PERANCANGAN MACHINE LEARNING PADA SISTEM MANAJEMEN BANJIR BERBASIS SENSOR IOT

1st Muhammad Raihan Al Hasani
Fakultas Teknik Elektro
Universitas Telkom
Bandung, Indonesia
mraihanalhasani@student.telkomuniversity.ac,id

2nd Ahmad Tri Hanuranto
Fakultas Teknik Elektro
Universitas Telkom
Bandung, Indonesia
athanuranto@telkomuniversity.ac.id

3rd Fardan
Fakultas Teknik Elektro
Universitas Telkom
Bandung, Indonesia
fardanfnn@telkomuniversity.ac.id

***Abstrak* — Penelitian ini bertujuan untuk mengembangkan dan mengimplementasikan sistem *machine learning* menggunakan algoritma *Random Forest* dalam manajemen bencana banjir di sungai Citarum yang meliputi wilayah Dayeuhkolot dan Bojongsoang, Kabupaten Bandung. Sistem ini memanfaatkan sensor IoT berupa sensor Ultrasonik dan sensor Waterflow yang terhubung ke mikrokontroler ESP32 WROOM *Wi-Fi* untuk mengukur tinggi air dan kuat arus air sebagai data *input*. Proses machine learning dengan algoritma *Random Forest* digunakan untuk mengklasifikasikan tinggi air dan kuat arus air ke dalam tiga kelas status banjir: aman, siaga, dan waspada. Hasil klasifikasi disampaikan melalui sistem informasi berupa Website kepada pengguna dan masyarakat sekitar, sehingga pengguna dapat mengambil tindakan yang tepat untuk mengurangi risiko banjir. Implementasi *machine learning* dalam sistem manajemen bencana banjir ini diharapkan dapat meningkatkan efektivitas dan responsivitas dalam menghadapi potensi banjir di wilayah terkait.**

Kata kunci* — Banjir, *Machine Learning, Internet of Things

I. PENDAHULUAN

Bencana banjir merupakan permasalahan serius di Indonesia, terutama dalam musim curah hujan tinggi. Bandung, sebagai salah satu kota dengan curah hujan yang tinggi, sering mengalami bencana banjir, khususnya di wilayah Dayeuhkolot, Baleendah, dan Bojongsoang di Kabupaten Bandung. Bencana banjir ini dipicu oleh faktor curah hujan yang tinggi pada hulu sungai dan kurangnya kesadaran masyarakat dalam menjaga lingkungan sekitar sungai Citarum dan Cikapundung.

Dalam upaya mengatasi masalah banjir, implementasi teknologi *machine learning* menjadi salah satu solusi yang menarik. Penelitian ini bertujuan untuk mengembangkan sistem manajemen bencana banjir berbasis machine learning di wilayah tersebut. Sistem ini akan menggunakan sensor IoT, seperti sensor Ultrasonik dan sensor *Waterflow* yang terhubung ke mikrokontroler ESP32 WROOM *Wi-Fi*, untuk mengukur tinggi air dan kuat arus air sebagai data *input*. Proses *machine learning* dengan algoritma *Random Forest* akan digunakan untuk mengklasifikasikan tinggi air dan kuat arus air ke dalam tiga kelas status banjir: aman, siaga, dan waspada.

Hasil klasifikasi tersebut akan disampaikan melalui sistem informasi berupa *Website* kepada pengguna, sehingga mereka dapat lebih responsif dalam menghadapi potensi banjir. Diharapkan implementasi *machine learning* dalam sistem manajemen bencana banjir ini akan meningkatkan efektivitas dan ketepatan dalam mengatasi bencana banjir, serta membantu dalam pengambilan keputusan yang lebih tepat untuk mengurangi risiko dan kerugian yang ditimbulkan oleh banjir di wilayah terkait.

II. KAJIAN TEORI

A. Internet of Things

Internet of Things (IoT) merujuk pada jaringan yang menghubungkan perangkat fisik, kendaraan, peralatan rumah tangga, dan objek lainnya yang dilengkapi dengan komponen elektronik, perangkat lunak, sensor, aktuator, dan kemampuan konektivitas. Melalui ini, pertukaran data dapat terjadi secara mutual antar elemen-elemen ini. IoT membuka peluang besar untuk mengintegrasikan dunia fisik secara langsung ke dalam sistem berbasis komputer, menghasilkan peningkatan efisiensi, manfaat ekonomi, serta pengurangan keterlibatan manusia.

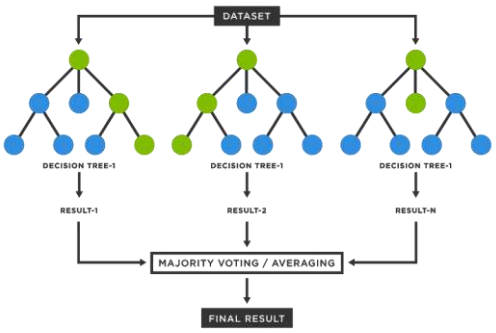
B. Machine Learning

Pembelajaran mesin merupakan penerapan teknologi komputer dan algoritma matematika yang mengadopsi metode pembelajaran dari data untuk menghasilkan prediksi di masa depan, sebagaimana yang dijelaskan oleh Goldberg & Holland (1988). Proses pembelajaran ini melibatkan dua tahapan, yaitu latihan (*training*) dan pengujian (*testing*), yang merupakan upaya dalam memperoleh kecerdasan, sesuai dengan pandangan Huang, Zhu, & Siew (2006).

Dalam ranah *machine learning*, isu terkait adalah bagaimana membangun program komputer agar secara otomatis meningkat berdasarkan pengalaman, sebagaimana dipertanyakan oleh Mitchell (1997).

C. Random Forest

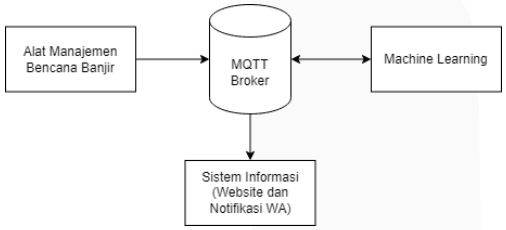
Random Forest adalah konsep di mana terbentuk suatu *ansambel* yang terdiri dari beberapa pohon keputusan secara acak, sesuai dengan definisi yang diberikan oleh Breiman (2001). Model klasifikasi *Random Forest* memiliki struktur umum seperti yang diilustrasikan dalam Gambar 1.



GAMBAR 1.
Model klasifikasi Random Forest
Sumber: TIBCO Software

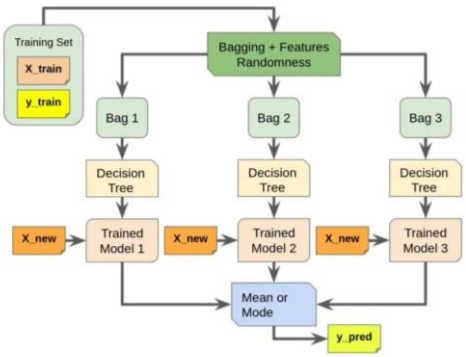
III. METODE

Dapat dilihat pada gambar 2 merupakan sistem desain *Classification data on server* yang dibangun menggunakan *machine learning*. Pada sistem ini dimulai dengan pengiriman data ketinggian air dan kuat arus kedalam MQTT Broker, selanjutnya *machine learning* akan mengambil data dari MQTT secara *real-time* dan melakukan klasifikasi untuk menentukan status bencana banjir, daerah yang akan terdampak dan kerugian yang dihasilkan dan dilanjutkan untuk ditampilkan pada sistem informasi berupa *website* dan pesan notifikasi WA.



GAMBAR 2.
Sistem Desain Machine Learning

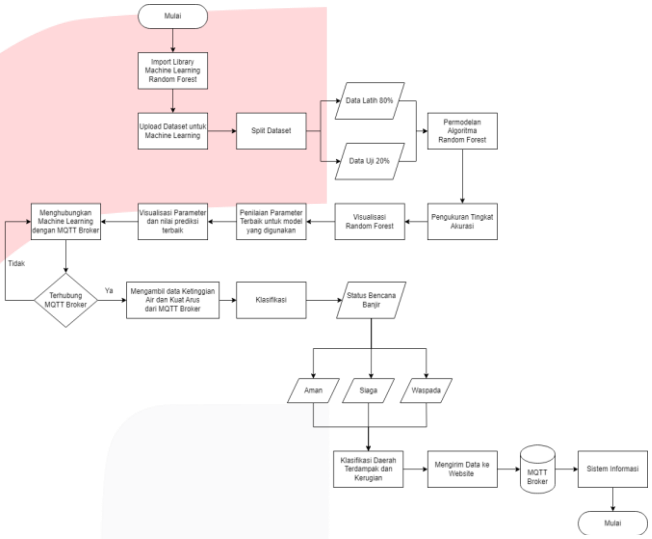
Machine learning yang dirancangan pada sistem manajemen bencana banjir menggunakan algoritma *Random Forest* untuk menentukan status bencana, daerah terdampak dan kerugian yang terjadi.



GAMBAR 3.
Proses Algoritma Random Forest

Pada gambar 3 menjelaskan mengenai algoritma RF yang digunakan untuk *machine learning*. Algoritma RF yang digunakan pada sistem manajemen banjir dibangun menggunakan bahasa pemrograman python. Perancangan *machine learning* dimulai dari pembuatan *library* yang akan digunakan, memasukan input dataset yang terdiri *variabel* yang digunakan untuk memprediksi hasil keluaran machine learning, pemisahan dan penyusunan *dataset*, pelatihan 80% data dan pengujian 20% data dari *dataset*. Setelah proses pelatihan dan pengujian data, algoritma RF pada machine learning melakukan prediksi yang didapatkan dari data set dalam bentuk format y_pred yang dapat dihitung tingkat akurasi dari algoritma RF, visualisasi dari setiap bag atau *Decision Tree* dan *variabel* yang paling penting untuk menentukan hasil klasifikasi *machine learning*.

A. Cara kerja Machine Learning



GAMBAR 4.
Alur Kerja Machine Learning

Pada gambar 4 merupakan *flowchart* dari sub-sistem ketiga, algoritma yang digunakan pada machine learning adalah *Random Forest* dengan menggunakan bahasa pemrograman python pada platform Google Collab. Langkah pertama yang dilakukan oleh *machine learning* adalah *import library* yang akan digunakan oleh *machine learning*, lalu *upload dataset* yang digunakan untuk *machine learning*, setelah itu *dataset* akan di *split* untuk menentukan data yang akan dilatih dan data yang akan di uji, pelatihan 80% dan pengujian *dataset* sebesar 20%. Setelah didapatkan data yang dilatih dan diuji, data akan proses menggunakan algoritma RF untuk menentukan nilai y_pred atau prediksi berdasarkan tingkat akurasinya. Selain tingkat akurasi *machine learning* dapat menampilkan visualisasi dari algoritma RF yang digunakan dan menentukan nilai parameter yang paling berpengaruh terhadap output keluaran machine learning.

Setelah melakukan pemodelan pada *machine learning*, *machine learning* dapat melakukan pengambilan data dari MQTT *Broker* berupa data dari sensor ultrasonik dan sensor *waterflow*. Data tersebut akan dinormalisasi menjadi tinggi air dan kuat arus air. Setelah proses normalisasi, dilanjutkan pada klasifikasi status bencana banjir dan daerah yang akan terdampak berdasarkan data yang telah diinput pada dataset di permodelan *machine learning*. Data yang dihasilkan akan ditampilkan pada website secara real-time.

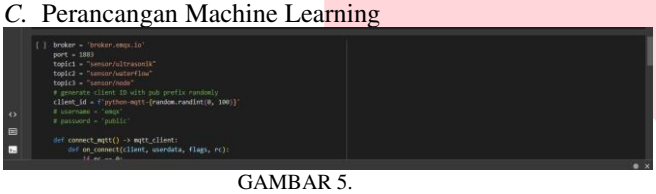
B. Software yang digunakan

Software penyusun yang digunakan dalam merancang Sistem Manajemen Banjir beserta fungsinya masing-masing dapat dilihat pada tabel berikut:

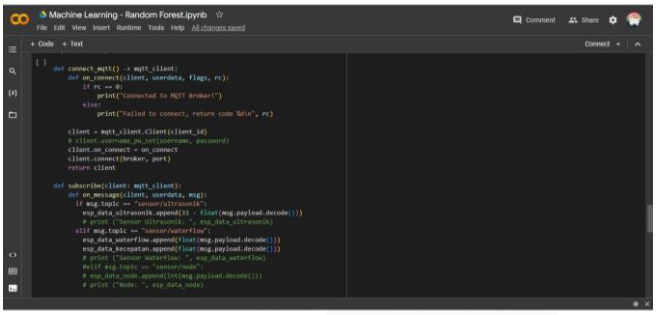
TABEL 1.
Software yang digunakan

No	Software	Fungsi
1.	Google Colab	*Platform* berbasis *cloud* yang disediakan oleh Google, yang memungkinkan untuk menjalankan dan mengembangkan kode Python di lingkungan *Jupyter Notebook*
2.	Pemrograman Python	Sebagai bahasa pemrograman untuk membangun *machine learning*

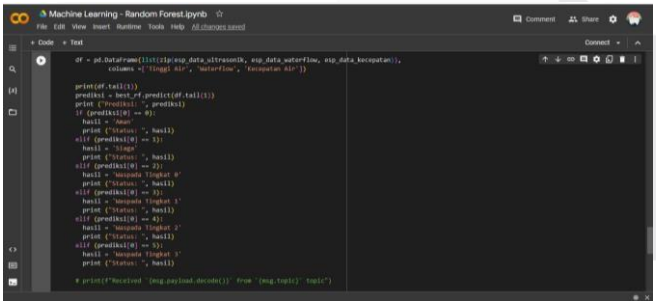
C. Perancangan Machine Learning



GAMBAR 5.
Perancangan Machine Learning



GAMBAR 6.
Perancangan Machine Learning

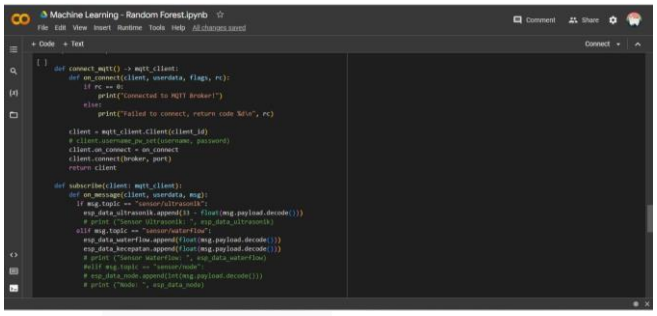


GAMBAR 7.
Perancangan Machine Learning

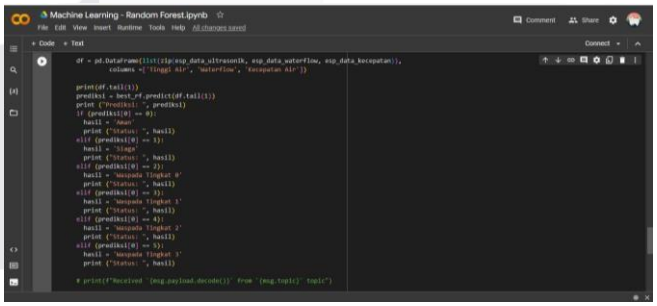
Gambar 8.
Perancangan Machine Learning

Pada gambar 5 hingga gambar 8 merupakan implementasi sub-sistem *machine learning* dimulai dari perancangan *machine learning* menggunakan bahasa pemrograman python pada *platform* Google Colab. Pada perancangan ini machine learning melakukan *import library* yang akan digunakan pada machine learning, selanjutnya melakukan *upload* file dalam bentuk CSV (*Comma Separated Value*) sebagai *dataset* dan menentukan *variabel* yang akan digunakan untuk klasifikasi pada machine learning. Selanjutnya *dataset* dipecah menjadi data uji sebesar 20% dan data latih sebesar 80% dari *dataset*. Setelah didapatkan data latih dan data uji, permodelan *machine learning* dari data latih dilakukan menggunakan algoritma *Random Forest*. Dari permodelan algoritma *random forest* dapat diketahui tingkat akurasi dari *machine learning*, selain tingkat akurasi *machine learning* menampilkan visualisasi dari *Decision Tree* yang digunakan pada *Random Forest*. *Machine learning* juga menampilkan nilai dan grafik dari parameter terbaik dari permodelan *machine learning*.

GAMBAR 9.
Proses Klasifikasi Machine Learning



GAMBAR 10.
Proses Klasifikasi Machine Learning



GAMBAR 11.
Proses Klasifikasi Machine Learning

GAMBAR 12.
Proses Klasifikasi Machine Learning

Pada gambar 9 hingga 12, merupakan implementasi pada proses klasifikasi yang dilakukan *machine learning*. Proses klasifikasi dimulai dari pengambilan data ketinggian air dan kuat arus air dari MQTT *Broker*. Data sensor ultrasonik dan sensor *waterflow* yang diambil dari MQTT *Broker* dinormalisasi menjadi Ketinggian Air dan Debit Air. Hasil kualifikasi berupa status bencana banjir, daerah terdampak dan prediksi kerugian yang akan terjadi.

IV. HASIL DAN PEMBAHASAN

Pada Pengujian *machine learning* kami melakukan pengujian keakuratan kinerja dari algoritma RF agar dapat menampilkan visualisasi dari algoritma RF, menampilkan nilai dari *accuracy*, *precision* dan *recall* dan juga memvisualisasikan nilai tersebut dalam bentuk label, machine learning dapat menentukan nilai parameter terbaik yang digunakan pada algoritma *Random Forest*.

A. Pengujian Random Forest

Pada pengujian *Random Forest* kami melakukan proses pencarian nilai variabel yang paling optimal untuk data pelatihan yang terdiri dari nilai Tinggi Air, nilai kuat arus yaitu *Waterflow* dan nilai Kecepatan air. Untuk skema pengujiannya digunakan model *random forest* dengan cara membuat variabel paling baik untuk model RF pada gambar 13.

```
# Create a variable for the best model
best_rf = rand_search.best_estimator_

# Print the best hyperparameters
print('Best hyperparameters:',  rand_search.best_params_)
```

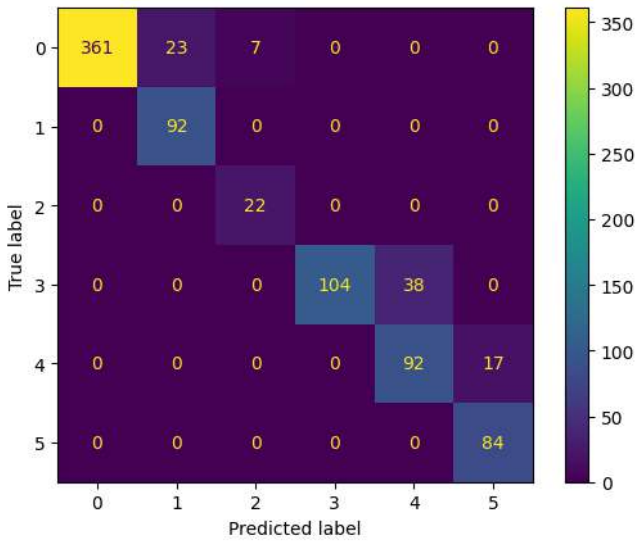
```
Best hyperparameters: {'max_depth': 8, 'n_estimators': 359}
```

GAMBAR 13.
Proses Pengujian Best Hyperparametes

```
# Generate predictions with the best model
y_pred = best_rf.predict(X_test)

# Create the confusion matrix
cm = confusion_matrix(y_test, y_pred)

ConfusionMatrixDisplay(confusion_matrix=cm).plot();
```

GAMBAR 14.
Proses Ganerate Prediksi

Pada gambar 14 merupakan proses *generate* nilai prediksi terhadap model terbaik yang digunakan model *random forest*. Nilai yang didapatkan ditampilkan pada gambar 15 dalam bentuk *confusion matrix* berdasarkan data uji (y_test) dan nilai prediksi (y_pred). Secara keseluruhan proses pada gambar 5.18 digunakan untuk mengambil model terbaik dan hyperparameter nya setelah melakukan pencarian acak.



GAMBAR 15.
Confusion Matrix

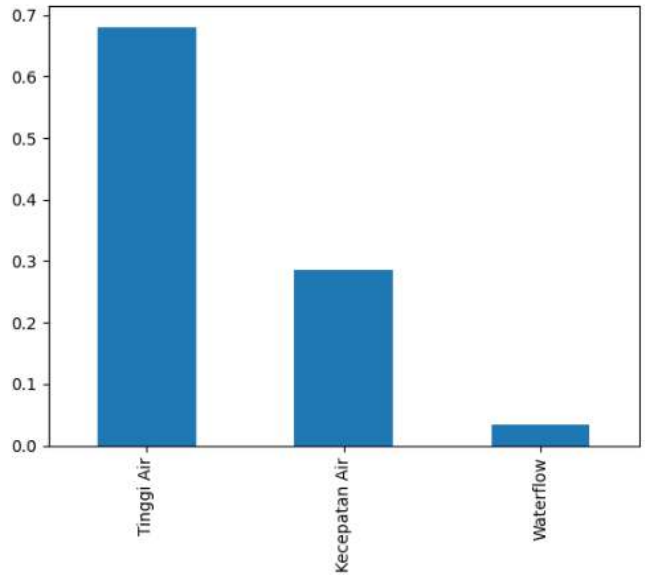
Plot yang dihasilkan pada *Confusion Matrix* diatas memberikan representasi visual dari kinerja model dalam hal contoh yang diklasifikasikan dengan benar dan salah.

```
y_pred = best_rf.predict(X_test)

accuracy = accuracy_score(y_test, y_pred)
precision = precision_score(y_test, y_pred, average='macro')
recall = recall_score(y_test, y_pred, average='macro')

print("Accuracy:", accuracy)
print("Precision:", precision)
print("Recall:", recall)
```

```
Accuracy: 0.8964285714285715
Precision: 0.8642555222760857
Recall: 0.9077825670498084
```

GAMBAR 16.
Proses Pengukuran Nilai

Metode yang digunakan pada pengukuran nilai prediksi (y_pred) adalah mengklasifikasi KNN (*K-Nearest Neighbor)* digunakan untuk menghasilkan label prediksi (y_pred) untuk set pengujian (X_test). Selanjutnya, tiga metrik kinerja dihitung menggunakan label sebenarnya (y_test) dan label prediksi (y_pred).



GAMBAR 17.
Tingkat Kepentingan Variabel

Pada gambar 17 merupakan hasil visualisasi dari kepentingan fitur dari model best_rf (yang dianggap sebagai model hutan acak terlatih) dan nama fitur dari data pelatihan (X_train.columns). Metode yang digunakan sort_values() untuk mengurutkan kepentingan fitur dalam urutan menurun. Berdasarkan hasil visualisasi kepentingan variabel dalam *Random Forest,* dapat disimpulkan bawah Tinggi Air merupakan variabel paling berpengaruh dalam melakukan proses klasifikasi untuk menentukan status bencana, daerah terdampak dan kerugian banjir, yang diikuti oleh kecepatan air dan *waterflow* atau debit air.

B. Tabel

TABEL 2.
Parameter Terbaik

Parameter	Nilai
Max_depth	8
N-estimator	359

Pada Tabel 5.9 merupakan *hyperparameters* untuk model *random forest* yang digunakan pada machine learning sistem manajemen banjir. Max_depth merupakan jumlah pemisahan yang diperbolehkan untuk dibuat oleh setiap pohon keputusan dan N_estimator merupakan jumlah pohon keputusan di *Random Forest*.

TABEL 3.
Nilai Accuracy, Precision, dan Recall

Pengukuran	Hasil
Accuracy	0.8964285714285715
Precision	0.8642552522760857
Recall	0.9077825670498084

V. KESIMPULAN

Pada kesimpulan pengujian *machine learning*, model algoritma yang digunakan pada *machine learning* yaitu algoritma *Random Forest*. Pada hasil pengujian *Random Forest*, ditunjukan bahwa variabel terpenting yang digunakan pada model *machine learning* adalah Ketinggian air yang ikuti dengan kecepatan air dan debit air. Dalam proses permodelan *machine learning*, *hyperparameters* untuk model *random forest*. Max_depth merupakan jumlah pemisahan yang diperbolehkan untuk dibuat oleh setiap pohon keputusan sebanyak 8 dan N_estimator merupakan jumlah pohon keputusan sebanyak 359. *Confusion Matrix* digunakan untuk mengambil model terbaik dan *hyperparameter* nya setelah melakukan pencarian acak. Pada penilaian *accuracy*, *precision*, dan *recall* model random forest memiliki akurasi sebesar 0.89, nilai presisi sebesar 0.86 dan *recall* sebesar 0.90. Hal ini disebabkan karena data pelatihan dari cluster 4 yang memiliki nilai status bencana berbeda dengan data cluster 1, 2 dan 3

REFERENSI

[1] Anbarasan, M., Muthu, B. A., Sivaparthipan, C. B., Sundarasekar, R., Kadry, S., Krishnamoorthy, S., Samuel, D. J. R., & Dasel, A. A. (2020). Detection of flood disaster system based on IoT, big data and convolutional deep neural network. Computer Communications, 150, 150–157. doi: 10.1016/j.comcom.2020.02.029

[2] Goldberg, D. E., & Holland, J. H. (1988). Machine learning. Journal of the ACM, 55(1), 279-292.

[3] Huang, G. B., Zhu, Q. Y., & Siew, C. K. (2006). Extreme learning machine: theory and applications. Neurocomputing, 70(1-3), 489-501.

[4] Mitchell, T. M. (1997). Machine Learning. McGraw-Hill.

[5] Zahir, S. B., Ehkan, P., Sabapathy, T., Jusoh, M., Osman, M. N., Yasin, M. N., Abdul Wahab, Y., Hambali, N. A. M., Ali, N., Bakhit, A. S., Husin, F., Md.Kamil, M. K., & Jamaludin, R. (2019). Smart IoT Flood Monitoring System. Journal of Physics: Conference Series, 1339, 012043. doi:10.1088/1742-6596/1339/1/012043.

[6] Ma, J., Tan, X., & Zhang, N. (2010). FLOOD MANAGEMENT AND FLOOD WARNING SYSTEM IN CHINA. Irrigation and Drainage, 59, 17–22. doi: 10.1002/ird.513

[7] Ufuoma, G., Sasanya, B. F., Abaje, P., & Awodutire, P. (2021). EFFICIENCY OF CAMERA SENSORS FOR FLOOD MONITORING AND WARNINGS. Scientific African, DOI: https://doi.org/10.1016/j.sciaf.2021.e00887