

# MENUJU PREDIKSI KEJADIAN LONGSOR AKIBAT HUJAN MENGUNAKAN PENDEKATAN KECERDASAN BUATAN DI BANJARNEGARA

## Towards rainfall-Induced Landslides Prediction by Utilizing Artificial Intelligence Approach in Banjarnegara

**Anistia Malinda Hidayat<sup>1)</sup>, Adi Mulsandi<sup>2)</sup>, Hastuardi Harsa<sup>3)</sup>, Bambang Supriyadi<sup>2)</sup>, Aly Ilyas<sup>1)</sup>**

<sup>1)</sup> Badan Meteorologi Klimatologi dan Geofisika, Jl. Angkasa I, No.2 Kemayoran, Jakarta Pusat 10610.

<sup>2)</sup> Sekolah Tinggi Meteorologi Klimatologi dan Geofisika, Jl. Meteorologi No. 5, Tangerang 15119.

<sup>3)</sup> Badan Riset dan Inovasi Nasional, Jl. M.H. Thamrin No. 8, Jakarta Pusat 10340.

\*e-mail: anistia.malinda.hidayat@bmet.go.id

### Intisari

Data Badan Nasional Penanggulangan Bencana (BNPB) menyebutkan adanya tren peningkatan kejadian longsor di Banjarnegara dari 2014-2018. Faktor curah hujan menyebabkan 78 % kejadian longsor dari total 108 kejadian. Pola akumulasi curah hujan tinggi yang diamati terjadi pada bulan November hingga Maret (NDJFM) di Banjarnegara. Akumulasi curah hujan tersebut memiliki hubungan linier dengan kejadian longsor pada periode NDJFM. Penelitian ini dilakukan untuk memprediksi kejadian longsor akibat curah hujan di Banjarnegara dengan menggunakan Jaringan Syaraf Tiruan (JST) berbasis probabilistik, yaitu Probabilistic Neural Network (PNN). Data curah hujan diturunkan dari data Global Satellite Mapping of Precipitation (GSMaP) dan terdiri dari akumulasi, intensitas, serta durasi curah hujan dalam empat skala waktu yang berbeda, yang digunakan sebagai input untuk model JST. Dua skenario yang diterapkan untuk input data, dengan menerapkan transformasi data menggunakan transformasi log dan Principal Component Analysis (PCA). Sepuluh konfigurasi diuji coba dengan menggunakan model PNN. Pada skenario pertama, konfigurasi keempat menunjukkan performa terbaik di mana nilai akurasi mencapai 0,854, nilai Probability of Detection (POD) tertinggi 0,344, dan nilai False Alarm Ratio (FAR) yang relatif rendah dibandingkan yang lain, yaitu 0,46. Berdasarkan parameter akurasi dan statistik POD, kinerja PNN meningkat pesat ketika skenario kedua diterapkan. Analisis performa PNN menunjukkan bahwa konfigurasi ketujuh pada skenario ini dengan nilai  $\sigma$  0,7 memiliki performa terbaik dalam memprediksi kejadian longsor. Nilai akurasi meningkat hingga 0,863 dan POD meningkat signifikan hingga 0,905.

**Kata Kunci:** Kecerdasan Buatan, Longsor Akibat Hujan, Probabilistic Neural Network, Cuaca Berdampak Tinggi.

### Abstract

National Agency for Hazard Mitigation (BNPB) data stated an inclining trend of landslides in Banjarnegara from 2014 to 2018. The rainfall factor caused 78 % of landslides out of 108 cases. A large amount of rainfall accumulation was observed in November up to March (NDJFM) in Banjarnegara. Rainfall accumulation has a linear relationship with landslide events during that period. This research was conducted to better predict a rainfall-induced landslide in Banjarnegara using the Probabilistic-based Artificial Neural Network (ANN), which is Probabilistic Neural Network (PNN). Rainfall data derived from Global Satellite Mapping of Precipitation (GSMaP) data comprise rainfall accumulation, intensity, and duration in four different time scales, used as input for ANN models. Two scenarios were applied to the input data: log transformation and Principal Component Analysis (PCA). Ten configurations were performed using the PNN model. In the first scenario, the fourth configuration shows the best performances where the accuracy value reached up to 0.854, the highest Probability of Detection (POD) value of 0.344, and a relatively low False Alarm Ratio (FAR) value compared to the others, which is 0.46. Based on the accuracy and POD statistics parameter, the performance of PNN increased remarkably when the second scenario was applied. Analysis of the performance of PNN showed that the seventh configuration of this scenario, where its  $\sigma$  value is 0.7, has the best performance in predicting landslide events. The accuracy value increased to 0.863, and POD significantly increased to 0.905.

**Keywords:** *Artificial Intelligence, Rainfall-induced Landslides, Probabilistic Neural Network, High-impact Weather.*

## 1. PENDAHULUAN

Bentuk topografi yang berbukit dengan karakteristik hujan tinggi menjadikan Indonesia sebagai lokasi yang ideal terjadinya longsor (Gustono, 2018). Data Badan Nasional Penanggulangan Bencana (BNPB) menunjukkan bahwa tren kejadian longsor di Indonesia meningkat dalam kurun waktu 10 tahun terakhir. Provinsi Jawa Tengah menjadi daerah dengan tingkat kerawanan tertinggi di Indonesia. Sementara itu, Kabupaten Banjarnegara merupakan daerah paling rentan longsor kedua setelah Wonogiri di Provinsi Jawa Tengah. Dalam rentang tahun 2009–2018, bencana longsor di Banjarnegara telah mengakibatkan setidaknya 119 orang meninggal dunia, 5180 orang terdampak, 68 hektar area persawahan rusak, dan berbagai kerusakan fisik lainnya, baik pribadi maupun fasilitas umum. Ditinjau dari aspek kependudukan, jumlah populasi penduduk di Kabupaten Banjarnegara jauh lebih padat dibandingkan Wonogiri, yaitu mencapai 916.875 jiwa dengan kepadatan penduduk berkisar 857,12 jiwa/km<sup>2</sup>. Mengacu pada pedoman prakiraan berbasis dampak, eksposur bencana longsor terhadap keselamatan penduduk di Banjarnegara memiliki risiko yang tinggi (WMO, 2015).

Sebagai negara beriklim tropis, Indonesia memiliki karakteristik curah hujan rata-rata yang tinggi (Ramage, 1968; Nakamura *et al.*, 1994; Soenarmo, 2007) dengan nilai rata-rata terendah 100 mm/bulan dan tertingginya mencapai 300 mm/bulan di mana monsun menjadi faktor paling berpengaruh terhadap curah hujan (Aldrian & Susanto, 2003). Sementara itu, kejadian longsor seringkali terjadi selama atau setelah terjadinya hujan lebat (Guzzetti *et al.*, 2008; Che *et al.*, 2012). Bencana tanah longsor yang dipicu oleh hujan merupakan masalah utama di daerah dataran tinggi dan pegunungan (Crozier *et al.*, 1992; Brooks *et al.*, 2002). Fenomena hujan yang memicu terjadinya longsor dapat dikategorikan sebagai kejadian cuaca berdampak berat atau yang lebih dikenal dengan istilah *high-impact weather* (Papagiannaki *et al.*, 2015).

Pemodelan yang dapat digunakan untuk merujuk permasalahan tersebut dapat berupa model empiris atau model deterministik. Shaw dan Johnson (1995) meneliti kejadian tanah longsor yang disebabkan oleh curah hujan pada lereng bukit yang curam menggunakan model empiris, yaitu SMORPH. Sebaliknya, model-model berbasis mekanisme atau model deterministik sangat sering diterapkan pada area spesifik, di mana pendekatannya dikenal sebagai penilaian stabilitas lereng geoteknik kopling ke dalam pendekatan geoteknik. Beberapa contoh model deterministik adalah SHALSTAB

(Montgomery & Dietrich 1994), dSLAM (Wu & Sidle, 1995), LISA (Hammond *et al.*, 1992), dan TRIGRS (Baum *et al.*, 2002). Model-model tersebut dapat memiliki performa yang baik jika didukung oleh data pengamatan lokasi longsor yang lengkap. Namun dalam banyak kasus, observasi langsung di lapangan untuk mendapatkan data yang digunakan untuk pemodelan seringkali tidak memungkinkan karena terkendala oleh beberapa hal seperti sulitnya akses menuju lokasi atau jauh dari fasilitas umum seperti jaringan internet dan listrik. Kendala-kendala tersebut menyebabkan sulitnya pengaplikasian model prediksi di keseharian. Oleh karena itu, perlu upaya yang lebih kreatif dan mumpuni untuk memungkinkan pengembangan model prediksi longsor terutama untuk input data yang independen dari pengamatan lokal, misalnya data satelit, dan jenis model harus mudah serta dapat diimplementasikan di berbagai platform aplikasi komputer yang dapat diakses secara bebas.

Hujan merupakan salah satu parameter dinamika atmosfer yang bersifat nonlinear, pemodelan nonlinear dengan memanfaatkan Kecerdasan Buatan/*Artificial Intelligence* (AI) merupakan alternatif solusi dalam menentukan probabilitas kejadian longsor di Banjarnegara. Jaringan Saraf Tiruan (JST) merupakan salah satu bidang penelitian AI paling populer dengan model komputasi yang didasarkan pada struktur organisasi otak manusia. Belakangan, JST telah dimanfaatkan untuk studi terkait longsor, dengan mengacu pada penentuan secara tidak langsung dari parameter pemicu (Mayoraz *et al.*, 1996; Aleotti *et al.*, 1999, Harsa *et al.*, 2023) dan juga pemetaan daerah rentan longsor (Lee *et al.*, 2001; Fernandez-Steege *et al.*, 2002). Jenis JST yang akan digunakan dalam penelitian ini adalah JST berbasis probabilistik dengan tujuan klasifikasi, yaitu *Probabilistic Neural Network* (PNN). Keberhasilan pemanfaatan JST dalam penentuan dan evaluasi kejadian longsor pada beberapa penelitian sebelumnya membuka peluang pengembangan penelitian terkait kejadian longsor. Penelitian ini menyajikan analisis karakteristik curah hujan dan kaitannya dengan kejadian longsor, uji performa model PNN dalam memprediksi kejadian longsor di Banjarnegara, serta menentukan konfigurasi skema terbaiknya.

## 2. METODE

Parameter yang digunakan dalam penelitian ini adalah akumulasi curah hujan, intensitas curah hujan, dan durasi hujan yang diperoleh dari citra satelit *Global Satellite Mapping of Precipitation* (GSMaP). GSMaP adalah produk curah hujan global multisatelit, yang dikembangkan dengan

penggunaan gabungan pengamatan satelit gelombang mikro dan inframerah (IR) pada resolusi temporal dan spasial yang tinggi. Informasi detail produk GSMaP didapatkan dari Pusat Penelitian Observasi Bumi Badan Eksplorasi Luar Angkasa Jepang. Produk GSMaP diperoleh dari website Japan Aerospace Exploration Agency (JAXA), melalui alamat <http://hokusai.eorc.jaxa.jp/> dengan menggunakan *username* rainmap dan kata sandi *Niskur+1404*. Penelitian ini juga menggunakan data kejadian longsor yang tersedia di laman BNPB (<https://dibi.bnpb.go.id/>). Keluaran model JST kemudian dibandingkan dengan data sebenarnya, yang kemudian dikenal sebagai *ground truth data*. Penelitian ini dilakukan di Kecamatan Wanayasa, sebagian kecil Kecamatan Kalibening, Karangobar, Pejawaran, dan Pagentan yang termasuk dalam wilayah *grid* dengan koordinat 7° 15' S dan 109° 45' BT. Data BNPB dari tahun 2014–2018 menunjukkan bahwa lokasi penelitian dikategorikan sebagai daerah rawan longsor di mana kejadian longsor paling mungkin terjadi di Banjarnegara. Penelitian ini dilakukan selama 5 tahun dari tahun 2014–2018.

Tahapan penelitian dimulai dengan mengumpulkan data akumulasi, intensitas, dan durasi curah hujan yang memicu terjadinya longsor dari tahun 2014–2018. Data GSMaP diklasifikasikan menjadi 3 data yang berbeda yaitu, data akumulasi curah hujan (*A*), intensitas curah hujan (*I*) dan durasi hujan (*D*). Ketiga variabel tersebut kemudian ditentukan sebagai prediktor atau fitur. Sementara itu, data kejadian longsor ditentukan sebagai variabel respon. Sebelum menentukan data sebagai input model JST, ketiga prediktor tersebut dibedakan berdasarkan skala waktunya, yaitu pada hari kejadian (*H*), 1 hari sebelum kejadian tanah longsor (*H-1*), 2 hari sebelum kejadian (*H-2*), 3 hari sebelum kejadian (*H-3*). Hasil keluaran model JST digunakan untuk memprediksi longsor pada hari kejadian (*H*) mulai pukul 00.00 UTC hingga keesokan harinya (*H+1*) pukul 00.00 UTC. Kombinasi antara data fitur dan data respons dibagi menjadi dua bagian, yaitu data latih dan data uji. Data selama periode 2014–2017 digunakan sebagai data latih untuk model PNN; sedangkan data selama tahun 2018 digunakan sebagai data uji. Jumlah sampel untuk data latih harus lebih banyak dibandingkan data uji agar pola data latih dapat dipahami sehingga memudahkan model dalam memprediksi data uji (Ryan, 2018).

Sebelum digunakan sebagai input untuk PNN, data latih akan dinormalisasi menggunakan persamaan 2.1–2.3. Langkah ini disebut langkah *preprocessing* yang ditujukan pada semua prediktor agar memiliki skala nilai yang sama (Ryan, 2018), tanpa mendistorsi perbedaan antarvariabel rentang nilai. Normalisasi data sangat penting untuk mendapatkan hasil yang lebih baik dan untuk mengurangi kesalahan estimasi selama perhitungan yang signifikan dan mengurangi waktu perhitungan yang dibutuhkan selama fase pelatihan (Sola dan Sevilla, 1997).

$$x_{norm} = \frac{x - x_{rata}}{x_{std}} \quad (2.1)$$

dengan,

$$x_{rata} = \frac{\sum_i^n x}{n} \quad (2.2)$$

dan,

$$x_{std} = \frac{\sum_i^n (x_{rata} - x)^2}{n - 1} \quad (2.3)$$

Dimana  $x_{norm}$  merupakan nilai masukan yang sudah dinormalisasi dan siap digunakan dalam pemodelan JST,  $x$  merupakan parameter input atau masukan yang akan dinormalisasi,  $x_{rata}$  rata-rata nilai parameter  $x$ ,  $x_{std}$  standar deviasi dari parameter  $x$  dan  $n$  adalah jumlah data pada parameter  $x$ .

Setelah dilakukan normalisasi, akan ada dua skenario untuk menghitung nilai probabilitas kejadian longsor di daerah penelitian. Skenario 1 ditentukan ketika data input yang dinormalisasi dimasukkan sebagai input data langsung (tanpa transformasi data). Sedangkan, skenario 2 ditentukan dengan mentransformasikan data GSMaP terlebih dahulu menggunakan transformasi log dan *Principal Components Analysis* (PCA) sebelum dijadikan data input. Data curah hujan merupakan data dengan variabilitas yang sangat tinggi sehingga transformasi log digunakan untuk mengurangi pengaruh perbedaan nilai antar data input yang signifikan tanpa mengubah variasi antar data. Transformasi log dianggap baik untuk data yang dinyatakan dengan probabilitas dengan bentuk distribusi data cenderung seragam (Reimann *et al.*, 2008). Setelah melakukan transformasi log dengan pertimbangan bahwa total data yang digunakan terlalu banyak dan tidak semua data mendukung informasi penting untuk model JST, maka PCA diterapkan. PCA berperan dalam menyaring fitur-fitur penting dalam data sebelum digunakan sebagai data input untuk model PNN.

Pembelajaran algoritma PNN dapat dimulai dengan asumsi bahwa terdapat himpunan data latih  $S = \{X_{C_{ij}}\}$  untuk klasifikasi masalah yang mengandung masukan  $X_{C_{ij}}$  (misal data ke- $j$  yang termasuk dalam kelas  $C_j$ ). Jika  $S$  termasuk dalam data masukan untuk kelas  $NC$ , maka  $S$  dapat dibagi menjadi  $NC$  himpunan bagian  $C_i$  yang mengandung kelas data masukan (misal  $C_i = \{X_{C_{i,1}}, X_{C_{i,2}}, \dots, X_{C_{i,|C_i|}}\}$ ). Dalam teknik pembelajaran klasifikasi supervisi, vektor input baru  $X_{new}$  seharusnya diklasifikasikan ke salah satu kelas  $NC$  yang memiliki probabilitas posterior tertinggi (Bishop, 2000). Menurut Specht (1990), penentuan nilai Gaussian kernel dengan data berdimensi- $d$  dihitung berdasarkan persamaan 2.4.

$$\omega_{i,j} = \frac{1}{(2\pi)^{d/2} \sigma^2} \cdot \exp\left(-\frac{\|x - x'_n\|^2}{2\sigma^2}\right) \quad (2.4)$$

Dimana  $\omega_{i,j}$  merupakan nilai *Gaussian kernel* dan  $\sigma$  merupakan parameter distribusi (*spread*

parameter) atau lebar fungsi *parzen window* yang nilainya dapat ditentukan berdasarkan eksperimen untuk vektor input yang telah dinormalisasi dengan nilai berada pada rentang 0-1. Sedangkan  $x$  merupakan nilai titik pusat pada perhitungan *parzen window*, dan  $\|x - x'_n\|^2$  jarak *Euclidian* kuadrat antara data pusat (data yang ingin dicari probabilitasnya) dan data dalam suatu kategori dalam lapisan masukan.

Estimasi untuk mendapatkan *pdf* yang baik dapat diperoleh dengan merata-ratakan *multivariate normal Gaussian kernel* yang merupakan produk dari masing-masing *kernel univariate* dan berfokus pada setiap sampel yang diketahui. Berdasarkan teknik tersebut di atas, probabilitas kelas bersyarat (*class-conditional probability*) yang merupakan bentuk persamaan dari *parzen window* dalam PNN dihitung berdasarkan persamaan 2.5.

$$P(X_{new} | C_i) = P_i = \frac{1}{|C_i|} \sum_{j=1}^{|C_i|} \omega_{i,j} \quad (2.5)$$

Dimana  $P_i$  merupakan peluang kelas bersyarat (*class-conditional probability*),  $|C_i|$  merupakan jumlah elemen dari himpunan bagian  $C_i$  dan  $\omega_{i,j}$  merupakan nilai *Gaussian kernel* dari  $X_{C_i,j}$

**Tabel 1.** Konfigurasi nilai  $\sigma$  yang diujikan pada PNN

Konfigurasi	$\sigma$	Konfigurasi	$\sigma$
1	0,1	6	0,6
2	0,2	7	0,7
3	0,3	8	0,8
4	0,4	9	0,9
5	0,5	10	1,0

**Tabel 2.** Tabel kontingensi (Lauritzen, 1982)

		Observasi		Total
		Iya	Tidak	
Prediksi	Iya	<i>Hit</i>	<i>False Alarm</i>	Prediksi Iya
	Tidak	<i>Miss</i>	<i>Correct negative</i>	Prediksi tidak
		Observasi Iya	Observasi tidak	Total

Konfigurasi *hyperparameter* PNN dalam penelitian ini divisualisasikan pada Tabel 1. Tabel tersebut menunjukkan 10 skema konfigurasi yang akan diuji menggunakan model PNN. Hasil prediksi kejadian longsor secara keseluruhan menggunakan dua skenario PNN yang berbeda kemudian dibandingkan untuk menentukan skenario JST yang memiliki performa terbaik dalam memprediksi kejadian longsor di daerah penelitian. Hasil prediksi kejadian longsor menggunakan PNN pada semua skenario dan konfigurasi kemudian dianalisis dan disimpulkan.

Performa kinerja PNN ditentukan melalui tabel kontingensi pada Tabel 2. Prediksi longsor dikategorikan *Hit* jika prediksi longsor diikuti dengan kejadian longsor yang terjadi selama periode

pengamatan dalam durasi 24 jam dari pukul 00.00 UTC pada hari kejadian (H) hingga hari berikutnya (H+1) pukul 00.00 UTC.

Berdasarkan tabel kontingensi di atas, nilai *Probability of Detection* (POD) dan *False Alarm Ratio* (FAR) dapat ditentukan. Persamaan untuk menghitung akurasi, POD, dan FAR adalah sebagai berikut:

$$Akurasi = \frac{Hit + Correct\ negative}{Total}$$

$$POD = \frac{Hit}{Hit + Miss}$$

$$FAR = \frac{False\ alarm}{False\ alarm + Hit}$$

Dimana *Hit* merupakan prediksi Iya observasi Iya, *Correct negative* merupakan prediksi tidak observasi tidak. Sedangkan *Miss* adalah prediksi tidak observasi Iya dan *False alarm* merupakan prediksi Iya observasi tidak

### 3. HASIL DAN PEMBAHASAN

#### 3.1. Transformasi Data Masukan Model JST

Transformasi data merupakan salah satu teknik analisis data grafis penting untuk meningkatkan visibilitas bagian-bagian tertentu dari data. Transformasi logaritmik, selanjutnya disebut sebagai transformasi-log, merupakan salah satu contoh transformasi yang digunakan untuk mengurangi pengaruh adanya data dengan nilai tinggi yang tidak biasa (*outliers*). Nilai tinggi tersebut dapat diakibatkan karena adanya data-data *outliers* atau pencilan data yang terletak jauh dari konsentrasi data utama. Data pencilan berbeda dengan data hujan dengan intensitas tinggi, data pencilan dalam satelit atau pengamatan digital biasanya memiliki nilai 9999 yang merupakan data tidak wajar. Sehingga transformasi yang dilakukan pada data tidak akan mengubah variabilitas atau jarak antardata. Transformasi dengan menggunakan log (*log-transformation*) dinilai baik untuk data-data yang dinyatakan dalam proporsi atau probabilitas dan memiliki bentuk distribusi data yang cenderung seragam (Reimann et al., 2008). Dalam penelitian ini, proses *re-scaling* dengan transformasi-log diterapkan pada semua fitur yang digunakan sebagai data masukan model JST.

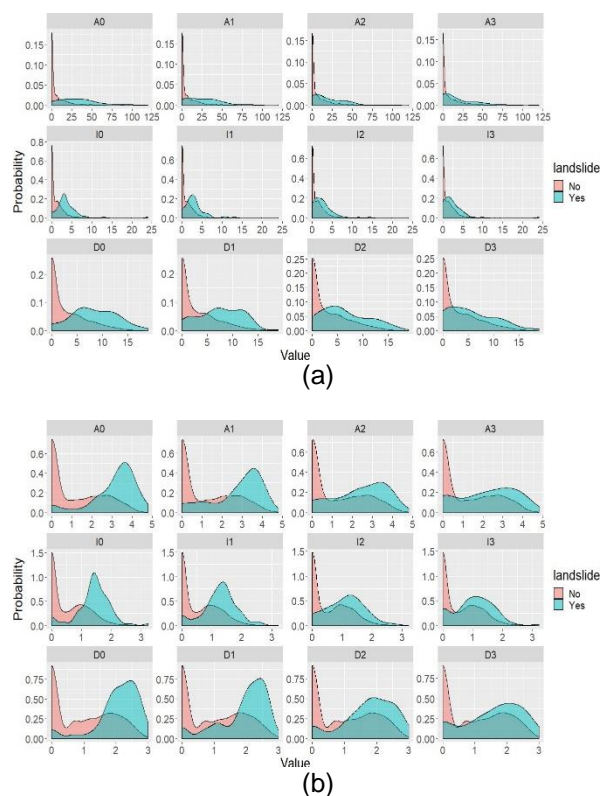
Penerapan transformasi-log pada semua fitur menyebabkan distribusi grafik miring ke kanan cukup kuat, utamanya pada ketiga fitur curah hujan pada kelas longsor. Kondisi yang sama juga dijelaskan oleh Reimann et al. (2008), di mana penerapan transformasi-log menyebabkan histogram miring ke kanan secara kuat (*strongly right-skewed*). Jika dibandingkan dengan analisis grafik sebelum dilakukan transformasi (Gambar 1a), rentang atau *range*

(misal nilai maksimum dan minimum) dari semua fitur curah hujan pada semua skala waktu berkurang secara drastis karena pengaruh penciran (*outliers*) telah dikurangi, artinya penerapan transformasi-log memperkecil rentang antardata pada setiap fitur data masukan tanpa mengubah variasi nilai dalam data (Gambar 1b). Selain itu, nilai peluang pada masing-masing fitur curah hujan dalam semua skala waktu mengalami peningkatan jika dibandingkan dengan sebaran nilai peluang sebelum dilakukan transformasi-log. Penerapan transformasi-log memberikan visualisasi yang lebih berguna dari data-data utama yang digunakan.

Transformasi-log sederhana pada data-data lingkungan (*environmental data*), termasuk meteorologi, cukup baik untuk divisualisasikan sebagai distribusi normal. Data-data lingkungan umumnya dicirikan dengan keberadaan data bernilai sangat tinggi yang menyimpang secara signifikan dari konsentrasi data utama. Dalam kondisi tersebut, transformasi data logaritmik tidak cukup untuk membantu memposisikan sebaran data sebagai distribusi normal. Oleh karena itu, diperlukan metode statistik lain untuk mendapatkan hasil yang dapat lebih diandalkan (Reimann et al., 2008). Metode statistik lain yang digunakan setelah dilakukan transformasi-log pada semua fitur data masukan adalah PCA.

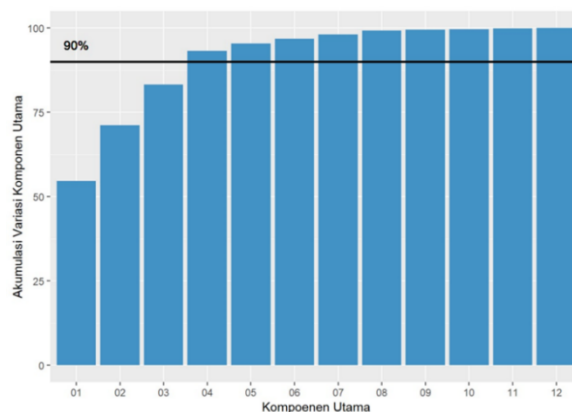
PCA merupakan teknik yang bertujuan untuk mengurangi dimensi data (Hotelling, 1933; Jolliffe, 2002; Johnson dan Wichern, 2002; Jackson, 2003). Teknik PCA memiliki peran penting dalam penelitian ini, mengingat terdapat 12 fitur curah hujan (matriks data 1826 x 12) yang digunakan sebagai data masukan dan tentu tidak semuanya dapat memberikan informasi yang penting bagi model JST. Dalam hal ini, akan ada dimensi data yang jauh lebih penting keberadaannya dibandingkan dimensi data lain dalam satu himpunan data tersebut. PCA berperan untuk “memeras” data tersebut dan mengambil beberapa dimensi data yang mengandung informasi penting yang dapat menjelaskan sekumpulan data yang banyak tersebut sehingga dapat meningkatkan kemampuan prediksi atau klasifikasi data (Reimann et al., 2008).

Dalam metode PCA, pemilihan ambang batas didasarkan pada kurva variabilitas total (Wang & Jin, 2005). Cheng et al. (2019) menggunakan ambang batas 90 % dan 95 % untuk menguji akurasi beberapa algoritma. Hasil penelitian tersebut menunjukkan bahwa penggunaan ambang batas 90 % menghasilkan nilai akurasi yang lebih tinggi.



**Gambar 1.** Sebaran nilai peluang sebelum (1a) dan sesudah (1.b) dilakukan transformasi-log pada masing-masing variabel dipisahkan berdasarkan kejadian longsor ( $Y_a$ ) yang ditunjukkan dengan grafik yang berwarna hijau dan tidak longsor (Tidak) yang ditunjukkan dengan grafik berwarna merah muda.

Oleh karena itu, nilai ambang batas persentase 90 % digunakan dalam penelitian ini. Gambar 2 menunjukkan bahwa kontribusi komponen utama dengan ambang batas 90 % dicapai pada komponen utama ke empat (PC 4). Sehingga akan terdapat empat fitur (4 KU/PC) dari total 12 fitur yang digunakan sebagai masukan model JST setelah metode PCA dilakukan.



**Gambar 2.** Kontribusi komponen utama dengan ambang batas 90% dicapai pada komponen utama ke empat.

### 3.2. Performa Prediksi Longsor Berbasis PNN

#### 3.2.1 Performa PNN sebelum dilakukan transformasi pada data masukan

Prediksi kejadian longsor yang diakibatkan hujan (*rainfall-induced landslide*) yang dilakukan oleh beberapa peneliti sebelumnya menunjukkan hasil yang cukup memuaskan. Pemanfaatan JST dalam membuat peta kerentanan longsor di Inje, Korea menunjukkan akurasi prediksi mencapai 80,1 %, relatif lebih tinggi dibandingkan dengan pemanfaatan model regresi logistik yang berkisar 77,05 % (Saro *et al.*, 2016). Penelitian Chae *et al.* (2017) menunjukkan bahwa JST merupakan metode yang efektif untuk mengevaluasi berbagai faktor yang memicu distribusi longsor secara spasial di area yang luas. Ermini *et al.* (2005) menggunakan PNN untuk memprediksi kejadian longsor dan mendapatkan hasil yang memuaskan. Dalam penelitian ini, kejadian longsor diprediksi terjadi apabila hasil keluaran probabilitas model PNN bernilai lebih dari 50 %.

**Tabel 3** Performa model PNN sebelum dilakukan transformasi data

Konfigurasi	$\sigma$	Akurasi	POD	FAR
1	0,1	0,099	0,095	0,048
2	0,2	0,863	0,278	0,762
3	0,3	0,858	0,320	0,619
4	0,4	0,854	0,344	0,476
5	0,5	0,844	0,324	0,476
6	0,6	0,840	0,324	0,429
7	0,7	0,830	0,326	0,333
8	0,8	0,797	0,288	0,286
9	0,9	0,802	0,294	0,286
10	1,0	0,802	0,309	0,190

Tabel 3 menunjukkan performa PNN pada 10 konfigurasi berbeda dalam memprediksi kejadian longsor di wilayah penelitian. Jika ditinjau berdasarkan hasil perhitungan nilai akurasi, hasil penelitian menunjukkan bahwa nilai akurasi tertinggi prediksi kejadian longsor dengan menggunakan PNN adalah 0,863. Nilai tersebut merupakan hasil perhitungan berdasar Konfigurasi 2 dengan nilai  $\sigma$  0,2. Hasil perhitungan akurasi tersebut relatif lebih baik jika dibandingkan dengan penelitian yang dilakukan oleh Saro *et al.* (2016) di Inje, Korea yang menghasilkan nilai akurasi 80,1 %.

Secara umum, nilai akurasi rendah pada Konfigurasi 1, kemudian meningkat secara signifikan pada Konfigurasi 2. Pada Konfigurasi 3 hingga 10, nilai akurasi tersebut cenderung konstan (memiliki variasi yang kecil). Akurasi pada dasarnya merupakan ketiadaan bias, sehingga hasil analisis bisa saja memiliki nilai akurasi tinggi namun belum tentu menjadi "benar" (Reimann *et al.*, 2008). Oleh karena itu, analisis parameter statistik lain, yaitu POD, juga digunakan untuk memberikan sudut pandang lain terkait hasil prediksi model PNN.

Beberapa penelitian menyebutkan bahwa analisis nilai POD lebih representatif dalam kaitannya dengan prediksi. Semakin tinggi nilai POD maka hasil prediksi semakin mendekati sempurna, sebaliknya semakin rendah nilai POD maka hasil prediksi semakin buruk. Perhitungan nilai POD pada semua konfigurasi menunjukkan hasil yang jauh lebih rendah dibandingkan rentang nilai akurasi, yaitu berkisar antara 0,095 sampai 0,344. Nilai POD tertinggi mencapai 0,344 didapatkan pada konfigurasi keempat, di mana nilai sigmanya ( $\sigma$ ) adalah 0,4. Hasil perhitungan POD pada semua konfigurasi memiliki pola yang serupa dengan fluktuasi nilai akurasi, di mana nilai POD sangat rendah pada Konfigurasi 1, kemudian meningkat dan relatif konstan (memiliki variasi yang kecil) pada Konfigurasi 2 sampai 10.

Ketika menampilkan beberapa perbandingan dalam perhitungan statistik, FAR dihitung sebagai rasio antara jumlah kejadian negatif yang salah diklasifikasikan sebagai positif dan jumlah total kejadian sebenarnya. Nilai FAR merepresentasikan bagian dari prakiraan kejadian yang tidak terjadi berdasarkan hasil observasi. Semakin besar nilai FAR (mendekati 1), semakin buruk hasil prediksi, hasil prakiraan banyak yang salah atau salah semua. Sebaliknya, semakin kecil nilai FAR (mendekati 0), semakin baik hasil prediksi model, hasil prakiraan banyak yang benar atau benar semua.

Dalam penelitian ini, nilai FAR bervariasi antara 0,048 (Konfigurasi 1) sampai 0,762 (Konfigurasi 2). Hasil perhitungan FAR pada Konfigurasi 1 menunjukkan nilai yang sangat rendah, kemudian nilai ini meningkat drastis pada Konfigurasi 2 hingga Konfigurasi 3. Nilai FAR menurun secara bertahap pada Konfigurasi 4 hingga 10, mencapai 0,190 pada Konfigurasi 10. Merujuk pada penjelasan ketiga parameter statistik untuk menguji hasil prediksi PNN yang dilakukan dalam penelitian ini, maka dapat disimpulkan bahwa model PNN dengan Konfigurasi 4 memiliki performa terbaik dalam memprediksi kejadian longsor di wilayah penelitian. Hal tersebut dikarenakan konfigurasi ini memiliki nilai akurasi tinggi mencapai 0,854, nilai POD paling tinggi dibanding konfigurasi lain mencapai 0,344, serta nilai FAR yang relatif lebih rendah dibandingkan Konfigurasi 2 dan 3. Keseluruhan proses prediksi kejadian longsor tanpa melakukan modifikasi pada data masukan model kemudian disebut sebagai skenario PNN 1.

#### 3.2.2 Performa PNN sesudah dilakukan transformasi pada data masukan

Penerapan transformasi-log memperkecil rentang antardata pada masing-masing fitur yang dijadikan sebagai data masukan model JST sehingga data lebih terkonsentrasi pada interval tertentu. Secara umum, kondisi tersebut mempermudah proses klasifikasi. Setelah dilakukan

transformasi-log pada semua fitur data masukan, selanjutnya teknik PCA diterapkan. Setelah tahap pre-processing data dilakukan dengan penerapan transformasi-log dan PCA pada semua fitur data masukan, data tersebut kemudian siap dijadikan sebagai data input model PNN. Keseluruhan proses ini kemudian disebut sebagai skenario PNN 2. Performa model PNN dengan skenario 2 ditampilkan pada Tabel 4.

**Tabel 4** Performa model PNN setelah dilakukan transformasi data.

Konfigurasi	$\sigma$	Akurasi	POD	FAR
1	0,1	0,929	0,190	0,692
2	0,2	0,910	0,333	0,731
3	0,3	0,888	0,524	0,738
4	0,4	0,871	0,619	0,750
5	0,5	0,860	0,810	0,734
6	0,6	0,855	0,857	0,735
7	0,7	0,863	0,905	0,716
8	0,8	0,860	0,905	0,721
9	0,9	0,852	0,905	0,732
10	1,0	0,847	0,905	0,740

Secara umum, penerapan transformasi-log dan PCA pada setiap fitur data masukan model meningkatkan nilai akurasi dan POD secara signifikan. Peningkatan nilai akurasi tertinggi jika dibandingkan skenario PNN 1 mencapai 837,58 % diamati pada Konfigurasi 1, dimana nilai akurasi pada skenario sebelum dilakukan transformasi data adalah 0,095, sementara nilai akurasi setelah dilakukan transformasi data mencapai 0,929. Selain Konfigurasi 1, peningkatan akurasi juga terjadi pada konfigurasi lainnya namun tidak cukup signifikan berkisar antara 1,81–7,92 %. Secara umum, hasil luaran model PNN dengan Skenario 2 menunjukkan adanya peningkatan nilai POD yang jauh lebih signifikan pada semua konfigurasi dibandingkan dengan rasio peningkatan nilai akurasi prediksi. Peningkatan nilai POD tertinggi pada Skenario 2 mencapai 213,65 % diamati pada Konfigurasi 8, di mana nilai POD meningkat dari 0,288 menjadi 0,905. Selain itu, peningkatan nilai POD pada kategori lainnya juga sangat signifikan berkisar antara 20 % sampai 207,62 %.

Namun, peningkatan nilai akurasi dan POD yang signifikan pada skenario PNN 2 diiringi dengan peningkatan nilai FAR pada semua konfigurasi. Nilai FAR pada semua konfigurasi berkisar antara 0,692 sampai 0,750. Merujuk pada penjelasan di atas, Konfigurasi 7 merupakan konfigurasi yang dinilai paling baik karena memiliki nilai akurasi yang tinggi mencapai 0,863 dan nilai POD sangat tinggi mencapai 0,905. Selain itu, nilai FAR pada Konfigurasi 7 merupakan nilai terendah kedua di antara hasil luaran semua konfigurasi yang diuji coba, yaitu bernilai 0,716. Nilai FAR yang relatif tinggi mengindikasikan cukup besarnya potensi kesalahan model PNN dengan Skenario 2 ketika memberikan peringatan longsor namun pada kenyataannya tidak terjadi longsor (*false alarm*).

Ditinjau dari segi praktis, kondisi tersebut memiliki sisi positif dan negatif. Berdasar sudut pandang positif, tingginya nilai FAR dapat meningkatkan kewaspadaan masyarakat meskipun bisa jadi banyak terdapat peringatan longsor yang salah. Sementara dari sisi negatif, banyaknya kesalahan peringatan longsor yang dihasilkan dapat menurunkan kepercayaan masyarakat terhadap hasil prediksi model. Oleh karena itu, perlu dilakukan studi lebih lanjut terkait dengan implementasi model secara praktis di lapangan untuk mengetahui jenis performa model yang paling nyaman dirasakan oleh masyarakat. Secara garis besar berdasarkan pertimbangan ketiga parameter statistik dan analisis hasil performa model PNN Skenario 1, penerapan Konfigurasi 7 dengan menggunakan PNN Skenario 2 memiliki performa prediksi paling baik di wilayah penelitian.

#### 4. KESIMPULAN

Tren kejadian longsor yang terjadi akibat faktor curah hujan meningkat dari tahun ke tahun, utamanya di Banjarnegara, Jawa Tengah. Performa PNN sebagai salah satu metode kecerdasan buatan akan diuji untuk mendeteksi longsor yang dipicu hujan di Banjarnegara. Prediksi longsor tanpa melakukan transformasi pada data input menunjukkan nilai akurasi yang baik dengan menggunakan nilai  $\sigma$  lebih dari 0,2 dengan berkisar antara 0,802–0,863. Namun nilai FAR relatif lebih rendah saat menggunakan nilai  $\sigma$  lebih dari 0,4, yaitu berkisar antara 0,190–0,476. Merujuk pada penjelasan ketiga parameter statistik untuk pengujian hasil prediksi PNN yang dilakukan dalam penelitian ini, dapat disimpulkan bahwa model PNN dengan Konfigurasi 4 memiliki performa terbaik untuk memprediksi kejadian longsor di daerah penelitian. Hal tersebut dikarenakan penggunaan Skenario 1 dengan Konfigurasi 4 memiliki nilai akurasi yang tinggi (0,854), nilai POD tertinggi (0,344), dan nilai FAR yang relatif lebih rendah (0,476).

Secara umum implementasi Skenario 2 dengan melakukan transformasi pada semua fitur data input meningkatkan nilai akurasi dan POD secara signifikan pada semua konfigurasi. Namun peningkatan nilai akurasi dan POD ini diikuti dengan peningkatan nilai FAR. Nilai FAR pada Skenario 1 berkisar antara 0,048–0,762 dan meningkat dari 0,692–0,750 pada Skenario 2. Dibandingkan berbagai konfigurasi pada Skenario 2, Konfigurasi 7 merupakan konfigurasi terbaik karena memiliki nilai akurasi yang tinggi 0,863 dan nilai POD yang sangat tinggi mencapai 0,905. Selain itu, nilai FAR pada Konfigurasi 7 merupakan nilai terendah kedua di antara semua keluaran hasil konfigurasi yang telah diuji, yaitu 0,716. Dalam sudut pandang praktisi, kondisi tersebut memiliki sisi positif dan negatif. Di sisi positif, nilai FAR yang lebih tinggi dapat meningkatkan kewaspadaan masyarakat

terhadap adanya peringatan meskipun mungkin terdapat kesalahan dalam prediksi longsor. Sedangkan di sisi negatif, runtutan kesalahan prediksi longsor dapat menurunkan kepercayaan masyarakat terhadap keluaran model prediksi. Penelitian lebih lanjut perlu dilakukan untuk mengkaji model kecerdasan buatan yang juga mempertimbangkan aspirasi publik. Penggunaan data lain seperti kemiringan lahan, jenis tanah, peta kerawanan juga dapat menjadi bahan pertimbangan mengingat kejadian longsor membutuhkan dimensi keilmuan yang luas, tidak hanya dari perspektif cuaca saja. Hasil penelitian ini diharapkan dapat menjadi landasan atau kerangka awal prediksi bencana longsor yang dipicu oleh aktivitas hujan di Banjarnegara, Jawa Tengah. Ke depannya, variasi data input menggunakan output model prediksi perlu dilakukan sebagai lanjutan penelitian untuk menguji keandalan model PNN dalam memprediksi longsor yang dipicu hujan di Banjarnegara.

## 5. DAFTAR PUSTAKA

- Aldrian, E., & Susanto, R.D. (2003). Identification of three dominant rainfall regions within Indonesia and their relationship to sea surface temperature. *International Journal of Climatology*, 23, 1435-1452.
- Aleotti, P., & Chowdhury, R. (1999). Landslide hazard assessment: summary review and new perspectives. *Bulletin of Engineering Geology and the environment*, 58, 21-44.
- Baum, R.L., Savage, W.Z., & Godt, J.W. (2002). *TRIGRS: a Fortran program for transient rainfall infiltration and grid-based regional slope-stability analysis, version 2.0* (p. 75). Reston, VA, USA: US Geological Survey.
- Bishop, C. (2000). *Neural networks for pattern recognition*. United States, Oxford University Press Inc., Inggris, 23-25.
- Brooks, S.M., Crozier, M.J., Preston, N.J., & Anderson, M.G. (2002). Regolith stripping and the control of shallow translational hillslope failure: application of a two-dimensional coupled soil hydrology-slope stability model, Hawke's Bay, New Zealand. *Geomorphology*, 45, 165– 179.
- Chae, B. G., Park, H. J., Catani, F., Simoni, A., & Berti, M. (2017). Landslide prediction, monitoring and early warning: a concise review of state-of-the-art. *Geosciences Journal*, 21(6), 1033-1070.
- Cheng, X., Jing, W., Song, X., & Lu, Z. (Eds.). (2019). *Data Science: 5th International Conference of Pioneering Computer Scientists, Engineers and Educators, ICPCSEE 2019, Guilin, China, September 20–23, 2019, Proceedings, Part I* (Vol. 1058). Springer Nature.
- Crozier, M.J., Gage, M., Pettinga, J.R., Selby, M.J., & Wasson, R.J. (1992). The stability of hillslopes. *Landforms of New Zealand*, 2, 63–90.
- Ermioni, L., Catani, F., & Casagli, N. (2005). Artificial neural network applied to landslide susceptibility assessment. *Geomorphology*, 66, 327 – 343.
- Fernández-Steeger, T. M., Rohn, J., & Czurda, K. (2018). Identification of landslide areas with neural nets for hazard analysis. In *Landslides*, 163-168. Routledge.
- Gustono, S.T. (2018). Landslide susceptibility assessment in Garut, West Java, Indonesia. *Thesis*, Science in Physical Land Resources, Vrije Universiteit Brussel, Belgium, 82.
- Guzzetti, F., Peruccacci, S., Rossi, M., & Stark, C.P. (2008). The Rainfall Intensity–Duration Control of Shallow Landslides and Debris Flows: An Update. *Landslides*, 5, 3 – 17.
- Hammond, C., Hall, D., Miller, S., & Swetik, P. (1992). *Level I stability analysis (LISA) documentation for version 2.0* (Vol. 285). US Department of Agriculture, Forest Service, Intermountain Research Station.
- Harsa, H., Anistia, M. H., Mulsandi, A., Suprihadi, B., Kurniawan, R., Habibie, M. N., ... & Praja, A. S. (2023). Machine learning and artificial intelligence models development in rainfall-induced landslide prediction. *IAES International Journal of Artificial Intelligence*, 12(1), 262.
- Hotelling, H. (1933). Analysis of A Complex of Statistical Variables into Principal Components. *Journal of Educational Psychology*, 24, 417–441.
- Jackson, J. E. (2005). *A user's guide to principal components* (p. 33). John Wiley & Sons.
- Japan Aerospace Exploration Agency Earth Observation Research Center, 2013. *User's Guide for Global Rainfall Map in Near-Real-Time by JAXA Global Rainfall Watch (GSMaP\_NRT)*, Version 2.4.
- Johnson, R. A., & Wichern, D. W. (2002). *Applied multivariate statistical analysis*. Prentice Hall: Upper Saddle River, New Jersey, USA, 430.
- Jolliffe, I. T. (2002). *Principal component analysis for special types of data* (pp. 78-80). Springer New York.
- Lauritzen, S. L. (1979). *Lectures on contingency tables*. University of Copenhagen.
- Lee, S., Ryu, J., Min, K., & Won, J. (2001). Development of two artificial neural network methods for landslide susceptibility analysis. In *IGARSS 2001. Scanning the Present and Resolving the Future. Proceedings. IEEE 2001 International Geoscience and Remote Sensing Symposium (Cat. No. 01CH37217)* (Vol. 5, pp. 2364-2366). IEEE.



- Mayoraz, F., Cornu, T., & Vulliet, L. (1996). Using neural networks to predict slope movements. In *Proc. 7th Int. Symp. on Landslides* (Vol. 1, pp. 295-300).
- Montgomery, D. R., & Dietrich, W. E. (1994). A physically based model for the topographic control on shallow landsliding. *Water resources research*, 30(4), 1153-1171.
- Noerdjito, W. A., & HASYIM, A. (1994). Regional Difference and Seasonality of Rainfall in Java, with Special Reference to Bogor. *Tropics*, 4(1), 93-103.
- Nakamura, K., Noerdjito, W.A., & Hasyim, A. (1994). Regional difference and seasonality of rainfall in Java, with special reference to Bogor. *Tropics*, 4(1), 93–103.
- Papagiannaki, K., Lagouvardos, K., & Kotroni, V. (2013). A database of high-impact weather events in Greece: a descriptive impact analysis for the period 2001–2011. *Natural Hazards and Earth System*.
- Ramage, C. S. (1968). Role of a tropical "maritime continent" in the atmospheric circulation. *Monthly Weather Review*, 96(6), 365-370.
- Reimann, C., Filzmoser, P., Garrett, R., & Dutter, R. (2011). *Statistical data analysis explained: applied environmental statistics with R*. John Wiley & Sons.
- Ryan, M. (2018). Pemanfaatan Jaringan Saraf Tiruan (ANN) Untuk Prediksi Curah Hujan Berbasis Probabilistik Dan Optimalisasi Keluaran Model Cuaca Numerik. Graduate Thesis, STMKG, 81 pp.
- Saro, L., Woo, J. S., Kwan-Young, O., & Moungh-Jin, L. (2016). The spatial prediction of landslide susceptibility applying artificial neural network and logistic regression models: A case study of Inje, Korea. *Open Geosciences*, 8(1), 117-132.
- Shaw, S. C., & Johnson, D. H. (1995). Slope morphology model derived from digital elevation data. *Washington Department of Natural Resources. Washington*.
- Specht, D. F. (1990). Probabilistic neural networks. *Neural networks*, 3(1), 109-118.
- Wang, L., & Jin, Y. (Eds.). (2005). *Fuzzy Systems and Knowledge Discovery: Second International Conference, FSKD 2005, Changsha, China, August 27-29, 2005, Proceedings* (Vol. 3613). Springer Science & Business Media.
- Wu, W., & Sidle, R. C. (1995). A distributed slope stability model for steep forested basins. *Water resources research*, 31(8), 2097-2110.