

Prediksi Banjir menggunakan ANFIS-PCA sebagai Peringatan Dini Bencana Banjir

AGUSTINA RACHMAWARDANI^{1,2}, SASTRA KUSUMA WIJAYA¹,
PRAWITO¹, ARSHASENA SOPAHELWAKAN³

¹Departmen Fisika, FMIPA, Universitas Indonesia, Depok, Indonesia

²Sekolah Tinggi Meteorologi Klimatologi dan Geofisika, STMKG, Jakarta, Indonesia

³Badan Meteorologi Klimatologi dan Geofisika, BMKG Jakarta, Indonesia

E-mail: agustina.rahmawardani@stmkg.ac.id

Received 13 Oktober 2023 | Revised 24 November 2023 | Accepted 3 Desember 2023

ABSTRAK

Di antara kejadian bencana yang terjadi di Indonesia, 76 persen terdiri dari bencana hidrometeorologi seperti banjir, badai, longsor, dan kebakaran hutan. Provinsi DKI Jakarta sebagai daerah perkotaan sangat rentan terhadap banjir. Persamaan matematis yang kompleks dapat digunakan untuk memodelkan kejadian banjir secara fisik. Sistem pembelajar (machine learning) adalah sistem yang merancang dan mengembangkan algoritma yang menggunakan data historis untuk melakukan prediksi banjir. Dengan menggunakan data ini, sistem pembelajar dapat menghasilkan nilai probabilitas dasar, yang sangat membantu sistem prediksi, memberikan solusi yang lebih hemat biaya dan kinerja yang lebih baik. Prediksi yang akurat dan tepat dapat membantu strategi pengelolaan sumber daya air, analisis kebijakan dan rekomendasi serta pemodelan evakuasi lebih lanjut. Penelitian ini akan dibahas tentang Perancangan Sistem Peringatan Dini Banjir berbasis Ensemble Machine Learning sebagai mitigasi bencana banjir. Hasil dari penelitian menunjukkan nilai RMSE dari algoritma ANFIS – PCA adalah sebesar 0.12 dan koefisien korelasi (R^2) sebesar 0.856.

Kata kunci: Prediksi Banjir, Machine Learning, ANFIS, ANFIS – PCA

ABSTRACT

The nation of Indonesia is prone to disaster, with 76% of natural disasters being hydrometeorological, such as floods, landslides, tropical cyclones, and droughts. Flood occurrences can be physically modeled using complex mathematical equations. Machine Learning serves as a system for designing and developing algorithms that can predict flood events using historical data. Machine learning systems can leverage existing data to produce underlying probability values, making significant contributions to prediction systems that offer better performance and cost-effective solutions. Accurate predictions contribute to water resource management strategies, policy recommendations, and further evacuation modeling. This research will discuss an Early Warning Flood System design based on Ensemble Machine Learning as a flood disaster mitigation measure. The research results show that the RMSE value and coefficient correlation (R^2) for the ANFIS - PCA algorithm are 0.12 and 0.856, respectively.

Keywords: Flood Early Warning, Machine Learning, ANFIS, ANFIS – PCA

1. PENDAHULUAN

Perubahan pola dan intensitas elemen iklim selama periode waktu tertentu yang dapat dibandingkan (biasanya lebih dari 30 tahun) dikenal sebagai perubahan iklim. Perubahan iklim berlangsung dalam waktu yang cukup lama (*slow pace*), atau lambat, dan berubah secara perlahan. Perubahan perlahan ini disebabkan oleh berbagai peristiwa ekstrem yang terjadi pada variabilitas iklim secara konsisten. Salah satu konsekuensi yang timbul akibat adanya perubahan iklim adalah terjadinya curah hujan yang cukup ekstrem. Disaat terjadi perubahan alam ekstrem daerah yang mendapat surplus air bisa mengalami bencana hidrometeorologi **(Edvin, dkk, 2011)** .

Bencana hidrometeorologi, seperti banjir, badai, longsor, siklon tropis, dan kebakaran hutan, menyumbang 76% dari bencana yang terjadi di Indonesia. Banjir terjadi ketika volume air suatu badan air seperti sungai atau danau meluap atau melimpah dari bendungan, sehingga air keluar dari sungai dan merendam daratan **(Giarno, dkk, 2022)**. Banjir adalah kejadian ketika air meluap dan melimpah ke daerah daratan karena faktor seperti hujan lebat, salju yang meleleh, atau peristiwa alam seperti badai tropis atau letusan gunung berapi. Hal ini dapat mengakibatkan tergenangnya air di wilayah yang biasanya kering, yang berpotensi menyebabkan kerusakan pada properti, lingkungan, serta mengancam keselamatan manusia dan hewan.

DKI Jakarta adalah salah satu daerah perkotaan yang rentan terhadap banjir karena merupakan cekungan banjir yang dikelilingi oleh tiga belas sungai. Dimana tanah di bagian utara Jakarta telah mengalami pengangkatan tanah akibat proses tektonik **(Mauliate, dkk, 2017)**. Berdasarkan kondisi klimatologinya, DKI Jakarta memiliki tingkat curah hujan yang tinggi. Kementerian Pekerjaan Umum dan Perumahan Rakyat (PUPR) menyatakan bahwa berdasarkan data pengamatan BMKG di stasiun Tanjung Priuk menyatakan bahwa rata – rata curah hujan per tahun di DKI Jakarta adalah 2.500 mm – 3.000 mm **(Gustari, dkk, 2012)**. Jumlah curah hujan pada tahun 2020 mencapai 2.819,4 mm (Badan Penanggulangan Bencana Daerah **(BPBD, 2013)**). Kondisi tersebut membuat DKI Jakarta menjadi kawasan yang rawan terjadi banjir. Unit Pengelola Statistik DKI Jakarta (2020) mencatat pada tanggal 1 Januari 2020 terdapat sebanyak 157 kelurahan dari total 261 kelurahan di DKI Jakarta yang terdampak banjir hingga memakan 19 korban jiwa dan sebanyak 31.232 orang harus mengungsi (Bappenas, 2020). Kondisi tersebut disebabkan oleh tingginya curah hujan yang membuat meluapnya beberapa sungai yang melintasi Jakarta, dan adanya banjir kiriman dari daerah lain, selain itu kurang lengkapnya sistem peringatan dini banjir juga menjadi salah satu sebab banyaknya kerugian akibat bencana tersebut.

Curah hujan yang tinggi merupakan salah satu faktor terjadinya banjir di kota-kota besar terutama di Jakarta. Air sungai meluap terus membanjiri permukiman karena banyaknya bangunan yang berdiri di bantaran sungai membuat ruang badan sungai menyempit. Faktor tambahan adalah kekurangan Ruang Terbuka Hijau, juga dikenal sebagai RTH, yang mengurangi area resapan air dan menyebabkan banjir. Selain itu, pembangunan bangunan baru dan hotel-hotel di daerah Jakarta menyebabkan penggunaan air tanah yang tidak sesuai dengan kebutuhan.

Banjir yang terjadi di Jakarta membawa dampak negatif bagi masyarakat di antara aktifitas masyarakat terhenti sehingga akan membawa dampak kerugian ekonomi. Korban banjir juga mengalami kerugian ekonomi yang signifikan karena mereka kehilangan banyak aset

mereka. Masyarakat di daerah yang sering terkena banjir akan sulit untuk berkembang dan menjadi lebih produktif karena kerugian ekonomi yang disebabkan oleh banjir. Akibatnya, penghambatan peningkatan kesejahteraan yang disebabkan oleh banjir yang tidak segera ditangani malah dapat mengakibatkan peningkatan kemiskinan warga sekitar. **(Rachmawardani, dkk, 2022)**. Sistem peringatan dini merupakan suatu hal yang penting dalam upaya mengurangi dampak buruk dan kerugian akibat banjir. BNPB menjelaskan sistem peringatan dini merupakan rangkaian pemberian informasi dan peringatan kepada masyarakat akan adanya potensi bencana untuk mengurangi kerugian dan dampak buruknya. Kerugian secara langsung dari bencana banjir yang terjadi dapat dilihat dari banyaknya kerusakan bangunan-bangunan rumah tinggal, terhambatnya fasilitas umum dan infrastruktur, serta adanya rasa trauma. Dengan adanya suatu sistem peringatan dini banjir diharapkan dapat berperan dalam mitigasi dan evakuasi dini sehingga dapat meminimalisir kerugian akibat bencana banjir, baik kerugian harta maupun jiwa.

Persamaan matematis yang kompleks dapat digunakan untuk memodelkan kejadian banjir secara fisik. Sistem pembelajar (*machine learning*) adalah sistem yang merancang dan mengembangkan algoritma yang dapat melakukan prediksi kejadian banjir dengan menggunakan data historis. Dengan menggunakan data ini, sistem pembelajar dapat menghasilkan nilai probabilitas dasar, yang sangat membantu sistem prediksi, memberikan solusi yang lebih hemat biaya dan kinerja yang lebih baik. Prediksi yang akurat dan tepat dapat membantu strategi pengelolaan sumber daya air, pemodelan evakuasi serta saran dan analisis kebijakan yang bermanfaat bagi masyarakat. Dalam penelitiannya **(Ortiz-García, dkk, 2014)** menjelaskan bagaimana penggunaan *Machine Learning (ML)* dapat digunakan untuk memodelkan sistem hidrologi yang kompleks seperti banjir. Banyak algoritma *Machine Learning (ML)* berhasil memprediksi banjir baik dalam jangka pendek maupun jangka panjang. Algoritma seperti Jaringan Syaraf Tiruan (JST) **(Kim, dkk, 2016)**, *neurofuzzy* **(Mosavi & Edalatifar, 2019)** **(Dineva, dkk, 2014)**, *Support Vector Machine (SVM)* **(Premalatha, dkk, 2022)**, dan *Support Vector Regresi (SVR)* **(Taherei, dkk, 2018)**. Selain itu, hibridisasi dengan metode ML lainnya, teknik komputasi lunak, simulasi numerik, dan/atau model fisik dapat meningkatkan kinerja ML.

Model berbasis data – Data Driven Model (DDM) seringkali perlu dilatih dengan *dataset* yang mewakili sistem **(Solomatine, dkk, 2008)**. Diantara berbagai DDM yang sudah dibahas dalam berbagai literatur model, *Neuro-Fuzzy Models* atau NFM baru-baru ini telah mendapatkan popularitas karena kemampuan penalaran dan pembelajarannya. Penerapan awal *Neuro Fuzzy Model* dalam prediksi aliran sungai dapat dilihat pada penelitian **(Chang, dkk, 2001)**. Penelitian-penelitian berikut menggunakan algoritma ANFIS dalam pemodelannya. **(Wang, dkk, 2009)** membandingkan model ANFIS, *Autoregressive Moving Average (ARMA)*, *Artificial Neural Network (ANN)*, *Genetic Programming (GP)* dan *Support Vector Machine (SVM)* untuk memprediksi debit bulanan untuk PLTA Manwan di Sungai Lancangjiang dan PLTA Hongjiadu di Sungai Wujiang, Cina pada tahun 2009. ANFIS, GP dan SVM mencapai hasil terbaik. **(Firat & Turan, 2010)** menerapkan model ANFIS dalam memprediksi debit sungai bulanan dengan menggunakan data dari Sungai Göksu, Turki bagian selatan. **(Talei, dkk, 2010)** menggunakan ANFIS untuk pemodelan kejadian curah hujan-limpasan dan menunjukkan keunggulan ANFIS dibandingkan model fisik SWMM (*Storm Water Management Model*). **(Talei & Chua, 2012)** juga membahas pengaruh jeda waktu dan pemilihan input untuk meningkatkan kinerja model ANFIS. **(Ghalkhani, dkk, 2013)** menggunakan model ANFIS untuk sebuah sungai di Iran dan menyarankan penggunaan ANFIS sebagai sistem prakiraan banjir dan sistem peringatan dini. **(Nguyen, dkk, 2018)** menggunakan ANFIS untuk memperkirakan tingkat ketinggian di Hilir Sungai

Mekong. Dalam penelitian yang lebih baru, **(Chang, dkk, 2016)** menggunakan ANFIS dalam meramalkan curah hujan dan aliran sungai di Waduk Shihmen, Taiwan.

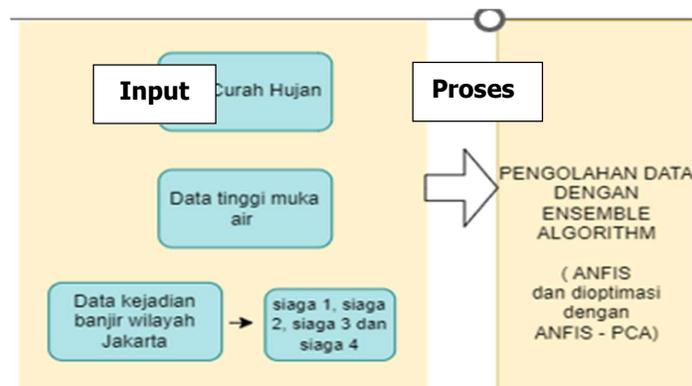
Principal Component Analysis (PCA) dapat mengurangi besarnya dimensi dari data yang diobservasi menjadi dimensi yang lebih kecil tanpa kehilangan informasi yang signifikan dalam menggambarkan keseluruhan data. Banyak penelitian yang mengeksplorasi berbagai metode untuk mengurangi dimensi data banjir sebelum menetapkan model prediksi banjir **(Chang, dkk, 2001)**. Analisis komponen utama (PCA) adalah metode statistik yang diterapkan untuk memproyeksikan data ke variabel (komponen) linier dan tidak berkorelasi menggunakan transformasi ortogonal. Komponen utamanya (PC) dapat menyimpan informasi sebanyak mungkin dalam data, dan oleh karena itu sering digunakan untuk analisis data dan reduksi dimensi **(Chang, dkk, 2001)**. Lin mengembangkan jaringan *backpropagation* berdasarkan beberapa data aliran masuk untuk memprediksi genangan banjir maksimum di wilayah perkotaan, dengan dimensi data pelatihan (kurva debit 50 dimensi) dikurangi menjadi tujuh PC (dengan varian kumulatif menjelaskan 97% dari varians total) oleh PCA. **(Lin, dkk, 2020)**

Jadi, penelitian ini akan membahas perancangan sistem peringatan dini banjir yang menggunakan pembelajaran berbasis *hybrid Machine Learning* ANFIS – PCA sebagai mitigasi untuk bencana banjir. Salah satu cara untuk mengatasi banjir adalah dengan membangun sistem peringatan dini yang akan memberi tahu orang-orang tentang bencana tersebut. Dengan memberikan informasi terkini dan aktual, sistem peringatan dini ini diharapkan dapat mengurangi kerugian ekonomi akibat kematian. Sesuai dengan Perpres Nomor 87 tahun 2020, yang menetapkan visi penanggulangan bencana "Mewujudkan Indonesia Tangguh Bencana untuk Pembangunan Berkelanjutan", sistem peringatan dini diperlukan untuk kesiapsiagaan (persediaan) sebelum bencana dan mitigasi banjir **(Presiden Republik Indonesia, 2020)**.

2. METODE

Pada bagian metode penelitian ini penulis akan menjelaskan konsep sistem secara keseluruhan yang meliputi blok diagram sistem, *dataset*, algoritma sistem dengan *ensemble machine learning* dan diagram alir.

2.1 Blok Diagram Sistem



Gambar 1. Blok Diagram Sistem

Gambar 1 di atas menunjukkan blok diagram sistem yang terdiri atas:

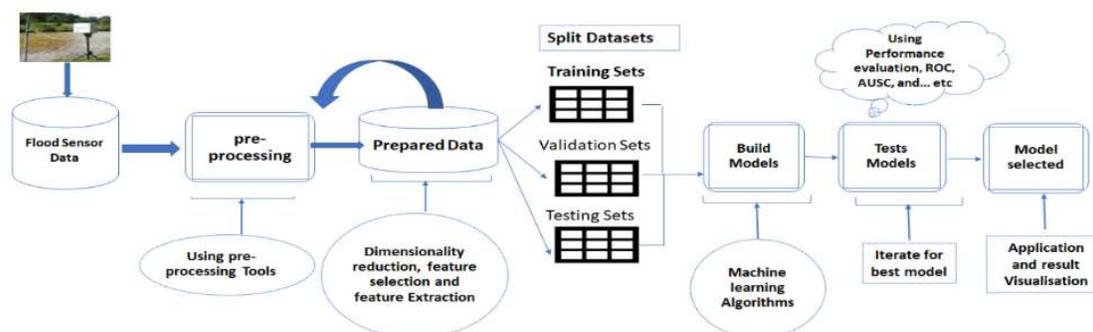
1. *Input* sistem
Input sistem terdiri atas data curah hujan dan data tinggi muka air wilayah / daerah Katulampa, Depok dan Manggarai mulai tahun 2014 – 2020, serta data kejadian banjir wilayah Bidara Cina mulai tahun 2014 – 2020.
2. Proses
Proses pengolahan atau algoritma yang digunakan dalam prediksi banjir ini adalah menggunakan metode *Ensemble Machine Learning* yang terdiri atas ANFIS (*Adaptive Neuro Fuzzy Inference System*) dan dioptimasi menggunakan PCA-ANFIS.
3. *Output* Sistem
Output sistem peringatan dini banjir ini adalah menyampaikan informasi tentang peristiwa banjir dan apabila nilai kejadian banjir melebihi *threshold* yang ada (curah hujan - hujan lebat 100 mm) dan tinggi muka air pada siaga 1 - Katulampa : 200 cm, Depok : 350 cm dan Manggarai : 950 cm) maka akan ditampilkan peringatan dini banjir.

2.2 Alur Penelitian

Dataset yang digunakan adalah data tinggi muka air dan curah hujan kemudian dilakukan *preprocessing* data dengan teknik *Interquartile Range (IQR)*. *IQR* disebut juga sebagai teknik untuk mendeteksi ketidaknormalan data atau *Outlier* dengan cara membagi *dataset* menjadi empat bagian sama besar menggunakan *quartile range* Q1, Q2 dan Q3. Data yang berada di luar batas bawah dan atas *interquartile range* disebut dengan *Outlier*. *IQR* disebut juga dengan perbedaan antara Q1 dan Q3 ($Q3 - Q1$). Dari nilai *IQR* dapat ditentukan *Outlier* data (Agarwal, dkk, 2014) (Alabrah, dkk, 2023). *Outlier* dapat dibagi menjadi *lower Outlier* dan *upper Outlier*, dengan perhitungan sebagai berikut :

- a. ***Outlier* batas bawah kuartil 1**
 $BBP = Q1 - 1.5 IQR$
- b. ***Outlier* batas atas kuartil 1**
 $BAP = Q3 + 1.5 IQR$
- c. ***Outlier* batas bawah kuartil 3**
 $BBPJ = Q1 - 3 IQR$
- d. ***Outlier* batas bawah kuartil 3**
 $BAPJ = Q3 + 3 IQR$

Selanjutnya Data yang dimasukkan ke dalam model terdiri dari data pelatihan dan penilaian. Data pelatihan dan penilaian ini akan dipisahkan melalui validasi split dengan perbandingan 80 : 20, di mana 80 adalah data pelatihan dan 20 adalah data penilaian. Untuk memeriksa *dataset* penelitian, algoritma ANFIS dan ANFIS-PCA digunakan. Metode yang paling efektif menghasilkan nilai Mean Absolut Error dan nilai Root Mean Square Error terkecil dari model yang dipelajari dalam penelitian. Kemudian *outputnya* adalah nilai prediksi tinggi muka air (Caesarendra, dkk, 2017). Alur penelitian prediksi dapat dilihat pada Gambar 2.



Gambar 2. Alur Penelitian (Khalaf, dkk, 2020)

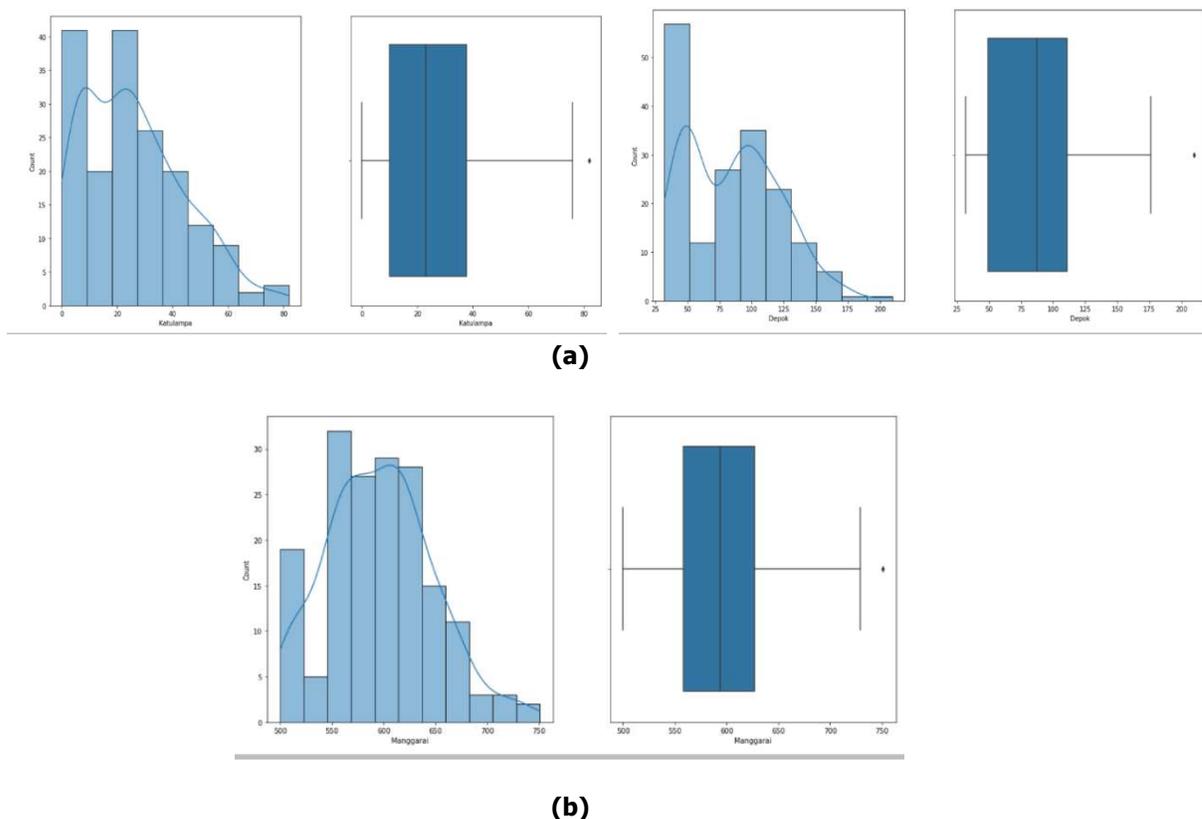
2.3 Dataset

Dataset merupakan sekumpulan data yang disusun secara terstruktur. Secara umum *dataset* direpresentasikan dalam bentuk tabel yaitu baris dan kolom. Tiap baris dan kolom mewakili variabel tertentu. *Dataset* yang digunakan dalam penelitian sistem prediksi dini banjir ini adalah:

- Data ketinggian tinggi muka air wilayah DAS Ciliwung antara lain TMA Katulampa, TMA Depok dan TMA Manggarai mulai tahun 2014 - 2020 yang bisa didapat dari (<https://pantaubanjir.jakarta.go.id>, n.d.)
Ketinggian muka air Katulampa dibagi menjadi empat kategori yaitu kategori siaga 1 ketinggian di atas 200 cm, siaga 2 dengan ketinggian 150 -199 cm, siaga 3 dengan ketinggian 80 – 149 cm, dan siaga 4 dengan ketinggian kurang dari 79 cm. Tinggi muka air DAS Wilayah Depok dibagi menjadi empat kategori yaitu siaga 4 dengan nilai < 199, siaga 3 (200 – 269) , siaga 3 (270 – 349) dan siaga 1 > 350. Data DAS Manggarai dibagi menjadi empat kategori yaitu Siaga 4 dengan ketinggian < 749, siaga 3 dengan ketinggian 750 – 849, siaga 2 dengan ketinggian 850 – 949 dan siaga 1 dengan ketinggian > 950 (**Geofisika et al., 2012**).
- Data Curah hujan wilayah Katulampa, Depok dan Manggarai mulai tahun 2014 - 2020 yang didapat dari web BMKG (awscenter@bmgk.go.id, BMKG). Gambar 4 menunjukkan *site* alat ukur curah hujan di wilayah Katulampa, Depok dan Manggarai.
- Data kejadian banjir dari Jakarta Opendata
Data kejadian banjir wilayah Jakarta mulai tahun 2014 – 2020 dapat didownload dari web Jakarta Opendata dengan alamat :(<https://data.jakarta.go.id/dataset/rekapitulasi-kejadian-banjir-pertahun/resource>), Pemprov DKI Jakarta)

2.4 Transformasi Data

Proses hasil filtering menggunakan *Interquartile Range* (IQR) ditunjukkan pada Gambar 3.



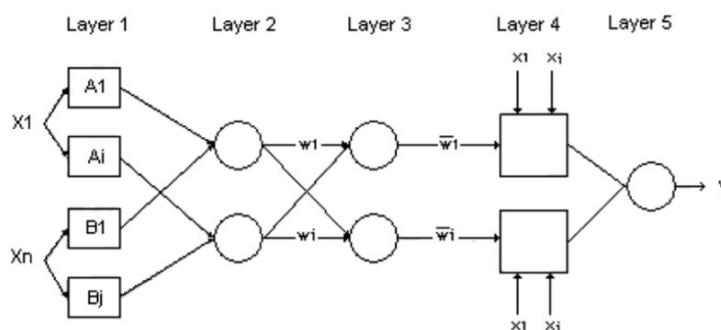
Gambar 3. Data Tinggi Muka Air Katulampa, Depok dan Manggarai (a) Grafik Distribusi Data Setelah Dihilangkan *Outlier* (b) Distribusi Data dalam *Interquartile Range* (IQR)

Dari Gambar 3 dapat dilihat bahwa data hasil *filtering* dengan menggunakan IQR sudah berada pada *range* distribusi normal dan data *Outlier* sudah tidak ada.

2.5 Adaptive Neuro Fuzzy Inference System (ANFIS)

Sistem *Inferensi Neuro Fuzzy Adaptif* (ANFIS) adalah kombinasi antara logika *fuzzy* dan Jaringan Syaraf Tiruan (JST). Kelebihan logika *fuzzy* terletak pada kemampuannya untuk memodelkan aspek kualitatif pengetahuan manusia dan proses pengambilan keputusan dengan menerapkan basis aturan. Jaringan Syaraf Tiruan, di sisi lain, memiliki keunggulan dalam mengenali pola dan memecahkan masalah tanpa bergantung pada pemodelan matematik. ANFIS, karena gabungan keduanya, mampu memanfaatkan data historis untuk melakukan prediksi terhadap peristiwa di masa depan yang menggabungkan kelebihan logika *fuzzy* dan kemampuan jaringan syaraf tiruan (Chatterjee, dkk, 2020).

Sistem ANFIS terdiri dari lima lapisan: lapisan fuzzifikasi, lapisan aturan, lapisan normalisasi, lapisan defuzzifikasi, dan hasil neuro tunggal (Ata, dkk, 2010). Struktur ANFIS ditunjukkan pada Gambar 4.



Gambar 4. Struktur ANFIS (Ata, dkk, 2010)

Disebut dengan *fuzzifikasi layer*. *Fuzzifikasi layer* menggunakan fungsi *membership* dari nilai *input*.

$$Z_{1,i} = \epsilon_{A_i}(x) \text{ untuk } i = 1,2 \quad (1)$$

$$Z_{1,i} = \epsilon_{B_{i-2}}(x) \text{ untuk } i = 3,4 \quad (2)$$

Simpul $Z_{1,i}$ menyatakan derajat keanggotaan masing-masing masukan terhadap himpunan fuzzy A dan B.

Lapisan kedua: *Layer* ini disebut *rule layer*. Pembobotan atau (w_1) dari rule ini didapat dari nilai *membership* yang dihitung di *layer fuzzifikasi*.

$$Z_{2,i} = w_i = \epsilon_{A_i}(x) \cdot \epsilon_{B_i}(y) \text{ untuk } i = 1,2 \quad (3)$$

Lapisan ketiga: *Layer* ini disebut *layer normalisasi*. Menghitung normalisasi pembobotan dari masing-masing *rule* :

$$Z_{3,i} = \frac{w_i}{w_1 + w_2 + w_3 + w} \text{ untuk } i = 1,2 \quad (4)$$

Lapisan keempat: merupakan *layer defuzzifikasi*. Bobot masing-masing rule dihitung pada masing-masing *node layer*.

$$Z_{4,i} = w_i f_i = (p_i x + q_i y + r_i) \quad (5)$$

$\{p_i, q_i, r_i\}$ adalah himpunan parameter yang disebut *consequent parameter*.

Lapisan kelima: *Layer* penjumlahan. Aktual *output* dari ANFIS didapat dengan menjumlahkan *output* masing-masing *rule* pada *layer defuzzifikasi*.

$$Z_{5,i} = \frac{\sum w_i f_i}{\sum w_i} = \sum_i w_i f_i \quad (6)$$

2.6 Principal Component Analysis (PCA)

Dalam statistika, Principal Component Analisis (PCA) atau Analisis Komponen Utama adalah teknik yang menyederhanakan data melalui transformasi linier untuk membuat sistem koordinat baru dengan varians yang paling besar. PCA juga dapat mengurangi dimensi data secara signifikan (**Smith, 2006**).

Menemukan *eigenvector* dan *eigenvalue* adalah hal utama PCA. Pada dasarnya, tujuan proses PCA adalah untuk menyederhanakan variabel yang diamati dengan mengurangi dimensinya. Ini dicapai dengan mengubah variabel bebas awal menjadi variabel baru yang tidak memiliki korelasi atau komponen utama sama sekali. Untuk proses transformasi, *eigenvector* dengan *eigenvalue* yang besar sangat penting. Oleh karena itu, tidak akan ada kehilangan data yang signifikan jika dimensi direduksi dengan menghilangkan *eigenvector* dan *eigenvalue* yang sangat kecil (**Smith, 2006**).

Berikut adalah langkah-langkah dalam PCA (**Karamizadeh, dkk, 2013**):

Langkah 1: vektor kolom atau baris berukuran $N \times 2$ mewakili kumpulan gambar M ($B_1, B_2, B_3, \dots, B_M$) dengan ukuran $N \times N$.

Langkah 2: rata-rata gambar *training set* (μ) digambarkan sebagai:

$$\mu = \frac{1}{m} \sum_{n=1}^m B_n \quad (7)$$

Langkah 3 : rata-rata *image vector* berbeda dengan masing-masing *image* pelatihan

$$W_i = B_i - \mu \quad (8)$$

Langkah 4 : Total *scatter matrix* atau *covarian matrix* dihitung dari Φ sebagaimana ditunjukkan di bawah :

$$C = \sum_{n=1}^M w_n w_n^T = A A^T \quad (9)$$

dimana $A = [W_1 W_2 W_3 \dots W_n]$

Langkah 5 : Hitung *eigenvector* UL dan *eigenvalue* λ_L dari kovarian matrik C

Langkah 6 : Untuk klasifikasi gambar, fitur ini dapat dimanfaatkan. Hitung *vector* bobot:

$$\Omega^T = [w_1, w_2, \dots, w_M^T], \quad (10)$$

dimana

$$H_k = U^T (B - \mu), \quad k = 1, 2, \dots, M'$$

3. HASIL DAN PEMBAHASAN

3.1 Deskripsi data

Dataset yang digunakan pada penelitian ini sebanyak 1383 data untuk masing-masing parameter. Data diambil tahun 2014 – 2020. Untuk mengetahui statistik data dapat dilihat pada Tabel 1.

Tabel 1. Deskripsi *Dataset*

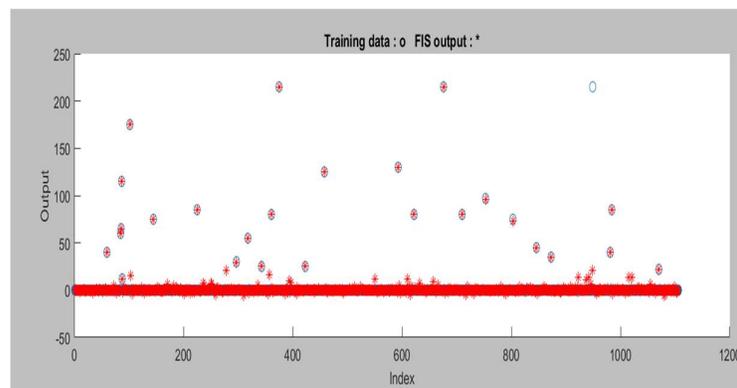
	Katulampa	Depok	Manggarai	ARG Depok	ARG Manggarai	ARG Katulampa	Bidara Cina
count	1383.000000	1383.000000	1383.000000	1383.000000	1383.000000	1383.000000	1383.000000
mean	38.221981	111.018800	609.906725	2.418221	0.872191	2.636443	1.873463
std	24.087889	39.006507	59.860275	4.935653	2.056546	4.960100	14.891864
min	0.000000	25.000000	495.000000	0.000000	0.000000	0.000000	0.000000
25%	20.000000	81.000000	569.000000	0.000000	0.000000	0.000000	0.000000
50%	36.000000	108.000000	607.000000	0.000000	0.000000	0.000000	0.000000
75%	54.000000	134.000000	647.000000	2.000000	0.400000	2.800000	0.000000
max	118.000000	236.000000	804.000000	22.000000	10.400000	21.000000	215.000000

Tabel 1 menunjukkan jumlah data yang digunakan dalam penelitian ini. Jumlah keseluruhan data yang digunakan adalah 1383 data, dan *mean* adalah nilai tengah dari masing-masing atribut penelitian. *Dataset* standar deviasi atau simpangan penelitian adalah *std*, dan nilai *min* adalah nilai paling kecil dari *dataset*. Nilai 25%, 50%, dan 75% menunjukkan jumlah data di kuartil 1, kuartil 2 dan kuartil 3, dan nilai maksimum adalah nilai paling tinggi dari *dataset* penelitian

3.2 Hasil Pelatihan dan Pengujian ANFIS

Data yang digunakan pada pelatihan dan pengujian ANFIS dibagi menjadi 80% data pelatihan dan 20% data pengujian dari total data sebanyak 1383 data.

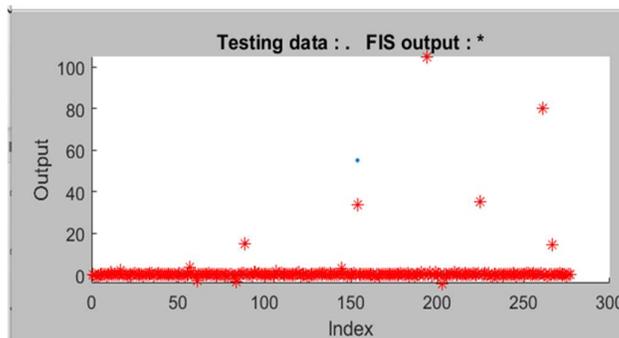
Parameter yang digunakan pada pelatihan ANFIS adalah sebagai berikut: *Membership function gaussian 3 3 3 3 3 3* dengan 6 *input* dan 1 *output*. Pemilihan tipe *membership function* tidak ada perbedaan hasil yang signifikan ketika menggunakan model yang berbeda, baik segitiga maupun trapezium. Jumlah *Epoch* yang digunakan 5, 10, 15 *Epoch* serta ada 729 *Rules*. Nilai RMSE terbaik dari proses pelatihan ANFIS adalah sebesar 6.12.



Gambar 5. Grafik Perbandingan Hasil Prediksi dan Data Aktual Data Training

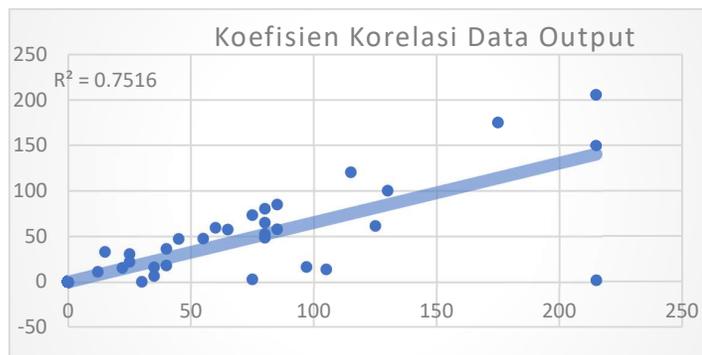
Kemudian setelah proses pelatihan maka selanjutnya melakukan pengujian pada data *testing*. Data *testing* yang digunakan adalah sebanyak 277 data. Parameter yang digunakan pada pelatihan ANFIS adalah sebagai berikut: *Membership function gaussian 3 3 3 3 3 3* dengan 6 *input* dan 1 *output*. Jumlah *Epoch* yang digunakan 5, 10, 15 *Epoch* serta ada 729

Rules. Nilai RMSE terbaik dari proses *testing* ANFIS adalah sebesar 1.64. Gambar 5 merupakan *output* data *testing*.



Gambar 6. Grafik Perbandingan Hasil Prediksi dan Data Aktual Data Testing

Pada Gambar 6 menerangkan koefien korelasi (R^2) antara data *output* actual dan data *output* prediksi. Nilai Koefisien korelasi data *output* actual dan data *output* prediksi dengan menggunakan algoritma ANFIS adalah sebesar 0.75.



Gambar 7. Korelasi Data *Output* Aktual dan *Output* Prediksi

3.3 Hasil pelatihan dan Pengujian ANFIS – PCA

Metode PCA merupakan salah satu metode reduksi dimensi pada *preprocessing machine learning*. Penerapan algoritma dalam proses PCA serta analisis dalam menentukan variabel PCA adalah sebagai berikut:

1. Menentukan matriks *dataset* dengan data yang telah distandarisasi secara keseluruhan ($X = [x_1, x_2, \dots, x_n]$)

Tabel 2. Matrik *Dataset* Standarisasi

	Katulampa	Depok	Manggarai	ARG Depok	ARG Manggarai	ARG Katulampa
0	0.385	-0.373	0.518	-0.272	-0.406	-0.585
1	0.167	-0.811	0.068	-0.129	-0.406	-0.585
2	0.210	-0.933	0.082	0.585	-0.406	-0.182
3	0.254	0.431	1.664	-0.558	-0.406	-0.585
4	1.996	1.648	2.277	1.157	-0.406	2.234

2. Menghitung nilai *covariance matrix* sebagai berikut:

Tabel 3. Covariance Matrix

	PC.1	PC.2	PC.3	PC.4	PC.5	PC.6
0	-0.59	-0.771	0.025	-0.235	-0.042	0
1	-0.493	0.533	0.067	-0.546	0.241	0.334
2	-0.48	0.286	-0.222	0.235	0.039	-0.763
3	-0.229	-0.015	0.551	0.596	0.503	0.188
4	-0.026	0.124	0.762	-0.158	-0.568	-0.237
5	-0.354	0.156	-0.247	0.459	-0.603	0.464

3. Nilai dari *eigenvector* dan *eigenvalue* dari hasil *decomposisi covariance matrix* yang telah diurutkan adalah sebagai berikut:

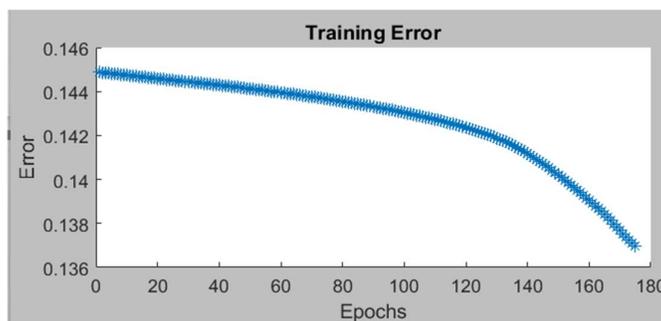
Tabel 4. Loading Factor (Eigenvector) yang Diurutkan Berdasarkan Eigenvalues

	0	1	2	3	4	5
0	1.004	0.621	0.51	0.185	0.079	0.357
1	0.621	1.004	0.3	0.175	-0.003	0.15
2	0.51	0.3	1.004	0.153	-0.086	0.297
3	0.185	0.175	0.153	1.004	0.101	0.025
4	0.079	-0.003	-0.086	0.101	1.004	0.012
5	0.357	0.15	0.297	0.025	0.012	1.004

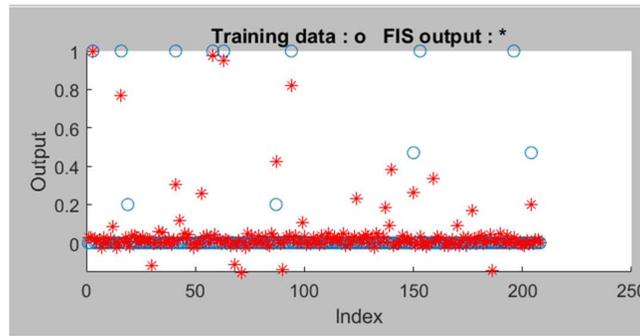
Tabel 5. Eigenvalue Yang Telah Diurutkan

	PC.1	PC.2	PC.3	PC.4	PC.5	PC.6
0	2.227	0.304	1.101	0.824	0.94	0.628

Setelah melakukan tahap seleksi variabel menggunakan teknik PCA dilakukan uji *training* dan *testing* model sistem prediksi peringatan dini banjir menggunakan algoritma ANFIS – PCA. Untuk proses training parameter yang digunakan adalah sebagai berikut *Membership function gaussian* 3 3 dengan 2 *input* dan 1 *output*. Jumlah *epoch* maksimal adalah 175 *epoch* dan nilai RMSE sebesar 0.13

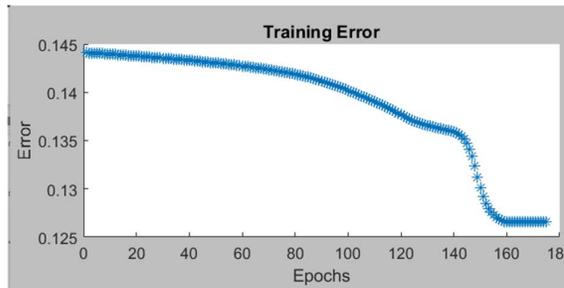


Gambar 8. Nilai RMSE Algoritma ANFIS – PCA dengan 175 Epoch

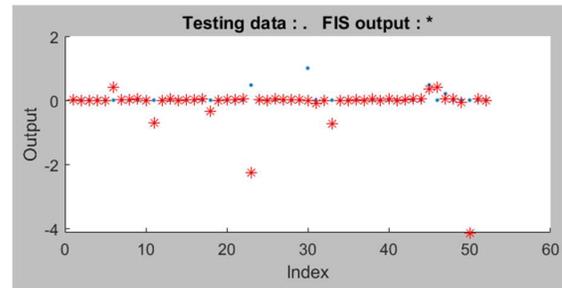


Gambar 9. Perbandingan Hasil Prediksi dan Data Aktual

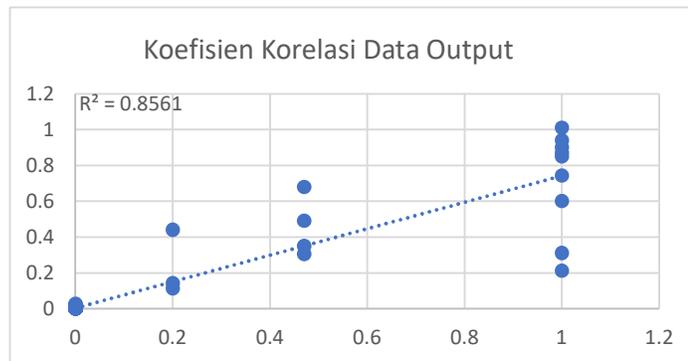
Gambar 8 dan 9 masing masing merupakan nilai RMSE dari data *training* yaitu sebesar 0.13 dan perbandingan dari data aktual banjir dan hasil prediksi banjir. Sedangkan nilai RMSE untuk data *testing* dapat dilihat pada Gambar 10 adalah sebesar 0.127 dan hasil prediksi dari data *testing* dapat dilihat pada Gambar 11.



Gambar 10. Nilai RMSE Data *Testing*



Gambar 11. Perbandingan Data Aktual dan Hasil Prediksi



Gambar 10. Korelasi Data *Output* Aktual dan *Output* Prediksi

Gambar 12 menunjukkan nilai koefisien korelasi data *output* yaitu data aktual dan data prediksi dari algoritma ANFIS – PCA sebesar 0.85.

4. KESIMPULAN

Telah berhasil dibuat model algoritma *ANFIS* dan diperbaiki dengan metode *ANFIS- PCA* untuk memprediksi banjir dengan menggunakan *dataset* yang sudah dilakukan *preprocessing*. Akurasi yang didapatkan pada algoritma *ANFIS* pada data *training* dan data *testing* dengan RMSE sebesar 6.12 dan 1.64. Akurasi yang didapatkan pada algoritma *ANFIS PCA* pada data *training* dan data *testing* dengan nilai RMSE-nya sebesar 0.13 dan 0.12. Dengan melakukan reduksi dimensi menggunakan teknik PCA, maka algoritma *ANFIS – PCA* menghasilkan nilai RMSE yang lebih kecil dibanding algoritma *ANFIS*. Hasil grafik koefisien korelasi antara Data aktual banjir dan hasil prediksi banjir dengan menggunakan algoritma *ANFIS* menunjukkan nilai 0.79 sedangkan dengan menggunakan algoritma *ANFIS- PCA* sebesar 0.856. Hal ini menunjukkan nilai koefisien korelasi algoritma *ANFIS – PCA* menunjukkan angka yang lebih bagus.

DAFTAR RUJUKAN

- Agarwal, S. (2014). Data mining: Data mining concepts and techniques. In *Proceedings - 2013 International Conference on Machine Intelligence Research and Advancement, ICMIRA 2013*. <https://doi.org/10.1109/ICMIRA.2013.45>
- Alabrah, A. (2023). An Improved CCF Detector to Handle the Problem of Class Imbalance with *Outlier* Normalization Using IQR Method. *Sensors*, 23(9). <https://doi.org/10.3390/s23094406>
- Ata, R., & Kocyigit, Y. (2010). An adaptive neuro-fuzzy inference system approach for prediction of tip speed ratio in wind turbines. *Expert Systems with Applications*, 37(7), 5454–5460. <https://doi.org/10.1016/j.eswa.2010.02.068>
awscenter@bmgk.go.id. (n.d.). *BMKG*.
- Badan Penanggulangan Bencana Daerah (BPBD). (2013). *Disaster Management Plan of DKI Jakarta Province 2013-2017*. 50–63.
- Caesarendra, W., & Pamungkas, D. (2017). *EMG based Classification of Hand Gestures using PCA and ANFIS*. 18–23.
- Chang, F. J., & Chen, Y. C. (2001). A counterpropagation fuzzy-neural network modeling approach to real time streamflow prediction. *Journal of Hydrology*, 245(1–4), 153–164. [https://doi.org/10.1016/S0022-1694\(01\)00350-X](https://doi.org/10.1016/S0022-1694(01)00350-X)
- Chatterjee, S., & Das, A. (2020). A novel systematic approach to diagnose brain tumor using integrated type-II fuzzy logic and *ANFIS* (adaptive neuro-fuzzy inference system) model. *Soft Computing*, 24(15), 11731–11754. <https://doi.org/10.1007/s00500-019-04635-7>

- Dineva, A., Várkonyi-Kóczy, A. R., & Tar, J. K. (2014). Fuzzy expert system for automatic wavelet shrinkage procedure selection for noise suppression. *INES 2014 - IEEE 18th International Conference on Intelligent Engineering Systems, Proceedings*, 163–168. <https://doi.org/10.1109/INES.2014.6909361>
- Edvin Aldrian, Mimin Karmini, B. (2011). *Adaptasi dan Mitigasi Perubahan Iklim di Indonesia*.
- Firat, M., & Turan, M. E. (2010). Monthly river flow forecasting by an adaptive neuro-fuzzy inference system. *Water and Environment Journal*, 24(2), 116–125. <https://doi.org/10.1111/j.1747-6593.2008.00162.x>
- Geofisika, D., Meteorologi, D., Matematika, F., Ilmu, D., & Alam, P. (2012). *PENENTUAN BATAS AMBANG CURAH HUJAN PENYEBAB BANJIR (Studi Kasus DAS Ciliwung Hulu) ARRIDHA DARA KOMEJI*.
- Ghalkhani, H., Golian, S., Saghafian, B., Farokhnia, A., & Shamseldin, A. (2013). Application of surrogate artificial intelligent models for real-time flood routing. *Water and Environment Journal*, 27(4), 535–548. <https://doi.org/10.1111/j.1747-6593.2012.00344.x>
- Giarno, G., Saputra, A. H., & Rachmawardani, A. (2022). Optimalisasi Edukasi Informasi Geohidrometeorologi Untuk Masyarakat Perkotaan (Studi Kasus: Kelurahan Jurang Mangu Timur, Kecamatan Pondok Aren, Kota Tangerang Selatan, Banten). *To Maega: Jurnal Pengabdian Masyarakat*, 5(3), 554. <https://doi.org/10.35914/tomaega.v5i3.1294>
- Gustari, I., Hadi, T. W., Hadi, S., & Renggono, F. (2012). Akurasi Prediksi Curah Hujan Harian Operasional Di Jabodetabek: Perbandingan Dengan Model Wrf. *Jurnal Meteorologi Dan Geofisika*, 13(2), 119–130. <https://doi.org/10.31172/jmg.v13i2.126>
<https://data.jakarta.go.id/dataset/rekapitulasi-kejadian-banjir-pertahun/resource>. (n.d.). *Pemprov DKI Jakarta*.
- Karamizadeh, S., Abdullah, S. M., Manaf, A. A., Zamani, M., & Hooman, A. (2013). An Overview of Principal Component Analysis. *Journal of Signal and Information Processing*, 04(03), 173–175. <https://doi.org/10.4236/jsip.2013.43b031>
- Kim, S., Matsumi, Y., Pan, S., & Mase, H. (2016). A real-time forecast model using artificial neural network for after-runner storm surges on the Tottori coast, Japan. *Ocean Engineering*, 122, 44–53. <https://doi.org/10.1016/j.oceaneng.2016.06.017>
- Mauliate, H., Gromiko, A., Unggul, U. E., Tomang, T., & Jeruk, K. (2017). Evaluasi ketangguhan kota terhadap bencana banjir. *Jurnal Planesa*, 8(1), 41–53.

- Mosavi, A., & Edalatifar, M. (2019). A Hybrid Neuro-Fuzzy Algorithm for Prediction of Reference Evapotranspiration. In *Lecture Notes in Networks and Systems* (Vol. 53). Springer International Publishing. https://doi.org/10.1007/978-3-319-99834-3_31
- Nguyen, P. K. T., Chua, L. H. C., Talei, A., & Chai, Q. H. (2018). Water level forecasting using neuro-fuzzy models with local learning. *Neural Computing and Applications*, *30*(6), 1877–1887. <https://doi.org/10.1007/s00521-016-2803-9>
- Ortiz-García, E. G., Salcedo-Sanz, S., & Casanova-Mateo, C. (2014). Accurate precipitation prediction with support vector classifiers: A study including novel predictive variables and observational data. *Atmospheric Research*, *139*, 128–136. <https://doi.org/10.1016/j.atmosres.2014.01.012>
- Premalatha, G., & Bai, V. T. (2022). Wireless IoT and Cyber-Physical System for Health Monitoring Using Honey Badger Optimized Least-Squares Support-Vector Machine. *Wireless Personal Communications*, *124*(4), 3013–3034. <https://doi.org/10.1007/s11277-022-09500-9>
- Presiden Republik Indonesia. (2020). Peraturan Presiden Republik Indonesia Nomor 87 tahun 2020 tentang Rencana Induk Penanggulangan Bencana Tahun 2020-2044. *Database Peraturan BPK RI*, *87*, 1–31. <https://peraturan.bpk.go.id/Home/Details/146481/perpres-no-87-tahun-2020>
- Rachmawardani, A., Wijaya, S. K., & Shopaheluwakan, A. (2022). Sistem Peringatan Dini Banjir Berbasis Machine Learning: Studi Literatur. *METHOMIKA Jurnal Manajemen Informatika Dan Komputerisasi Akuntansi*, *6*(6), 188–198. <https://doi.org/10.46880/jmika.vol6no2.pp188-198>
- Smith, L. (2006). A tutorial on PCSA. *Department of Computer Science, University of Otago.*, 12–28.
- Solomatine, D., See, L. M., & Abrahart, R. J. (2008). Data-Driven Modelling: Concepts, Approaches and Experiences. *Practical Hydroinformatics*, 17–30. https://doi.org/10.1007/978-3-540-79881-1_2
- Taherei Ghazvinei, P., Darvishi, H. H., Mosavi, A., Bin Wan Yusof, K., Alizamir, M., Shamshirband, S., & Chau, K. W. (2018). Sugarcane growth prediction based on meteorological parameters using extreme learning machine and artificial neural network. *Engineering Applications of Computational Fluid Mechanics*, *12*(1), 738–749. <https://doi.org/10.1080/19942060.2018.1526119>
- Talei, A., & Chua, L. H. C. (2012). Influence of lag time on event-based rainfall-runoff

Prediksi Banjir menggunakan ANFIS-PCA sebagai Peringatan Dini Bencana Banjir

modeling using the data driven approach. *Journal of Hydrology*, 438–439, 223–233.
<https://doi.org/10.1016/j.jhydrol.2012.03.027>