

Pendugaan Tingkat Risiko Banjir dengan Menggunakan *Extreme Learning Machine* dan *Extreme Value Theory*

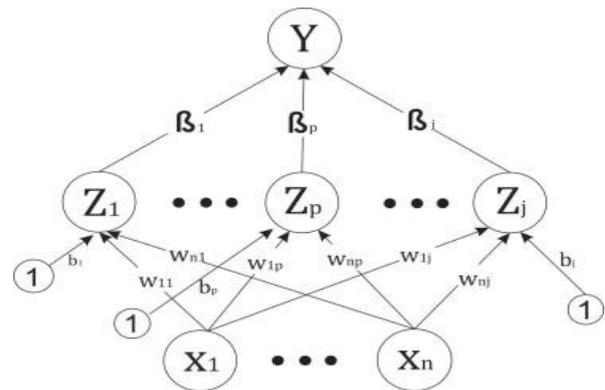
Rr. Andriana Ajeng Ayumurti, dan Galuh Oktavia Siswono
 Departemen Aktuaria, Institut Teknologi Sepuluh Nopember (ITS)
e-mail: galuh@aktuaria.its.ac.id

Abstrak—Banjir merupakan salah satu permasalahan yang sering terjadi di Indonesia, khususnya Surabaya. Banjir yang terjadi baik dalam skala kecil maupun skala besar membawa dampak negatif bagi lingkungan sekitar. Surabaya merupakan salah satu kota dengan tingkat banjir tertinggi di Indonesia, akibatnya beberapa wilayah di Surabaya terendam banjir yang cukup dalam dan menghambat aktivitas warga sekitar. Pada penelitian ini, digunakan data dasarian curah hujan dari salah satu stasiun di Surabaya dengan periode waktu dari Januari 2017 hingga Desember 2021. Pendugaan tingkat risiko pada penelitian ini menggunakan *Value at Risk* (VaR) dengan pendekatan *Extreme Value Theory* (EVT). Data penelitian berupa curah hujan akan dilakukan pra-pemrosesan data dengan mengidentifikasi missing value, observasi penciliran (outlier), dan observasi yang tidak sesuai dari data curah hujan di Surabaya. Kemudian mengidentifikasi karakteristik data curah hujan dengan statistika deskriptif dan pola sebaran curah hujan. Setelah didapatkan karakteristik data curah hujan, dilakukan peramalan dengan ELM yaitu data dibagi menjadi beberapa fitur dan target terlebih dahulu, setelah itu dilakukan normalisasi data. Data kemudian dibagi menjadi data training dan data testing untuk proses training dan testing. Kemudian dilakukan pengambilan sampel data ekstrim dengan metode *Peaks Over Threshold* dan *Block Maxima*. Lalu dilakukan perhitungan risiko dengan *Value at Risk* (VaR). Penelitian ini bertujuan untuk menduga tingkat risiko banjir serta menganalisis pengaruh yang dimiliki antara curah hujan dan banjir. Hasil penelitian didapat bahwa model terbaik didapat dengan MAPE data pengujian sebesar 9,81230 dibawah 10%. Data hasil ramalan menunjukkan bahwa curah hujan tertinggi terjadi di bulan Februari 2022. Tingkat risiko banjir dapat dilihat dari hasil VaR pada tingkat kepercayaan 90%, 95%, dan 99% yaitu pada GEV secara berturut-turut sebesar 143,9767, 145,1391118, 147,1209043 dan pada GPD secara berturut-turut sebesar 334,98, 340,3271661, 354,6074338 sehingga pemerintah Surabaya dapat membuat kebijakan terkait dengan kapasitas drainase atau penampungan air hujan sesuai dengan nilai VaR yang telah diperoleh.

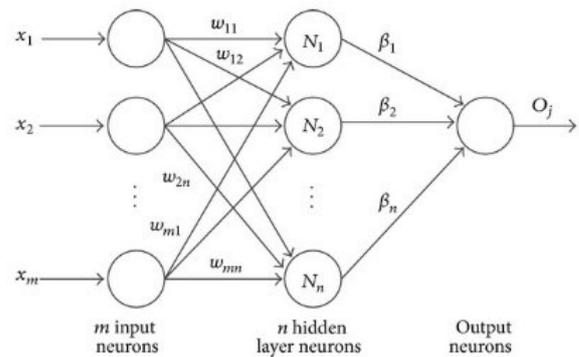
Kata Kunci—Banjir, *Block Maxima*, *Extreme Learning Machine*, *Peak Over Threshold*, *Value at Risk*..

I. PENDAHULUAN

CURAH hujan adalah tinggi rendahnya air yang terkumpul di tempat yang datar, tidak meresap dan tidak mengalir. Intensitas curah hujan tiap waktu berbeda-beda, curah hujan dengan intensitas yang tinggi atau biasa disebut dengan curah hujan ekstrim. Curah hujan ekstrem telah diketahui sejak lama menjadi faktor utama berkontribusi pada bencana banjir di Indonesia. Beberapa laporan oleh Word Meteorological Organization menjelaskan bahwa sebagian besar bencana banjir yang dipicu oleh curah hujan di Indonesia berkaitan dengan curah hujan ekstrem [1].



Gambar 1. Arsitektur jaringan syaraf tiruan.

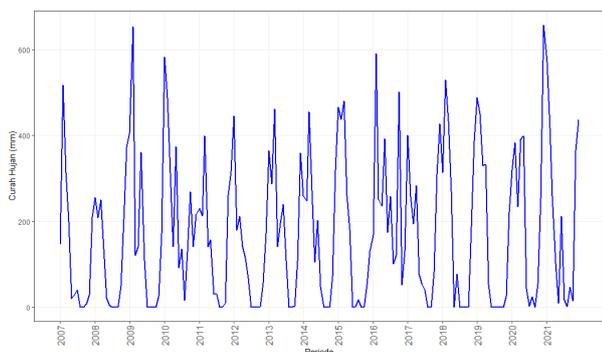


Gambar 2. Arsitektur jaringan ELM.

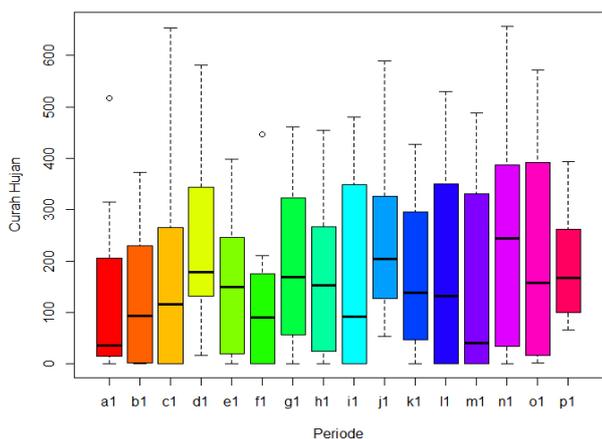
Banjir yang disebabkan oleh luapan sungai telah menjadi salah satu isu penting bagi sebagian besar waterfront cities di Asia. Secara umum banjir adalah suatu kejadian dimana air di dalam saluran meningkat sehingga melampaui batas kapasitas daya tampungnya [2]. Banjir terjadi disebabkan curah hujan dengan intensitas tinggi, dengan kata lain curah hujan mempengaruhi terhadap kejadian banjir.

Dalam pendugaan tingkat risiko pada banjir diperlukan peramalan terhadap curah hujan untuk mendapatkan model curah hujan mendatang. Salah satu metode peramalan yang telah digunakan adalah metode *Extreme Learning Machine* (ELM). ELM adalah salah satu metode jaringan syaraf tiruan yang digunakan untuk memprediksi peramalan. Metode ini merupakan pembelajaran baru dari *Artificial Neural Network* (ANN) dan dikembangkan untuk meningkatkan efisiensi (SLFNs) Single Hidden Layer Feedforward. ELM sepenuhnya otomatis diimplementasikan tanpa penyetelan berulang, dan secara teori tidak ada intervensi yang diperlukan dari pengguna [3].

Untuk mengetahui risiko kerugian akibat banjir, yang harus dilibatkan ialah curah hujan yang ekstrem dimana



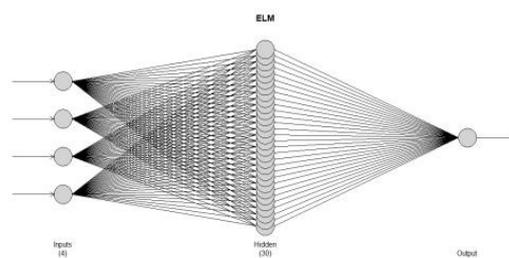
Gambar 3. Karakteristik curah hujan di Surabaya periode 2007-2021.



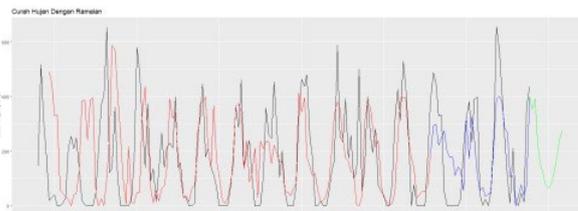
Gambar 4. Boxplot statistika deskriptif curah hujan periode tahunan.

kondisi jumlah curah hujan melebihi keadaan wajar. Salah satu cara yang dapat mengidentifikasi dan mengukur adanya kejadian ekstrem yaitu dengan *Extreme Value Theory* (EVT). EVT dapat meramalkan terjadinya kejadian ekstrem pada data heavytail. EVT juga dapat menjelaskan kerugian kejadian ekstrem yang tidak dapat dimodelkan dengan pendekatan biasa. Tujuan dari metode ini adalah untuk menentukan estimasi peluang kejadian ekstrem pada data curah hujan dengan memperhatikan ekor (tail) fungsi distribusi berdasarkan nilai-nilai ekstrem yang diperoleh. Identifikasi nilai ekstrem dengan EVT dilakukan dengan dua metode yaitu metode *Block Maxima* (BM) dan metode *Peaks Over Threshold* (POT). Metode *Block Maxima* (BM) dilakukan dengan nilai maksimum yang didapat dalam satu periode yang disebut sebagai blok dan metode sedangkan *Peaks Over Threshold* (POT) dilakukan dengan mengambil nilai yang melewati suatu nilai *threshold*.

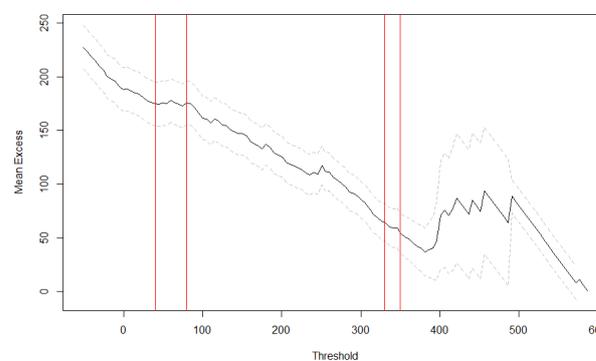
Peramalan curah hujan diperlukan untuk memberi gambaran keadaan curah hujan di masa mendatang yang dapat digunakan sebagai acuan dalamantisipasi menghadapi risiko terhadap banjir. Dengan melakukan peramalan curah hujan ini, risiko banjir dapat diminimalisir. Surabaya merupakan salah satu Kota terbesar kedua di Indonesia dengan tingkat banjir yang tinggi, oleh sebab itu peramalan curah hujan dilakukan untuk mendapatkan informasi mengenai jumlah curah hujan yang ada di Kota Surabaya. Dengan informasi mengenai jumlah curah hujan yang didapat diharapkan berguna untuk mengurangi dampak negatif yang ditimbulkan oleh intensitas curah hujan yaitu banjir sehingga dapat meminimalkan risiko banjir. Penelitian ini berfokus pada peramalan curah hujan menggunakan metode *Extreme Learning Machine* (ELM). Untuk mengetahui risiko kerugian



Gambar 5. Arsitektur jaringan *extreme learning machine*.



Gambar 6. Grafik aktual, latih, uji, dan ramalan.



Gambar 7. Grafik *mean residual plot*.

akibat banjir dengan metode *Value at Risk* (VaR) dengan pendekatan *Extreme Value Theory* (EVT).

II. TINJAUAN PUSTAKA

A. Curah Hujan dan Banjir

Curah Hujan (mm) adalah ketinggian air hujan yang terkumpul dalam penakar hujan pada tempat yang datar, tidak menyerap, tidak meresap dan tidak mengalir. Curah hujan adalah tinggi air yang terkumpul pada suatu tempat yang datar, tidak meresap dan tidak mengalir. Curah hujan dengan intensitas tinggi atau biasa disebut dengan curah hujan ekstrim telah lama diketahui menjadi faktor utama penyebab banjir di Indonesia.

Banjir adalah kejadian alam di mana suatu daerah atau daratan yang biasanya kering menjadi terendam air. Banyaknya air hujan yang turun ini tidak semuanya bisa cepat terserap dan mengalir, sehingga banyak air hujan yang menggenang. Air genangan dengan volume cukup besar ini sangat berbahaya karena dapat menyebabkan kerusakan properti seperti rumah, jembatan jalan, dan lain sebagainya. Selain itu, air genangan atau banjir dapat mengganggu kegiatan masyarakat sehari-hari.

B. Jaringan Syaraf Tiruan

Jaringan Syaraf Tiruan atau biasa disingkat JST adalah model matematika atau model komputasi yang terinspirasi dari struktur fungsional jaringan syaraf biologis. Arsitektur jaringan syaraf tiruan dapat dilihat pada Gambar 1. Jaringan

Tabel 1.
Variabel data penelitian

Variabel	Keterangan	Skala
Z	Curah hujan stasiun Juanda Surabaya	Rasio

Tabel 2.
Struktur data penelitian

Tahun	Bulan	Curah Hujan
2007	Januari	Z ₁
2007	Februari	Z ₂
2007	Maret	Z ₃
2007	April	Z ₄
.	.	.
.	.	.
.	.	.
2021	Oktober	Z ₂₅₀
2021	November	Z ₂₅₁
2021	Desember	Z ₂₅₂

Tabel 3.
Struktur data penelitian dengan metode *extreme learning machine*

Observasi	Variabel	Lag 1	Lag 2	Lag 3	Lag 4
	Z ₁	-	-	-	-
	Z ₂	Z ₁	-	-	-
	Z ₃	Z ₂	Z ₁	-	-
	Z ₄	Z ₃	Z ₂	Z ₁	-
1	Z ₅	Z ₄	Z ₃	Z ₂	Z ₁
2	Z ₆	Z ₅	Z ₄	Z ₃	Z ₂
3	Z ₇	Z ₆	Z ₅	Z ₄	Z ₃
⋮	⋮	⋮	⋮	⋮	⋮
n	Z _n	Z _{n-1}	Z _{n-2}	Z _{n-3}	Z _{n-4}

Tabel 4.
Statistika deskriptif curah hujan periode tahunan

Periode	Mean	Median	St.Dev	Maksimal
2007	126,280	34,850	160,8067	517,4
2008	125,242	92,650	129,8014	372,3
2009	165,580	115,200	205,9881	652,5
2010	241,200	177,800	169,9740	581,7
2011	149,090	148,800	136,6897	398,5
2012	115,792	90,850	129,5107	445,9
2013	189,200	168,300	155,7816	461,1
2014	165,020	153,250	149,0025	455,1
2015	168,720	91,250	194,4648	479,8
2016	248,100	204,600	166,1738	589,6
2017	177,000	138,600	154,9308	427,3
2018	183,200	132,500	197,0597	529,0
2019	158,620	40,150	193,4057	487,8
2020	230,090	244,200	209,1108	657,1
2021	204,530	156,900	201,8364	572,0
2022	190,950	166,760	109,4722	392,9

Syaraf Tiruan (JST) adalah sekumpulan neuron yang terorganisir dalam lapisan-lapisan (*layers*), diantaranya:

1. *Input Layer*
2. *Hidden Layer*
3. *Output Layer*

C. *Normalisasi dan Denormalisasi*

Dalam proses penerapan jaringan syaraf tiruan, sering kali kumpulan data yang tersedia tidak dapat langsung diproses. Dilakukan normalisasi, penskalaan, atau transformasi data yang ada sebelum dapat diproses di jaringan saraf.

1) *Normalisasi*

Sebagian besar jaringan syaraf tiruan dan kecerdasan buatan lainnya, memerlukan data untuk ditransformasikan ke skala tertentu sebelum diproses lebih lanjut. Biasanya dataset akan ditransformasikan menjadi range 0 hingga 1 atau -1 hingga 1. Proses transformasi ini sering juga disebut dengan normalisasi. Setiap data harus ditransformasikan ke dalam range yang sama. Untuk melakukan transformasi dataset,

Tabel 5.
Pembagian data latihan dan data uji

Pembagian	Presentase	Total Data
Data Latih	80%	144
Data Uji	20%	36
Total	100%	180

Tabel 6.
Hasil MAPE data latihan

Repetisi	<i>Hidden Layer Neuron</i>				
	10	15	20	25	30
5	12,8205	11,5441	11,4227	10,5714	10,4009
10	12,7852	11,7535	11,0914	10,4921	9,9697
15	12,2063	11,6222	11,1832	10,3694	10,1009
20	12,6106	11,6343	10,9338	10,5075	9,8123

Tabel 7.
Hasil MAPE data uji

Repetisi	<i>Hidden Layer Neuron</i>				
	10	15	20	25	30
5	13,4384	12,6866	11,7638	7,8503	3,4066
10	13,3166	13,1439	9,8455	6,1380	2,9967
15	13,0676	12,4662	10,2005	4,8139	2,9338
20	13,1898	12,7551	10,3883	6,2643	2,2883

penelitian ini menggunakan metode *Min Max Normalization*. Normalisasi dapat dilakukan dengan persamaan berikut:

$$X_{norm} = \frac{Xi - X_{min}}{X_{max} - X_{min}}$$

2) *Denormalisasi*

Denormalisasi adalah kebalikan dari proses normalisasi. Jika dalam proses normalisasi data yang ada (data sebenarnya) ditransformasikan ke dalam range tertentu, maka proses denormalisasi adalah proses yang menghasilkan output berupa data real world. Dengan kata lain, proses denormalisasi mengubah nilai jaringan saraf kembali ke domain aslinya. Denormalisasi dapat dilakukan dengan persamaan berikut:

$$X = X_{norm}(X_{max} - X_{min}) + X_{min}$$

D. *Extreme Machine Learning*

Extreme Learning Machine (ELM) merupakan metode pembelajaran baru dari jaringan saraf tiruan. Metode ELM adalah suatu jaringan saraf tiruan *feedforward* dengan *n* *single hidden layer feedforward neural networks* (SLFNs), metode ELM memiliki beberapa kelebihan dibandingkan dengan metode SLFNs terdahulu yaitu:

1. ELM memiliki kecepatan belajar yang jauh lebih baik. Hal ini dikarenakan dalam prosesnya tidak diperlukan iterasi sebagaimana yang diperlukan dalam metode SLFNs konvensional.
2. Tidak seperti SLFNs konvensional yang mudah terjebak pada minima lokal, ELM memiliki kinerja generalisasi yang lebih baik.
3. Pada metode ELM dapat digunakan fungsi aktivasi yang tidak terdiferensiasi.

Metode ELM mempunyai model matematis yang berbeda dari jaringan syaraf tiruan *feedforward*. Model matematis dari ELM lebih sederhana dan efektif. Untuk *N* jumlah sampel yang berbeda $x_i, [x_{i1}, x_{i2}, \dots, x_{in}]^T \in R^m$ SLFNs standar dengan jumlah layer tersembunyi sebanyak dengan fungsi aktivasi [4]. Arsitektur jaringan syaraf ELM dapat dilihat pada Gambar 2.

Tabel 8.
Hasil ramalan

Bulan, Tahun	Peramalan Data Curah Hujan Kota Surabaya
Januari 2022	354,5077474
Februari 2022	392,9100449
Maret 2022	226,4794518
April 2022	155,0761591
Mei 2022	117,9077542
Juni 2022	74,53179729
Juli 2022	66,09237333
Agustus 2022	81,08760055
September 2022	122,2082288
Oktober 2022	178,440924
November 2022	245,8306274
Desember 2022	276,3682656

Tabel 9.

Hasil parameter <i>block maxima</i>		
Karakteristik	Nilai	
Banyaknya Blok	47	15
Pengamatan tiap blok	4	12
Bentuk (ξ)	0,2166	0,0476
Lokasi (μ)	239,9944	380,1639
Skala (σ)	121,4777	45,2967
AIC	594,4545	166,0256

Tabel 10.

Hasil estimasi <i>value at risk</i> pada GEV			
Banyak Blok	Tingkat Kepercayaan		
	90%	95%	99%
15	332,8736	332,9022	332,9482

E. Mean Absolute Percentace Error (MAPE)

Menurut Zainun dan Majid (2003), MAPE adalah proses evaluasi kinerja suatu ramalan dengan menghitung nilai kesalahan (*error*) [5]. Persamaan MAPE yang digunakan antara lain sebagai berikut.

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{X_t - F_t}{X_t} \right| \times 100\%$$

dengan,

X_t : data aktual periode t

F_t : peramalan periode t

n : banyaknya data

F. Extreme Value Theory

Extreme Value Theory (EVT) adalah salah satu metode yang digunakan untuk menganalisis kejadian ekstrem seperti pada bidang klimatologi, hidrologi, keuangan, ekonomi dan asuransi. Metode ini mampu menjelaskan kerugian kejadian ekstrem yang tidak dapat dimodelkan dengan pendekatan biasa, seperti distribusi normal, karena data finansial tidak bersifat normal, lebih bersifat ekor gemuk. Ekor gemuk secara umum menjelaskan bahwa nilai esktrim terjadi lebih sering jika dibandingkan dengan data normal. Sebagian besar data iklim seperti curah hujan memiliki ekor distribusi yang gemuk, yaitu ekor distribusi turun secara lambat bila dibandingkan dengan distribusi normal [6]. Dalam mengidentifikasi pergerakan nilai ekstrim, ada dua pendekatan yang digunakan. Pendekatan pertama adalah metode *Block maxima* (BM), yang mengambil nilai maksimum dalam satu periode, sedangkan metode kedua adalah metode *Peaks Over Threshold* (POT), yang mengambil nilai yang melebihi nilai ambang batas.

Tabel 11.

Uji kolmogorov-smirmov	
Data	Threshold
	MRLP
Curah Hujan Surabaya	350(32)

Tabel 12.

Estimasi parameter <i>peaks over threshold</i>	
Karakteristik	Nilai
Threshold (u)	3530
Jumlah Pengamatan (n)	188
Jumlah Pengamatan diatas Threshold (N_u)	328
Bentuk (ξ)	-0,00345
Skala (σ)	56,8816

Tabel 13.

Hasil estimasi <i>value at risk</i> GPD			
Data	Tingkat Kepercayaan		
	90%	95%	99%
Curah Hujan	379,9782	4418,2294	503,5961

G. Estimasi Parameter Generalized Extreme Value (GEV)

Block Maxima, metode BM adalah metode yang dapat mengidentifikasi nilai ekstrem berdasarkan nilai maksimum dari data observasi yang dikelompokkan berdasarkan periode tertentu atau yang disebut blok [7]. Langkah-langkah pengambilan sampel dengan metode Block Maxima adalah sebagai berikut.

- Membagi data ke dalam blok dengan periode waktu tertentu, seperti bulanan, triwulan, atau tahunan.
- Setelah blok terbentuk, kemudian mengambil sampel nilai ekstrim dengan nilai maksimum atau nilai tertinggi dari data pada setiap blok. Sampel tersebut digunakan dalam analisis

H. Estimasi Parameter Generalized Pareto Distribution (GPD)

1) Peak Over Threshold

Peak Over Threshold (POT) merupakan salah satu metode untuk mengidentifikasi nilai ekstrim dengan menggunakan *threshold* (u). Data yang berada di atas *threshold* tersebut akan diidentifikasi sebagai nilai ekstrim. Semakin tinggi nilai *threshold* maka data ekstrim akan semakin mengikuti distribusi General Pareto [8].

2) Penentuan Ambang Batas

Untuk mendapatkan nilai ekstrim dengan metode *Peaks Over Threshold*, terlebih dahulu menentukan nilai *threshold*(u) sebagai nilai acuan. Ada beberapa cara dalam menentukan *threshold* diantaranya adalah dengan metode *Mean Residual Life Plot* (MRLP) dan metode persentase. Metode MRLP ini didasarkan pada rata-rata pelampauan [9].

Penentuan ambang batas meliputi uji kesesuaian distribusi. Menurut Daniel (1989) pemeriksaan distribusi dilakukan dengan uji *Kolmogorov-Smirnov*. Pengujian ini dilakukan dengan menyesuaikan fungsi distribusi empiris (berdasarkan sampel).

$F_n(x)$ dengan distribusi teoritis tertentu (sesuai yang dihipotesiskan) $F_0(x)$. Uji Hipotesis terdiri dari (a) $H_0 : F_n(x) = F_0(x)$ (Data mengikuti distribusi teoritis $F_0(x)$). (b) $H_1 : F_n(x) \neq F_0(x)$ (Data tidak mengikuti dist. teoritis $F_0(x)$). (c) Statistik Uji : $D = \text{Sup}|F_n(x) - F_0(x)|$. (d) Daerah Kritis : Tolak H_0 jika $D_{hitung} < D_\alpha$ [10].

I. Value at Risk

Value at Risk didefinisikan sebagai kerugian maksimum yang diharapkan dari nilai suatu data dalam periode tertentu dan pada tingkat kepercayaan tertentu. VaR merupakan konsep perhitungan risiko yang terjadi akibat pengembangan konsep kurva normal, nilai VaR dapat diperoleh dengan menggunakan metode GEV dengan rumus sebagai berikut:

$$VaR_{p(GEV)} = \hat{\mu} + \frac{\hat{\sigma}}{\xi} [(-\ln(1-p))^{-\xi} - 1]$$

sedangkan untuk nilai VaR dengan metode GPD dapat diperoleh menggunakan rumus sebagai berikut:

$$VaR_{p(GPD)} = \hat{\mu} + \frac{\hat{\sigma}}{\xi} \left[\left(\frac{n}{Nu} \times p \right)^{-\xi} - 1 \right]$$

dimana Nu merupakan banyaknya pengamatan yang melebihi *threshold* dan p merupakan persen kuantil [11].

III. METODOLOGI PENELITIAN

A. Sumber Data dan Variabel Penelitian

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari Badan Stasiun Meteorologi, Klimatologi, dan Geofisika (BMKG). Data tersebut berupa data bulanan curah hujan di Surabaya dengan periode selama 14 tahun, yaitu dari Januari 2007 hingga Desember 2021. Tabel 1 merupakan variabel data penelitian, sedangkan Tabel 2 merupakan struktur data penelitian.

Tabel 3 merupakan struktur data penelitian yang digunakan dalam peramalan menggunakan metode Extreme Learning Machine.

B. Langkah Analisis

Berikut merupakan langkah analisis yang akan dilakukan pada penelitian ini:

1. Melakukan studi literatur
2. Mengumpulkan data curah hujan di stasiun Juanda Surabaya pada periode Januari 2007 - Desember 2021.
3. Melakukan *preprocessing* data sebelum dianalisis meliputi identifikasi *missing value* dan imputasi data.
4. Melakukan eliminasi data untuk data yang tidak sesuai (seperti tanggal pencatatan yang tidak sesuai) serta imputasi data pada *missing* menggunakan rata-rata curah hujan pada tanggal dan bulan yang sama dengan tanggal dan bulan data *missing*
5. Mengidentifikasi pola curah hujan dengan statistika deskriptif dan pola sebaran hujan.
6. Melakukan perancangan model *Extreme Learning Machine* (ELM) dengan langkah-langkah sebagai berikut: (a) Membagi data menjadi data training dengan presentase 80% dan testing 20%. (b) Melakukan perancangan dan pelatihan model ELM untuk mendapatkan bobot dan bias dengan tingkat kesalahan rendah. (c) Melakukan pengujian terhadap model jaringan syaraf tiruan yang telah dibangun. (d) Melakukan proses dan hasil prediksi data dengan kriteria MAPE.
7. Mendapatkan hasil peramalan untuk dicari titik extreme
8. Melakukan pemodelan nilai ekstrem dari data actual dan hasil peramalan
9. Melakukan perancangan model *Block Maxima* dengan langkah-langkah sebagai berikut: (a) Mengambil sampel

ekstrem dengan membuat blok untuk data curah hujan 2007 – 2021 dengan pendekatan GEV. (b) Memeriksa kesesuaian distribusi menggunakan pengujian hipotesis dengan uji Kolmogorov-Smirnov untuk masing-masing blok.

10. Melakukan perancangan model *Peak Over Threshold* dengan langkah-langkah sebagai berikut: (a) Menentukan nilai threshold dengan plot MRL. (b) Mengambil sampel ekstrem dimana data yang diambil merupakan data yang berada diatas nilai threshold dengan pendekatan GPD. (c) Memeriksa kesesuaian distribusi menggunakan pengujian hipotesis dengan uji Kolmogorov-Smirnov.
11. Melakukan perhitungan *Value at Risk* untuk mengetahui hasil dari masing-masing model (GPD dan GEV)
12. Menarik kesimpulan

IV. ANALISIS DAN PEMBAHASAN

A. Karakteristik Curah Hujan

Karakteristik data digunakan untuk melihat gambaran umum sebuah data. Penelitian ini berfokus pada curah hujan Kota Surabaya periode Januari 2007 hingga Desember 2021 Karakteristik data curah hujan dapat dilihat dengan menggunakan grafik dan analisis statistika deskriptif yang ditunjukkan pada Gambar 3 dan Tabel 4

Pada Gambar 3 dapat dilihat bahwa kondisi curah hujan yang terjadi di wilayah Surabaya terindikasi terdapat pola musiman. Pada tahun 2007 hingga 2021 curah hujan cenderung ekstrem dan tidak stabil dengan tingkat paling tinggi terjadi di tahun 2009 dan 2021. Selain itu, dapat dilihat pada statistika deskriptif mengenai data curah hujan di Surabaya, seperti yang ditampilkan pada Tabel 4. Dapat diketahui bahwa masing-masing periode memiliki tingkat tinggi rendahnya curah hujan yang berbeda. Karakteristik data dari masing-masing periode statistika deskriptif data curah hujan dapat dilihat pada Tabel 4.

Tabel 4 menunjukkan statistika deskriptif pada masing-masing periode curah hujan. Pada tabel 4, dapat diketahui curah hujan tertinggi terjadi di tahun 2020 yaitu sebesar 657,1. Hal ini terjadi dikarenakan pada tahun 2020 Indonesia masuk dalam lima tahun musim hujan terbasah sepanjang 40 tahun kebelakang. BMKG Juanda menyatakan, berdasarkan pantauan terakhir kondisi atmosfer dan muka laut yang ada menyebabkan peningkatan intensitas curah hujan

Pada Tabel 4 tahun 2010 memiliki rata-rata tertinggi dibandingkan dengan tahun lainnya, hal ini disebabkan karena pada tahun 2010 hujan di tiap bulannya memiliki intensitas curah hujan yang tinggi. Hal tersebut dapat dilihat pada Gambar 4.

Nilai tengah untuk semua periode curah hujan (2007 – 2022) berada pada interval 30-245, dengan nilai tengah tertinggi pada periode 2016 dan 2020. Sementara nilai tengah terendah pada periode 2007 dan 2019. Dari sebaran di atas, terlihat bahwa ada curah hujan yang outlier karena memiliki nilai yang sangat tinggi di atas normal keragaman di perodenya. Curah hujan tersebut adalah pada periode 2007 dan 2012.

B. Identifikasi Model Peramalan Extreme Learning Machine (ELM)

1. Pemeriksaan data *missing*

2. Mengubah data menjadi time series
3. Normalisasi data. Data dirubah menjadi interval ke yang lebih kecil $[0,1]$ sehingga dapat digunakan sebagai input pada jaringan.
4. Pembagian data latih dan data uji disajikan pada tabel 5.
5. Arsitektur jaringan extreme learning machine. Gambar 5 merupakan gambaran asitektur data yang digunakan dalam proses prediksi dari metode Extreme Learning Machine data curah hujan Kota Surabaya. Neuron input pada asitektur data yaitu sebanyak 4 neuron yang diasumsikan merupakan perwakilan dari 4 bulan. Neuron output pada asitektur data yaitu sebanyak 1 Neuron yang merupakan hasil dari peramalan. Pengujian jumlah neuron pada hidden layer menunjukkan bahwa dengan jumlah neuron yang banyak maka akan mendapatkan nilai error yang lebih kecil. Pada penelitian ini dibandingkan beberapa jumlah *hidden layer* neuron, yaitu 10, 15, 20, 25, dan 30 hidden neuron. Hal ini bertujuan untuk mengetahui jumlah neuron yang terbaik berdasarkan nilai MAPE (*Mean Absolut Percentage Error*).
6. Output Training dan Testing Data. Tabel 6 menunjukkan tingkat kesalahan hasil peramalan model *Extreme Learning Machine* terbaik yang dihitung dengan menggunakan nilai MAPE pada training ditunjukkan pada *hidden layer neuron* 30 dengan kombinasi rep 20 yaitu sebesar 9,81230 yang kurang dari 10%. Tabel 7 menunjukan tingkat kesalahan hasil peramalan model Extreme Learning Machine pada data uji. Model terbaik ditentukan dengan menggunakan perhitungan nilai MAPE model. Sehingga dapat dilihat bahwa nilai MAPE terkecil ditunjukkan oleh model dengan hidden layer neuron 30 dan kombinasi rep 20 yaitu sebesar 2,28831, kurang dari 10%. Hasil peramalan pada data pelatihan cenderung memiliki tingkat kesalahan yang lebih besar daripada data pengujian.
7. Hasil Ramalan. Berdasarkan hasil peramalan yang telah diperoleh pada Tabel 8 dapat diketahui bahwa curah hujan pada bulan Januari 2022 diprediksi 354,5077474. Gambar 6 merupakan grafik hasil peramalan data curah hujan Kota Surabaya untuk periode Januari 2022. Gambar 6 tersebut menunjukkan bahwa curah hujan Kota Surabaya memiliki curah hujan menurun pada Maret 2022 hingga Juli 2022 dan kemudian diprediksi memiliki curah hujan naik sampai akhir Desember 2022. Curah hujan Kota Surabaya diperkirakan memiliki rata-rata sebesar 190,9534 dengan curah hujan tertinggi mencapai 392,91 dan curah hujan terendah 66,09.

C. Identifikasi Model Peramalan Extreme Value Theory (EVT)

1) Pemodelan Block Maxima

Perhitungan estimasi parameter metode *Block Maxima* adalah menggunakan data ekstrem yang telah diperoleh sebelumnya dan telah dilakukan pengujian kesesuaian distribusi. Hasil estimasi parameter dengan menggunakan Maximum Likelihood disajikan pada Tabel 9.

Tabel 9 menunjukkan bahwa banyaknya blok yang terbentuk adalah 47 dan 15 blok dengan banyaknya masing-masing pengamatan tiap blok adalah 4 dan 12 pengamatan. Hasil estimasi parameter lokasi tiap blok menyatakan letak titik pemusatan data, parameter skala menyatakan keragaman

data, dan parameter ekor menyatakan perilaku ekor kanan (maksimum). Hasil pengujian tipe distribusi menunjukkan bahwa data tidak berdistribusi *Gumbel* ataupun *Frechet*. Berdasarkan nilai parameter ekor yang dihasilkan menunjukkan nilai yang kurang dari 0, sehingga dapat disimpulkan bahwa distribusi curah hujan merupakan dari kelas distribusi *Weibull*.

2) Estimasi Value at Risk (VaR) GEV

Perhitungan estimasi *Value at Risk* menggunakan pendekatan GEV dihitung dengan selang kepercayaan tertentu yaitu 90%, 95%, dan 99%. Hasil dari estimasi VaR ditampilkan pada Tabel 10.

Berdasarkan Tabel 10 terlihat estimasi VaR pada data curah hujan dengan tingkat kepercayaan berbeda. Masing-masing memiliki interpretasi sebagai berikut (pada banyak blok 15 (12 pengamatan)):

- a. Tingkat Kepercayaan 90%. Curah hujan Surabaya memiliki nilai VaR sebesar 332,8736. Artinya terdapat kemungkinan sepuluh persen curah hujan akan turun lebih dari 332,8736 pada esok hari.
- b. Tingkat Kepercayaan 95%. Curah hujan Surabaya memiliki nilai VaR sebesar 332,9022. Artinya terdapat kemungkinan lima persen curah hujan akan turun lebih dari 332,9022 pada esok hari.
- c. Tingkat Kepercayaan 99%. Curah hujan Surabaya memiliki nilai VaR sebesar 332,9482. Artinya terdapat kemungkinan satu persen curah hujan akan turun lebih dari 332,9482 pada esok hari.

3) Pemodelan Peaks Over Threshold

Pendekatan lainnya digunakan untuk mengidentifikasi nilai ekstrem dalam suatu data adalah pendekatan *Peaks Over Threshold*. Konsep dari pendekatan ini adalah mengidentifikasi nilai ekstrem dengan cara menetapkan ambang batas atau threshold. Data yang melebihi nilai threshold dianggap sebagai nilai ekstrem. Penentuan threshold sangat penting karena jika penentuan nilai threshold terlalu tinggi maka data ekstrim terlalu sedikit dan mengakibatkan variansi yang tinggi. Sedangkan jika threshold terlalu kecil maka akan mengakibatkan bias.

Untuk penentuan Threshold, pada penelitian ini penentuan *threshold* menggunakan MRLP. Nilai *threshold* ditentukan berdasarkan data curah hujan bulanan tahun 2007 hingga 2022. Pada metode MRLP penentuan threshold memperhatikan banyaknya amatan yang melampui threshold dan pendekatan garis linear yang konsisten. Gambar 5 menunjukkan grafik MRL pada data curah hujan Kota Surabaya menggunakan data tahun 2007 hingga 2022 dengan selang *threshold* $u \approx 600$ pada selang kepercayaan 95%. Grafik MRL menunjukkan garis relatif linier dari $u \approx 40$ sampai $u \approx 80$. Pada $u \approx 330$ sampai $u \approx 350$ terdapat beberapa pola linear yang terbentuk. Berdasarkan uraian tersebut, grafik MRL akan dibatasi pada *threshold* $u \approx 40$ sampai $u \approx 350$. Gambar 7 memperlihatkan *threshold* yang perlu diperhatikan karena garis relatif linier dan konsisten diatas setelah *threshold* yaitu *threshold* $u \approx 40$ sampai $u \approx 80$, $u \approx 230$ sampai $u \approx 250$ $u \approx 330$ sampai $u \approx 350$. Pada $u \approx 350$ hanya sebanyak 32 amatan sehingga jumlah tersebut dinilai belum cukup untuk melakukan pendugaan GPD.

Selang nilai *threshold* diambil sesuai dengan plot yang membentuk menyerupai garis linier. Selang nilai tersebut

dilakukan fitting GPD yang dapat dilihat pada Lampiran 2. Nilai calon threshold untuk data curah hujan Kota Surabaya dengan metode MRLP dipilih berdasarkan hasil fitting terhadap selang nilai threshold yang menghasilkan p-value < alpha (0,05) pada MK Test, nilai AIC terkecil, dan jumlah data diatas threshold tidak kurang dari 10% dari total data. Pada disajikan pada Tabel 9, angka dalam kurung menunjukkan banyaknya nilai curah hujan yang berada diatas nilai threshold, selanjutnya calon threshold yang diperoleh pada Tabel 9 akan diuji kesesuaian sebarannya apakah mengikuti sebaran GPD atau tidak.

Berdasarkan hasil pengujian pada Tabel 11, dapat diketahui bahwa distribusi dengan threshold 350 menyebar mengikuti sebaran GPD yang artinya bahwa data ekstrem yang digunakan telah mengikuti distribusi GPD.

4) Estimasi Parameter Peaks Over Threshold

Perhitungan estimasi parameter metode POT adalah menggunakan data ekstrem yang berada diatas data ekstrem. Hasil estimasi parameter dengan menggunakan *Maximum Likelihood* disajikan pada Tabel 12

Tabel 12 menghasilkan estimasi parameter bentuk yang menunjukkan perilaku ekor dari data ekstrem, dimana semakin besar nilai bentuk maka peluang terjadinya nilai ekstrem akan semakin besar pula, sedangkan untuk parameter skala menunjukkan keragaman nilai ekstrem.

5) Estimasi Value at Risk (VaR) GPD

Perhitungan estimasi Value at Risk menggunakan pendekatan GPD dihitung dengan selang kepercayaan tertentu yaitu 90%, 95%, dan 99%. Hasil dari estimasi VaR ditampilkan pada Tabel 13

Berdasarkan Tabel 13 terlihat estimasi VaR pada data curah hujan dengan tingkat kepercayaan berbeda. Masing-masing memiliki interpretasi sebagai berikut:

1. Curah hujan Surabaya memiliki nilai VaR sebesar 379,9782. Artinya terdapat kemungkinan sepuluh persen curah hujan akan turun lebih dari 379,9782 pada esok hari.
2. Curah hujan Surabaya memiliki nilai VaR sebesar 418,2294. Artinya terdapat kemungkinan lima persen curah hujan akan turun lebih dari 418,2294 pada esok hari.
3. Curah hujan Surabaya memiliki nilai VaR sebesar 503,5961. Artinya terdapat kemungkinan satu persen curah hujan akan turun lebih dari 503,596 pada esok hari.

V. KESIMPULAN/RINGKASAN

Kesimpulan yang didapatkan dari penelitian ini adalah sebagai berikut: (1) Model terbaik pada proses peramalan *Extreme Learning Machine* (ELM) diperoleh dengan jumlah

repetisi 20 dan *hidden layer* 30. MAPE yang dihasilkan pada data pengujian adalah sebesar 9,8123. Hasil peramalan yang didapat menunjukkan curah hujan tertinggi di tahun 2022 terjadi pada bulan februari. (2) Estimasi risiko banjir dapat dilihat dari hasil estimasi *Value at Risk* (VaR) pada tingkat kepercayaan 90%, 95%, dan 99% yaitu pada GEV secara berturut-turut 332,8736, 332,9022, 332,9482 secara berturut-turut sebesar dan pada GPD secara berturut-turut sebesar 379,9782, 418,2294, 503,5961 sehingga pemerintah Surabaya dapat membuat kebijakan terkait dengan kapasitas *drainase* atau penampungan air hujan sesuai dengan nilai VaR yang telah diperoleh.

Saran untuk penelitian selanjutnya yang dapat diterapkan berdasarkan penelitian ini adalah dapat menambah jumlah data dari data yang digunakan peneliti. Penambahan data dilakukan dengan tujuan untuk mendapatkan hasil peramalan dan grafik *Mean Residual Plot* yang lebih baik. Penelitian selanjutnya dapat menambahkan metode peramalan pembandingan dari *Extreme Learning Machine* (ELM). Penambahan metode dilakukan dengan tujuan untuk membandingkan ke efektivitasan hasil akurasi peramalan.

DAFTAR PUSTAKA

- [1] A. B. Sekaranom. *Kejadian Hujan Ekstrem Wilayah Tropis: Kombinasi Observasi Permukaan dan Satelit Meteorologi serta Karakteristik Lingkungan Pembentukannya*. Yogyakarta: Gadjah Mada University Press, 2020. ISBN: 978-602-386-923-7.
- [2] M. Hasbi. *Nilai Viskositas Aliran Sungai Sebagai Salah Satu Indikasi Potensi Banjir Bndang*. Yogyakarta: Deepublish, 2020. ISBN: 978-623-02-2050-0.
- [3] S. Kot and S. Nadarajah. *Extreme Value Distributions*. London: Imperial College Press, 2000. ISBN: 1860942245.
- [4] J. Tang, C. Deng, and G-B. Huang, "Extreme learning machine for multilayer perceptron," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 4, pp. 809-821, 2016. <https://doi.org/10.1109/TNNLS.2015.2424995>.
- [5] A. Raharja, "Penerapan Metode Exponential Smoothing untuk Peramalan Penggunaan Waktu Telepon di PT. Telkomsel DIVRE3 Surabaya," Departemen Sistem Informasi, Institut Teknologi Sepuluh Nopember, 2010.
- [6] K. Dharmawan, "Estimasi nilai VaR dinamis indeks saham menggunakan peak-over threshold dan block maxima," *Jurnal Matematika*, vol. 2, no. 2, p. 24, 2012.
- [7] L. F. Amalia, "Estimasi Parameter pada Pemodelan Spatial Extreme Value dengan Pendekatan Copula," Departemen Statistika, Institut Teknologi Sepuluh Nopember, 2017.
- [8] A. Rahayu, "Estimasi Parameter Distribusi Generalized Extreme Value (GEV) (Studi Kasus: Identifikasi Perubahan Iklim di Jakarta)," Departemen Statistika, Institut Teknologi Sepuluh Nopember, 2012.
- [9] G. Prayogo, "Change Point Analysis untuk Nilai Ekstrem dengan Pendekatan Sebaran Pareto Terampat.," Departemen Statistika, Institut Pertanian Bogor, 2014.
- [10] W. W. Daniel. *Statistika Nonparametrik Terapan*. Jakarta: Gramedia, 1989. ISBN: 9794036404.
- [11] V. Marimoutou, B. Raggad, and A. Trabelsi, "Extreme Value Theory and Value at Risk: Application to Oil Market," *Energy Economics*, vol. 31, no. 4, pp. 519-530. doi: 10.1016/j.eneco.2009.02.005.