

Penerapan Pembelajaran Mesin dalam Deteksi Cuaca: Sebuah Tinjauan Pustaka Sistematis

Application of Machine Learning in Weather Detection: A Systematic Literature Review

Mario Patrick Kalikit Pattisina¹, Achmad Bayu Agung Saputra¹, Fajar Wahyu
Santoso¹, Teguh Wahyono^{2*}

¹Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Kristen Satya
Wacana, Salatiga 50715, Indonesia

²Program Studi Bisnis Digital, Fakultas Teknologi Informasi, Universitas Kristen Satya
Wacana, Salatiga 50715, Indonesia

(*Email Korespondensi: teguh.wahyono@uksw.edu)

Abstrak: Penelitian ini merupakan tinjauan pustaka sistematis (*Systematic Literature Review*) yang membahas penerapan *machine learning* dalam deteksi cuaca. Studi ini mengidentifikasi tren, tantangan, serta efektivitas berbagai model *machine learning* dalam memprediksi kondisi cuaca, termasuk fenomena cuaca ekstrem. Berdasarkan analisis terhadap lima belas penelitian terbaru dua tahun terakhir (2024–2025), ditemukan bahwa faktor utama yang mempengaruhi akurasi prediksi cuaca meliputi variabel cuaca yang digunakan, kualitas dataset, serta pemilihan algoritma *machine learning*. Model *XGBoost* dan teknik *deep learning* seperti *Multi-Layer Perceptron* (MLP) dan *BiLSTM* terbukti memberikan akurasi tertinggi dalam mendeteksi cuaca ekstrem. Selain itu, penggunaan data *real-time* dari sensor IoT dan teknik *balancing* data meningkatkan keakuratan prediksi. Hasil penelitian ini menunjukkan bahwa *machine learning* memiliki potensi besar dalam meningkatkan kualitas prediksi cuaca, namun masih menghadapi tantangan seperti ketidakpastian cuaca dan keterbatasan data berkualitas tinggi.

Kata Kunci: *machine learning*, *deep learning*, deteksi cuaca, tinjauan pustaka sistematis.

Abstract: This study is a Systematic Literature Review (SLR) that examines the application of machine learning in weather detection. The study aims to identify trends, challenges, and the effectiveness of various machine learning models in predicting weather conditions, including extreme weather phenomena. Based on an analysis of fifteen recent studies (2024–2025), the main factors influencing weather prediction accuracy include the weather variables used, dataset quality, and the choice of machine learning algorithms. Models such as *XGBoost* and deep learning techniques like *Multi-Layer Perceptron* (MLP) and *BiLSTM* have proven to provide the highest accuracy in detecting extreme weather. Additionally, the use of real-time data from IoT sensors and data balancing techniques can improve prediction accuracy. The results of this study indicate that machine learning has significant potential to enhance weather prediction quality, although challenges remain, such as limited high-quality data and the uncertainty of climate change.

Keywords: machine learning, deep learning, weather detection, systematic literatur review.



1. Pendahuluan

Systematic literature review (SLR) adalah metode penelitian yang digunakan untuk mengumpulkan, menilai, dan menganalisis semua bukti penelitian yang relevan terkait dengan pertanyaan ilmiah (Azevedo, 2024). Pendekatan ini bertujuan untuk menyajikan rangkuman penelitian sebelumnya secara objektif. SLR sangat penting dalam berbagai bidang, termasuk kesehatan, pendidikan, dan ilmu sosial. Metode ini merupakan tinjauan penting bagi para peneliti yang sudah memastikan bahwa keputusan mereka didasarkan pada bukti terbaik yang tersedia (Siswanto, 2024). Berbeda dengan tinjauan pustaka tradisional, SLR mengikuti prosedur yang ketat mulai dari perumusan pertanyaan penelitian, penentuan kriteria inklusi dan eksklusi, pencarian literatur secara menyeluruh, hingga analisis dan sintesis data dari studi-studi terpilih. Dengan menggunakan SLR diharapkan dapat memberikan gambaran menyeluruh dan obyektif terhadap status terkini suatu bidang ilmu, mengungkapkan temuan-temuan yang konsisten, mengidentifikasi kesenjangan penelitian, serta memberikan landasan bagi penelitian lanjutan yang lebih mendalam. Penelitian ini bertujuan untuk melakukan SLR khususnya dalam ranah penerapan *machine learning* untuk kasus deteksi cuaca.

Prediksi cuaca merupakan hal yang sangat penting dalam kehidupan manusia. Ketika meramalkan cuaca, ada banyak kondisi yang dapat diamati, seperti suhu atmosfer, kelembaban, kekuatan matahari, angin, hujan, dan kondisi udara lainnya. Perubahan iklim global dan meningkatnya frekuensi kejadian cuaca ekstrem telah menjadi tantangan besar bagi berbagai sektor kehidupan, mulai dari pertanian, transportasi, energi, hingga mitigasi bencana (Rochmawati, 2024). Kemampuan untuk mendeteksi dan memprediksi kondisi cuaca secara akurat dan *real-time* menjadi sangat krusial dalam menghadapi tantangan tersebut. Dalam beberapa dekade terakhir, kemajuan teknologi informasi dan kecerdasan buatan, khususnya *machine learning* (ML), telah membuka peluang baru dalam meningkatkan akurasi dan efisiensi sistem deteksi cuaca (Intan, 2021).

Beberapa prediksi cuaca dilakukan melalui cara komputasi konvensional seperti pemodelan numerik berbasis persamaan fisika yang kompleks dan memerlukan sumber daya komputasi tinggi. Metode tradisional dalam prediksi cuaca tersebut dikenal sebagai *Numerical Weather Prediction* (NWP), yaitu pendekatan komputasi yang menggunakan model matematika berdasarkan hukum-hukum fisika atmosfer, seperti hukum termodinamika, dinamika fluida, dan hukum konservasi massa serta energi. Karena kompleksitas sistem atmosfer sangat tinggi dan jumlah variabel yang dihitung sangat besar, proses ini membutuhkan daya komputasi yang sangat besar serta infrastruktur perangkat keras yang canggih. Adapun metode ini memiliki keterbatasan, terutama dalam menangkap pola non linier dan lokal yang kompleks, serta dalam hal kecepatan pemrosesan untuk prediksi jangka pendek atau *real-time*. Oleh karena itu diperlukan pendekatan *machine learning* dalam mengatasi permasalahan tersebut. ML mampu mempelajari pola dari data historis dan observasi sensor, serta menyesuaikan model secara adaptif untuk menghasilkan prediksi yang lebih cepat dan presisi (Wiwaha, 2024).

Sejumlah studi telah menunjukkan bahwa algoritma ML seperti *Support Vector Machine* atau SVM (Sulistiyowati, 2024), *Random Forest* (Rahman, 2024), *Artificial Neural Networks* atau ANN (Panggabean, 2021), dan model *Deep Learning* seperti *Long Short-Term Memory* atau LSTM (Pitaloka, 2024) memberikan hasil menjanjikan dalam berbagai aplikasi prediksi dan deteksi cuaca. Prediksi yang dilakukan mencakup beberapa variabel cuaca seperti hujan, suhu, kelembaban, serta peringatan dini bencana seperti badai dan banjir. Meskipun demikian, terdapat beragam pendekatan, jenis data, dan evaluasi performa yang digunakan dalam penelitian-penelitian tersebut, sehingga diperlukan pemetaan secara sistematis untuk memahami arah perkembangan dan celah penelitian yang masih terbuka.

Penelitian ini bertujuan untuk melakukan *Systematic Literature Review* terhadap penerapan *machine learning* dalam deteksi cuaca. Dengan melakukan kajian mendalam terhadap literatur yang relevan dalam beberapa tahun terakhir, penelitian ini diharapkan dapat memberikan pemahaman yang komprehensif mengenai tren, algoritma yang paling banyak digunakan, jenis data yang umum digunakan, serta tantangan yang dihadapi dalam implementasi ML untuk deteksi cuaca. Dalam penelitian ini terdapat dua pertanyaan penelitian yang akan dibahas. Pertanyaan penelitian pertama adalah faktor apa saja yang mempengaruhi akurasi prediksi cuaca menggunakan *machine learning*, sedangkan pertanyaan penelitian kedua adalah seberapa akurat model *machine learning* dalam mendeteksi fenomena cuaca ekstrem. Hasil kajian dari penelitian ini bisa menjadi rekomendasi bagi peneliti dan praktisi dalam

mengembangkan model prediksi cuaca yang lebih baik di masa depan, baik dalam konteks pemilihan model atau algoritma, maupun dalam pemilihan variabel-variabel prediksinya.

2. Kajian Pustaka

2.1 Deteksi Cuaca

Deteksi cuaca merupakan proses pengamatan dan identifikasi kondisi atmosfer berdasarkan parameter-parameter seperti suhu, kelembaban, tekanan udara, kecepatan angin, serta curah hujan. Informasi deteksi cuaca ini sangat penting untuk berbagai sektor, termasuk pertanian, penerbangan, logistik, dan mitigasi bencana. Secara konvensional, deteksi dan prediksi cuaca dilakukan melalui pemodelan numerik berbasis fisika, seperti *Numerical Weather Prediction* (NWP). Meskipun cukup akurat, metode ini membutuhkan data yang sangat besar dan sumber daya komputasi tinggi serta tidak selalu mampu menangkap pola non linier yang kompleks (Putri, 2021). Sehingga pada perkembangannya dikembangkan banyak metode, diantaranya berkembang pesat adalah metode *machine learning* dan *deep learning*.

2.2 Machine Learning

Machine Learning (ML) adalah cabang dari kecerdasan buatan (*Artificial Intelligence*) yang berfokus pada pengembangan algoritma yang memungkinkan sistem untuk belajar dari data dan membuat prediksi atau keputusan tanpa diprogram secara eksplisit (Intan, 2021; Avezedo, 2024). Dalam konteks deteksi cuaca, ML digunakan untuk membangun model prediktif berdasarkan data historis cuaca. ML dapat mengidentifikasi pola tersembunyi dan hubungan kompleks antar variabel yang sulit ditangkap oleh metode konvensional. Beberapa pendekatan ML yang umum digunakan antara lain (Pessach, 2022):

- a. *Supervised Learning*: Model dilatih menggunakan data berlabel. Contoh dari algoritma ini adalah *Support Vector Machine* (SVM), *Random Forest* (RF), *K-Nearest Neighbors* (KNN), dan *Artificial Neural Networks* (ANN).
- b. *Unsupervised Learning*: Digunakan ketika data tidak berlabel, misalnya dalam pengelompokan pola cuaca. Contoh: K-Means Clustering, Principal Component Analysis (PCA).
- c. *Deep Learning*: Pendekatan lanjutan dari ML yang menggunakan arsitektur jaringan saraf dalam, seperti *Convolutional Neural Network* (CNN) dan *Long Short-Term Memory* (LSTM), yang sangat efektif dalam menangani data spasial dan temporal seperti citra satelit atau deret waktu cuaca (Camacho, 2025).

2.3 Penerapan Machine Learning dalam Deteksi Cuaca

Penerapan ML dalam deteksi cuaca telah menunjukkan potensi besar dalam meningkatkan akurasi dan efisiensi prediksi. Model ML dapat digunakan untuk berbagai tugas, seperti prediksi hujan dan intensitas curah hujan, deteksi pola badai atau siklon tropis, estimasi suhu udara harian atau jangka panjang, klasifikasi jenis awan berdasarkan citra satelit dan peringatan dini terhadap kondisi cuaca ekstrem. Model-model ini memanfaatkan beragam sumber data, seperti data historis dari stasiun cuaca, citra radar dan satelit, serta sensor lingkungan. Keunggulan ML terletak pada kemampuannya untuk memproses data besar (*big data*), belajar secara adaptif, serta menangani hubungan non linier antar variabel cuaca yang kompleks (Intan, 2021; Panggabean, 2021).

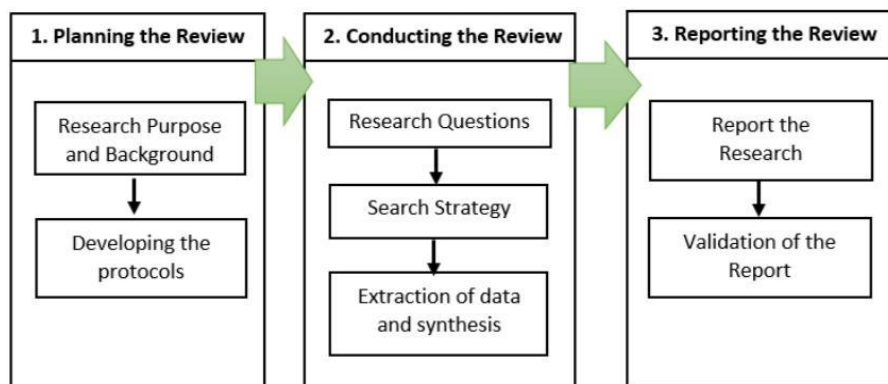
Machine learning dalam ramalan cuaca memungkinkan analisis data skala besar menggunakan model yang dapat belajar dari data historis dan memprediksi pola cuaca di masa depan. Algoritma yang sering digunakan termasuk XGBoost (Wiwaha, 2024). Ini adalah akurasi tinggi pengenalan cuaca ekstrem hingga 98.62%. *Support Vector Machine* (SVM) ini efektif ketika mengenali perubahan kelembaban pada $r^2 = 0,723$ (Rifqi, 2024). *Random Forest* ini lebih baik daripada *Decision Tree* dan *Naive Bayes* untuk mengklasifikasikan curah hujan ekstrem dengan skor $F1 = 65,5\%$ (Maheswara, 2025).

Dalam beberapa tahun terakhir, teknologi kecerdasan buatan dan *machine learning* telah menjadi populer untuk meningkatkan akurasi peramalan cuaca (Rochmawati, 2024; Wiwaha, 2024). Kemampuan untuk menganalisis dan memprediksi data historis tanpa pemrograman eksplisit berulang menjadikan *machine learning* sebagai alat yang sangat berguna di bidang meteorologi. Algoritma seperti *Support Vector Machine*, *Random Forest*, dan pendekatan *Ensemble Learning* diterapkan untuk memproses data cuaca di berbagai tingkat kompleksitas. Penelitian sebelumnya menunjukkan bahwa penggunaan *machine learning* memberikan hasil yang menjanjikan untuk pengenalan cuaca. Sebagai contoh, ketika menggunakan *Deep Learning* berbasis *Convolutional Neural Network* (CNN) menggunakan fragmen lipatan untuk menganalisis pola awan dan meningkatkan keakuratan prediksi cuaca. Selain itu, ketika menggunakan model *Random Forest* untuk mendeteksi pola suhu dan kelembaban yang berkontribusi pada prediksi fenomena gelombang panas. Sementara itu, pendekatan *Recurrent Neural Network* (RNN) yang menggunakan data deret waktu lebih efektif dalam meningkatkan keakuratan ramalan cuaca ekstrem.

Masalah dalam penerapan *machine learning* untuk peramalan cuaca mencakup beberapa tantangan dalam implementasi pembelajaran mesin untuk peramalan cuaca. Beberapa masalah tersebut diantaranya adalah ketidakpastian cuaca dan perubahan iklim mempengaruhi keakuratan model, serta ketersediaan catatan data berkualitas tinggi, terutama di daerah di mana sensor cuaca terbatas (Wiwaha, 2024). Hal lain yang menjadi tantangan adalah model Interpretasi dengan algoritma kompleks seperti pembelajaran mendalam sulit untuk dijelaskan dalam keputusan.

3. Metode Penelitian

Penelitian ini dilakukan dengan pendekatan *Systematic Literature Review* (SLR) yang bertujuan untuk mengidentifikasi, mengevaluasi, dan menginterpretasikan penelitian terkait penerapan *machine learning* dalam deteksi cuaca dalam tiga tahun terakhir (2023 - 2025). Metode SLR digunakan karena memungkinkan identifikasi tren, tantangan, dan peluang penelitian yang lebih sistematis dan terstruktur (Azevedo, 2024). Pada Gambar 1 metode SLR menggunakan tiga tahapan utama yang meliputi *Planning the review* (perencanaan tinjauan), *Conducting the review* (pelaksanaan tinjauan), *Reporting the review* (pelaporan hasil).



Gambar 1. Metode *Systematic Literature Review*

3.1 Tahap *Planning The Review*

Tahap *planning the review* dalam SLR merupakan langkah awal yang krusial untuk memastikan bahwa proses kajian literatur dilakukan secara sistematis, transparan, dan dapat di replikasi. Pada tahap ini, peneliti merumuskan tujuan dan pertanyaan penelitian yang spesifik, mengembangkan protokol *review* yang mencakup kriteria inklusi dan eksklusi, serta menentukan strategi pencarian literatur dan sumber basis data yang akan digunakan (Siswanto, 2024). Selain itu, tahap ini juga mencakup perencanaan metode seleksi studi dan teknik analisis data yang akan diterapkan. Dengan perencanaan yang matang, tahap ini membantu meminimalkan bias dan memastikan bahwa

kajian literatur yang dilakukan fokus, relevan, dan berkualitas tinggi.

3.2 Tahap *Conducting The Review*

Setelah tahap perencanaan langkah selanjutnya adalah melakukan *review*. Pada bagian ini, rumusan *Research Question (RQ)* dilakukan untuk menunjukkan permasalahan utama yang akan dibahas dan di analisis. Dilanjutkan dengan perumusan strategi pencarian dan ekstraksi data. Analisis dan sintesis yang dihasilkan dari SLR memberikan jawaban terhadap RQ yang telah ditentukan. *Research Question* untuk penelitian ini sebagai berikut:

- RQ1: Faktor apa saja yang mempengaruhi akurasi prediksi cuaca menggunakan *machine learning*?
- RQ2: Seberapa akurat model *machine learning* dalam mendeteksi fenomena cuaca ekstrem?

Strategi pencarian melibatkan identifikasi sumber yang akan dijadikan acuan. Tahapan proses seleksi dan *review* artikel menggunakan metode Prisma dapat dilihat pada gambar . Dalam penelitian ini satu sumber *database* dipilih *Google Scholar* dan *Scopus* dengan tujuan agar ada keterwakilan penelitian-penelitian berbahasa Indonesia dan berbahasa inggris (Internasional). Kemudian selanjutnya adalah menentukan *string* pencarian literatur yang telah ditentukan. Pemilihan kata kunci yang tepat akan menentukan keakuratan literatur yang ditemukan. *String* pencarian berikut digunakan dalam penelitian ini yaitu (“*machine learning*” AND “prediksi cuaca”) untuk artikel dalam *Google Scholar* dan (“*machine learning*” AND “*weather prediction*”) untuk artikel *Scopus*.

Setelah data ditemukan, maka data akan dipilih dengan menggunakan mekanisme eksklusi/inklusi yang diterapkan. Untuk kriteria inklusi, artikel akan dimasukkan jika memenuhi kriteria seperti: topik sesuai dengan fokus penelitian prediksi cuaca dengan *machine learning*, tipe publikasinya adalah jurnal ilmiah nasional berbahasa Indonesia (untuk *database Google Scholar*) dan jurnal internasional ter indeks *Scopus* dan berbahasa inggris (untuk *database Scopus*). Waktu publikasi yang diambil adalah dua tahun terakhir antara tahun 2024–2025, serta aksesibilitas teks lengkap dapat diakses (*full paper*). Sedangkan untuk kriteria eksklusi diantaranya adalah tidak membahas metode atau topik yang relevan, tidak ada *full-text*, duplikasi dari sumber lain, serta studi pada domain lain yang tidak relevan.

3.3 Tahap *Reporting The Review*

Tahap *Reporting the Review* dalam *Systematic Literature Review (SLR)* adalah tahap akhir yang berfokus pada penyusunan dan penyampaian hasil kajian secara sistematis, transparan, dan dapat dipertanggungjawabkan (Siswanto, 2024). Dalam tahap ini, peneliti menyusun laporan atau artikel ilmiah yang mencakup seluruh proses SLR, mulai dari perumusan pertanyaan penelitian, strategi pencarian literatur, proses seleksi studi, ekstraksi dan sintesis data, hingga temuan utama dari kajian. Laporan harus menyajikan informasi secara jelas dan terstruktur, biasanya dalam bentuk narasi yang disertai tabel, grafik, atau visualisasi lain untuk memperkuat analisis. Selain menyampaikan hasil, bagian ini juga membahas keterbatasan studi, implikasi praktis dan teoretis dari temuan, serta rekomendasi untuk penelitian selanjutnya. Tujuan utama dari tahap ini adalah untuk memastikan bahwa hasil kajian dapat dipahami, digunakan, dan dijadikan rujukan oleh peneliti maupun praktisi di bidang terkait.

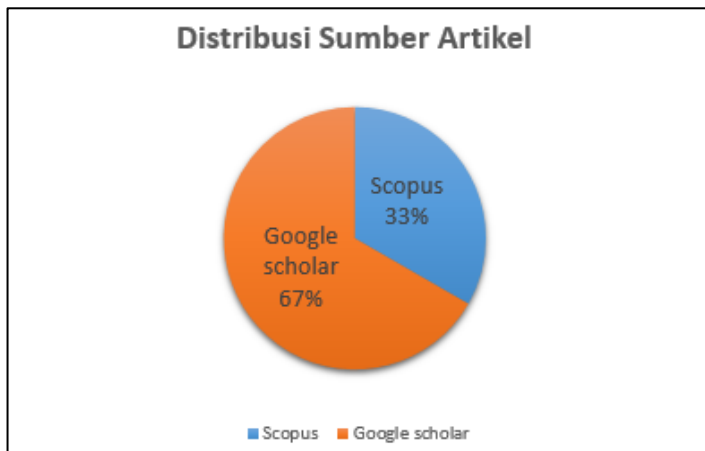
4. Hasil dan Pembahasan

Bagian ini akan membahas tentang temuan penelitian terkait dengan dua RQ yang telah dituliskan di awal penelitian. Kedua RQ tersebut adalah faktor-faktor apa saja yang mempengaruhi akurasi prediksi cuaca menggunakan *machine learning*, serta seberapa akurat model *machine learning* dalam mendeteksi fenomena cuaca ekstrem.

4.1 Deskripsi Artikel

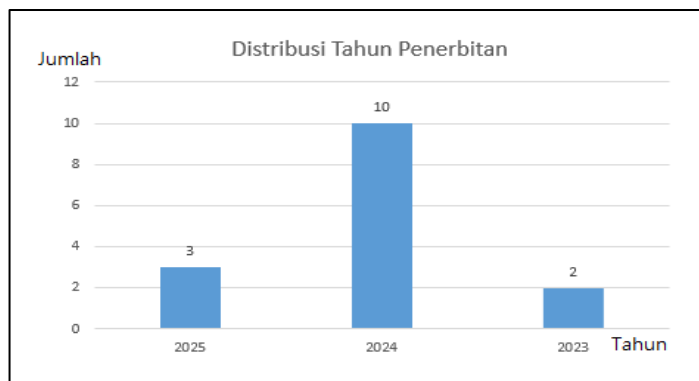
Seperti yang telah dijelaskan sebelumnya, artikel yang dicari bersumber dari dua *database* yaitu *Google Scholar* dan *Scopus*. *Google Scholar* digunakan untuk mencari artikel berbahasa Indonesia, yang merepresentasikan

penelitian di Indonesia. Sedangkan *Scopus* digunakan untuk mendapatkan artikel berbahasa Inggris dalam periode 3 tahun terakhir. Gambar 2 di bawah menunjukkan bahwa temuan artikel ini dari sisi distribusi sumber artikel menunjukkan bahwa 67% artikel bersumber dari *Google Scholar* dan 33% artikel didapatkan dari *database Scopus*.



Gambar 2. Distribusi Sumber Artikel

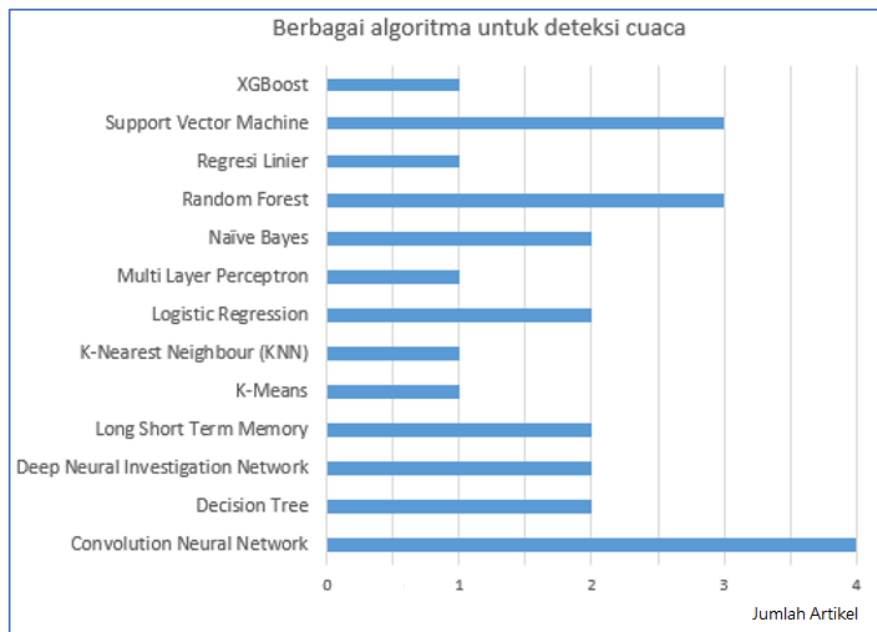
Selanjutnya dari sisi distribusi tahun publikasi atau tahun penerbitan artikel, diperoleh seperti pada Gambar 3 di bawah. Dari gambar tersebut bisa diperhatikan bahwa distribusi artikel menunjukkan 3 artikel diterbitkan pada tahun 2025, 10 artikel diterbitkan di tahun 2024 dan 2 artikel diterbitkan di tahun 2023.



Gambar 3. Distribusi Tahun Publikasi

4.2 Metode Machine Learning untuk Deteksi Cuaca

Berdasarkan temuan penelitian ini, terdapat 13 (tiga belas) algoritma yang digunakan untuk deteksi cuaca. Ketiga algoritma tersebut adalah *Convolution Neural Network*, *Decision Tree*, *Deep Neural Investigation Network*, *Long Short Term Memory*, *K-Means*, *K-Nearest Neighbour (KNN)*, *Logistic Regression*, *Multi Layer Perceptron*, *Naïve Bayes*, *Random Forest*, *Regresi Linier*, *Support Vector Machine* dan *XGBoost*. Ketiga belas algoritma tersebut dapat dilihat pada Gambar 4 di bawah ini.



Gambar 4. Berbagai algoritma deteksi cuaca

Dari Gambar 4 dapat dilihat bahwa diantara ketiga belas algoritma tersebut, *Convolution Neural Network* adalah metode yang paling sering digunakan (Liu, 2023; Bone, 2023; Pitaloka, 2024; Camacho, 2025), disusul kemudian *Support Vector Machine* (Sulistyowati, 2024; Mahendra, 2024; Rifqi, 2024) dan *Random Forest* (Jones, 2024; Rahman, 2024; Maheswara, 2025).

4.3 Faktor yang Mempengaruhi Akurasi Prediksi Cuaca

Prediksi cuaca yang dimaksud pada bagian ini adalah proses memperkirakan kondisi atmosfer di masa depan (bisa secara periodik dalam jam, hari, atau minggu ke depan) berdasarkan analisis data variabel-variabel cuaca yang tersedia sebelumnya. Misalnya untuk mengetahui apakah akan terjadi cuaca ekstrem seperti hujan, panas, berawan, badai, angin kencang, dan sebagainya. Berdasarkan analisis artikel-artikel yang ditemukan, terdapat beberapa faktor yang dapat mempengaruhi prediksi cuaca menggunakan *machine learning*, yang diantaranya adalah variabel cuaca yang digunakan dalam prediksi, kualitas dan kelengkapan *dataset* dan pemilihan metode *machine learning*.

a. Variabel cuaca yang digunakan dalam prediksi

Studi menunjukkan bahwa akurasi model *machine learning* sangat dipengaruhi oleh variabel cuaca yang digunakan sebagai *input*. Mengutip dari beberapa penelitian terdahulu (Liu, 2023; Rochmawati, 2024; Intan, 2024; Maheswara, 2025; Camacho, 2025; Adiyasa et al, 2025) maka terdapat beberapa variabel utama yang sering digunakan dalam prediksi cuaca yang diantaranya adalah sebagai berikut.

- Suhu: variabel ini merupakan salah satu variabel penting dalam prediksi cuaca, karena memiliki korelasi yang tinggi dengan cuaca ekstrem (Camacho et al., 2025; Intan et al., 2021; Maheswara, 2025; Rochmawati, 2024; Sutaryani, 2024).
- Curah hujan: variabel ini juga merupakan faktor penting dalam mendeteksi potensi hujan dan banjir (Maheswara, 2025; Intan et al, 2021).
- Tekanan udara: variabel ini penting karena mempengaruhi dalam membentuk pola cuaca tertentu

(Camacho et al. 2025; Maheswara, 2025; Rochmawati, 2024).

- Kelembaban udara: variabel ini berpengaruh dalam pembentukan awan dan potensi terjadinya hujan (Sulistiyowati, 2024; Mahendra, 2024; Rifqi, 2024).
- Kecepatan dan arah angin: variabel ini berperan dalam menentukan kemungkinan dalam terjadinya badai atau perubahan cuaca mendadak (Rochmawati, 2024; Maheswara, 2025; Intan et al, 2021).

Penggunaan variabel cuaca yang tepat sangat penting dalam meningkatkan keakuratan model prediksi. Sebagai contoh, studi yang dilakukan oleh Sulistiyowati et al. (2024) menunjukkan bahwa kelembapan udara dapat diprediksi dengan akurasi tinggi menggunakan *Support Vector Machine* (SVM), dengan nilai $R^2 = 0,723$.

b. Kualitas dan kelengkapan dataset

Dataset yang digunakan harus memiliki resolusi yang tinggi dan data yang lengkap untuk memberikan hasil prediksi yang lebih akurat dibandingkan dengan *dataset* yang memiliki banyak *missing values* (Jose, 2024, Rifqi, 2025). Beberapa *dataset* yang digunakan dalam berbagai penelitian yang ditemukan diantaranya adalah data Meteorologi konvensional. Data ini digunakan oleh sebagian besar penelitian untuk memprediksi atau mengklasifikasikan kondisi cuaca. Beberapa penelitian yang menggunakan data meteorologi konvensional ini diantaranya adalah Mahendra et al. (2024) yang menggunakan data suhu dan kelembapan dari BMKG, Sulistiyowati et al. (2024) yang melakukan prediksi kelembapan dengan data cuaca harian serta Camacho et al. (2025) yang memanfaatkan data *irradiansi* matahari.

Dataset lain yang digunakan dalam berbagai penelitian yang ditemukan adalah data yang berasal dari sensor dan *IoT* serta *Wireless Sensor Network* (WSN). Beberapa penelitian yang menggunakan *dataset* tersebut adalah Wiwaha et al. (2024) yang menggunakan *dataset* dari WSN dan *IoT* untuk prediksi cuaca *real-time* serta Pitaloka et al. (2024) tentang sistem prediksi banjir dengan sensor. Penelitian Wiwaha (2024) menunjukkan bahwa penggunaan *Wireless Sensor Network* (WSN) dan *Internet of Things* (*IoT*) dalam pengumpulan data cuaca secara *real-time* dapat meningkatkan akurasi prediksi. Teknik *data balancing* seperti SMOTE juga meningkatkan keakuratan dalam *dataset* yang tidak seimbang, terutama untuk mendeteksi fenomena cuaca ekstrem.

Selanjutnya adalah penggunaan *dataset* yang berasal dari data simulasi atau sintesis yang biasanya digunakan untuk model *Deep Learning* serta digunakan dalam studi berbasis simulasi. Berbagai data sintesis tersebut diantaranya adalah data sintesis untuk pelatihan model klasifikasi awan, deteksi *noise* sinyal, dan sebagainya yang berasal dari *dataset open-source*. Contoh yang menggunakan *dataset* ini adalah Liu et al. (2023) menggunakan citra satelit dan peta iklim lokal yang digunakan untuk segmentasi zona iklim lokal. Putri & Kristianto (2021) menggunakan data *Global Forecast System* (GFS) sebagai simulasi untuk prediksi cuaca dalam 7 hari ke depan.

c. Pemilihan metode *machine learning*

Setiap model *machine learning* memiliki keunggulan dan kelemahan dalam menangani *dataset* untuk mendeteksi cuaca. Model XGBoost menunjukkan akurasi tertinggi (98,62%) dalam mendeteksi cuaca ekstrem (Wiwaha, 2024). *Support Vector Machine* (SVM) efektif dalam mendeteksi perubahan kelembapan udara tetapi memiliki keterbatasan dalam menangani *dataset* yang besar (Rifqi, 2024). *Random Forest* lebih unggul dibandingkan *Decision Tree* dan *Naive Bayes* dalam mengklasifikasikan hujan ekstrem dengan *F1-score* 65,5% (Maheswara, 2025). Dari hasil ini, dapat disimpulkan bahwa pemilihan model yang sesuai dengan jenis data dan fenomena cuaca yang akan diprediksi merupakan faktor kunci dalam meningkatkan akurasi prediksi cuaca.

4.4 Akurasi Model Machine Learning dalam Mendeteksi Fenomena Cuaca Ekstrem

Berbagai metode *machine learning* telah digunakan untuk mendeteksi fenomena cuaca ekstrem seperti hujan deras, badai, dan perubahan kelembapan udara ekstrem. Berikut adalah beberapa hasil evaluasi dari model yang telah dianalisis.

4.4.1 Perbandingan Akurasi Model Dalam Deteksi Cuaca Ekstrem

Tabel 1 menunjukkan akurasi model dalam deteksi cuaca ekstrem. Hasil pada tabel tersebut menunjukkan bahwa *XGBoost* dan model *deep learning* seperti MLP dan BiLSTM memberikan hasil terbaik dalam deteksi cuaca ekstrem, sementara model seperti SVM dan *Logistic Regression* lebih cocok untuk prediksi variabel spesifik seperti kelembaban udara atau curah hujan.

Table 1. Akurasi Model Dalam Deteksi Cuaca Ekstrem

Metode Machine Learning	Akurasi(%)	Kesimpulan
<i>XGBoost</i> (Wiwaha, 2024)	98,62%	Model terbaik dalam mendeteksi hujan ekstrem dan badai.
<i>Logistic Regression Neural Network</i> (LRNN) (Adiyasa, 2025; Jones, 2024)	79%	Akurat dalam prediksi banjir tetapi kurang akurat untuk cuaca ekstrem lainnya.
<i>Naive Bayes, Decision Tree, Random Forest</i>	65,5% (Random Forest tertinggi)	<i>Random Forest</i> lebih baik daripada <i>Naive Bayes</i> dan <i>Decision Tree</i> .
<i>Support Vector Machine</i> (SVM) (Sulistiyowati, 2024; Mahendra, 2024; Rifqi, 2024)	$R^2 = 0,723$	Cukup akurat dalam memprediksi kelembaban udara.
<i>K-Nearest Neighbour</i> (KNN), <i>Logistic Regression, Random Forest, Naive Bayes, Multi-Layer Perceptron</i> (MLP) (Jones, 2024)	99,03% (MLP tertinggi)	<i>Deep Learning</i> lebih unggul dalam menangani kompleksitas data.
CNN, BiLSTM, DNIN (<i>Deep Neural Investigation Network</i>) (Pitaloka, 202)	94,96%	Model <i>Deep Learning</i> memiliki akurasi tinggi tetapi membutuhkan lebih banyak data.

4.4.2 Keunggulan dan Kelemahan Model Machine Learning

Tabel 2 berikutnya menunjukkan bahwa *XGBoost* dan *Deep Learning* menjadi pilihan utama dalam mendeteksi cuaca ekstrem terutama jika data yang digunakan memiliki volume besar dan beragam sumber. *XGBoost* (*Extreme Gradient Boosting*) merupakan algoritma yang sangat populer dalam prediksi cuaca karena kemampuannya memberikan akurasi tinggi serta mampu menangani *dataset* berukuran besar dengan efisien. Model ini bekerja dengan membangun sejumlah pohon keputusan secara berurutan untuk meminimalkan kesalahan prediksi. Namun, kekuatan *XGBoost* ini datang dengan tantangan, yaitu kebutuhan untuk melakukan *tuning* parameter yang kompleks agar hasil prediksi bisa lebih optimal (Wiwaha, 2024).

Sementara itu *Support Vector Machine* (SVM) dikenal efektif dalam memodelkan data non-linear dan sangat baik dalam memprediksi parameter cuaca seperti kelembaban udara (Sulistiyowati, 2024; Mahendra, 2024; Rifqi, 2024). Algoritma ini bekerja dengan mencari *hyperplane* terbaik untuk memisahkan data dalam ruang berdimensi tinggi. Meski demikian, SVM kurang cocok digunakan pada *dataset* berukuran besar karena proses pelatihannya bisa menjadi sangat lambat.

Table 2. Keunggulan dan Kelemahan Model Machine Learning

Metode	Keunggulan	Kelemahan
<i>XGBoost</i>	Akurasi tinggi, mampu menangani <i>dataset</i> berukuran besar.	Membutuhkan <i>tuning</i> parameter yang kompleks.
<i>Support Vector Machine</i> (SVM)	Baik untuk memprediksi kelembaban udara	Kurang optimal untuk <i>dataset</i> berukuran besar.
<i>Random Forest</i>	Baik untuk mengklasifikasikan curah hujan ekstrem.	Kurang akurat dibandingkan model <i>Deep Learning</i> .
CNN, BiLSTM, DNIN	Sangat akurat dalam menganalisis pola cuaca yang kompleks.	Membutuhkan <i>dataset</i> yang berukuran besar dan komputasi yang tinggi.

Random Forest adalah metode *ensemble* yang membangun banyak pohon keputusan dan menggabungkannya untuk mendapatkan hasil klasifikasi yang lebih stabil dan akurat (Maheswara, 2024). Dalam konteks prediksi cuaca, *Random Forest* cukup andal dalam mengklasifikasikan kejadian curah hujan ekstrem. Namun, dibandingkan dengan model *Deep Learning*, akurasi *Random Forest* masih lebih rendah dalam mengenali pola-pola cuaca yang sangat kompleks (Rahman, 2024; Joses, 2024).

Sementara itu model *Deep Learning* seperti *Convolutional Neural Network* (CNN), *Bidirectional Long Short-Term Memory* (BiLSTM), dan *Deep Neural Inference Network* (DNIN) sangat unggul dalam mendeteksi pola cuaca yang kompleks dan dinamis. Model-model ini mampu menangkap hubungan spasial dan temporal dalam data cuaca dengan sangat baik. Meski demikian, kekurangannya adalah memerlukan *dataset* berukuran besar dan sumber daya komputasi tinggi untuk bisa dilatih secara optimal (Pitaloka, 2024; Liu, 2024; Katotch, 2024; Shin, 2025).

4.5 Implikasi dan Keterbatasan Studi

Secara praktis, hasil studi ini menegaskan bahwa *machine learning* memiliki potensi besar dalam mendeteksi dan memprediksi cuaca, terutama ketika digunakan bersama dengan data *real-time* yang dikumpulkan melalui WSN dan IoT. Hal ini dapat dimanfaatkan oleh lembaga meteorologi, lembaga kebencanaan, pertanian presisi, dan industri transportasi untuk meningkatkan akurasi peringatan dini terhadap cuaca ekstrem. Pemilihan model seperti *XGBoost* dan BiLSTM dapat menjadi acuan awal dalam merancang sistem prediksi cuaca berbasis ML yang lebih adaptif dan responsif terhadap dinamika atmosfer. Sedangkan dari sisi teoretis, temuan ini memperkaya pemahaman tentang bagaimana berbagai model *machine learning* berperan dalam pengenalan pola cuaca. Studi ini juga menyoroti pentingnya pemilihan variabel *input* (fitur cuaca) dan kualitas data sebagai penentu utama akurasi model. Hal ini menegaskan bahwa tidak hanya algoritma yang penting, tetapi juga bagaimana data diolah, direkam, dan dipilih sebagai *input* untuk pelatihan model. Temuan ini memberikan dasar bagi pengembangan teori integratif antara ilmu meteorologi dan teknik *data science*.

Selanjutnya penelitian ini memiliki keterbatasan utama pada jumlah literatur yang dianalisis, yakni hanya 15 artikel yang terdiri dari 10 artikel dari *Google Scholar* dan 5 artikel dari *Scopus*. Hal itu mengingat penelitian yang dilakukan memiliki keterbatasan waktu. Jumlah ini memang tergolong sedikit untuk sebuah *systematic literature review*, yang idealnya mencakup literatur yang lebih luas dan beragam untuk meningkatkan generalisasi temuan. Selain itu, keterbatasan dalam cakupan basis data, diantaranya tidak mencakup indeks seperti IEEE, Springer, atau Web of Science juga dapat menyebabkan kurangnya representasi studi terbaru dan berkualitas tinggi dalam bidang ini. Meski demikian, dengan keterbatasan tersebut, penelitian ini dapat memberikan gambaran yang cukup lengkap

sesuai dengan pertanyaan penelitian yang diajukan.

5. Simpulan dan Saran

Berdasarkan hasil *systematic literature review* dalam penelitian ini dapat ditarik simpulan bahwa penggunaan *machine learning* menunjukkan potensi yang cukup besar untuk pengenalan cuaca. Keakuratan peramalan cuaca menggunakan *machine learning* dipengaruhi oleh beberapa faktor: variabel cuaca yang digunakan, kualitas dan integritas catatan data, dan pilihan metode *machine learning* yang benar. Model seperti *XGBoost* dan *Deep Learning* (MLP, BILSTM) telah menunjukkan kinerja yang cocok untuk pengakuan fenomena cuaca ekstrem, sedangkan model seperti SVM lebih efektif untuk memprediksi variabel cuaca spesifik dan kelembaban. Penggunaan WSN dan IoT dalam mengumpulkan data cuaca waktu nyata juga berkontribusi pada peningkatan akurasi prediksi. Saran untuk penelitian ke depan perlu menambah jumlah literatur baik dari jurnal maupun *conference*, terutama yang terindeks oleh pengindeks bereputasi agar hasil dan pembahasan lebih komprehensif.

Untuk penelitian ke depan, disarankan untuk memperluas jumlah dan sumber literatur dengan memasukkan database seperti *IEEE Xplore*, *ScienceDirect*, *SpringerLink*, dan *Web of Science* untuk memperkaya dan memvalidasi temuan. Penelitian ke depan bisa juga melakukan analisis kuantitatif meta-analisis untuk membandingkan performa berbagai algoritma berdasarkan metrik yang sama (RMSE, MAE, *Accuracy*, dan lain-lain). Saran lain adalah dengan mengelompokkan studi berdasarkan jenis cuaca yang diprediksi (seperti hujan, kelembaban, angin, badai) agar pemilihan model *machine learning* lebih kontekstual.

Daftar Pustaka

- Azevedo, B. F., Rocha, A. M. A., & Pereira, A. I. (2024). Hybrid approaches to optimization and machine learning methods: A systematic literature review. *Machine Learning*, 113(7), 4055–4097.
- Adiyasa, B., Dewi, C., & Adikara, P. P. (2025). Deteksi bencana banjir berdasarkan data curah hujan di Jakarta menggunakan logistic regression neural network. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*. Retrieved from <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/14436>
- Bône, C., Gastineau, G., Thiria, S., Gallinari, P., & Mejia, C. (2023). Detection and attribution of climate change using a neural network. *Journal of Advances in Modeling Earth Systems*, 15(10), e2022MS003475.
- Camacho, M., et al. (2025). Short-medium-term solar irradiance forecasting with a CEEMDAN-CNN-ATT-LSTM hybrid model using meteorological data. *Applied Sciences*, 15(3), 1275.
- Intan, I., Rismayani, R., Nurdin, N., & Koswara, A. T. (2021). Performance analysis of weather forecasting using machine learning algorithms. *Pekommas*, 6(2), 1–8.
- Pendekatan metode ensemble learning untuk prakiraan cuaca menggunakan soft voting classifier. Retrieved from <https://www.wunderground.com>
- Joses, S., Yulvida, D., & Rochimah, S. (2024). Pendekatan Metode Ensemble Learning untuk Prakiraan Cuaca menggunakan Soft Voting Classifier, *J. Appl. Comput. Sci. Technol.*, vol. 5, no. 1, pp. 72 - 8.
- Katoch, A., Shree, A., Sharma, R., Brijbasi, A., Sharma, M., & Verma, S. S. (2024, June). Enhancing cloud detection performance: A comparative study of CNN models and architectures. In *2024 IEEE Students Conference on Engineering and Systems (SCES)* (pp. 1–6). IEEE.
- Liu, L., Lin, M., Du, Z., Liu, J., Chen, G., & Du, J. (2023). Developing a CNN-based, block-scale oriented local climate zone mapping approach: A case study in Guangzhou. *Building and Environment*, 240, 110414.
- Mahendra, M. F. R., Azizah, N. L., & Sumarno. (2024). Implementasi machine learning untuk memprediksi cuaca menggunakan support vector machine. *Jurnal Komputasi*. Retrieved from <https://ejournal.jak-stik.ac.id/index.php/komputasi/article/view/3499>
- Maheswara, I. D. G. L., & Al'aziz, A. H. (2025). Perbandingan model machine learning pada klasifikasi curah hujan di Bogor. *Jurnal Teknologi Informasi (INTI)*, 19(2), 202–210. <https://doi.org/10.33480/inti.v19i2.6296>
- Pitaloka, E., Hartanto, T. B. A., & Sandiwarno, S. (2024). Penerapan machine learning untuk prediksi bencana banjir. *Jurnal Sistem Informasi Bisnis*.

- Panggabean, D. A. H., Sihombing, F. M., & Aruan, N. M. (2021). Prediksi tinggi curah hujan dan kecepatan angin berdasarkan data cuaca dengan penerapan algoritma artificial neural network (ANN). *Prosiding Seminastika*, 3(1), 1–7.
- Putri, D. M., & Kristianto, A. (2021). Prediksi debit Sungai Bengawan Solo menggunakan numerical weather model global forecast system dan integrated flood analysis system. *Jurnal Penelitian Pengelolaan Daerah Aliran Sungai*, 5(1), 41–50.
- Pessach, D., & Shmueli, E. (2022). A review on fairness in machine learning. *ACM Computing Surveys (CSUR)*, 55(3), 1–44.
- Rahman, F. D., Zulfa, M. I., & Taryana, A. (2024). Clustering dan klasifikasi data cuaca Kota Cilacap menggunakan K-Means dan Random Forest. *Jurnal Sains dan Teknologi*, 1(2), 90–97.
- Rifqi, M. N., & Aldisa, R. T. (2024). Penerapan metode support vector machine dalam memprediksi cuaca. *Jurnal Sistem dan Informatika*. 5(2), 368–379.
- Rochmawati, D. R. (2024). Prediksi cuaca dengan jaringan syaraf tiruan menggunakan Python. *Jurnal Teknologi Komputer dan Informatika*, 2(2), 176–186.
- Siswanto, E., & Meiliasari, M. (2024). Kemampuan pemecahan masalah pada pembelajaran matematika: Systematic literature review. *Jurnal Riset Pembelajaran Matematika Sekolah*, 8(1), 45–59.
- Sulistyowati, I. D., Sunarno, S., & Djuniadi, D. (2024). Penerapan machine learning dengan algoritma support vector machine untuk prediksi kelembapan udara rata-rata. *Just IT, : Jurnal Sistem Informasi, Teknologi Informasi dan Komputer*, 15(1), 284–290.
- Shin, Y., et al. (2024). LSTM-autoencoder based detection of time-series noise signals for water supply and sewer pipe leakages. *Water*, 16(18), 2631.
- Sutaryani, A., Sunarno, S., & Djuniadi, D. (2024). Perbandingan performa model machine learning dalam prediksi suhu di Semarang. *Jurnal Informatika dan Teknik Elektro Terapan*, 12(3), 234–240.
- Wiwaha, D. D., Setiawan, D. P., & Pramudita, B. A. (2024). Sistem prediksi cuaca wireless sensor network dan teknologi IoT dengan machine learning. *e-Proceeding of Engineering*, 11(4), 2831–2840.