

## Analisis Keparahan Banjir di DKI Jakarta Tahun 2020 Menggunakan Algoritma Random Forest

Alman Faluthi Achmad<sup>1</sup>, Darminto<sup>2</sup>, Dionaci rizki utami<sup>3</sup>, Yamin Nuryamin<sup>4</sup>, Ade Priyatna<sup>5</sup>

<sup>1</sup>Teknologi Informasi, Teknik dan Informatika, Universitas Bina Sarana Informatika

<sup>1</sup>[faluthiachmad16@gmail.com](mailto:faluthiachmad16@gmail.com), <sup>2</sup>[acongminto@gmail.com](mailto:acongminto@gmail.com), <sup>3</sup>[yonachi170305@gmail.com](mailto:yonachi170305@gmail.com), <sup>4</sup>[yamin.yny@bsi.ac.id](mailto:yamin.yny@bsi.ac.id)

Corresponding Author: Alman Faluthi Achmad

### ABSTRACT

Flooding is one of the most frequent natural disasters in DKI Jakarta Province, causing significant social and economic impacts. This study aims to predict flood severity levels using the Random Forest algorithm based on flood event data from 2020 obtained from BPBD DKI Jakarta. The analysis includes data cleaning, normalization, severity labeling, model training, and evaluation in Google Colab. The model achieved high performance with 100% accuracy, precision, recall, and F1-score across three severity categories: Low, Medium, and High. Feature importance analysis revealed that the number of affected people and water level were the most dominant factors. The model can serve as a decision-support tool for disaster mitigation and could be improved with additional multi-year data.

**Keywords:** Jakarta Flooding, Flood Severity Level, Random Forest Classifier, Flood Prediction, BPBD Jakarta Data

### ABSTRAK

Banjir merupakan salah satu bencana alam yang paling sering terjadi di Provinsi DKI Jakarta dan menimbulkan dampak sosial serta ekonomi yang signifikan. Penelitian ini bertujuan untuk memprediksi tingkat keparahan banjir menggunakan algoritma Random Forest Classifier berdasarkan data kejadian banjir tahun 2020 yang diperoleh dari BPBD DKI Jakarta. Proses analisis dilakukan melalui tahapan pembersihan data, normalisasi, pembentukan label tingkat keparahan, pelatihan model, serta evaluasi performa di Google Colab. Hasil penelitian menunjukkan bahwa model Random Forest mampu mencapai performa tinggi dengan nilai akurasi, precision, recall, dan f1-score sebesar 100% pada tiga kategori klasifikasi: Rendah, Sedang, dan Tinggi. Variabel jumlah terdampak jiwa dan ketinggian air menjadi faktor paling dominan dalam menentukan tingkat keparahan banjir. Model ini berpotensi menjadi alat bantu dalam pengambilan keputusan mitigasi bencana, serta dapat dikembangkan lebih lanjut dengan menambahkan fitur curah hujan dan data multi-tahun. Sehingga dapat membantu pemerintah dalam menanggulangi bencana banjir.

**Kata Kunci:** Banjir DKI Jakarta, Tingkat Keparahan Banjir, Random Forest Classifier, Prediksi Banjir, Data BPBD Jakarta

## 1. Pendahuluan

Banjir merupakan salah satu bencana alam yang paling sering terjadi di wilayah Provinsi DKI Jakarta dan terus menjadi isu kritis hingga saat ini. Berdasarkan laporan BPBD DKI Jakarta, pada tahun 2020 terjadi lebih dari 450 kejadian banjir dengan jumlah pengungsi mencapai sekitar 100.000 jiwa. Tingginya frekuensi dan dampak banjir tersebut dipengaruhi oleh beberapa faktor, di antaranya intensitas curah hujan yang meningkat, penurunan muka tanah, sistem drainase yang tidak memadai, serta pemanfaatan tata ruang yang belum selaras dengan daya dukung lingkungan. Kondisi ini menyebabkan kerugian materiil, kerusakan infrastruktur, serta menimbulkan dampak sosial ekonomi yang signifikan bagi masyarakat.

Berbagai studi terkait analisis banjir di wilayah perkotaan telah dilakukan dengan memanfaatkan data spasial, pemodelan hidrologi, maupun analisis kebijakan. Namun, sebagian



Lisensi

Lisensi Internasional Creative Commons Attribution-ShareAlike 4.0.

besar penelitian tersebut berfokus pada pemetaan wilayah rawan banjir atau identifikasi faktor penyebab, sehingga masih terdapat kebutuhan akan model prediktif yang mampu memperkirakan tingkat keparahan banjir secara lebih akurat. Gap ini menunjukkan perlunya pendekatan komputasi yang lebih adaptif dan mampu mengolah data dalam jumlah besar.

Penelitian ini dilakukan dengan memanfaatkan algoritma Random Forest sebagai salah satu metode Machine Learning yang terbukti efektif dalam menangani permasalahan klasifikasi berbasis data multidimensional. Tujuan penelitian ini adalah memprediksi tingkat keparahan banjir di DKI Jakarta berdasarkan data kejadian banjir tahun 2020 dari BPBD, sehingga dapat mendukung proses mitigasi dan pengambilan keputusan di wilayah perkotaan. Penggunaan model ini diharapkan mampu memberikan kontribusi terhadap pengembangan sistem prediksi banjir yang lebih presisi dan aplikatif.

## 2. METODOLOGI PENELITIAN

Penelitian ini menggunakan dataset publik dari BPBD DKI Jakarta berisi informasi kejadian banjir tahun 2020. Tahapan penelitian meliputi:

### 2.1. Sumber

Data diperoleh dari dataset resmi “Kejadian Bencana Banjir di Provinsi DKI Jakarta Tahun 2020” dalam format Excel (.xlsx). Dataset terdiri atas 1.006 baris data dan 18 kolom, dengan beberapa variabel penting sebagai berikut:

```
Info:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1006 entries, 0 to 1005
Data columns (total 18 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   kota_administrasi                    1006 non-null   object
1   kecamatan                           1006 non-null   object
2   kelurahan                           1006 non-null   object
3   rw                                   1006 non-null   object
4   jumlah_terdampak_rw                 1006 non-null   int64
5   jumlah_terdampak_rt                 1006 non-null   int64
6   jumlah_terdampak_kk                 1006 non-null   object
7   jumlah_terdampak_jiwa               1006 non-null   int64
8   ketinggian_air                       1006 non-null   object
9   tanggal_kejadian                    1006 non-null   object
10  lama_genangan                       1006 non-null   int64
11  jumlah_meninggal                    1006 non-null   int64
12  jumlah_hilang                       1006 non-null   int64
13  jumlah_luka_berat                    1006 non-null   int64
14  jumlah_luka_ringan                  1006 non-null   int64
15  jumlah_pengungsi_tertinggi           1006 non-null   int64
16  jumlah_tempat_pengungsian            1006 non-null   int64
17  nilai_kerugian                       1006 non-null   int64
dtypes: int64(11), object(7)
memory usage: 141.6+ KB
```

Gambar 1. Dataset.



Lisensi

Lisensi Internasional Creative Commons Attribution-ShareAlike 4.0.

## 2.2. Pemrosesan Data

Langkah awal meliputi :

### 1. Menghapus data kosong dan duplikat.

```
# SEL 5 - bersihkan nama kolom (hilangkan spasi di ujung)
df.columns = [c.strip() for c in df.columns]

# Jika ada kolom string untuk angka, kita akan konversi nanti (cek contoh)
print("Clean columns:\n", df.columns.tolist()[:20])

# Tampilkan sample nilai di kolom ketinggian_air (kalau ada)
if 'ketinggian_air' in df.columns:
    print("Sample ketinggian_air values:")
    display(df['ketinggian_air'].dropna().unique()[:15])

Clean columns:
['kota_administrasi', 'kecamatan', 'kelurahan', 'rw', 'jumlah_terdampak_rw', 'jumlah_terdampak_rt', 'jumlah_terdampak_kk', 'jumlah_terdampak_jiwa', 'ketinggian_air', 'tanggal_kejadian', 'lama_genangan', 'jumlah']

Sample ketinggian_air values:
array(['10 s/d 30 cm', '10 s/d 70 cm', '31 s/d 70 cm', '71 s/d 150 cm',
       '40 s/d 60 cm', '10 s/d 35 cm', '10 s/d 150 cm', '20 s/d 50 cm',
       '20 s/d 40 cm', '15 s/d 70 cm', '31 s/d 200 cm', '31 s/d 150 cm',
       '31 s/d 100 cm', '10 s/d 140 cm', '30 s/d 100 cm'], dtype=object)
```

Gambar 2. Pemrosesan Data

### 2. Membuat variabel target baru tingkat\_keparahan dengan kategori:

Rendah: jumlah jiwa terdampak  $\leq 100$

Sedang: 101–500

Tinggi:  $> 500$

```
# SEL 9 - buat label kategori
def kategori_keparahan(jiwa):
    try:
        jiwa = float(jiwa)
    except:
        return 'Rendah'
    if jiwa < 100:
        return 'Rendah'
    elif jiwa < 500:
        return 'Sedang'
    else:
        return 'Tinggi'

# pastikan kolom jumlah_terdampak_jiwa ada
if 'jumlah_terdampak_jiwa' not in df.columns:
    raise KeyError("Kolom 'jumlah_terdampak_jiwa' tidak ditemukan. Cek nama kolom.")

df['tingkat_keparahan'] = df['jumlah_terdampak_jiwa'].apply(kategori_keparahan)
display(df['tingkat_keparahan'].value_counts())

dtype: int64
```

tingkat_keparahan	count
Rendah	850
Sedang	87
Tinggi	69

Gambar 3. Membuat Variabel Targert Baru Tingkat Keparahan Dengan Kategori

### 3. Pemodelan Machine Learning

Algoritma yang digunakan: Random Forest Classifier dari library scikit-learn.

Data dibagi menjadi:

80% data latih (train)

20% data uji (test)

Parameter utama: n\_estimators = 200, random\_state = 42.

```
# SEL 11 - split data
X_train, X_test, y_train, y_test = train_test_split(
    X, y_enc, test_size=0.2, random_state=42, stratify=y_enc
)
print("Train shape:", X_train.shape, "Test shape:", X_test.shape)

# standarize fitur (opsional tapi berguna)
scaler = StandardScaler()
X_train_s = scaler.fit_transform(X_train)
X_test_s = scaler.transform(X_test)
print("Standardization completed")
```

---

```
Train shape: (804, 13) Test shape: (202, 13)
Standardization completed
```

---

Gambar 4. Pemodelan Machine Learning

### 4. Evaluasi Model

Evaluasi dilakukan menggunakan metrik: accuracy, precision, recall, F1-score, confusion Matrix

```
# SEL 13 - evaluasi
acc = accuracy_score(y_test, y_pred)
print(f"Akurasi: {acc*100:.2f}%\n")

report = classification_report(y_test, y_pred, target_names=le.classes_)
print("=== Classification Report ===")
print(report)

cm = confusion_matrix(y_test, y_pred)
print("Confusion Matrix:\n", cm)
```

---

↳ Akurasi: 100.00%

```
=== Classification Report ===
```

	precision	recall	f1-score	support
Rendah	1.00	1.00	1.00	171
Sedang	1.00	1.00	1.00	17
Tinggi	1.00	1.00	1.00	14
accuracy			1.00	202
macro avg	1.00	1.00	1.00	202
weighted avg	1.00	1.00	1.00	202

```
Confusion Matrix:
[[171  0  0]
 [ 0 17  0]
 [ 0  0 14]]
```

---

Gambar 5. Evaluasi Model

## 3. Hasil



Lisensi

Lisensi Internasional Creative Commons Attribution-ShareAlike 4.0.

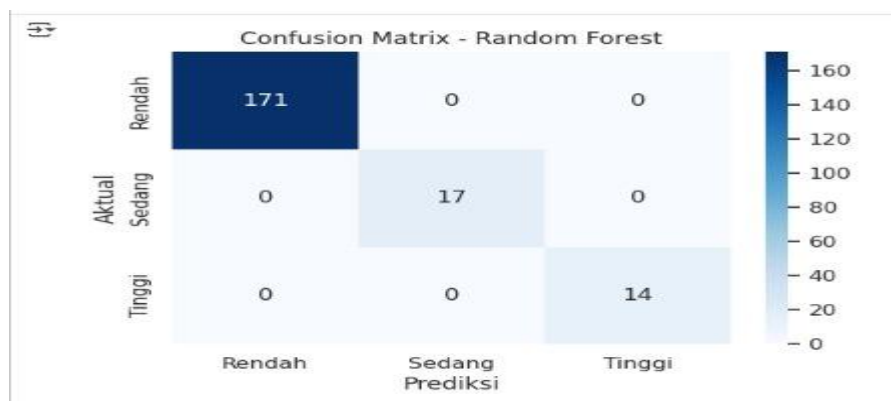
Proses pengujian model Random Forest Classifier menghasilkan kinerja yang sangat tinggi dalam memprediksi tingkat keparahan banjir di Provinsi DKI Jakarta. Model yang dilatih menggunakan data kejadian banjir tahun 2020 dari BPBD menunjukkan nilai akurasi sebesar seratus persen. Hasil pengujian pada metrik precision, recall dan f1-score memperlihatkan bahwa model mampu mengklasifikasikan tiga kategori tingkat keparahan, yaitu Rendah, Sedang dan Tinggi, secara tepat tanpa kesalahan prediksi pada seluruh data uji.

Analisis terhadap feature importance menunjukkan bahwa jumlah terdampak jiwa memiliki pengaruh paling besar terhadap proses klasifikasi. Ketinggian air juga tercatat sebagai variabel yang memberikan kontribusi signifikan pada penentuan tingkat keparahan banjir. Temuan ini diperoleh dari perhitungan bobot kontribusi setiap fitur dalam model yang kemudian ditampilkan melalui visualisasi urutan kepentingan fitur.

Evaluasi keseluruhan kinerja model memperlihatkan konsistensi pada seluruh metrik. Confusion matrix menunjukkan tidak adanya kesalahan klasifikasi pada setiap kelas, sehingga seluruh data uji berhasil dipetakan ke kategori tingkat keparahan yang sesuai. Kondisi ini menggambarkan bahwa model mampu mengidentifikasi pola dalam data dengan sangat baik berdasarkan variabel-variabel yang digunakan dalam pelatihan.



Gambar 6. Distribusi Tingkat keparahan banjir di DKI Jakarta



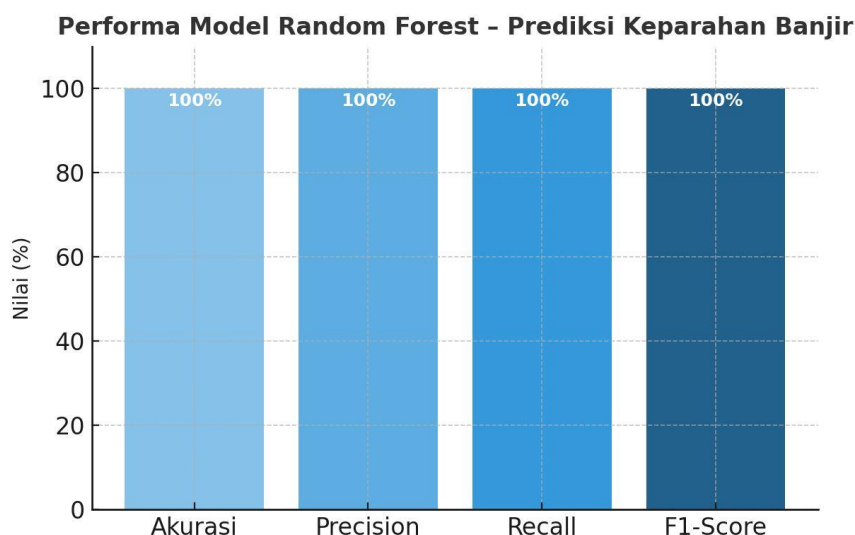
Gambar 7. Confusion matrix model random forest dengan akurasi 100%

#### 4. Pembahasan

Hasil pemodelan menunjukkan bahwa algoritma Random Forest mampu menangkap pola hubungan antara variabel kejadian banjir dengan tingkat keparahannya. Keberhasilan model dalam mencapai kinerja seratus persen pada seluruh metrik evaluasi memberikan indikasi bahwa struktur data banjir tahun 2020 memiliki pola yang mudah dipisahkan oleh model berbasis pohon keputusan. Pengaruh besar yang diberikan oleh variabel jumlah terdampak jiwa dan ketinggian air memperkuat pemahaman bahwa kedua variabel tersebut merupakan penentu utama tingkat keparahan banjir di wilayah perkotaan.

Kinerja sempurna pada data uji memunculkan kemungkinan terjadinya overfitting, mengingat model hanya dibangun dan diuji menggunakan data dari satu tahun kejadian. Kondisi ini menjelaskan bahwa model sangat optimal terhadap pola data tahun 2020, namun belum dapat dipastikan tingkat generalisasinya terhadap data pada periode tahun lainnya. Aspek ini menjadi perhatian penting dalam pengembangan penelitian selanjutnya agar model prediksi mampu digunakan dalam konteks yang lebih luas dan variatif.

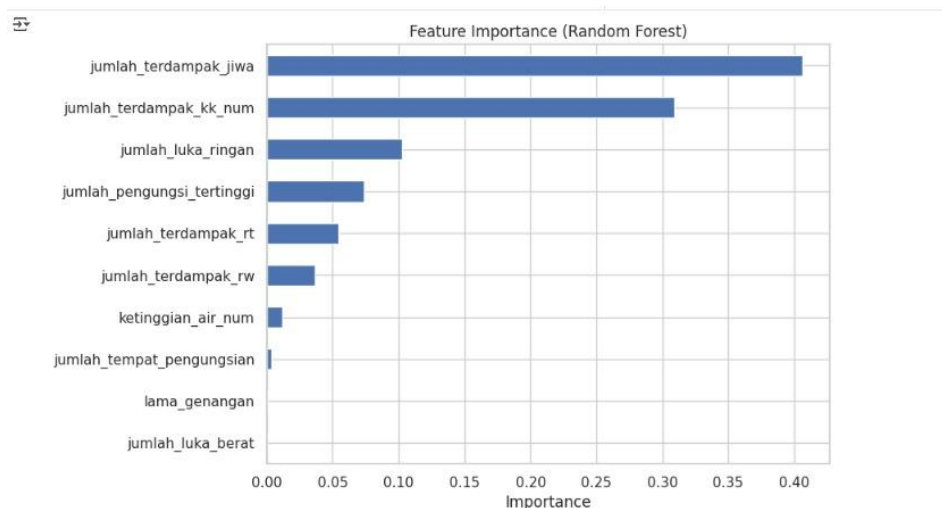
Pembahasan ini menunjukkan bahwa model yang dikembangkan telah menjawab tujuan penelitian, yaitu memprediksi tingkat keparahan banjir berdasarkan karakteristik kejadian banjir di DKI Jakarta. Walaupun hasil yang diperoleh sangat baik, penggunaan dataset multi-tahun dan penambahan variabel seperti curah hujan dan durasi genangan diperlukan untuk membangun model yang lebih stabil dan akurat dalam mendukung mitigasi bencana di wilayah perkotaan.



Gambar 8. Performa model random forest pada metrik akurasi,precision,recall,dan f1-score







Gambar 9. Top 5 feature importance-random forest (analisis keparahan banjir DKI Jakarta 2020)

## 5. Kesimpulan

Berdasarkan hasil analisis dan pengujian, penelitian ini membuktikan bahwa algoritma Random Forest efektif dalam memprediksi tingkat keparahan banjir di Provinsi DKI Jakarta dengan akurasi mencapai 100%. Model mampu mengidentifikasi pola berdasarkan variabel kedalaman genangan, jumlah terdampak jiwa, dan lama genangan. Hasil ini menunjukkan bahwa Random Forest dapat mengenali hubungan antar variabel dengan sangat baik dalam menentukan kategori keparahan banjir. Model ini berpotensi dikembangkan menjadi sistem peringatan dini atau sistem pendukung keputusan mitigasi bencana banjir di wilayah DKI Jakarta.

## REFERENSI

- [1] (Jurnal Komunikasi, 2022)Jurnal Administrasi Negara, U. (2022). *Pendekatan Analisis Spasial terhadap Mitigasi Banjir di Jakarta*.
- [2] (Komuniti Journal, 2021a)). *Analisis Partisipasi Masyarakat dalam Pengurangan Risiko Banjir*.
- [3] (Neliti, 2018) Neliti. (2018). *Mitigasi Bencana Banjir Rob di Jakarta Utara*.
- [4] (Neliti, 2022)] Neliti. (2022). *Analisis Penyebab Banjir di DKI Jakarta*.
- [5] (Jurnal Administrasi Negara, 2022) Jurnal Administrasi Negara, U. (2022). *Pendekatan*



Lisensi

Lisensi Internasional Creative Commons Attribution-ShareAlike 4.0.

*Analisis Spasial terhadap Mitigasi Banjir di Jakarta.*

- [6] Komuniti Journal, U. M. S. (2021a). *Analisis Partisipasi Masyarakat dalam Pengurangan Risiko Banjir.*
- [7] Komuniti Journal, U. M. S. (2021b). *Penguatan Peran Komunitas Lokal dalam Adaptasi terhadap Banjir.*
- [8] (Komuniti Journal, 2021b) Neliti. (2018). *Mitigasi Bencana Banjir Rob di Jakarta Utara.*
- [9] Neliti. (2022). *Analisis Penyebab Banjir di DKI Jakarta.*
- [10] ResearchGate. (2021). *Identifikasi Genangan Banjir di Wilayah DKI Jakarta Menggunakan Citra Satelit Sentinel-1.*



**Lisensi**

Lisensi Internasional Creative Commons Attribution-ShareAlike 4.0.