tinggi. Dalam praktiknya, algoritma ini memiliki kemampuan prediksi yang baik dan juga memberikan beberapa ukuran pentingnya variabel sehubungan dengan prediksi variabel hasil.

Teknik pembelajaran mesin penting untuk mengembangkan sistem pertanian presisi. Sekarang, pemodelan matematika telah diusulkan untuk mempromosikan modernisasi pertanian untuk meningkatkan keberlanjutan dan ketahanan pangan secara signifikan. Pembelajaran mesin dan pembelajaran Statistik digunakan untuk menganalisis data pertanian untuk membuat keputusan yang cerdas [20]. Teknik ini meningkatkan pertanian berkelanjutan dan ketahanan pangan dengan menjadikannya lebih andal, mampu, dan membantu meningkatkan produktivitas. Random forest adalah bagian dari pembelajaran mesin dan pembelajaran statistik.

Tujuan Random Forest adalah untuk mengurangi dimensi [21]. Sederhana untuk diterapkan, memberikan prediksi yang akurat, dan dapat menangani sejumlah besar variabel tanpa overfitting [22]. Algoritma ini sangat cocok untuk kumpulan data sedang hingga besar atau data berdimensi sangat tinggi. Penelitian ini menarik dalam pembelajaran mesin karena menghasilkan model regresi yang akurat. Random Forest diterima secara luas untuk memilih variabel dalam data besar karena biasanya lebih membantu daripada yang lain [23].

IV. KESIMPULAN DAN SARAN

A. Kesimpulan

Penelitian ini mengevaluasi metrik kinerja berbagai algoritma pembelajaran mesin secara kuantitatif menggunakan MAPE, MSE, R-square, dan SSE. Semua ukuran model validasi menunjukkan bahwa hasil yang lebih baik diperoleh ialah algoritma Random Forest dengan nilai MAPE (9.545), MSE (6.764), R-square (0.8328), dan SSE (87,570.9).

Pertanian presisi adalah solusi untuk mengatasi tantangan petani dan industri dalam menghasilkan informasi yang diperlukan. Penelitian sebelumnya [24] menemukan bahwa alat pertanian presisi dapat mendukung petani untuk mendapatkan informasi yang relevan dan membuat keputusan yang akurat. Kemampuan tersebut pada dasarnya diprakarsai oleh sejumlah besar dataset yang terdiri dari berbagai variabel dan variabel dependen, termasuk hubungannya [25].

B. Saran

Adapun saran pada penelitian ini adalah:

1. Agar melakukan pengumpulan data untuk variabel independen lain yang mempengaruhi produktivitas jagung selain luas lahan antara lain curah hujan, kelembaban, temperatur, intensitas cahaya matahari, dan jumlah petani
2. Perlu mengimplementasikan metode hybrid dari algoritma mesin pembelajaran untuk mengatasi data dalam jumlah besar.

DAFTAR PUSTAKA

[1] R. Taghizadeh-Mehrjardi, K. Nabiollahi, L. Rasoli, R. Kerry, and T. Scholten, "Land Suitability Assessment and Agricultural Production Sustainability Using Machine Learning Models," *Agronomy*, vol. 10, no. 4, p. 573, Apr. 2020, doi: 10.3390/agronomy10040573.

[2] K. Akpoti, A. T. Kabo-bah, and S. J. Zwart, "Agricultural land suitability analysis: State-of-the-art and outlooks for integration of climate change analysis," *Agric Syst*, vol. 173, pp. 172–208, Jul. 2019, doi: 10.1016/j.agsy.2019.02.013.

[3] S. Gholami *et al.*, "Food security analysis and forecasting: A machine learning case study in southern Malawi," *Data* Policy, vol. 4, p. e33, Oct. 2022, doi: 10.1017/dap.2022.25.

[4] H. C. J. Godfray *et al.*, "Food Security: The Challenge of Feeding 9 Billion People," *Science (1979)*, vol. 327, no. 5967, pp. 812–818, Feb. 2010, doi: 10.1126/science.1185383.

[5] N. N. Misra, Y. Dixit, A. Al-Mallahi, M. S. Bhullar, R. Upadhyay, and A. Martynenko, "IoT, Big Data, and Artificial Intelligence in Agriculture and Food Industry," *IEEE Internet Things J*, vol. 9, no. 9, pp. 6305–6324, 2022, doi: 10.1109/JIOT.2020.2998584.

[6] Md. T. Shakoor, K. Rahman, S. N. Rayta, and A. Chakrabarty, "Agricultural production output prediction using Supervised Machine Learning techniques," in *2017 1st International Conference on Next Generation Computing Applications (NextComp)*, 2017, pp. 182–187. doi: 10.1109/NEXTCOMP.2017.8016196.

[7] U. Grote, A. Fasse, T. T. Nguyen, and O. Erenstein, "Food Security and the Dynamics of Wheat and Maize Value Chains in Africa and Asia," *Front Sustain Food Syst*, vol. 4, Feb. 2021, doi: 10.3389/fsufs.2020.617009.

[8] C. Xiao, "USING MACHINE LEARNING FOR EXPLORATORY DATA ANALYSIS AND PREDICTIVE MODELS ON LARGE DATASETS," Thesis, UNIVERSITY OF STAVANGER, 2015. doi: http://hdl.handle.net/11250/299600.

[9] Mark Menagie, "A comparison of machine learning algorithms using an insufficient number of labeled observations," Scince, Vrije Universiteit Amsterdam, 2018. doi: chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://vu-business-analytics.github.io/internship-office/reports/report-menagie.pdf.

[10] L. Zhou, S. Pan, J. Wang, and A. V. Vasilakos, "Machine learning on big data: Opportunities and challenges," *Neurocomputing*, vol. 237, pp. 350–361, May 2017, doi: 10.1016/j.neucom.2017.01.026.

[11] A. Farrell *et al.*, "Machine learning of large-scale spatial distributions of wild turkeys with high-dimensional environmental data," *Ecol Evol*, vol. 9, no. 10, pp. 5938–5949, May 2019, doi: 10.1002/ece3.5177.

[12] K. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis, "Machine Learning in Agriculture: A Review," *Sensors*, vol. 18, no. 8, p. 2674, Aug. 2018, doi: 10.3390/s18082674.

[13] L. Breiman, "Random Forests," *Mach Learn*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

[14] S. Han and H. Kim, "On the Optimal Size of Candidate Feature Set in Random forest," *Applied Sciences*, vol. 9, no. 5, p. 898, Mar. 2019, doi: 10.3390/app9050898.
[15] R. M. Freund, P. Grigas, and R. Mazumder, "A new perspective on boosting in linear regression via subgradient optimization and relatives," *The Annals of Statistics*, vol. 45, no. 6, Dec. 2017, doi: 10.1214/16-AOS1505.
[16] D. Opitz and R. Maclin, "Popular Ensemble Methods: An Empirical Study," *Journal of Artificial Intelligence Research*, vol. 11, pp. 169–198, Aug. 1999, doi: 10.1613/jair.614.
[17] Z. P. Brodeur, J. D. Herman, and S. Steinschneider, "Bootstrap Aggregation and Cross-Validation Methods to Reduce Overfitting in Reservoir Control Policy Search," *Water Resour Res*, vol. 56, no. 8, Aug. 2020, doi: 10.1029/2020WR027184.
[18] H.-Y. Kim, "Statistical notes for clinical researchers: simple linear regression 2 – evaluation of regression line," *Restor Dent Endod*, vol. 43, no. 3, 2018, doi: 10.5395/rde.2018.43.e34.
[19] H. Lu and X. Ma, "Hybrid decision tree-based machine learning models for short-term water quality prediction," *Chemosphere*, vol. 249, p. 126169, Jun. 2020, doi: 10.1016/j.chemosphere.2020.126169.
[20] A. Cravero and S. Sepúlveda, "Use and Adaptations of Machine Learning in Big Data—Applications in Real Cases in Agriculture," *Electronics (Basel)*, vol. 10, no. 5, p. 552, Feb. 2021, doi: 10.3390/electronics10050552.
[21] P. Zhang, Z.-Y. Yin, Y.-F. Jin, and T. H. T. Chan, "A novel hybrid surrogate intelligent model for creep index prediction based on particle swarm optimization and random forest," *Eng Geol*, vol. 265, p. 105328, Feb. 2020, doi: 10.1016/j.enggeo.2019.105328.
[22] G. Biau and E. Scornet, "A random forest guided tour," *TEST*, vol. 25, no. 2, pp. 197–227, Jun. 2016, doi: 10.1007/s11749-016-0481-7.
[23] R. S and S. Kumar J, "Performance evaluation of random forest with feature selection methods in prediction of diabetes," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 10, no. 1, p. 353, Feb. 2020, doi: 10.11591/ijece.v10i1.pp353-359.
[24] P. K. A. Vinayak N. Malavade, "Role of IoT in Agriculture," *IOSR Journal of Computer Engineering (IOSR-JCE)*, vol. 4, no. June, p. 2016, 2016,[Online].Available:https://www.iosrjournals.org/iosr-jce/papers/Conf.16051/Volume-1/13. 56-57.pdf
[25] I. A. T. Hashem, I. Yaqoob, N. B. Anuar, S. Mokhtar, A. Gani, and S. Ullah Khan, "The rise of 'big data' on cloud computing: Review and open research issues," Inf Syst, vol. 47, pp. 98–115, Jan. 2015, doi: 10.1016/j.is.2014.07.006.