

## Prediksi Banjir Di Dki Jakarta Dengan Menggunakan Algoritma K-Means Dan Random Forest

Wasis Haryo Sasoko<sup>1</sup>, Eka Wahyu Pujiharto<sup>2</sup>, Ruby Haris<sup>3</sup>, Adela Yuza Kania<sup>4</sup>, Kusrini<sup>5</sup>, Kusnawi<sup>6</sup>  
<sup>1,2,3,4,5,6</sup>Magister Teknik Informatika, Universitas Amikom Yogyakarta

[yokowasis@students.amikom.ac.id](mailto:yokowasis@students.amikom.ac.id), [eka.wahyu.pujiharto@students.amikom.ac.id](mailto:eka.wahyu.pujiharto@students.amikom.ac.id), [ruby.haris@students.amikom.ac.id](mailto:ruby.haris@students.amikom.ac.id),  
[adelakania@students.amikom.ac.id](mailto:adelakania@students.amikom.ac.id), [kusrini@amikom.ac.id](mailto:kusrini@amikom.ac.id), [kusnawi@amikom.ac.id](mailto:kusnawi@amikom.ac.id)

### Abstract

*This research aims to develop a flood prediction method that can be used to implement effective prevention and mitigation measures in dealing with frequent natural disasters in DKI Jakarta. The approach used in this study involves the utilization of Machine Learning techniques with a combination of K-Means and Random Forest algorithms. Historical data on water gates, water levels, and other relevant factors are used as inputs for the development of the flood prediction model. The K-Means method is employed to cluster the water level data, and the results of the K-Means clustering process are then used as parameters in the Random Forest method. A total of 20 experiments were conducted, varying the value of k from 1 to 20 in the K-Means algorithm. The experimental results show that the best accuracy and f-1 score were achieved at k=14, with an accuracy rate of 95% and an f-1 score of 90%. This indicates that the developed flood prediction model is capable of providing accurate and reliable predictions in identifying flood potential. This research holds significant implications for flood management in vulnerable cities. With an effective flood prediction method, prevention and mitigation measures can be implemented more efficiently, thereby reducing the negative impacts caused by floods.*

*Keywords: Flood Prediction, K-Means, Random Forest*

### Abstrak

Penelitian ini bertujuan untuk mengembangkan metode prediksi banjir yang dapat digunakan untuk mengambil langkah-langkah pencegahan dan mitigasi yang efektif dalam menghadapi bencana alam yang sering terjadi di DKI Jakarta. Pendekatan yang digunakan dalam penelitian ini adalah menggunakan metode Machine Learning dengan kombinasi algoritma K-Means dan Random Forest. Data historis pintu air, tinggi air, dan faktor-faktor lain yang relevan digunakan sebagai input dalam pengembangan model prediksi banjir. Untuk mengelompokkan data tinggi air, digunakan metode K-Means. Hasil dari proses clustering K-Means tersebut kemudian dijadikan parameter dalam metode Random Forest. Sebanyak 20 percobaan dilakukan dengan variasi nilai k dari 1 hingga 20 dalam algoritma K-Means. Hasil eksperimen menunjukkan bahwa akurasi dan nilai f-1 yang paling baik ditemukan pada k=14, dengan tingkat akurasi mencapai 95% dan nilai f-1 mencapai 90%. Hal ini menunjukkan bahwa model prediksi banjir yang dikembangkan mampu memberikan prediksi yang akurat dan dapat diandalkan dalam mengidentifikasi potensi banjir. Penelitian ini memiliki implikasi penting dalam pengelolaan bencana banjir di kota-kota yang rentan terhadap bencana tersebut. Dengan adanya metode prediksi banjir yang efektif, langkah-langkah pencegahan dan mitigasi dapat diambil secara lebih efisien, sehingga dapat mengurangi dampak negatif yang ditimbulkan oleh banjir.

Kata kunci: Prediksi Banjir, K-Means, Random Forest.

### 1. Pendahuluan

Sebagai ibu kota negara Indonesia, DKI Jakarta memiliki populasi yang padat dan merupakan kawasan perkotaan yang sangat terurbanisasi. Kombinasi antara perubahan iklim global, pertumbuhan populasi, dan urbanisasi yang cepat telah meningkatkan risiko terjadinya banjir di wilayah ini. Oleh karena itu, diperlukan pendekatan yang efisien dan efektif untuk

mengidentifikasi faktor-faktor yang mempengaruhi banjir dan memprediksi potensi terjadinya banjir di masa depan [1].

Dalam beberapa tahun terakhir, penggunaan metode pemodelan seperti K-Means dan Random Forest telah menjadi topik penelitian yang menarik dalam bidang prediksi banjir [1]. Metode-metode ini mampu menggabungkan data historis dan variabel lingkungan

untuk membuat prediksi yang lebih akurat mengenai kemungkinan terjadinya banjir di suatu wilayah [1].

Pada penelitian terdahulu random forest (RF), k-nearest neighbor (KNN), multilayer perceptron (MLP), dan kombinasi algoritma genetika-fungsi radial basis gaussian-regresi vektor pendukung (GA-RBF-SVR) digunakan untuk memprediksi banjir pada wilayah pantai barat daya Bangladesh. Meskipun mencapai akurasi prediksi yang tinggi, perbedaan spasial dalam hasil model teramati, dengan koefisien korelasi piksel antara model yang berbeda berkisar antara 0,62 hingga 0,91. Model yang dioptimalkan menunjukkan akurasi prediksi yang tinggi dan meningkatkan kesepakatan spasial dengan mengurangi jumlah kesalahan klasifikasi [2].

Penelitian terhadap akurasi prediksi probabilitas banjir menggunakan tiga model pembelajaran mesin (MLM), termasuk Random Forest (RF), Artificial Neural Network (ANN), dan Generalized Linear Model (GLM) juga pernah dilakukan oleh [3] dimana Studi ini memilih 14 faktor penyebab banjir sebagai variabel independen, dan 220 lokasi banjir dipilih sebagai variabel dependen. Variabel dependen dibagi menjadi pelatihan (70%) dan validasi (30%) untuk pemodelan kecenderungan banjir. Kurva Karakteristik Operasi Penerima (ROC), indeks Kappa, akurasi, dan kriteria statistik lainnya digunakan untuk mengevaluasi akurasi model. Hasil penelitian menunjukkan bahwa memperbaiki DEM saja tidak signifikan mempengaruhi akurasi prediksi probabilitas banjir terlepas dari MLM yang digunakan dan independen dari model statistik yang digunakan untuk menilai akurasi kinerja.

Penelitian tentang banjir juga pernah dilakukan pada [4] Hasil penelitian menunjukkan bahwa indikator karakteristik curah hujan yang dikembangkan oleh model dapat secara efektif mengidentifikasi jenis-jenis banjir potensial, dan parameter model yang ditentukan berdasarkan tingkat banjir historis dapat disesuaikan dan digunakan untuk tugas ramalan baru berdasarkan kesamaan .

Penelitian tentang analisis kerentanan banjir juga pernah dilakukan dan dianggap sebagai pendekatan penting dalam pengelolaan banjir. Identifikasi daerah yang rentan terhadap banjir telah dilakukan dengan menerapkan algoritma pembelajaran mesin (ML) yang canggih (random forest (RF), support vector machine (SVM), dan extreme gradient boosting (XGBoost)) di bagian bawah DAS Sungai Raidak, Bengal Barat India. Peta kerentanan banjir telah dihasilkan berdasarkan 14 faktor kondisioning banjir yang berbeda. Model dievaluasi secara konvensional menggunakan kurva ROC (receiver operating characteristics). Nilai AUC dari ROC adalah di atas 0,80 untuk semua model dan XGBoost menunjukkan efikasi tertinggi (AUC = 0,92) [5].

Salah satu algoritma clustering adalah K-means, yang digunakan untuk mengelompokkan data ke dalam beberapa kelompok atau cluster berdasarkan kemiripan karakteristiknya [6]. Pada hasil dari Algoritma K-means ini menghasilkan data baru tinggi air yang sudah terklusterisasi dan diberikan kode tinggi air [7] untuk dapat diproses pada algoritma Random Forest dalam proses klasifikasi.

Random Forest adalah salah satu algoritma pembelajaran mesin yang populer dalam klasifikasi dan regresi [8], [9]. Algoritma ini memanfaatkan kombinasi dari beberapa pohon keputusan (decision tree) untuk menghasilkan prediksi yang lebih stabil dan akurat [10]. Dalam konteks prediksi banjir, Random Forest dapat memproses data hasil clustering dari Algoritma K-Means untuk menentukan prediksi banjir secara akurat berdasarkan klasifikasi tinggi air.

Dalam studi ini digunakan kombinasi dari Algoritma K-Means dan Random Forest untuk melakukan prediksi banjir di DKI Jakarta. Data historis pintu air, tinggi air, dan faktor-faktor lain yang relevan digunakan sebagai input untuk model prediksi. Melalui eksperimen dan analisis yang dilakukan, dihasilkan model prediksi yang akurat dan dapat membantu dalam upaya mitigasi banjir di DKI Jakarta.

Beberapa urgensi penelitian antara lain :

Mitigasi Dampak Banjir: Dengan adanya model prediksi banjir yang akurat, masyarakat dan pemerintah DKI Jakarta dapat mengambil langkah-langkah pencegahan yang lebih efektif. Informasi tentang kemungkinan terjadinya banjir di suatu wilayah dapat digunakan untuk menyusun rencana evakuasi penduduk, perlindungan aset, dan penanganan darurat, sehingga dapat mengurangi dampak negatif dari banjir [11].

Perencanaan Pembangunan Kota yang Tahan Banjir: Model prediksi banjir ini dapat digunakan sebagai panduan dalam perencanaan dan pengembangan kota yang lebih tahan banjir. Informasi tentang daerah-daerah yang berpotensi terkena banjir dapat digunakan untuk mengatur tata ruang, mengoptimalkan sistem drainase, dan mengimplementasikan infrastruktur yang sesuai guna mengurangi risiko banjir di masa depan [12].

Peningkatan Ketersediaan Data dan Informasi: Dalam konteks pengelolaan banjir, data dan informasi yang akurat merupakan aset yang berharga [9]. Penelitian ini dapat memberikan kontribusi dalam meningkatkan ketersediaan data tentang curah hujan, tinggi muka air sungai, dan faktor-faktor terkait lainnya [13].

Pengembangan Teknologi Prediksi Banjir yang Lebih Maju: Dengan menggunakan kombinasi K-Means dan Random Forest dalam prediksi banjir, penelitian ini dapat membuka jalan untuk pengembangan teknologi

prediksi banjir yang lebih maju di masa depan. Metode ini dapat menjadi dasar untuk pengembangan model prediksi yang lebih kompleks dan akurat, serta integrasi dengan teknologi lain seperti Internet of Things (IoT) dan sistem pemantauan real-time [14].

Penelitian Berkelanjutan tentang Mitigasi Banjir: Penelitian ini juga dapat menjadi bagian dari rangkaian penelitian berkelanjutan yang berkaitan dengan mitigasi banjir di DKI Jakarta [15]. Temuan dan hasil penelitian ini dapat digunakan sebagai dasar untuk penelitian berikutnya yang bertujuan untuk meningkatkan pemahaman tentang pola banjir, mengidentifikasi faktor-faktor yang mempengaruhi tingkat keparahan banjir, serta mengembangkan model prediksi yang lebih canggih.

## 2. Metode Penelitian

Metode penelitian menggunakan metode penelitian kuantitatif. Dalam penelitian ini, data historis pintu air, tinggi air, dan faktor-faktor lain yang relevan dikumpulkan secara numerik. Data tersebut kemudian dianalisis menggunakan kombinasi K-Means dan Random Forest.

### 2.1. Data Tinggi Air

Pada penelitian ini, dataset yang digunakan berasal dari Open Data Jakarta yang merupakan data public. Dataset yang diambil dari Open Data Jakarta menampilkan data tinggi muka air sungai, kali, dan laut Provinsi DKI Jakarta secara real-time dari Januari sampai dengan desember tahun 2020.

#### Variabel pada dataset ini:

nama pintu air = lokasi melakukan pengamatan,  
lokasi = tempat di mana anda melihat sesuatu (laut, nama sungai, atau kali,dll.)  
latitude = garis lintang,  
longitude = garis bujur,  
tanggal = tanggal dan jam data diambil,  
tinggi air = tinggi air pada saat data diambil dalam satuan CM,  
status siaga = status waspada terhadap ketinggian air.

#### Keterangan status siaga :

Siaga I : Bencana,  
Siaga II : Kritis,  
Siaga III : Waspada,  
Siaga IV/Normal : Aman.

### 2.2. Persiapan Data

Tahap persiapan data merupakan salah satu tahap penting dalam penelitian machine learning [16]. Pada tahap ini, data yang akan digunakan untuk melatih dan menguji model prediksi banjir di Provinsi DKI Jakarta perlu dipersiapkan secara cermat. Berikut adalah beberapa tahapan dalam persiapan data:

Pengumpulan Data: Tahap pertama adalah mengumpulkan data yang relevan dengan prediksi banjir. Data diambil dari web Open Data Jakarta.

Pembersihan Data: Pada tahap ini dilakukan pemilihan variabel penting yang akan digunakan untuk perhitungan prediksi banjir, sedangkan variabel yang tidak berpengaruh terhadap prediksi banjir dihapus.

Normalisasi Data: Pada proses ini dilakukan juga pengecekan nilai-nilai variabel yang abnormal, dan dilakukan *adjustment* pemberian nilai minimal normal [17].

Transformasi Data: Tahap ini melibatkan transformasi data ke format yang lebih sesuai untuk analisis machine learning. Data kategorikal banjir kita inisialisasi dengan numerik 0 atau 1.

Pembagian Dataset: Setelah data selesai dipersiapkan, dataset perlu dibagi menjadi set pelatihan (training set), set validasi (validation set), dan set pengujian (testing set). Set pelatihan digunakan untuk melatih model, set validasi digunakan untuk menyetel parameter dan memilih model terbaik, sedangkan set pengujian digunakan untuk menguji kinerja model secara objektif. Kita bagi data training 80% dan data testing 20%.

### 2.3. Metode Data Mining

#### 2.3.1. K-Means

Salah satu algoritma clustering adalah K-means, yang digunakan untuk mengelompokkan data ke dalam beberapa kelompok atau grup yang didasarkan pada kemiripan karakteristiknya [18]. Untuk menjelaskan proses K-means dalam konteks clusterisasi tinggi air pada pintu air, berikut adalah langkah-langkahnya:

Persiapan data: Kumpulkan data tinggi air pada pintu air yang akan di-cluster. Data ini harus mencakup nilai tinggi air pada setiap pintu air yang ingin dianalisis.

Menentukan jumlah cluster (K): Pilih jumlah cluster yang diinginkan. Jumlah cluster ini harus sesuai dengan jumlah kelompok yang ingin dibentuk berdasarkan karakteristik tinggi air pada pintu air.

Inisialisasi centroid: Centroid awal adalah titik-titik acak yang digunakan sebagai pusat setiap cluster. Pilih K titik awal secara acak dari data tinggi air sebagai centroid awal.

Perhitungan jarak: Hitung jarak antara setiap data tinggi air dengan centroid menggunakan metrik jarak seperti Euclidean, Manhattan, atau metrik jarak lainnya. Jarak ini mengukur seberapa mirip setiap data dengan centroid.

Penugasan data ke cluster: Setiap data tinggi air akan ditugaskan ke cluster yang memiliki centroid terdekat dengannya. Data yang memiliki jarak terkecil dengan centroid akan ditempatkan di cluster yang sesuai.

Perbarui centroid: Setelah data ditugaskan ke cluster, hitung kembali posisi centroid baru untuk setiap cluster.

Posisi centroid baru dihitung dengan mengambil rata-rata dari semua data yang termasuk dalam cluster tersebut.

Ulangi langkah 4-6: Ulangi langkah 4 hingga 6 secara iteratif sampai tidak ada perubahan dalam penugasan data ke cluster atau telah mencapai kriteria penghentian yang ditentukan sebelumnya. Kriteria penghentian dapat berupa jumlah iterasi maksimum atau konvergensi posisi centroid.

Hasil clusterisasi: Setelah iterasi selesai, Anda akan mendapatkan kelompok-kelompok atau cluster yang sesuai dengan karakteristik tinggi air pada pintu air. Setiap cluster akan memiliki centroid yang mewakili karakteristik rata-rata tinggi air dalam kelompok tersebut.

Evaluasi dan interpretasi: Evaluasi hasil clusterisasi untuk memahami pola dan karakteristik tinggi air pada pintu air di setiap cluster. Anda dapat menganalisis perbedaan antara cluster dan menarik kesimpulan mengenai faktor-faktor yang mempengaruhi tinggi air pada pintu air.

### 2.3.2. Random Forest

Algoritma Random Forest adalah metode pembelajaran mesin yang digunakan untuk melakukan prediksi atau klasifikasi dengan memanfaatkan sejumlah pohon keputusan (decision trees) [19]. Berikut adalah langkah-langkah dalam algoritma Random Forest:

Pengumpulan Data: mengumpulkan data tinggi air yang sudah dilakukan proses clusterisasi hasil perhitungan algoritma K-Means.

Pembuatan Set Data Latih: Data yang telah dikumpulkan kemudian perlu dibagi menjadi dua set, yaitu set data latih (training set) dan set data uji (test set). Set data latih akan digunakan untuk melatih model Random Forest, sedangkan set data uji akan digunakan untuk menguji kinerja model.

Pembentukan Pohon Keputusan: Setelah mendapatkan set data latih, langkah selanjutnya adalah membangun sejumlah pohon keputusan.

Menggabungkan Pohon Keputusan: Setelah sejumlah pohon keputusan terbentuk, langkah selanjutnya adalah menggabungkan hasil prediksi dari setiap pohon tersebut. Dalam kasus regresi, hasil prediksi dapat diambil dengan melakukan rata-rata dari prediksi semua pohon. Sedangkan dalam kasus klasifikasi, hasil prediksi dapat ditentukan dengan memilih kelas mayoritas dari semua pohon.

Prediksi: Setelah model Random Forest terbentuk, digunakan untuk memprediksi apakah terjadi banjir atau tidak pada situasi baru. Data yang digunakan untuk prediksi harus memiliki atribut yang sama dengan data yang digunakan untuk melatih model.

Evaluasi dan Penyetelan: Evaluasi kinerja model Random Forest dilakukan dengan membandingkan hasil prediksi dengan label asli pada set data uji. Jika diperlukan dapat dilakukan penyetelan pada model untuk meningkatkan kinerja.

## 3. Hasil dan Pembahasan

### 3.1. Pemrosesan Data

Penyederhanaan dilakukan pada dataset awal untuk memilih variabel yang paling relevan dan signifikan untuk memprediksi banjir dengan menggunakan python. Hasil penyederhanaan menyisakan 5 kolom dengan variabel penting, perhatikan Tabel.1.

### 3.2. Perhitungan K-Means

Proses perhitungan K-means adalah sebuah metode pengelompokan data yang sangat populer dalam analisis data dan machine learning. Tujuan utama dari K-means adalah untuk mengelompokkan data ke dalam beberapa kelompok yang disebut cluster, di mana setiap cluster memiliki kemiripan antara data yang ada di dalamnya. Dalam konteks ini, kita ingin mengklusterisasi tinggi air.

Langkah pertama dalam algoritma K-means adalah menentukan jumlah cluster yang diinginkan. Dalam kasus ini, kita akan mulai dengan  $k=3$  dan bertambah hingga  $k=20$ . Setelah jumlah cluster ditentukan, langkah selanjutnya adalah menginisialisasi posisi awal dari centroid, yaitu titik pusat setiap cluster. Posisi centroid awal dapat dipilih secara acak atau dengan menggunakan metode lainnya.

Setelah inisialisasi, langkah berikutnya adalah melakukan iterasi untuk mengupdate posisi centroid dan mengelompokkan data ke dalam cluster yang sesuai. Iterasi dimulai dengan menghitung jarak antara setiap data dengan setiap centroid dan mengassign data ke cluster dengan centroid terdekat. Kemudian, posisi centroid diupdate dengan menghitung rata-rata posisi data di setiap cluster.

Iterasi ini dilakukan secara berulang-ulang hingga tercapai kondisi berhenti, yang dapat berupa jumlah iterasi maksimum yang ditentukan sebelumnya atau saat tidak ada perubahan signifikan dalam posisi centroid. Saat berhenti, kita akan memiliki hasil klusterisasi dengan posisi centroid yang optimal.

Dalam kasus ini, proses diulang untuk setiap nilai  $k$  dari 3 hingga 20. Artinya, kita akan mengklusterisasi tinggi air menjadi 3 cluster, lalu 4 cluster, dan seterusnya hingga 20 cluster. Dengan melihat hasil klusterisasi untuk berbagai nilai  $k$ , kita dapat menganalisis perbedaan dan kesamaan dalam klusterisasi, serta memilih nilai  $k$  yang paling sesuai untuk tujuan kita. Contoh koding yang digunakan untuk machine learning random forest sebagai berikut:

Tabel .1 . Pengkodean Variabel

	kode_pi ntu	tinggi_a ir	kode_k ota	tanggal	status_s iaga
0	16	680	1	1/1/2020	Status : Normal
1	6	0	2	1/1/2020	Status : Normal
2	19	2050	3	1/1/2020	Status : Siaga 1
3	10	890	3	1/1/2020	Status : Normal
4	19	2230	3	1/1/2020	Status : Siaga 1
...	...	...	...	...	...
323185	17	1800	4	12/31/2020	Status : Normal
323186	18	1580	3	12/31/2020	Status : Normal
323187	19	1100	3	12/31/2020	Status : Normal
323188	21	1530	2	12/31/2020	Status : Normal
323189	1	1680	1	12/31/2020	Status : Normal

```
from sklearn.cluster import KMeans

kmeans = KMeans(n_clusters=kmeanscluster)
kmeans.fit(dfforest[['tinggi_air']])

dfforest['cluster'] = kmeans.labels_
```

Tabel.2  
 Hasil Eksperimen Random Forest

k	F1 Score	Akurasi
3	0.01	0.88
4	0.46	0.87
5	0.52	0.89
6	0.75	0.90
7	0.60	0.92
8	0.62	0.92
9	0.62	0.92
10	0.84	0.94
11	0.78	0.94
12	0.58	0.93
13	0.81	0.94
14	0.90	0.95
15	0.82	0.96
16	0.79	0.94
17	0.73	0.95
18	0.85	0.95
19	0.82	0.95
20	0.77	0.96

### 3.3. Perhitungan Random Forest

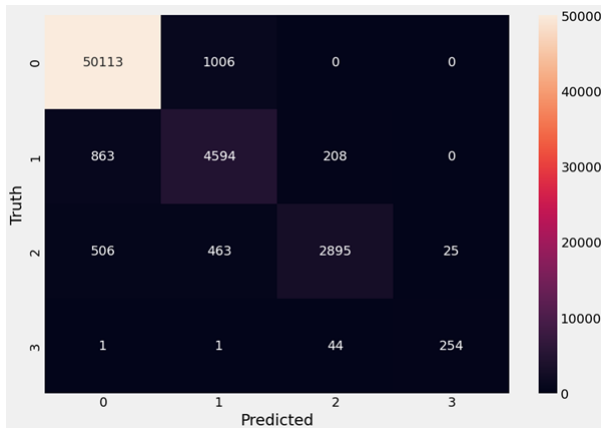
Hasil dari Klasterisasi K-Means 3 s/d 20, digunakan menjadi bahan untuk mengklasifikasikan tinggi air yang menjadi salah satu parameter pada algoritma random forest. Hasil dari random forest bisa dilihat pada dilihat pada tabel.2.

Hasil perhitungan prediksi banjir menggunakan Algoritma Random Forest pada aplikasi python menghasilkan nilai Confusion Matrix seperti pada Gambar. 1. Dataset baru dibagi menjadi 2 set data, yaitu data latih sebesar 80% dan data uji sebesar 20% kemudian di uji pada model python.

Kemudian dilakukan pengkodean variabel tinggi air berdasarkan clusterisasi pada K 14 menjadi dataset baru pada Tabel.3. Dataset baru hasil clusterisasi tersebut selanjutnya menjadi dataset untuk perhitungan prediksi banjir menggunakan Algoritma Random Forest.

Tabel.3  
 Dataset setelah klasterisasi tinggi air

	kode pintu	kode kota	status siaga	tinggi air
0	16	1	0	10
1	6	2	0	5
2	19	3	3	8
3	10	3	0	10
4	19	3	3	8
...	...	...	...	...
323185	17	4	0	4
323186	18	3	0	9
323187	19	3	0	0
323188	21	2	0	9
323189	1	1	0	4



Gambar. 1. Confusion Matrix Dataset K 14

Tabel. 4  
 Hasil Pengujian Algoritma Random Forest

	precision	recall	f1-score	support
0	0.99	0.97	0.98	51100
1	0.84	0.76	0.80	5671
2	0.70	0.99	0.82	3875
3	0.96	0.85	0.90	327
accuracy			0.95	60973
accuracy weighted	0.87	0.89	0.88	60973
avg	0.96	0.95	0.95	60973

Dari pengujian model Algoritma Random Forest, model memiliki nilai Recall dan F1 Score mendekati 1.00 dan nilai Akurasi yang baik sebesar 0.95 (Tabel. 4), maka model ini siap untuk diimplementasikan pada aplikasi prediksi banjir.

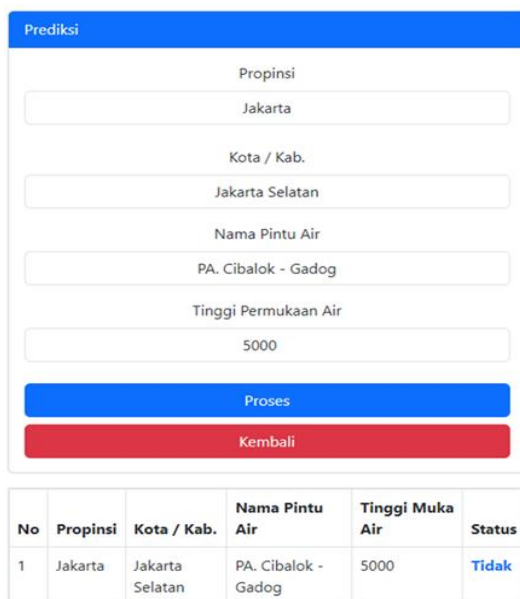
#### 4. Kesimpulan

Penerapan model K-Means dengan dataset yang telah dinormalisasi didapatkan nilai K yang paling optimal berada pada K=14 dimana nilai recall yang paling tinggi mendekati 1 yaitu 0,98 dan nilai akurasi 0,95. Dengan dataset baru yang dihasilkan model K-Means data dibagi menjadi data latih dan data uji dengan komposisi 80% untuk data latih dan 20% untuk data uji. Dengan menggunakan Model Random Forest model diuji dan dihasilkan nilai Recall dan F1 Score mendekati 1.00 dengan nilai akurasi yang baik sebesar 0,95. Dengan demikian model ini siap diimplementasikan pada aplikasi prediksi banjir.

Pada penelitian selanjutnya disarankan untuk menambah variabel lain yang mempengaruhi perhitungan prediksi banjir seperti curah hujan, kelembapan, arah angin dan lain lain.

#### 3.4. Implementasi

Model yang didapatkan diimplementasikan kedalam sebuah aplikasi prediksi banjir dengan menggunakan platform web. Tampilan hasil prediksi (Gambar. 2.), tampilan data histori prediksi (Gambar. 3.) sebagai berikut.



Gambar. 2. Tampilan Hasil Prediksi

21	Jakarta	Jakarta Pusat	P.A. Jembatan Merah	1000	Tidak
22	Jakarta	Jakarta Pusat	P.A. Jembatan Merah	200	Tidak
23	Jakarta	Jakarta Pusat	P.A. Jembatan Merah	10000	Tidak
24	Jakarta	Jakarta Selatan	PS. Pesanggrahan	400	Tidak
25	Jakarta	Jakarta Selatan	PS. Pesanggrahan	600	Tidak
26	Jakarta	Jakarta Selatan	PS. Pesanggrahan	6000	Ya
27	Jakarta	Jakarta Selatan	PS. Pesanggrahan	500	Tidak
28	Jakarta	Jakarta Selatan	PS. Pesanggrahan	5000	Ya
29	Jakarta	Jakarta Selatan	PS. Pesanggrahan	2000	Tidak
30	Jakarta	Jakarta Selatan	PS. Angke Hulu	2000	Tidak
31	Jakarta	Jakarta Selatan	PS. Angke Hulu	20000	Tidak
32	Jakarta	Jakarta Selatan	PS. Angke Hulu	6000	Ya
33	Jakarta	Jakarta Selatan	PS. Angke Hulu	3000	Tidak
34	Jakarta	Jakarta Selatan	PS. Angke Hulu	5000	Ya
35	Jakarta	Jakarta Selatan	PS. Angke Hulu	4000	Ya
36	Jakarta	Jakarta Selatan	PS. Angke Hulu	3000	Tidak

Gambar. 3. Tampilan Data Histori Prediksi

## Daftar Rujukan

- [1] O. Nixon Jimawan and D. Jayadi, "Analisis Statistika pengaruh curah hujan terhadap banjir di Jakarta melalui pemodelan Matematika," *Jurnal Meteorologi Klimatologi dan Geofisika*, vol. 5, no. 2, 2018.
- [2] M. S. G. Adnan *et al.*, "A novel framework for addressing uncertainties in machine learning-based geospatial approaches for flood prediction," *J Environ Manage*, vol. 326, Jan. 2023, doi: 10.1016/j.jenvman.2022.116813.
- [3] M. Avand, A. Kuriqi, M. Khazaei, and O. Ghorbanzadeh, "DEM resolution effects on machine learning performance for flood probability mapping," *Journal of Hydro-Environment Research*, vol. 40, pp. 1–16, Jan. 2022, doi: 10.1016/j.jher.2021.10.002.
- [4] Y. Tang *et al.*, "flood forecasting based on machine learning pattern recognition and dynamic migration of parameters," *J Hydrol Reg Stud*, vol. 47, Jun. 2023, doi: 10.1016/j.ejrh.2023.101406.
- [5] S. Ghosh, S. Saha, and B. Bera, "Flood susceptibility zonation using advanced ensemble machine learning models within Himalayan foreland basin," *Natural Hazards Research*, vol. 2, no. 4, pp. 363–374, Dec. 2022, doi: 10.1016/j.nhres.2022.06.003.
- [6] A. Nur Khomarudin, *Teknik Data Mining : Algoritma K-Means Clustering*. 2003. [Online]. Available: <https://agusnkhom.wordpress.com>
- [7] H. Xu, C. Ma, J. Lian, K. Xu, and E. Chaima, "Urban flooding risk assessment based on an integrated k-means cluster algorithm and improved entropy weight method in the region of Haikou, China," *J Hydrol (Amst)*, vol. 563, pp. 975–986, Aug. 2018, doi: 10.1016/J.JHYDROL.2018.06.060.
- [8] A. Primajaya and B. N. Sari, "Random Forest Algorithm for Prediction of Precipitation," *Indonesian Journal of Artificial Intelligence and Data Mining (IJAIDM)*, vol. 1, no. 1, pp. 27–31, 2018.
- [9] D. Husen, D. Sandi, and S. Bumbungan, "Analisis prediksi kebakaran hutan dengan menggunakan algoritma Random Forest Classifier," vol. 16, no. 1, 2022, [Online]. Available: <https://journal.uniku.ac.id/index.php/ilkom>
- [10] "fitctree, Fit binary decision tree for multiclass classification," *MathWorks*.
- [11] A. Taryana, M. Rifa, E. Mahmudi, and H. Becti, "Analisis kesiapsiagaan bencana banjir di Jakarta," *Jurnal Administrasi Negara*, Februari, vol. 13, pp. 302–311, 2022.
- [12] P. T. Juwono and A. Subagiyo, *Ruang Air dan Tata Ruang: Pendekatan teknis keairan dan pembangunan berkelanjutan dalam penanganan banjir perkotaan*. Universitas Brawijaya Press, 2017.
- [13] H. Rakuasa, Zainudin, W. Abdul Wahab, Kamiludin, A. Jaelani, and M. Rinaldi, "Dinas Sumber Daya Air DKI Jakarta," *Pengabdian kepada Masyarakat*, vol. 3, no. Maret, pp. 288–295, 2023, doi: 10.25008/altifani.v3i2.379.
- [14] J. M. Tien, "Internet of things, real-time decision making, and artificial intelligence," *Annals of Data Science*, vol. 4, pp. 149–178, 2017.
- [15] P. J. Ward, M. A. Marfai, F. Yulianto, D. R. Hizbaron, and J. Aerts, "Coastal inundation and damage exposure estimation: a case study for Jakarta," *Natural Hazards*, vol. 56, pp. 899–916, 2011.
- [16] F. N. Dhewayani *et al.*, "Implementasi K-Means Clustering untuk pengelompokan daerah rawan bencana kebakaran menggunakan model CRISP-DM," *Jurnal Teknologi dan Informasi*, doi: 10.34010/jati.v12i1.
- [17] A. Harmain, Paiman, H. Kurniawan, Kusrini, and D. Maulina, "Data normalization for K-Means efficiency on groups of areas with potential fores and /land fire based on heat spots distributions," *TEKNIMEDIA*, vol. 2, no. 2, pp. 83–89, 2021.
- [18] "Clustering Algoritma (K-Means) – School of Information Systems." <https://sis.binus.ac.id/2022/01/31/clustering-algoritma-k-means/> (accessed Jul. 02, 2023).
- [19] "Studi Kasus Random Forest Machine Learning untuk Pemula Data." <https://dqlab.id/studi-kasus-random-forest-machine-learning-untuk-pemula-data> (accessed Jul. 02, 2023).