IMPLEMENTATION OF MACHINE LEARNING ALGORITHM FOR FLOOD FORECASTING

Final (Draft) Report

Ilukkumbure S. P. M. Kaveesha W.

IT18022902

B.Sc. (Hons) Degree in Information Technology

Specialization in Information Technology

Department of Information Technology

Sri Lanka Institute of Information and Technology

Sri Lanka

September 2021

IMPLEMENTATION OF MACHINE LEARNING ALGORITHM FOR FLOOD FORECASTING

Final (Draft) Report

Ilukkumbure S. P. M. Kaveesha W.

IT18022902

B.Sc. (Hons) Degree in Information Technology

Specialization in Information Technology

Department of Information Technology

Sri Lanka Institute of Information and Technology

Sri Lanka

September 2021

Declaration

I declare that this is my own work, and this final report does not incorporate without acknowledgement any material previously submitted for a degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text. Also, I hereby grant to Sri Lanka Institute of Information Technology, the non-exclusive right to reproduce and distribute my dissertation, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature: - Date: 13 /10/ 2021

The above candidate has carried out research for the bachelor's degree Dissertation under my supervision.

Signature of Supervisor: ________________ Date: 11/5/2021

Acknowledgements

These achievements are of my fourth-year research for the module final year research project, as reported in this document. I'd want to express gratitude and credit to our supervisor, Mr. Samantha Rajapakse. This project is the result of the hard effort of all four members of the research team, as well as the encouragement, support, and advice of many others. I would want to offer our appreciation and gratitude to everyone who helped me to make my research a success.

Abstract

In Sri Lanka, flooding has been a highly dangerous problem. The ability to establish a mechanism for forecasting catastrophic weather conditions will be a huge help to those who have been impacted by disasters. In this study looks into the usage of Machine Learning (ML), Deep Learning (DL), IoT (internet of things), and crowdsourcing to help develop a pre- and post-flood management system as a way to control flood risk and mitigate potential flood risks. This flooding situation occurs for numerous numbers of reasons, these reasons can be categorized in to main two categories as human and natural causes. In order to address the above factors, we purpose development of early warning structure to mitigate the catastrophic impact it could have been on the public. The main research was performed on real-time weather monitoring and predictions, data mining-based flood forecasting with the aid of weather situations, crowdsourcing information collection techniques, and third-party API utilization of current weather situations. The solution monitors flooding and also rainfall detection, temperature, water level using IoT devices that allow the public to plan for any extreme weather conditions and take the required precautions.

For post flood risk solution, we provided an early warning to all users. Usually, end-users are public crowds, government officials, non-government officials. Overall flood predictions are for selected specific area of Sri Lanka with integrated components with IoT real time monitoring, deep learning algorithms and crowdsource to provide water level prediction above significance level.

Keywords – Deep Learning, Flood Prediction, Information Hub, Machine Learning, Water Level Prediction.

Table Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviations	Description
ML	Machine Learning
SLR	Simple Leaner Regression
ANN	Artificial Neutral Network
DCT	Decision Tree
IOT	Internet of Things
RRI	Rainfall Run off Inundation
SDLC	Software Development Life Cycle
IDE	Integrated Development Environment
API	Application Programming Interface
VM	Virtual Machine
ROI	Return on Investment
DMC	Disaster Management Centre
MCDT	Multi-Criteria Decision Tree
CART	Classification and Regression Tree

1 Introduction

Floods are a known threat to a number of locations around the world. Although floods are a natural danger induced by excessive rainfall, human-made factors have raised the risk of floods even in the event of minor rainfall. Those factors are generally linked to various unfriendly or obstructive uses of land resources, such as damming natural water flows to lowlands and the sea.

As a result, flood risk management incorporates both a short-term and long-term approach.

- Long-term Strategies

The long-term approach necessitates worldwide agreed-upon efforts to regulate human activities in order to conserve the environment and minimize the carbon footprint. The long-term approach necessitates worldwide agreed-upon efforts to restrict human activities to safeguard the environment and keep global warming below 1.5 degrees Celsius above pre-industrial levels, as well as other polluting human activities.

- Short-term Strategies

The short-term plan is to safeguard people from floods that are still there. This strategy necessitates flood prediction tools and a flood disaster management system that can issue alerts and evacuate people from flooded areas quickly.

The Kalu River basin in Sri Lanka was chosen as the study region because it is prone to flooding.[1]

According to a recent study, machine learning models and deep learning algorithms are commonly employed in the development of local weather prediction models.

The training data set is fed with historical flood and rainfall data.[2]

IoT sensors continuously monitor changing weather conditions, and the readings from the IoT devices are used to train the model for forecasting predictions.

Creation of Data-driven models to arrange and classify the final data sets, the acquired data sets are evaluated and confirmed using statistical analytic techniques.

In this report further discusses use of machine learning methods and artificial neural networks are used to predict flood occurrences and rainfall based on historical data sets.

1.1 Background & Literature Survey

Globally, the notion of flood risk management is rapidly expanding. Flood risk management is the framework for dealing with flooding emergencies.

The primary goal of flood risk management is to identify floods, activate before a flooding crisis develops, and reduce the potential consequences. Flood risk management offers a collection of theoretical foundations for before, during, and after a torrential downpour.

Prior research concluded that meteorological data provided a significant influence in flood forecasting. There are several research covered in area of flood damage mitigate methods [4].

Identify these issues of existing flood forecast system and optimizing those methods to develop a more accurate systematic model to provide predictions using computational approach, will be discussed in bellow onwards.

1.1.1 Current Flood Forecasting Systems

Upon identifying the pre-existing flood forecast systems, most of the data not-real-time and the existing data wasn’t sufficient for creating flood forecast systems. But machine-learning algorithms coupled to data mining methodologies can provide environmental forecasts based on historical data [5].

Current systems with low quality data input and machine learning models have a significance of existing rainfall gauges are incapable of accurately capturing the spatial distribution of precipitation. Estimates of rainfall can easily overstated and discharge data is only accessible during the flood season[5].

Another flood forecasting system conducted in the Kalu River basin area by Sri Lankan Irrigation Department and ICHARM Public Works research institute of Japan had concluded a simulation of Rainfall Run off Inundation (RRI) model has revealed that the river dimensions had a significant impact on model output. And even with the use of satellite data was used to validate the flood inundation [6]. In the same researchers had issue of not have enough the computational power of simulation in their existing systems. [5]– [12]

1.2 Research Gap

Similar selected number of studies are compared and contrasted to show the research gap of the new system.

The Flood water level prediction using machine learning adequate training data. Which will affect the decision-making models key in. The decision-making model will be a supervised learning algorithm. This will result in a model with water level prediction and classification with warning decision.

In the gap analysis the number of futures works available features with the system will explained in Table 1 Comparison of Research.

Table 1 Comparison of Research

Existing Research / System	Data	Flood Prediction Model	Decision Making Model
Research [7] **Intelligent Flood Management System**	Text Data, Genetic Data, Localize Data	No	No
Research [4]	Rainfall, Temperature	Yes	No
Proposed System	Temperature, Water Level, Rainfall	Yes	Yes

1.3 Research Problem

Flooding is a disastrous circumstance in Sri Lanka for a variety of causes. When there is a flooding issue, the water level rises quickly. The most affected individuals are who live around in areas that occurs flash flood and long-term flood prone areas. To avoid such incidents and to be properly prepared ahead of time must have a flood prediction to share the information to the individual who live in those areas.

Daily forecast report will only warn the individuals and heavy rains occurring in that area with forecast for longer times. The individual live aside river basins are prewarned but once the flooding happens the citizens are still not prepared. Providing the information for their personal devices it will significantly increase the chance of individuals acknowledgement of disaster alerts in the area.

There are no models created with active live data reading and utilizing crowdsourcing information to create models to predict the flooding. This is the issue that plans to be resolved in this research.

1.4 Research Objectives

1.4.1 Main Objective

Create the Flood Forecasting Model to predict the flooding for the selected specific area using historic data collected about past 3 years. Identify the main factors that mainly contribute to the flood prediction model. Research most effective algorithm for data sets that will result in increasing the accuracy of the model.

1.4.2 Sub Objective

1. Analysis on river basin flooding using Hydrological model and data-driven model.

   This model uses historic data to predict the long-time water level forecasting by monitoring the water levels.

2. Develop decision making model with rainfall data and crowdsourcing data.

   All the data gathered from IoT, crowdsourcing, rainfall models is used to develop the model for providing warning messages end users.

3. Provide method to overcome challenge of low performance and low accuracy.

   Most of the already existing solutions has problems with the accuracy and its performance. Use of new methodologies increase the accuracy of the models.

2 Methodology

2.1 Methodology

This section emphasizes the strategy for resolving the research problem through the use of appropriate methodologies.

The overall system architecture, crowdsourcing component overview, logical design of crowdsourcing, development process, data collection and requirement gathering, feasibility research, design components, and commercialization are all covered in this section.

2.1.1 Study Area

This study is aimed to prevent problems caused by recurrent flooding disaster due to lack of preparation. Due to continuous flooding in Kelaniya district near Kalu River area providing a suitable flood disaster management solution by forecasting flooding in riverside area and information the residents in that area.[1] The above-mentioned goal is accomplished identifying flood zones and analyze past and real-time flood data factors and then try to find a resolution using data analysis. With the data collected by metrological department, water levels be weather stations by the Kalu river and Irrigation Department. I will be performing water level prediction functionality in this research for forecasting longer time durations of flood forecasting, near Kalu river basin area as depicted in Figure 1.

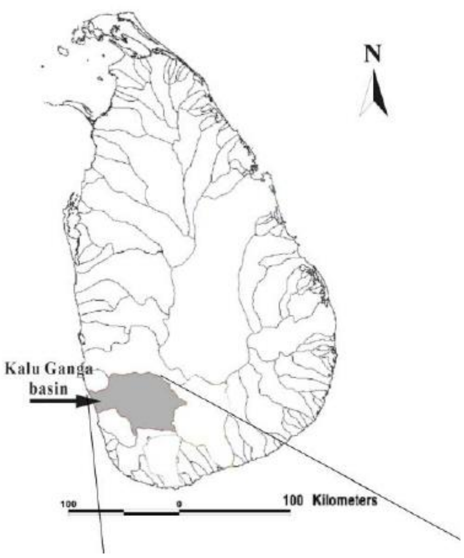


Figure 1 Study Area Kalu River Basin - Sri Lanka

The Kalu river basin area is spread across Western province, and Sabaragamuwa province which are two major provinces in Sri Lanka. River consists of Magura River, Kuda river and Kalu river.

There are five major water gauge stations with respectively with 3 branched rivers as in Figure 1.

Figure 2 Study Area Kalu River Gauge Stations

2.1.2 Development Process

The system was developed under the creation approach of the software development life cycle (SDLC) which is a method used for creating software applications and system. The development has six phases, Planning, Requirements gathering, Designing, Implementation, Testing, Evaluation and Maintenance. This software development cycle is in the use of each phase our development. For best approach we utilize the agile practices which are agile leaned approach of developing agile infrastructure and agile operations. As a research team and our complexity of our research components we value the collaboration between development and operations, the best approach was Agile based SDLC Figure 3. The development of the models for predictions of the flood situations utilizing weather and crowdsourced data, the requirements gathering was the initial stage. The next phase is analyzing the gathered requirements to create a design. As stages progress on to development to testing and deployment. Validating the system through white box testing stages. Since in the research type project the reviewing is extensively conducted to check all the outcomes.

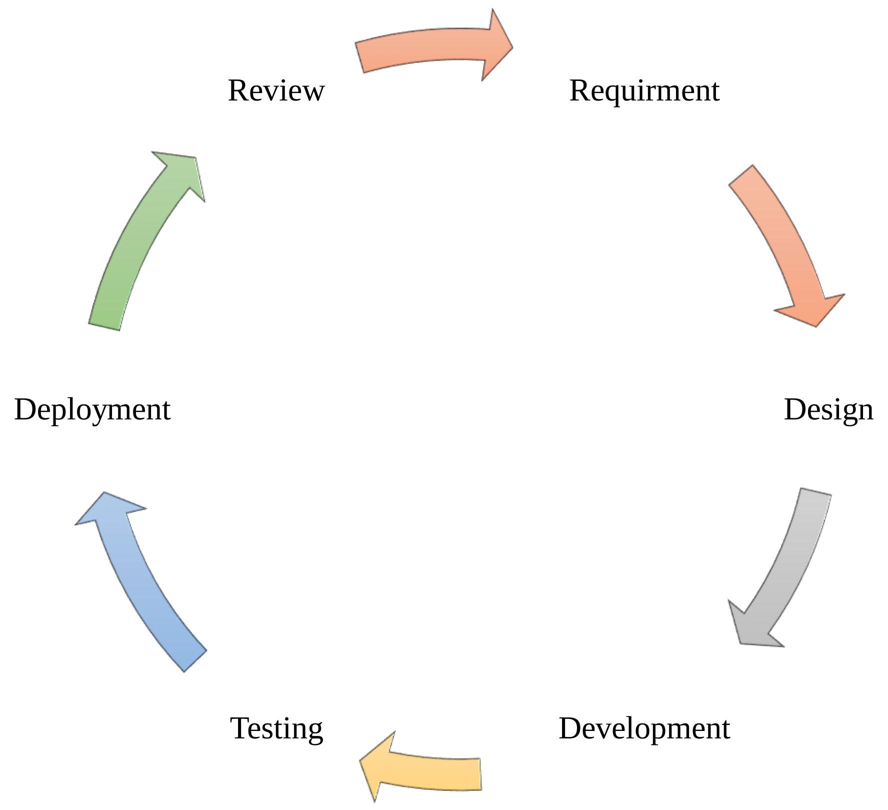


Figure 3 Agile Based Software Development Lifecycle

Requirement Gathering

Requirements are the overall functionalities and features of the end user expected to have in a system this system contains input and processed outputs of machine learning or mathematical models. Identified functional and nonfunctions requirements and nonfunctional requirements that are needed for flood forecasting and warning system are highlighted bellow.

Functional Requirements

- Access data from core databases
- Provide user access to accurate prediction
- Provide users API access forecast
- Present warning for the users in the area

Nonfunctional Requirements

- Easy use of the system
- Availability of the system
- Performance
- Manageability

2.1.2.1 System Architecture

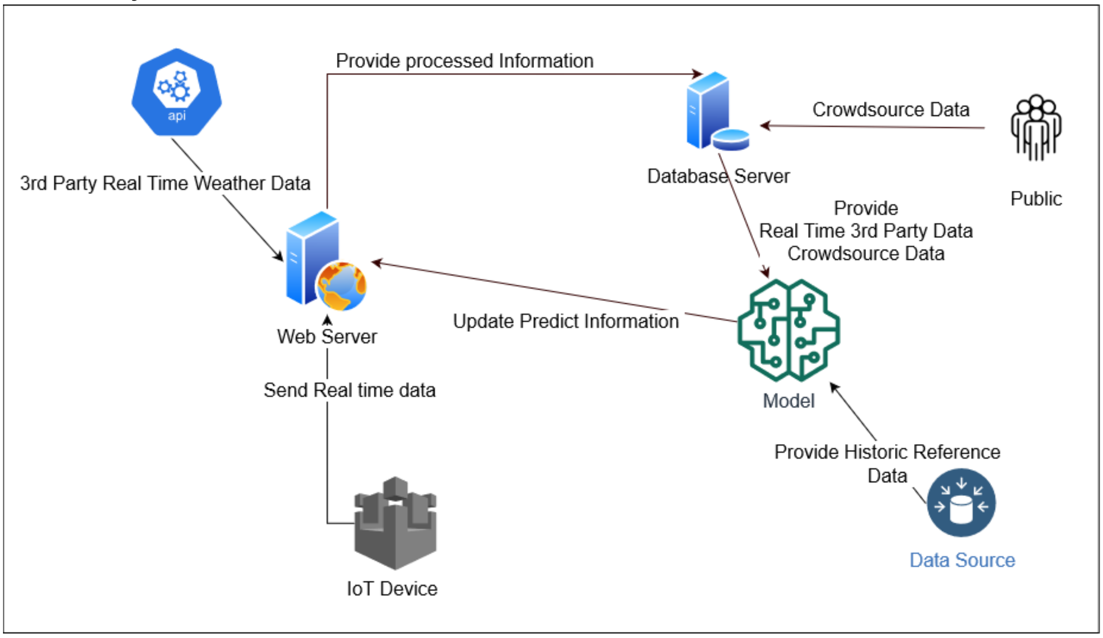


Figure 4 System Architectural Diagram

Development process consists of main three steps

1. Data Collection
2. Data Processing
3. Model Creation

Creation of flood forecasting model to predict the flooding for the selected area using the historic data from past years.

2.1.2.2 Data Collection

Data Sources are essential part of the building data-driven modes. The trusted data is collected by Metrology Department, Irrigation Department and Disaster Management Center in Sri Lanka. The data collected by past 3 years (2019 -2021) such as amount of rainfall occurrences and real-time basis, water levels change in river on time-basis, number of water levels discharged in a specific station, real-time water levels, rainfall durations, precipitation, elevation, flow directions of the river and rainfall are used in creation of machine learning models.

The data collected are sampled by twenty-four-hour intervals. And these collected data is collected from different source pre-processing of the data sets to normalize into common data frames was essential. The pre-processed datasets used in developing machine learning models for water level prediction models are presented in the Table 2 Water Model Processed Data Set Overview.

Table 2 Water Model Processed Data Set Overview

Factors	Time Interval	Type	Description
Tributary	2019 - 2021	Text	River / Sub River name
Station	2019 - 2021	Text	Gauge station name
Catchment Area	2019 - 2021	Km^2	Catchment area of the water near river gauge station.
Water Level Classification	24 hours	Remark {Normal, Alert, Minor, Major}	Four levels of water classifications
Water Level Before	24 hours	m	Water level 1hour before read
Water Level at Time	Present	m	Water level when the readings were taken.
Fluctuation	Present	Remark	Water level has a significant change since

		{Normal, Rising, Falling}	the 1hour reading to present.
Rainfall	24 hours	mm	Rainfall near the water gauge station.

```
In [3]: river_data = pd.read_csv('SeptDataSet4ANN_forColab.csv')
        river_data.head()
        river_data
```

Out[3]:

	Alert Level	Minor Level	Major Level	Level before	Water Level at time	Rising 2 or Falling 1 or Normal 0	RF in mm	F4, F3, Alert 2, Normal 1, Empty 0
0	3.0	6.0	7.5	2.18	6.84	0	0.8	1
1	5.0	6.0	5.5	2.14	6.29	0	2.0	1
2	5.0	4.5	5.0	4.62	2.88	2	40.0	1
3	5.0	10.7	7.0	4.14	2.12	1	9.8	1
4	10.0	10.7	12.2	8.45	8.92	0	6.1	2
...	...	...	...	...	...	...	...	...
9995	4.0	6.0	7.5	3.12	2.43	0	9.0	1
9996	3.0	4.5	8.0	8.45	8.28	0	9.0	1
9997	10.0	10.0	12.2	4.57	2.40	0	17.1	1
9998	10.0	10.7	12.2	3.31	8.92	1	38.5	2
9999	4.0	6.0	7.5	2.87	4.12	0	0.7	1

10000 rows × 8 columns

Figure 5 Station 4 - Rathnapura dataset

Real-time IoT devices sends data through the IoT devices are intended to be used in detecting the water level model and rainfall model. The bellow table provides an insight into data used in creation of decision tree model for decision making.

Table 3 Decision Tree Model Processed Data Set Overview

Factors	Unit	Type	Description
Distance	cm	Numeric	Distance from water level surface to the sensor
Humidity	percentage	Numeric	Relative Humidity from sensor
Temperature	Celsius	Numeric	Relative Temperature from sensor

Rain	Binary	Boolean {Rain, Not Rain}	The water droplet sensor reading about currently raining or not.
Rain Range	Class	Remark {No Rain, Rain, Heavy Rain}	Raining Classification by the Rain forecast model at in interval.
Water Level	Class	Remark {Normal, Alert, Minor, Major}	Water level classification by Water Level forecasting model.
Warn	Binary	Boolean {warn, not warn}	Decision Tree Decision.

```
df.head()
df
```

	timestamp	distance	Humidity	Temp	Current_Rain	Rain_Range	Water_Level_Class	Warn
0	1626886919	3	82	31	0	0	0	0
1	1629734639	32	78	31	0	0	0	0
2	1628334639	32	78	31	1	2	0	0
3	1626886919	32	77	31	0	0	0	0
4	1626887929	3	87	31	1	2	1	0
...	...	...	...	...	...	...	...	...
995	1629734639	32	78	31	1	0	0	0
996	1626887039	3	87	31	1	2	3	1
997	1626887020	3	90	31	1	0	3	0
998	1626886919	32	88	31	1	2	1	0
999	1626887039	3	90	31	1	1	1	0

1000 rows × 8 columns

Figure 6 IoT Pannipitiya dataset

2.1.2.3 Data Processing

Data collected from these sources are used to with multiple methods of analysis.

1. Mathematical based approach
2. Artificial Intelligence and machine learning based approach

Mathematical based approach

As a mathematical based there are several dominant data analysis techniques are used. There is serval analysis can be utilized such as Time Series analysis, Factor analysis and Regression analysis. Most suitable for use on identifying the correlation between a dependent variable and other independent variables factors it.

2.1.2.4 Model Solution

2.1.2.4.1 Water Level Prediction Model

Linear Regression Analysis provided dependent variable factor of fluctuations (water level normal, rising or falling) with accuracy levels correlating with the river alert level, water level before, and water level at the time, rainfall, water level classification.

The created linear regression model was evaluated using mean squared error, which in results.

The Figure 7 contains utilization of python SKLEARN and NUMPY libraries for the calculation of accuracy of liner regression model.

```
In [6]: linear = linear_model.LinearRegression()

        linear.fit(x_train, y_train)
        acc = linear.score(x_test, y_test)
        print(acc)

        0.7932012905714l306

In [7]: print('Coefficient : \n', linear.coef_)
        print('Intercept: \n', linear.intercept_)

        predictions = linear.predict(x_test)

        Coefficient :
         [[  0.83113332   0.46795598  -0.32625304]
         [  0.42533608  -1.16877498   1.22473926]
         [ -0.83744845 -77.75884296  78.60561722]]
        Intercept:
         [-3.33659835 -2.00999089 14.64257374]

In [8]: mse = metrics.mean_squared_error(y_test, predictions)

In [9]: print("Mean Squared Error {}".format(mse))

        Mean Squared Error 0.3134568743884474
```

Figure 7 Python SKLEARN to calculate Mean Squad Error Image

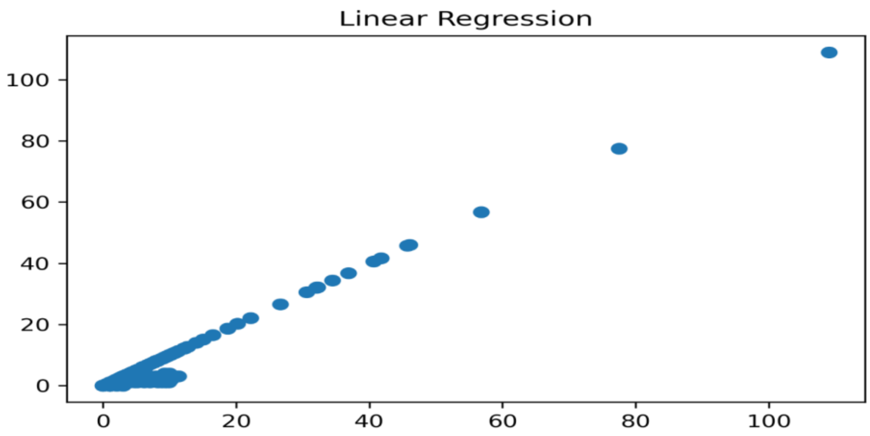


Figure 8 Water Model Leaner Regression

Artificial Intelligence and machine learning based approach.

Using of Artificial intelligence and machine learning to predict variables and identify significant relationships with the variables in modern approach in the creation of models as a solution. Serval machine learning methodologies are identified by Support Vector Machine (SVM) Algorithm, Fuzzy logic, Random Forest, and Artificial Neural Network (ANN). For the suitable cases as classification and regression challenges using SVM and ANN algorithms are adequate.

The bellow Figure 9 contains the flowchart of algorithm used in training the machine learning models.

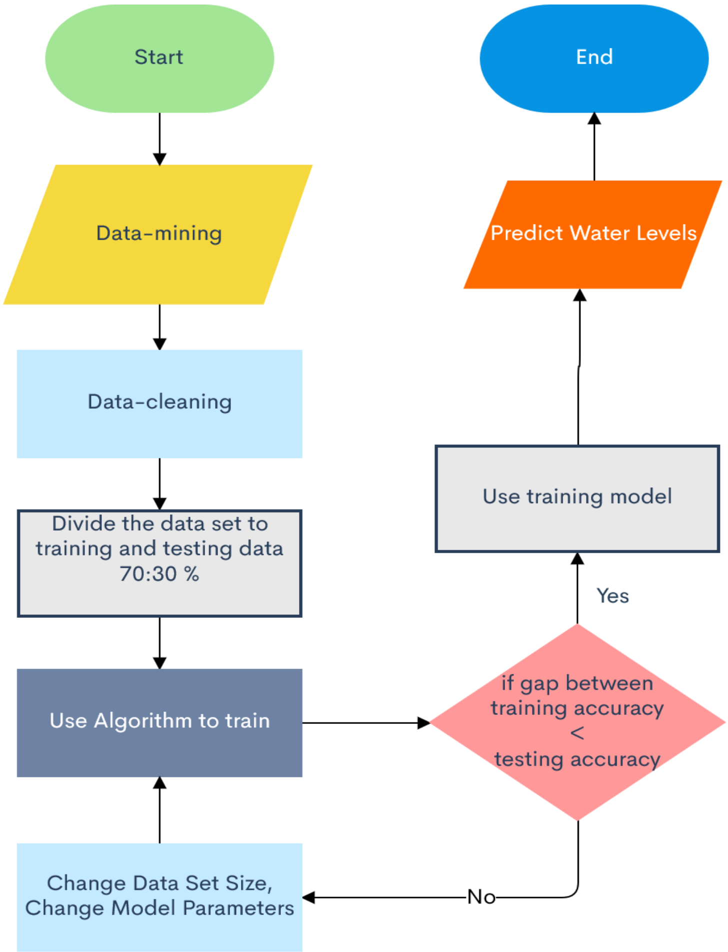


Figure 9 Models Building Algorithm

Support Vector Machines (SVM)

Support Vector Machines are a machine learning algorithm under supervised learning. And the plotted data in the dimensional space each number of features of values are in a particular coordinate. The SVM models are used in each 5 location data sets with attributes changed to identify contributions of the variables factor into main dependent variable.

For the accuracy of data, the data sets such as testing data and training data are normalized in to maintain the general distribution and ratio of the source data.

Artificial Neural Network (ANN)

Deep Learning approach of distinguishing other factors of dependent variable and classification in the use of utilizing multiple layers of combined neural layers with each other to provide a network of neurons. These networks of neurons are in hidden layer. As the input layer provides the inputs and output layer is number of dependent variables Figure 10 shows three inputs and two outputs.

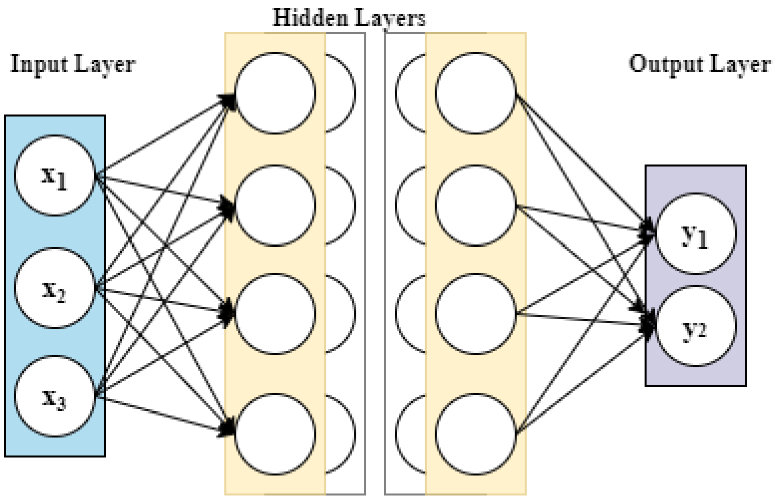


Figure 10 Artificial Neural Network Layers

As above Artificial network. The water model creation water level prediction initiated using TensorFlow and Keras Libraries.

2.1.2.4.2 Decision Making Model

Decision selection models are contained with decision chains. The data provided data has to calculate the operational possibilities of a decision. Most the systems as discussed in the Background & Literature Survey as more knowledgebase centre expert system in use. The approach of gathering data for build an expert system is diagramed in Figure 11.

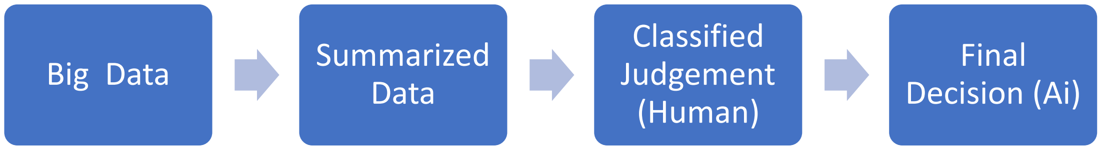


Figure 11 Optimal Expert System Flow Diagram

Decision Tree

The decision tree model is consisting of branches, edges, leaves. The model creation has two major steps induction and pruning. For this model it is utilizing the Gini index function because it is in the use of classification.

The data set used in this model creation is depicted in Table 3 Decision Tree Model Processed Data Set Overview. Using Classification and Regression Tree (CART) to perform classification.

A data-driven water level prediction model is used to provide accurate water level prediction.

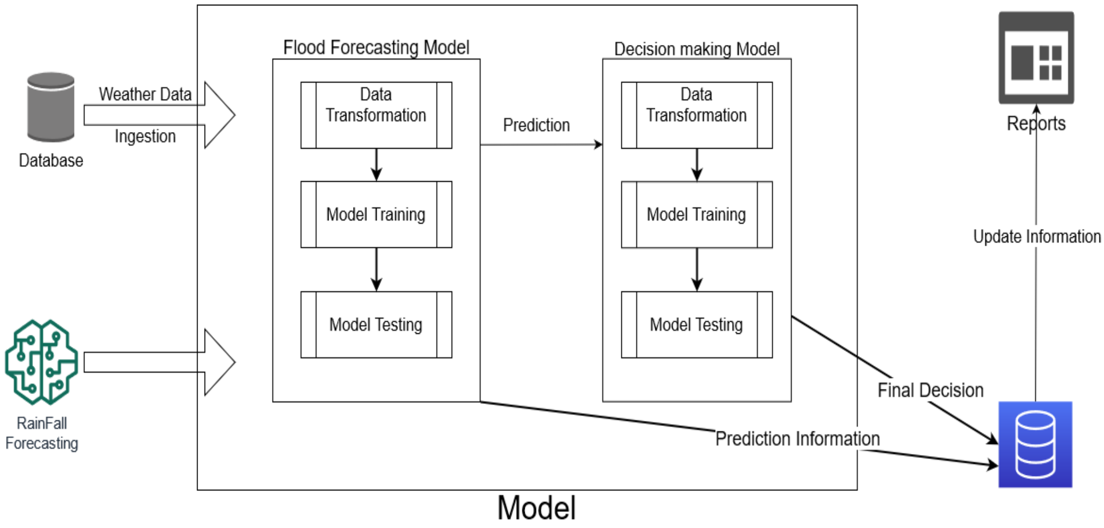


Figure 12 AI Model

As in Figure 12 these results of water level prediction model are be presented and passed on to the results in order for the decision-making model to make the final decision about the situation.

2.2 Feasibility Study

A feasibility analysis of such a research project assesses the system's viability for implementation. The primary purpose of a feasibility study is to establish whether or not the proposed system can be implemented. The feasibility study investigates the project's success rate. The feasibility study determines whether the production of service or goods is worthwhile to invest in. Following three categories of feasibility study, which are scheduled feasibility, economic feasibility, technical feasibility that we can proceed ahead of the development with limit resources we have.

The rapid development of the system we proposed with data-driven models and IoT requires specialized development tools and technologies.

2.3 Software specifications

2.3.1 Development Tools

- JetBrains PyCharm: Python IDE

Python programming language integrated development environment developed by a company Czech Republic named JetBrains. This tool has improved graphical debuggers, code analysis and seamless integration with other tools to version control and code analysis.

- GitLab: Git Repository

  GitLab is a Continuous integration lifecycle tool for the web that includes a Git repository manager. This tool includes continuous integration and development. Provided by GitLab Inc.

- Colab: Google Colaboratory
  Colab notebooks enables to merge executable code and rich text in a single document, as well as graphics, HTML, LaTeX, and other formats with running on the Jupyter Notebook environment. The tool is offered by Google and hosted on Cloud.

2.4 Commercialization aspects of the product

The risk management and prediction models are being created with initial startup cost of the creator's expense. These expenses are accounted for a capital endeavor. As the end development team has determined to expand to commercial operations.

Several methods are developed to provide the services to the government authorities and other third-party users, which will help on surpassing profit margins.

1. Subscription (monthly/ annually) based API access for forecasts.
2. One-time payment access for forecast in IoT device purchase

As the users are the government or consumer will be able to have dashboard data integrated to their system with subscription-based payment as a contract. The end users with IoT devices purchased to setup in the fields. And the selected areas will have the forecast on mobile app and web app dashboards.

All of the end users can categories as commercial user and non-commercial users. As an example, commercial users are the government authorities and researchers. And non-commercial users are the researchers, and nonprofit organizations work on relief efforts in an emergency situation.

Using the above methodologies, the developers will be recovering the capital spend for the development and users will have the return on investment (ROI) within the number of benefits they gain from employing systems functionalities.

2.5 Testing and Implementation

2.5.1 Testing

Software testing is a method of determining the accuracy of software by taking into account all of its characteristics (reliability, scalability, portability, reusability, and usability) along with reviewing the operation of software components to detect bugs, errors, and faults. Software testing has the ability to identify any errors and weaknesses during development.

Various types of testing enable us to detect flaws that are only noticeable during runtime. Using testing components summarized as Unit testing, Regression testing and Integration testing.

The machine learning models cannot be tested as a software regular testing. In the machine learning algorithms, the testing of the components includes two main data sets other than the training set,

1. Validation Set

   For hyperparameter tuning, having merely a training set and a testing set is insufficient. And that can result in overfitting. To avoid that, you can select a small validation data set to evaluate a model. Only after you get maximum accuracy on the validation set, you make the testing set come into the game.

2. Testing Test

   For assurance of the models guaranteed real world implications, choose samples for a testing set from training set instances that the machine has used in training.

   It is critical to remain objective during the selection process and to draw samples at random. Also, avoid training on testing data by not using the same set several times.

The diagram bellow Figure 13 Data Sets using for Model Validation show the process of validating the model.

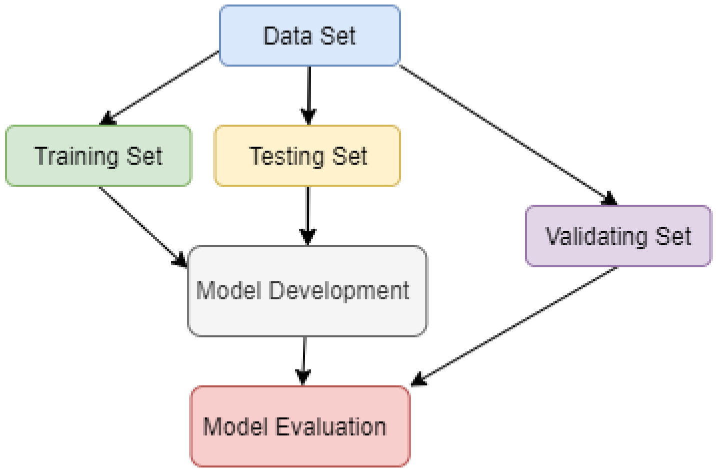


Figure 13 Data Sets using for Model Validation

There are other model testing approaches,

1. Pre-Train testing

   Testing the data sets labels, and missing labels and missing values.

2. Post-Train testing

   Checking the performance of trained models and check it performs accordingly. As an example, how much the classification of the models results. And minimal functionality testing of the model.

2.5.2 Implementation

The proposed model was created to be providing predicated classification as the daily records get updated.

The created model has been implemented as an API to provide predictions in Figure 14 API Demo Request.

Other API developed are listed in Appendices.

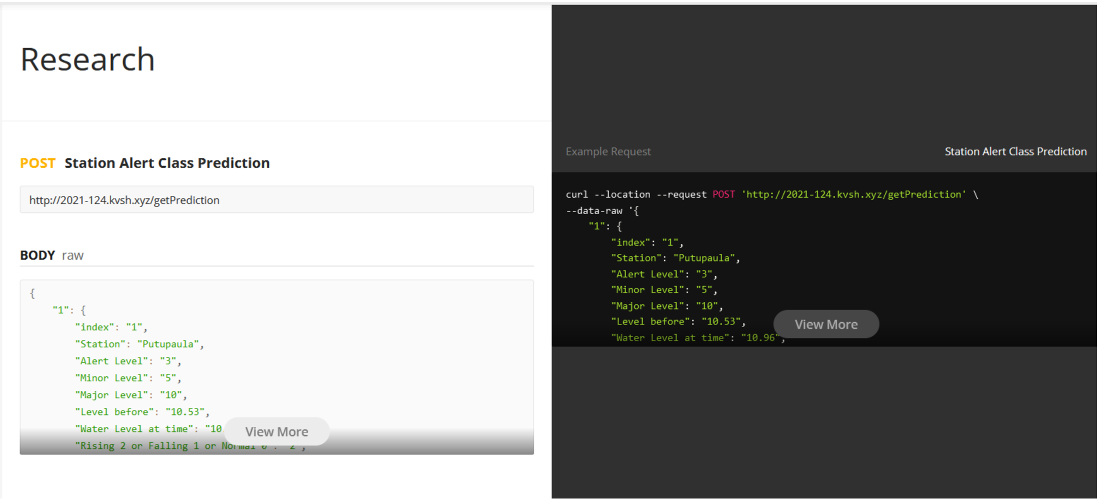


Figure 14 API Demo Request

The output of the model is displayed as bellow in Figure 15.

Figure 15 Forecast Prediction Response

The Table 4 Classification levels are displayed of the classification values and levels.

Table 4 Classification levels

Classification	Value
Normal	1
Alert Level	2
Minor Level	3
Major Level	4

The currently collected data are on firebase database so it can utilize in the future development and dashboarding the predictions.

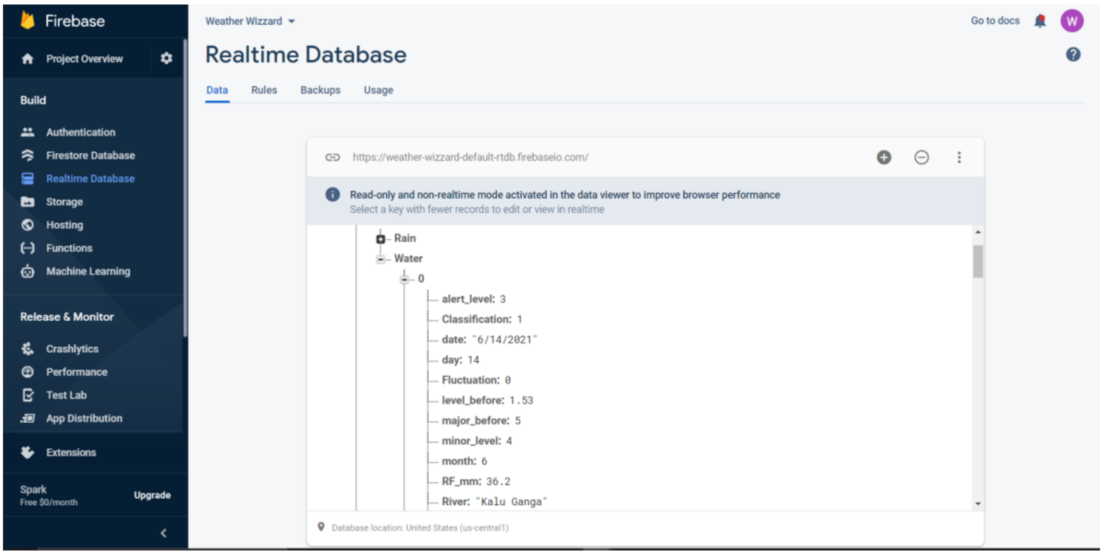


Figure 16 Firebase Snapshot of Flood Data

2.5.2.1 *Designs*

Dashboard view of the Water Level Reading, This User Interface provides the information of water level in meters Figure 17 Water level Dashboard View.

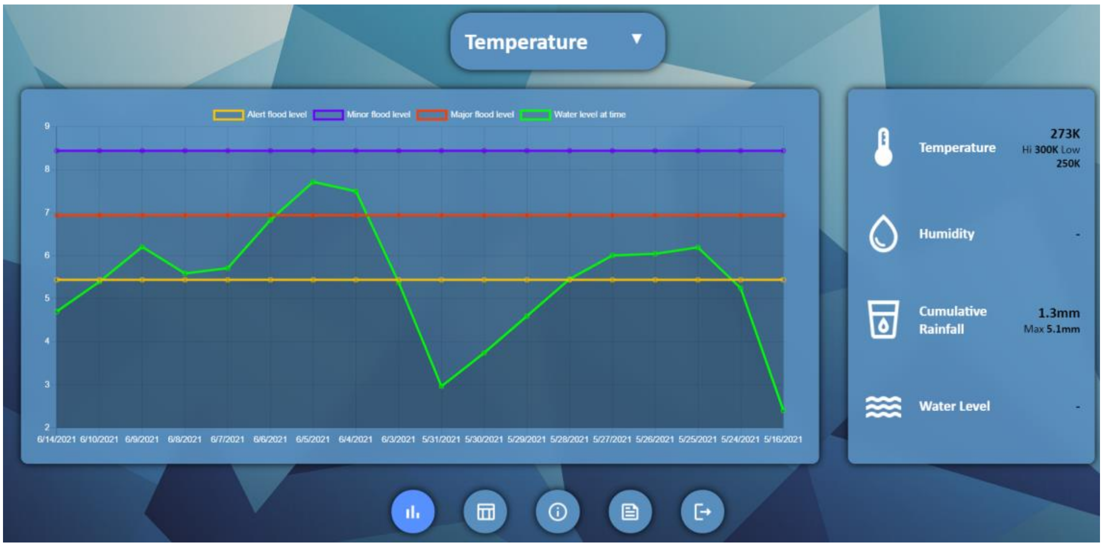


Figure 17 Water level Dashboard View

The Kalu River Basin Water Level Prediction is Displayed in this user interface.

For each selected station, the flood classification prediction will be displayed as in Figure 18 Station Prediction Classification Dashboard Values.

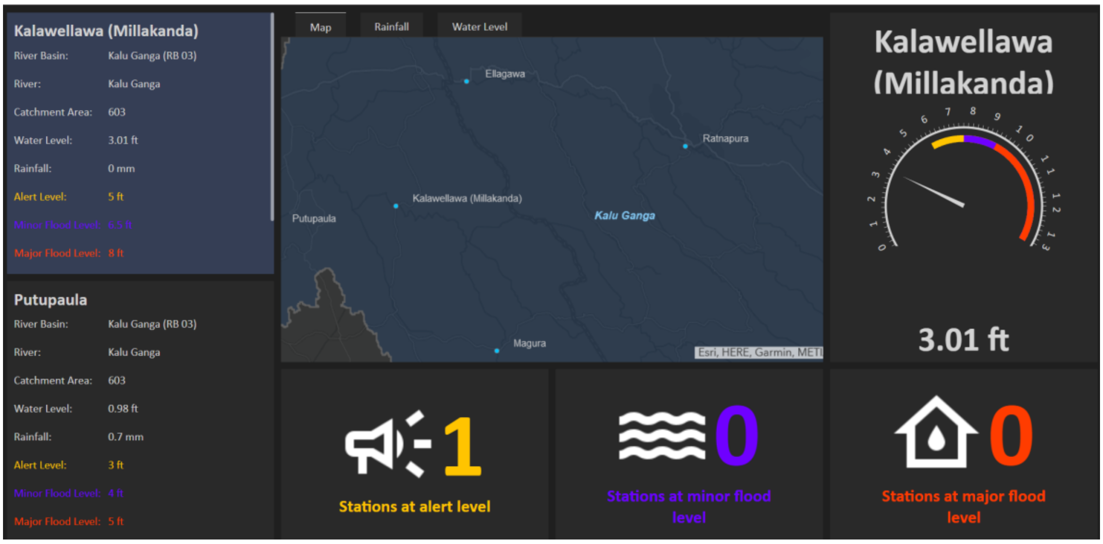


Figure 18 Station Prediction Classification Dashboard Values

3 Results and Discussion

3.1 Results

- Water Level Classification Model

  1. (SVM) Model Confusion Matrix in Figure 19 SVM Model Confusion Matrix

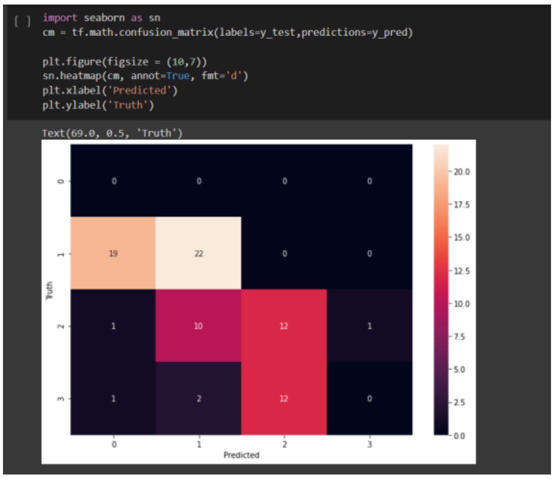


Figure 19 SVM Model Confusion Matrix

  2. ANN Model Confusion Matrix in Figure 20 ANN Model Confusion Matrix.

```python
import seaborn as sn
cm = tf.math.confusion_matrix(labels=y_test,predictions=y_pred)

plt.figure(figsize = (10,7))
sn.heatmap(cm, annot=True, fmt='d')
plt.xlabel('Predicted')
plt.ylabel('Truth')
```

```
Text(69.0, 0.5, 'Truth')
```

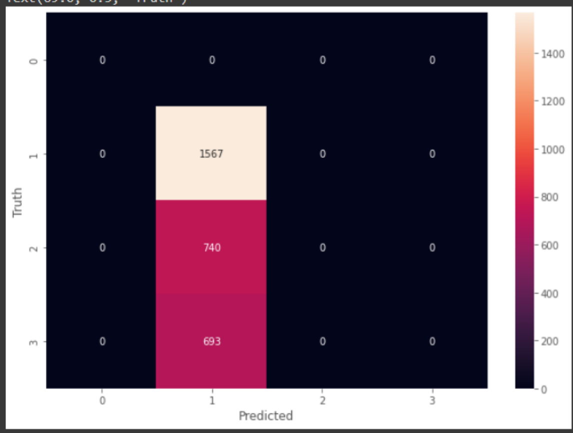


Figure 20 ANN Model Confusion Matrix

3. Value Loss Over Epoch

```python
plt.style.use('ggplot')
plt.plot(hist.history['loss'], label = 'loss')
plt.plot(hist.history['val_loss'], label='val loss')
plt.title("Loss vs Val_Loss")
plt.xlabel("Epochs")
plt.ylabel("Loss")
plt.legend()
plt.show()
```

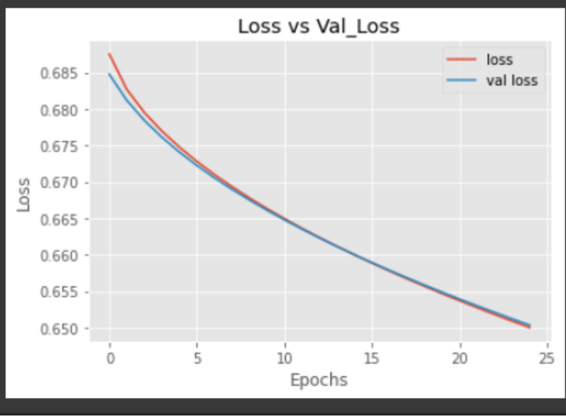


Figure 21 Value Lost Over Epoch

- Decision Tree Model

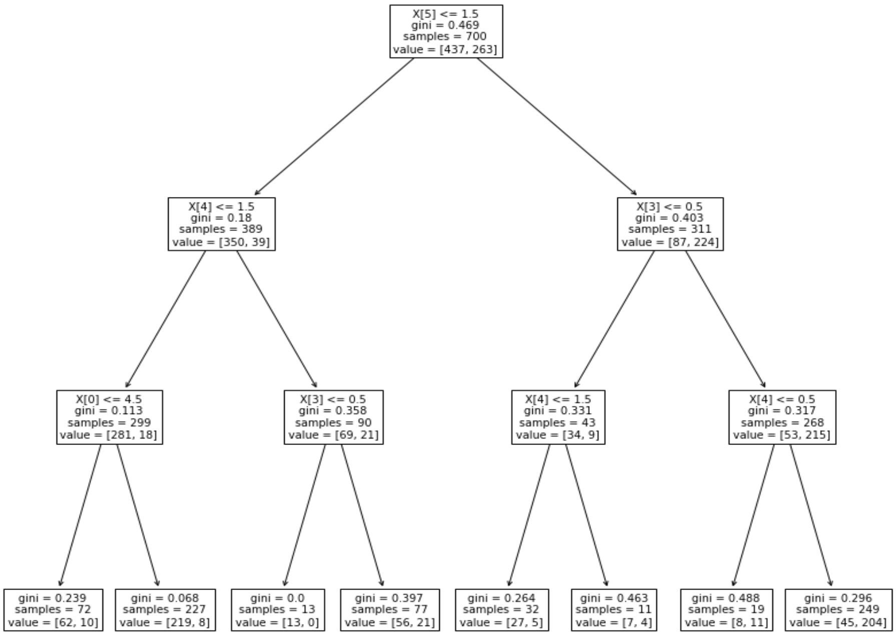


Figure 22 Decision Tree at alpha = 0.05

- X[0] = Distance
- X[1] = Humidity
- X[2] = Temp
- X[3] = Current Rain
- X[4] = Rain Rage
- X[5] = Water levels

Leaf Index Values

3.2 Research Finding

- Water Level Classification Model

  Most flood forecasting models use average results as shown in Table 5, yet the implementation of ANN has higher accuracy than all. Furthermore, results accuracy is check by conducting a metric of Mean Absolute Error for standardization. Regardless of the location of the gauge station, the accuracy of the test data does not vary.

- Decision Tree Model Evaluation

  Decision making model utilizes flood forecast of gauge stations, rainfall prediction model and IoT sensory data in Figure 22. It is a multi-criteria decision tree based (MCDT), model is most suitable for classification and can executed immediately and accurately Table 6.

Table 5 Accuracy table of water prediction models

Data	**Accuracy(percentage)%**		
	SLR	SVM	ANN
Train	81.2	79.4	89.01
Test	79	77.0	84.3

Table 6 Accuracy table of decision tree model

Data	**Decision Tree**
Train	50.4%
Test	52.4%

3.3 Discussion

The contribution for water level predictions, from the river basin data as each station has the daily fluctuations of the water levels, will not provide the whole predictions of an hourly occurrence. Providing a live data set from the gauge stations and remarks of the Disaster Management Centre (DMC) of Sri Lanka has a huge data change in the reports they provide. Building another IoT base device to water levels for River basin areas will provide continuous live river basin water readings.

The decision-making models currently uses decision tree has a lower accuracy than expected. As I can suggest using Decision Making Neural Decision Forests model will result in more accurate predictions.

4 Summary of Student's Contribution

Name	**Component**	**Task**
Ilukkumbure S. P. M. Kaveesha. W	1. Designing and developing machine Learning Model for water level prediction of the river station. 2. Design and develop the machine learning models for decision making.	1. Develop APIs to precent predictions. 2. Develop API to integrate system with other government systems. 3. Develop Dashboard to Precent Predictions. 4. Design Web Application and Fronted.

5 Budget

Component	**Amount (LKR)**
Azure VM: 1 vCPU 1GB $7.592/month	1526.13 *** 6 =** 9156.78
Irrigation Department Reports Two Years (2019-2021-July) Rs.40 per station / month 5 stations	40 * 5 * 12 * 3 = 7200.00
Total	**16356.78**

Table 7:Budget

6 Conclusion

The conducted research of experimental water level forecasting machine learning models with use of Linear Regression, Support Vector Machines and Artificial Neural Network based models provided accuracies did not affect regardless the station of the station so I conclude Kalu River Basin and slope has an equal distribution of water level and each of the factor of water level catchment area does not effectiveness of the flooding models because it’s a constant.

The natural factor of data does affect the area of Sabaragamuwa province and Western Province, and the Sea Level was not added to equation and not identify as a contributing factor.

The sudden variations of water model prediction really have contribution to the decision trees and the values really do affect the predictions accuracy.

Data driven solutions are highly dependable on data quality, therefore advance data collection and analyzing methods should be carried out. The final flood forecast, and predictions are displayed on the web application and mobile application to distribute predictions. Implementation of this kind of hybrid models require further research.

7 References

[1] Flood M and F. M, “Inundation Maps of the Kalu Ganga basin during the Flood in May 2003 4 . General Observations on Annual Extreme River Stages Annual Maximum Water Levels of Kaluganga in m MSL 7 . Flood Frequency Analysis using Annual Extremes,” , no. May, pp. 1–3, 2003.

[2] A. Journal, “Weather Forecasts Based on Rainfall Prediction Using Machine Learning Methodologies,” *Adalya Journal*, vol. 9, no. 6, pp. 72–80, 2020, doi: 10.37896/aj9.6/009.

[3] V. Bechtol and L. Laurian, “Restoring straightened rivers for sustainable flood mitigation,” *Disaster Prevention and Management: An International Journal*, vol. 14, no. 1, pp. 6–19, Jan. 2005, doi: 10.1108/09653560510583806.

[4] H. Thilakarathne and K. Premachandra, “Predicting Floods in North Central Province of Sri Lanka using Machine Learning and Data Mining Methods,” no. October, pp. 44–49, 2017.

[5] B. Li *et al.*, “On the operational flood forecasting practices using low-quality data input of a distributed hydrological model,” *Sustainability (Switzerland)*, vol. 12, no. 19, 2020, doi: 10.3390/su12198268.

[6] S. Kokularamanan, A. W. M. Rasmy, D. Perera, and T. Koike, “Development of a Flood Forecasting and Data Dissemination System for Kalu River Basin in Sri Lanka,” *Annual Sessions of IESL, The Institution of Engineers, Sri Lanka*, vol. 1, pp. 205–210, 2017.

[7] M. J. D. Fernando, D. A. K. K. Pathirana, W. J. K. T. D. Jayasooriya, S. A. H. Rathnaweera, and L. Rupasinghe, “Intelligent Flood Management System,” in *2019 International Conference on Advancements in Computing (ICAC)*, 2019, pp. 79–84. doi: 10.1109/ICAC49085.2019.9103407.

[8] M. Campolo, P. Andreussi, and A. Soldati, “River flood forecasting with a neural network model,” *Water Resources Research*, vol. 35, no. 4, pp. 1191–1197, 1999, doi: 10.1029/1998WR900086.

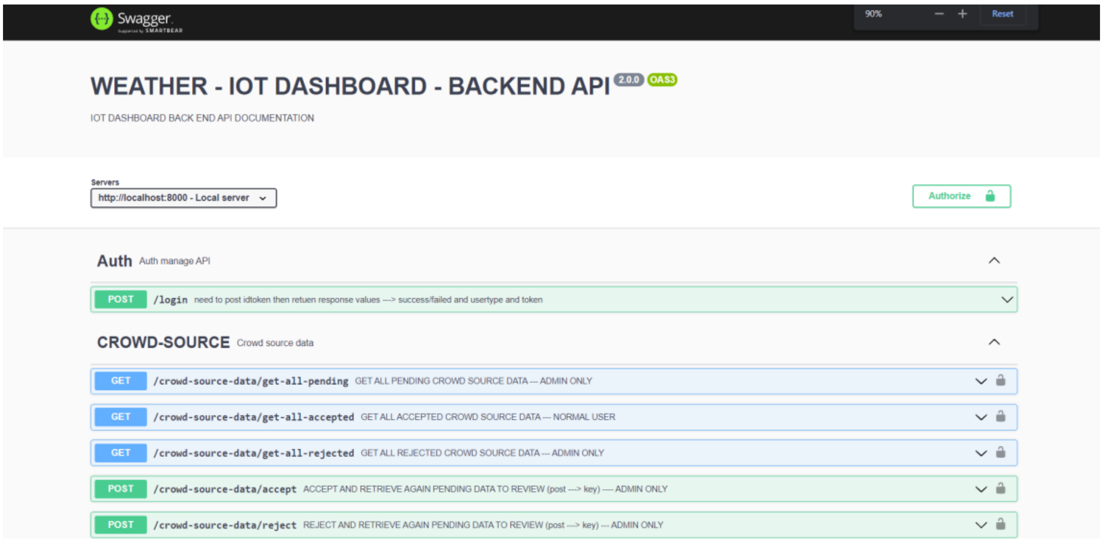
[9] A. Mosavi, P. Ozturk, and K. W. Chau, “Flood prediction using machine learning models: Literature review,” *Water (Switzerland)*, vol. 10, no. 11, pp. 1–40, 2018, doi: 10.3390/w10111536.

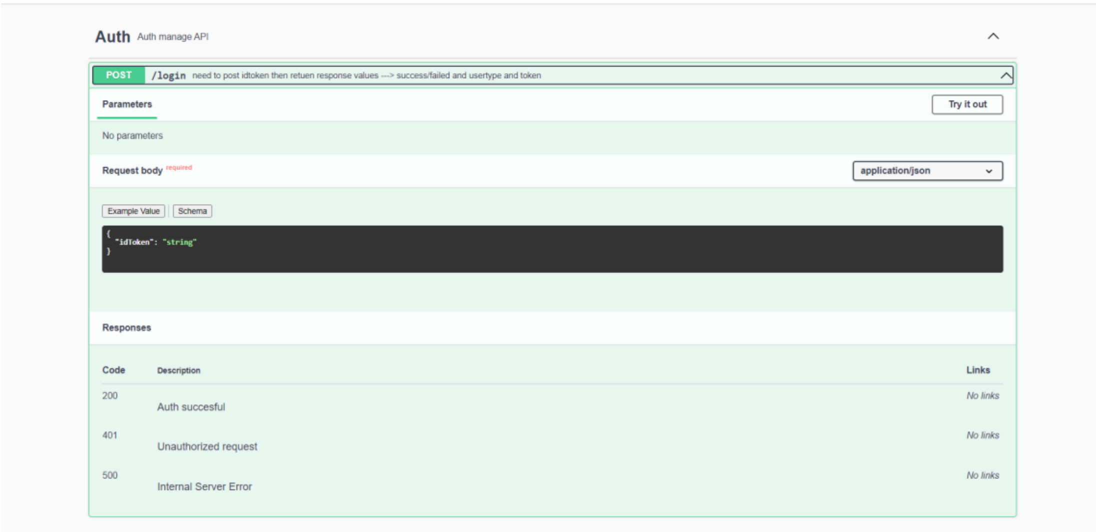
[10] S. Puttinaovarat and P. Horkaew, “Flood Forecasting System Based on Integrated Big and Crowdsource Data by Using Machine Learning Techniques,” *IEEE Access*, vol. 8, pp. 5885–5905, 2020, doi: 10.1109/ACCESS.2019.2963819.

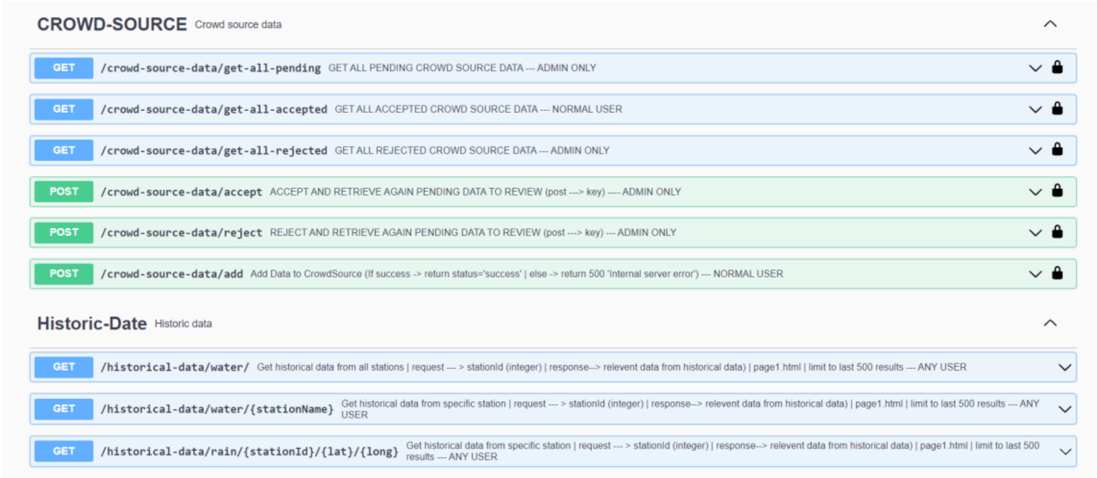
[11] S. N. Yang and L. C. Chang, “Regional inundation forecasting using machine learning techniques with the internet of things,” *Water (Switzerland)*, vol. 12, no. 6, 2020, doi: 10.3390/W12061578.

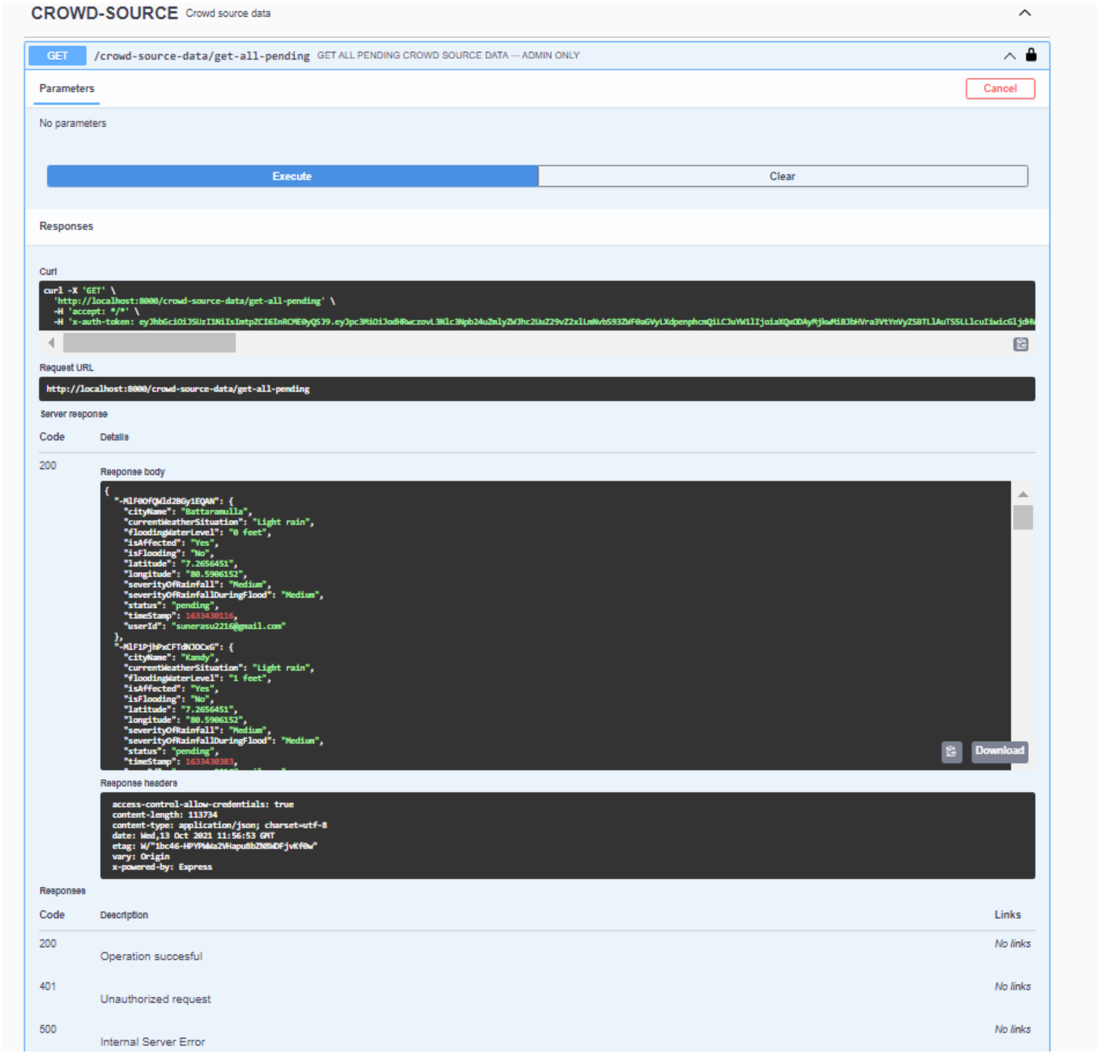
[12] R. J. Moore, V. A. Bell, and D. A. Jones, “Forecasting for flood warning,” *Comptes Rendus - Geoscience*, vol. 337, no. 1–2, pp. 203–217, 2005, doi: 10.1016/j.crte.2004.10.017.

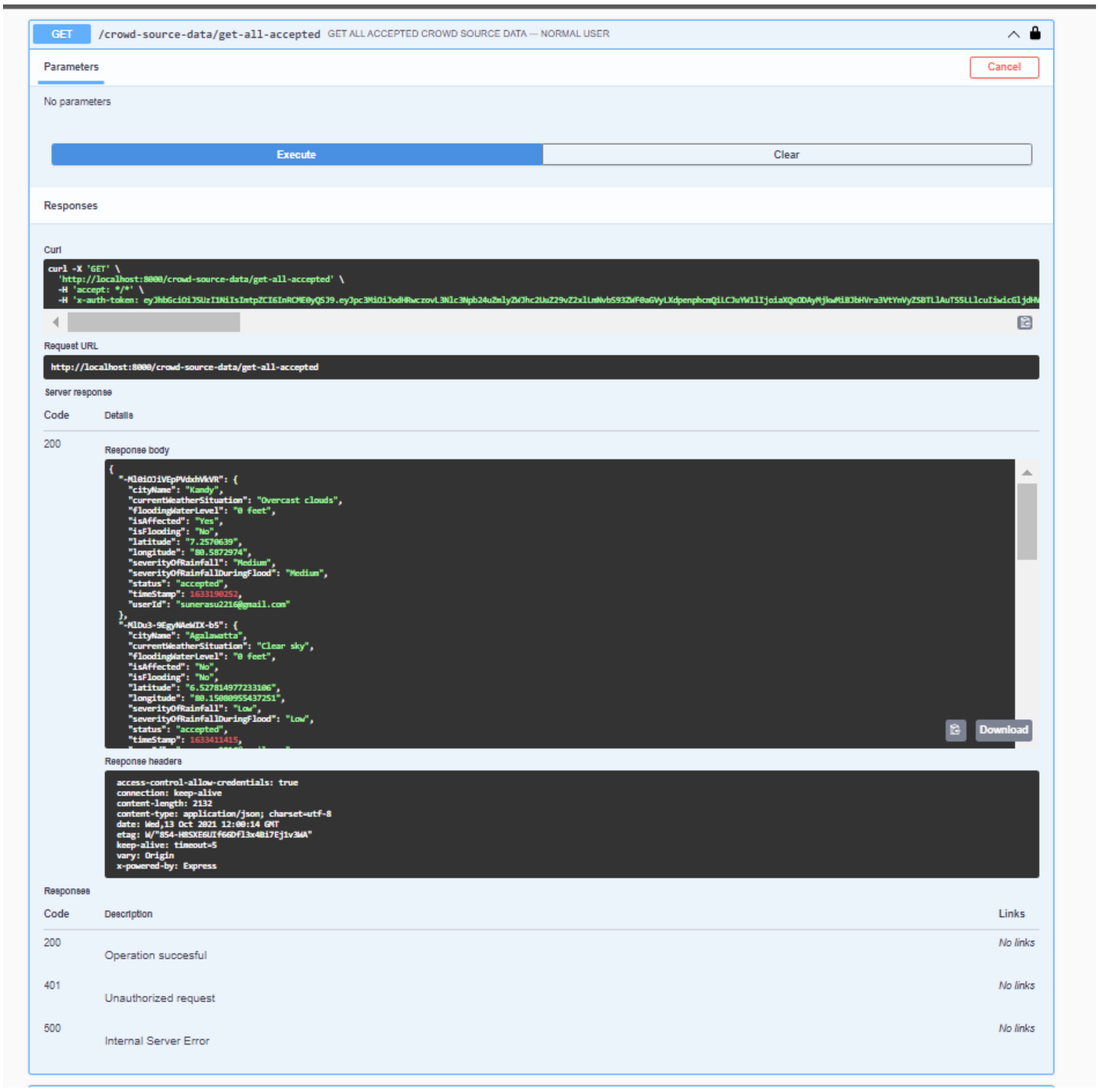
8 Appendices

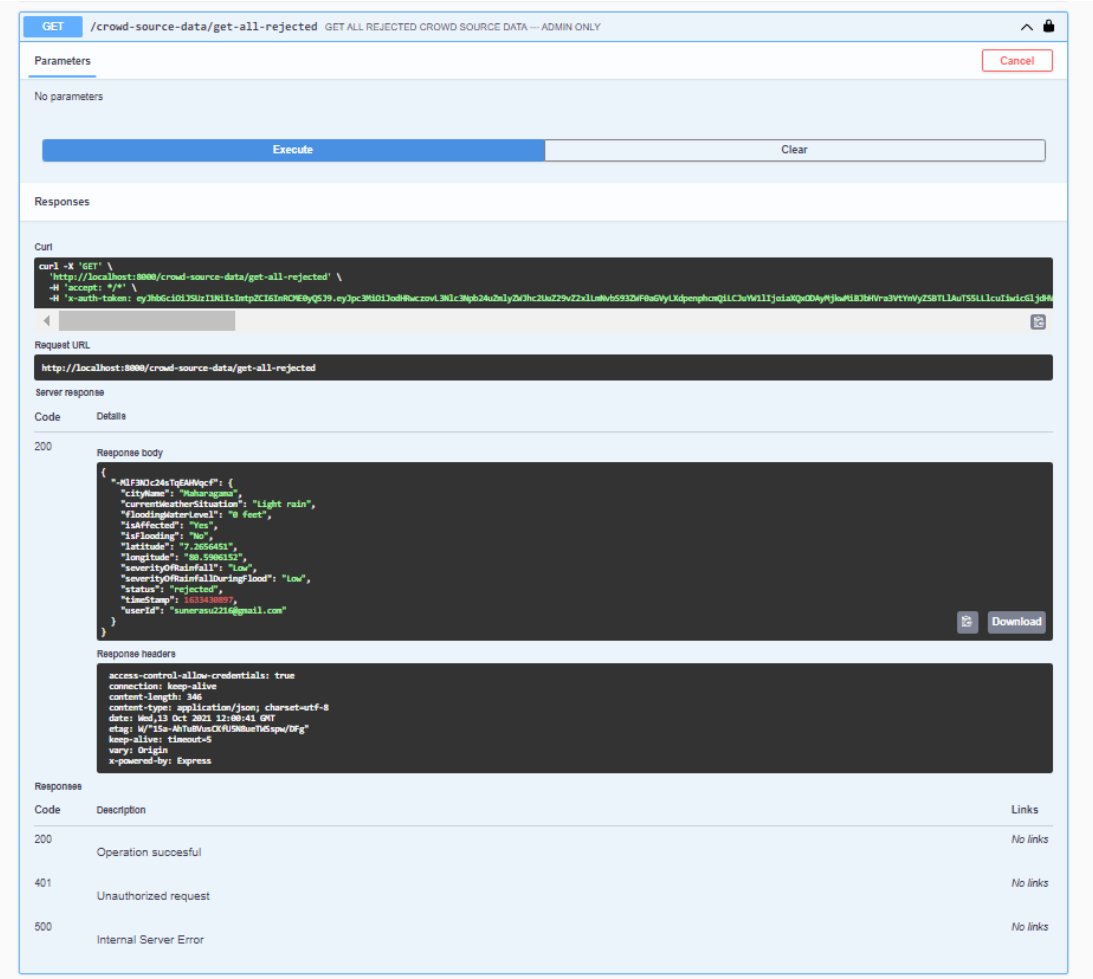


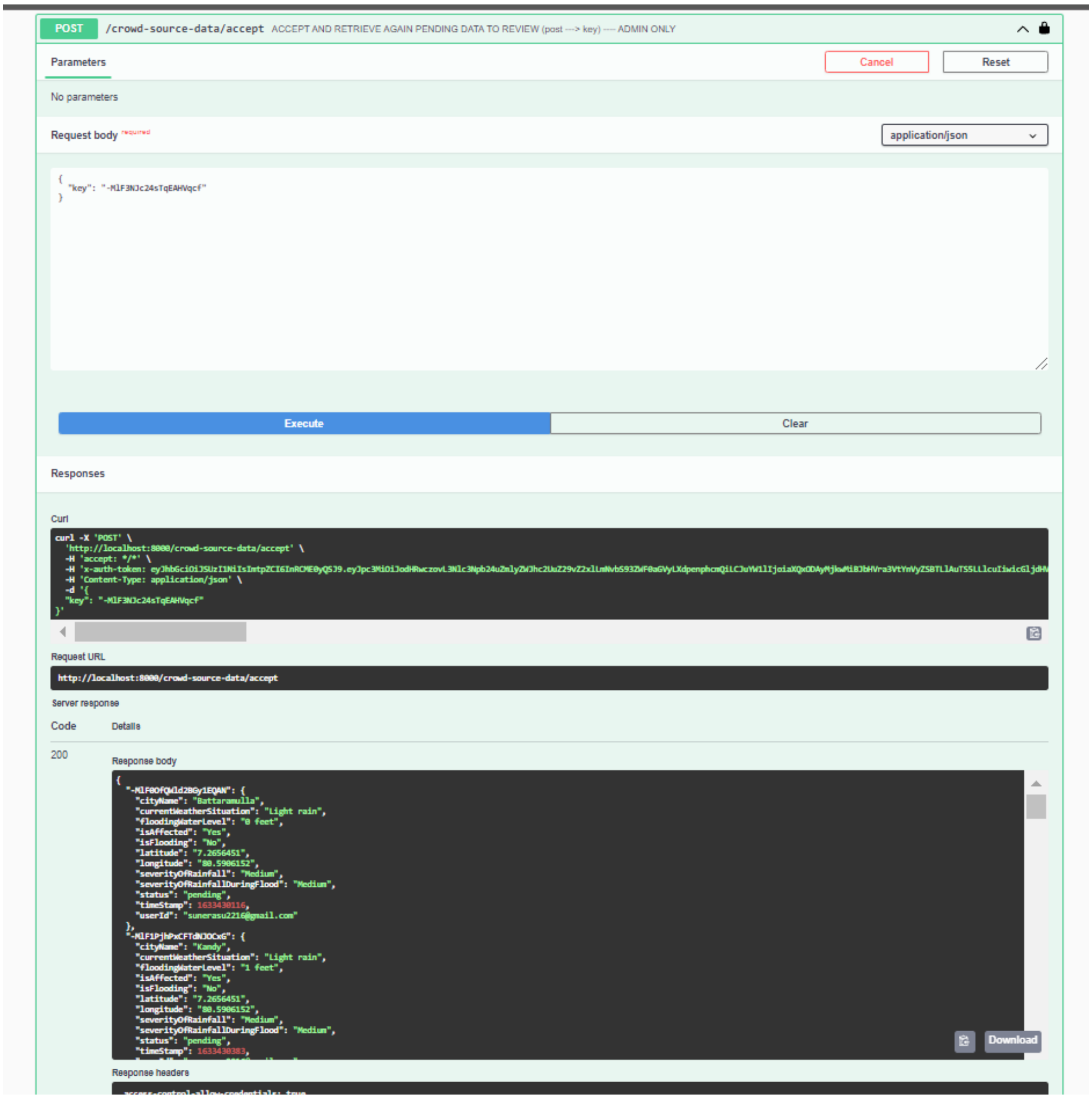


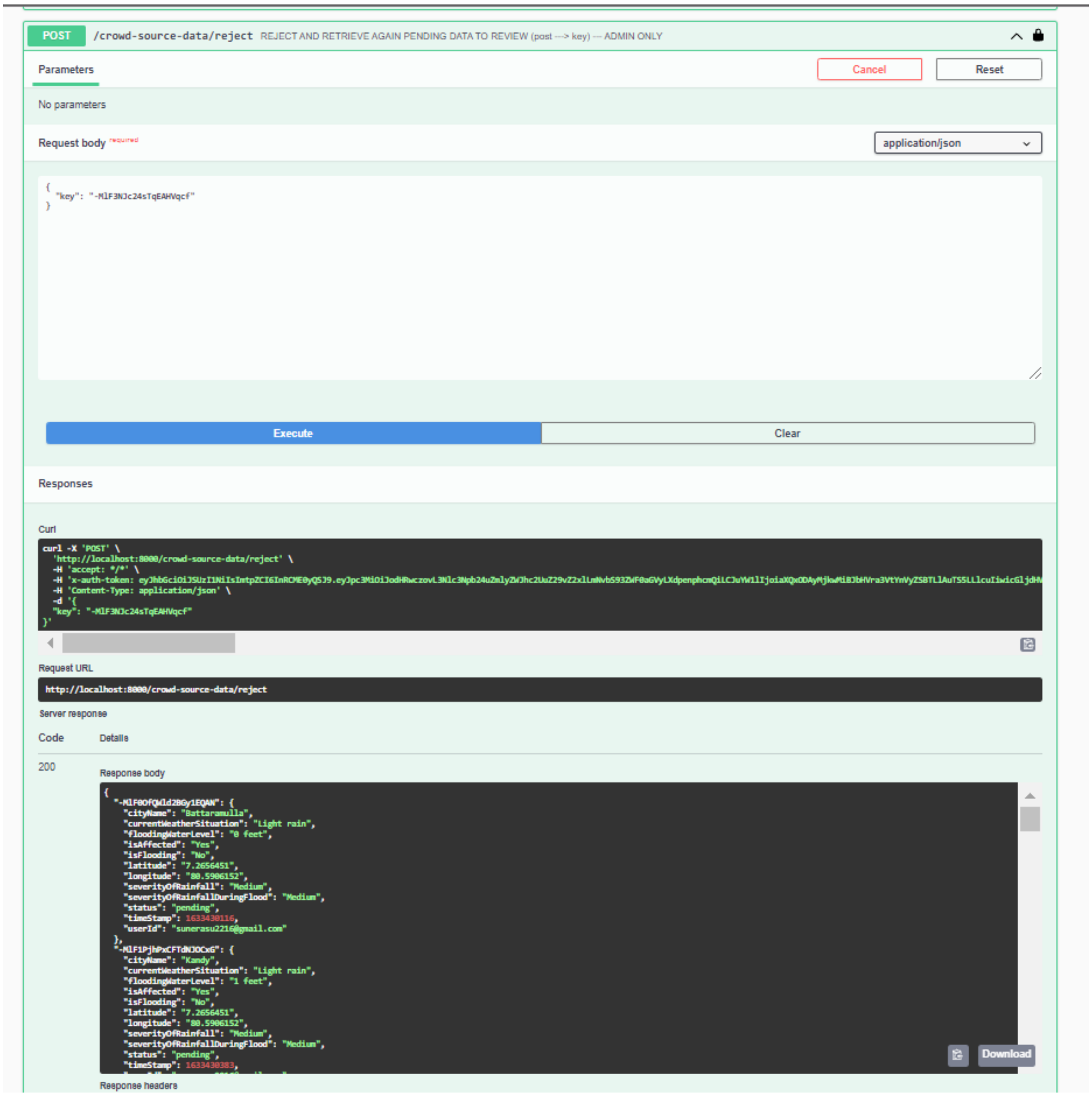

CROWD-SOURCE Crowd source data
GET /crowd-source-data/get-all-pending GET ALL PENDING CROWD SOURCE DATA --- ADMIN ONLY
GET /crowd-source-data/get-all-accepted GET ALL ACCEPTED CROWD SOURCE DATA --- NORMAL USER
GET /crowd-source-data/get-all-rejected GET ALL REJECTED CROWD SOURCE DATA --- ADMIN ONLY
POST /crowd-source-data/accept ACCEPT AND RETRIEVE AGAIN PENDING DATA TO REVIEW (post ---> key) ---- ADMIN ONLY
POST /crowd-source-data/reject REJECT AND RETRIEVE AGAIN PENDING DATA TO REVIEW (post ---> key) --- ADMIN ONLY
POST /crowd-source-data/add Add Data to CrowdSource (If success -> return status='success' | else -> return 500 'Internal server error') --- NORMAL USER
Historic-Date Historic data
GET /historical-data/water/ Get historical data from all stations | request --- > stationId (integer) | response--> relevent data from historical data) | page1.html | limit to last 500 results --- ANY USER
GET /historical-data/water/{stationName} Get historical data from specific station | request --- > stationId (integer) | response--> relevent data from historical data) | page1.html | limit to last 500 results --- ANY USER
GET /historical-data/rain/{stationId}/{lat}/{long} Get historical data from specific station | request --- > stationId (integer) | response--> relevent data from historical data) | page1.html | limit to last 500 results --- ANY USER

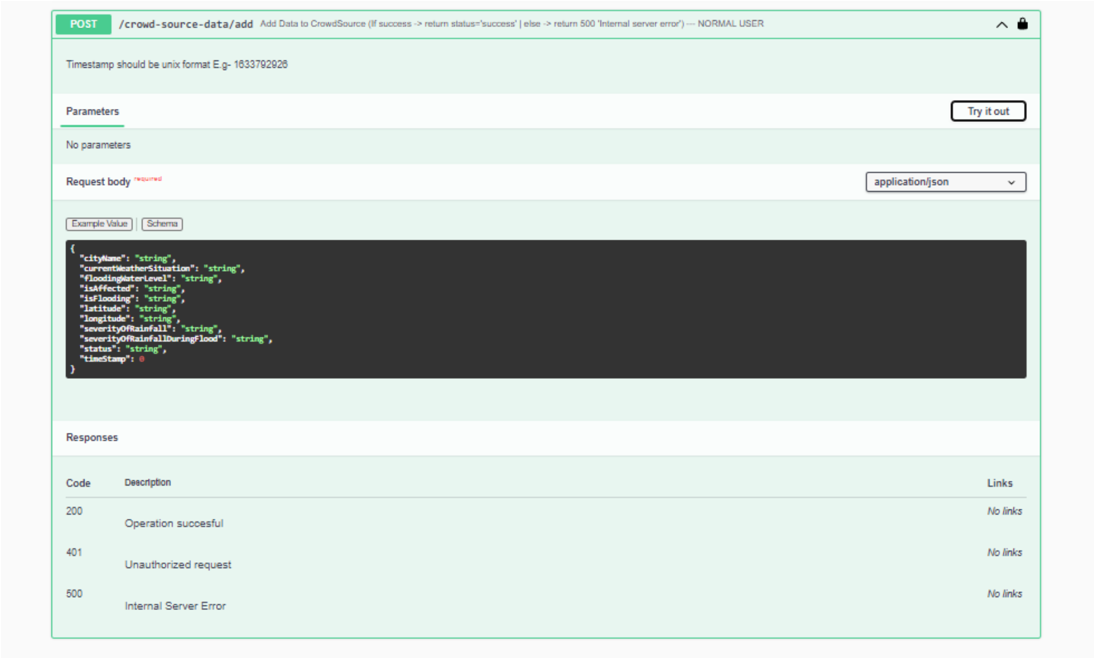

CROWD-SOURCE Crowd source data
GET /crowd-source-data/get-all-pending GET ALL PENDING CROWD SOURCE DATA --- ADMIN ONLY
Parameters
Cancel
No parameters
Execute
Clear
Responses
Curl
curl -X 'GET' \
'http://localhost:8000/crowd-source-data/get-all-pending' \
-H 'accept: */*' \
Request URL
http://localhost:8000/crowd-source-data/get-all-pending
Server response
Code
Details
200
Response body
{
"-MlF0OfQWld2BGy1EQAN": {
"cityName": "Battaramulla",
"currentWeatherSituation": "Light rain",
"floodingWaterLevel": "0 feet",
"isAffected": "Yes",
"isFlooding": "No",
"latitude": "7.2656451",
"longitude": "80.5906152",
"severityOfRainfall": "Medium",
"severityOfRainfallDuringFlood": "Medium",
"status": "pending",
"timeStamp": 1633430116,
"userId": "sunerasu2216@gmail.com"
},
"-MlF1PjhPxCFTdNJOCxG": {
"cityName": "Kandy",
"currentWeatherSituation": "Light rain",
"floodingWaterLevel": "1 feet",
"isAffected": "Yes",
"isFlooding": "No",
"latitude": "7.2656451",
"longitude": "80.5906152",
"severityOfRainfall": "Medium",
"severityOfRainfallDuringFlood": "Medium",
"status": "pending",
"timeStamp": 1633430383,
Download
Response headers
access-control-allow-credentials: true
content-length: 113734
content-type: application/json; charset=utf-8
date: Wed,13 Oct 2021 11:56:53 GMT
etag: W/"1bc46-HPYPWWa2VHapu8bZN8WDFjvKf0w"
vary: Origin
x-powered-by: Express
Responses
Code
Description
Links
200
Operation succesful
No links
401
Unauthorized request
No links
500
Internal Server Error
No links

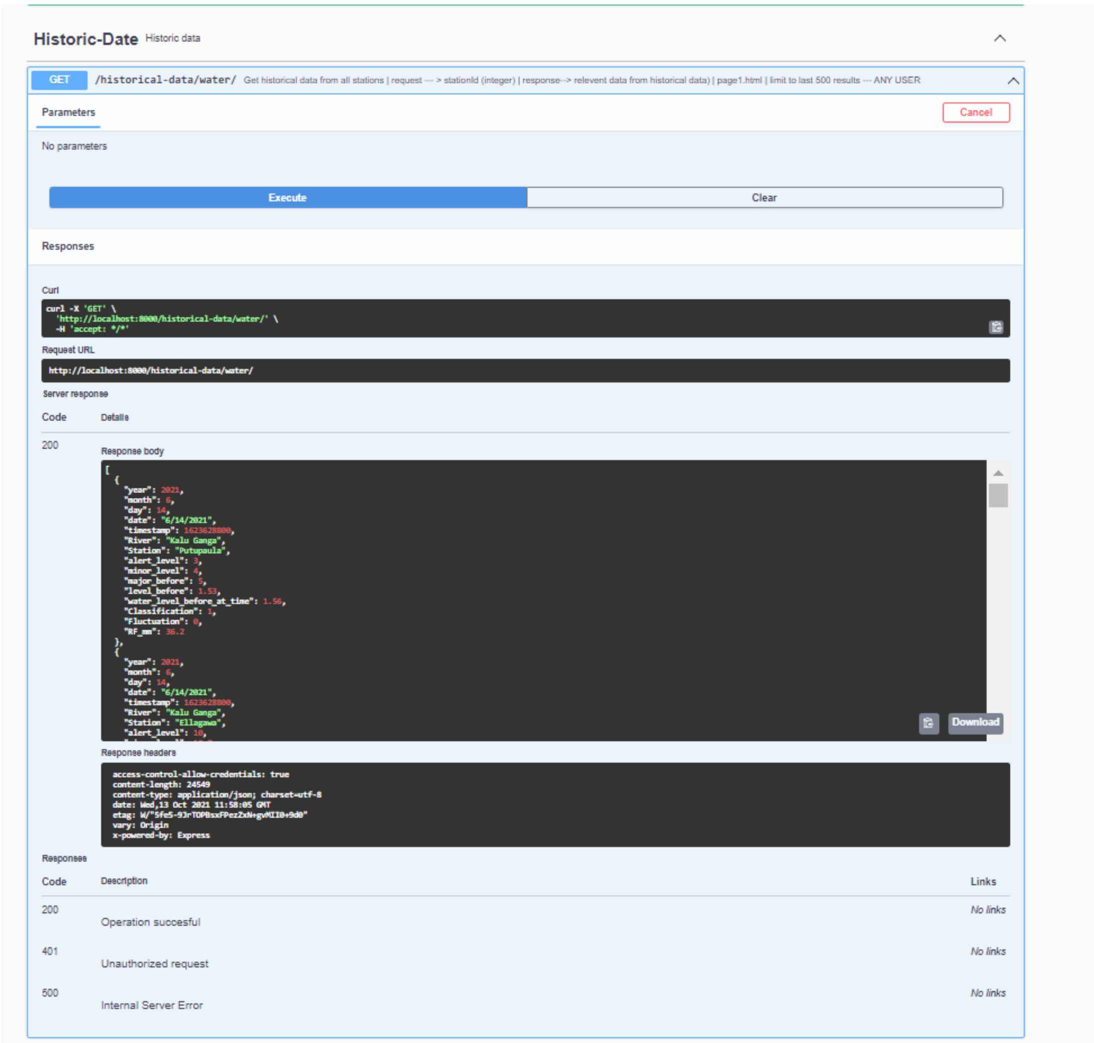

GET
/crowd-source-data/get-all-accepted
GET ALL ACCEPTED CROWD SOURCE DATA — NORMAL USER
Parameters
Cancel
No parameters
Execute
Clear
Responses
Curl
curl -X 'GET' \
'http://localhost:8000/crowd-source-data/get-all-accepted' \
-H 'accept: */*' \
Request URL
http://localhost:8000/crowd-source-data/get-all-accepted
Server response
Code
Details
200
Response body
{
"-Ml0iOJiVEpPVdxhVkVR": {
"cityName": "Kandy",
"currentWeatherSituation": "Overcast clouds",
"floodingWaterLevel": "0 feet",
"isAffected": "Yes",
"isFlooding": "No",
"latitude": "7.2570639",
"longitude": "80.5872974",
"severityOfRainfall": "Medium",
"severityOfRainfallDuringFlood": "Medium",
"status": "accepted",
"timeStamp": 1633190252,
"userId": "sunerasu2216@gmail.com"
},
"-MlDu3-9EgyNAeWIX-b5": {
"cityName": "Agalawatta",
"currentWeatherSituation": "Clear sky",
"floodingWaterLevel": "0 feet",
"isAffected": "No",
"isFlooding": "No",
"latitude": "6.527814977233106",
"longitude": "80.15080955437251",
"severityOfRainfall": "Low",
"severityOfRainfallDuringFlood": "Low",
"status": "accepted",
"timeStamp": 1633411415,
Download
Response headers
access-control-allow-credentials: true
connection: keep-alive
content-length: 2132
content-type: application/json; charset=utf-8
date: Wed,13 Oct 2021 12:00:14 GMT
etag: W/"854-H8SXE6UIf66Df13x48i7Ej1v3WA"
keep-alive: timeout=5
vary: Origin
x-powered-by: Express
Responses
Code
Description
Links
200
Operation succesful
No links
401
Unauthorized request
No links
500
Internal Server Error
No links

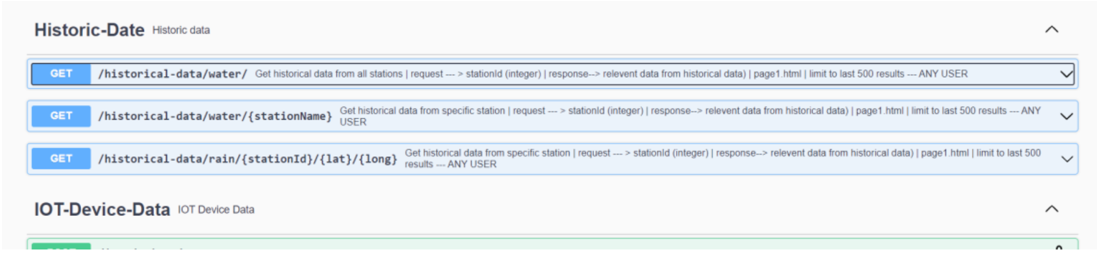

GET
/crowd-source-data/get-all-rejected
GET ALL REJECTED CROWD SOURCE DATA --- ADMIN ONLY
Parameters
Cancel
No parameters
Execute
Clear
Responses
Curl
curl -X 'GET' \
'http://localhost:8000/crowd-source-data/get-all-rejected' \
-H 'accept: */*' \
-H 'x-auth-token: eyJhbGciOiJSUzI1NiIsImtpZCI6InRCMEEyQSJ9.eyJpc3MiOiJodHRwczovL3Nlc3Npb24uZmlyZWJhc2UuZ29vZ2xlLmNvbS93ZWF0aGVyLXdpenphcmQiLCJuYW1lIjoiaXQxODAyMjkwMiBJbHVra3VtYmVyZSBTLlAuTS5LLlcuIiwicGljdHV
Request URL
http://localhost:8000/crowd-source-data/get-all-rejected
Server response
Code
Details
200
Response body
{
"-MlF3NJc24sTqEAHWqcf": {
"cityName": "Maharagama",
"currentWeatherSituation": "Light rain",
"floodingWaterLevel": "0 feet",
"isAffected": "Yes",
"isFlooding": "No",
"latitude": "7.2656451",
"longitude": "80.5906152",
"severityOfRainfall": "Low",
"severityOfRainfallDuringFlood": "Low",
"status": "rejected",
"timeStamp": 1633430897,
"userId": "sunerasu2216@gmail.com"
}
}
Download
Response headers
access-control-allow-credentials: true
connection: keep-alive
content-length: 346
content-type: application/json; charset=utf-8
date: Wed,13 Oct 2021 12:00:41 GMT
etag: W/"15a-AhTuBVusCKfUSN8ueTWSspw/DFg"
keep-alive: timeout=5
vary: Origin
x-powered-by: Express
Responses
Code
Description
Links
200
Operation succesful
No links
401
Unauthorized request
No links
500
Internal Server Error
No links

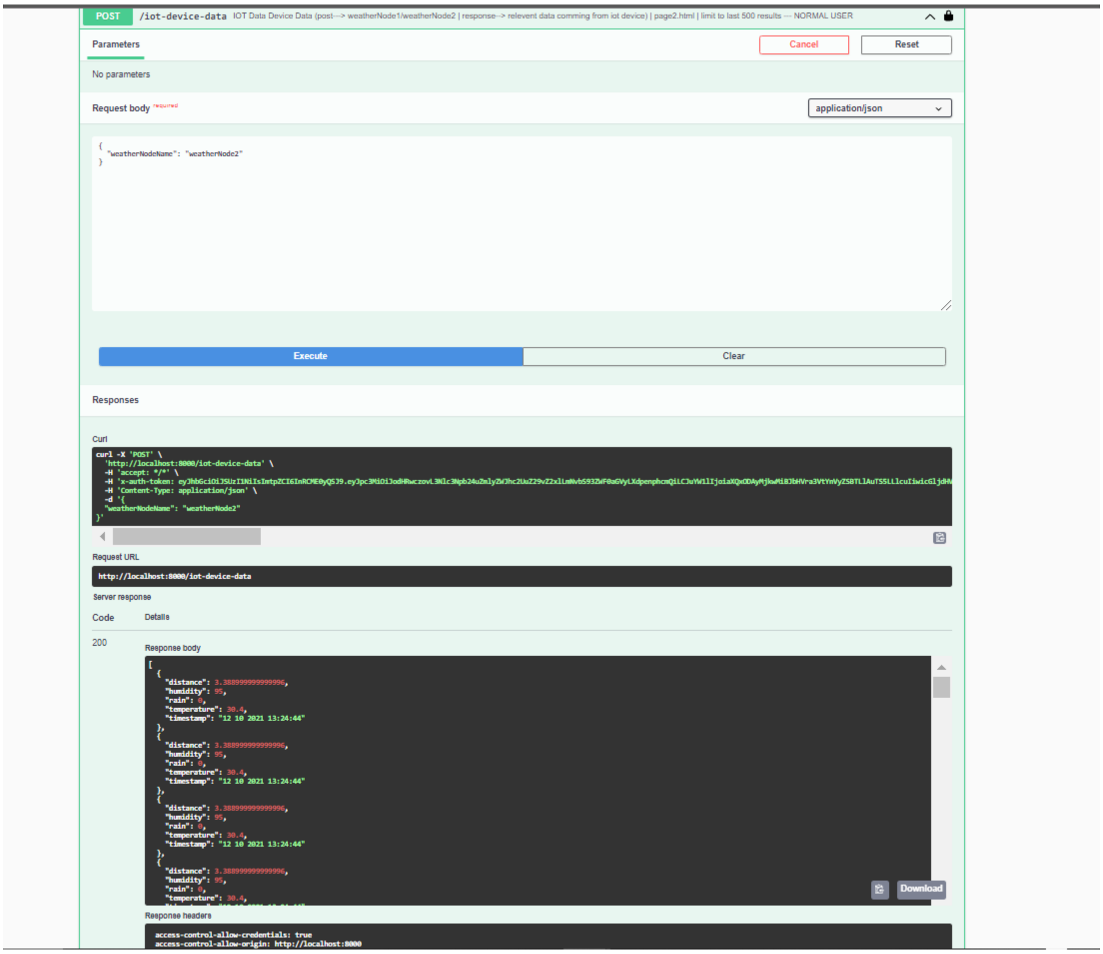

POST
/crowd-source-data/accept ACCEPT AND RETRIEVE AGAIN PENDING DATA TO REVIEW (post ---> key) ---- ADMIN ONLY
Parameters
Cancel
Reset
No parameters
Request body required
application/json
{
"key": "-MlF3NJc24sTqEAHVqcf"
}
Execute
Clear
Responses
Curl
curl -X 'POST' \
'http://localhost:8000/crowd-source-data/accept' \
-H 'accept: */*' \
-H 'Content-Type: application/json' \
-d '{
"key": "-MlF3NJc24sTqEAHVqcf"
}'
Request URL
http://localhost:8000/crowd-source-data/accept
Server response
Code
Details
200
Response body
{
"-MlF0OfQWld2BGy1EQAN": {
"cityName": "Battaramulla",
"currentWeatherSituation": "Light rain",
"floodingWaterLevel": "0 feet",
"isAffected": "Yes",
"isFlooding": "No",
"latitude": "7.2656451",
"longitude": "80.5906152",
"severityOfRainfall": "Medium",
"severityOfRainfallDuringFlood": "Medium",
"status": "pending",
"timeStamp": 1633430116,
"userId": "sunerasu2216@gmail.com"
},
"-MlF1PjhPxCFTdNJOCxG": {
"cityName": "Kandy",
"currentWeatherSituation": "Light rain",
"floodingWaterLevel": "1 feet",
"isAffected": "Yes",
"isFlooding": "No",
"latitude": "7.2656451",
"longitude": "80.5906152",
"severityOfRainfall": "Medium",
"severityOfRainfallDuringFlood": "Medium",
"status": "pending",
"timeStamp": 1633430383,
Download
Response headers


POST
/crowd-source-data/reject
REJECT AND RETRIEVE AGAIN PENDING DATA TO REVIEW (post --> key) -- ADMIN ONLY
Parameters
Cancel
Reset
No parameters
Request body
required
application/json
{
"key": "-MlF3NJc24sTqEAHVqcf"
}
Execute
Clear
Responses
Curl
curl -X 'POST' \
'http://localhost:8000/crowd-source-data/reject' \
-H 'accept: */*' \
-H 'x-auth-token: eyJhbGciOiJSUzI1NiIsImtpZCI6InRCMEEyQSJ9.eyJpc3MiOiJodHRwczovL3Nlc3Npb24uZmlyZWJhc2UuZ29vZ2xlLmNvbS93ZWF0aGVyLXdpenphcmQiLCJuYW1lIjoiaXQxODAyMjkwMiBJbHVra3VtYmVyZSBTLlAuTS5LLlcuIiwicGljdHV
-H 'Content-Type: application/json' \
-d '{
"key": "-MlF3NJc24sTqEAHVqcf"
}'
Request URL
http://localhost:8000/crowd-source-data/reject
Server response
Code
Details
200
Response body
{
"-MlF0OfQWld2BGy1EQAN": {
"cityName": "Battaramulla",
"currentWeatherSituation": "Light rain",
"floodingWaterLevel": "0 feet",
"isAffected": "Yes",
"isFlooding": "No",
"latitude": "7.2656451",
"longitude": "80.5906152",
"severityOfRainfall": "Medium",
"severityOfRainfallDuringFlood": "Medium",
"status": "pending",
"timeStamp": 1633430116,
"userId": "sunerasu2216@gmail.com"
},
"-MlF1PjhPxCFTdNJOCxG": {
"cityName": "Kandy",
"currentWeatherSituation": "Light rain",
"floodingWaterLevel": "1 feet",
"isAffected": "Yes",
"isFlooding": "No",
"latitude": "7.2656451",
"longitude": "80.5906152",
"severityOfRainfall": "Medium",
"severityOfRainfallDuringFlood": "Medium",
"status": "pending",
"timeStamp": 1633430383,
Download
Response headers


POST
/crowd-source-data/add
Add Data to CrowdSource (If success -> return status='success' | else -> return 500 'Internal server error') --- NORMAL USER
Timestamp should be unix format E.g- 1633792926
Parameters
Try it out
No parameters
Request body
required
application/json
Example Value
Schema
{
"cityName": "string",
"currentWeatherSituation": "string",
"floodingWaterLevel": "string",
"isAffected": "string",
"isFlooding": "string",
"latitude": "string",
"longitude": "string",
"severityOfRainfall": "string",
"severityOfRainfallDuringFlood": "string",
"status": "string",
"timeStamp": 0
}
Responses
Code
Description
Links
200
Operation succesful
No links
401
Unauthorized request
No links
500
Internal Server Error
No links


Historic-Date
Historic data
GET
/historical-data/water/
Get historical data from all stations | request --- > stationId (integer) | response--> relevent data from historical data) | page1.html | limit to last 500 results --- ANY USER
Parameters
Cancel
No parameters
Execute
Clear
Responses
Curl
curl -X 'GET' \
'http://localhost:8000/historical-data/water/' \
-H 'accept: */*'
Request URL
http://localhost:8000/historical-data/water/
Server response
Code
Details
200
Response body
[
{
"year": 2021,
"month": 6,
"day": 14,
"date": "6/14/2021",
"timestamp": 1623628800,
"River": "Kalu Ganga",
"Station": "Putupaula",
"alert_level": 3,
"minor_level": 4,
"major_before": 5,
"level_before": 1.53,
"water_level_before_at_time": 1.56,
"Classification": 1,
"Fluctuation": 0,
"RF_mm": 36.2
},
{
"year": 2021,
"month": 6,
"day": 14,
"date": "6/14/2021",
"timestamp": 1623628800,
"River": "Kalu Ganga",
"Station": "Ellagawa",
"alert_level": 10,
Download
Response headers
access-control-allow-credentials: true
content-length: 24549
content-type: application/json; charset=utf-8
date: Wed,13 Oct 2021 11:58:05 GMT
etag: W/"5fe5-9JrTOPBsxFPezZxN+gvMII0+9d0"
vary: Origin
x-powered-by: Express
Responses
Code
Description
Links
200
Operation succesful
No links
401
Unauthorized request
No links
500
Internal Server Error
No links


Historic-Date Historic data
GET /historical-data/water/ Get historical data from all stations | request --- > stationId (integer) | response--> relevent data from historical data) | page1.html | limit to last 500 results --- ANY USER
GET /historical-data/water/{stationName} Get historical data from specific station | request --- > stationId (integer) | response--> relevent data from historical data) | page1.html | limit to last 500 results --- ANY USER
GET /historical-data/rain/{stationId}/{lat}/{long} Get historical data from specific station | request --- > stationId (integer) | response--> relevent data from historical data) | page1.html | limit to last 500 results --- ANY USER
IOT-Device-Data IOT Device Data


POST /iot-device-data
Parameters
Cancel
Reset
No parameters
Request body
application/json
Execute
Clear
Responses
Curl
Request URL
http://localhost:8000/iot-device-data
Server response
Code
Details
200
Response body
Download
Response headers