Assignment 2

Unit: KIT717 Internet of Things and Web Applications

MD TANVIR RAHMAN
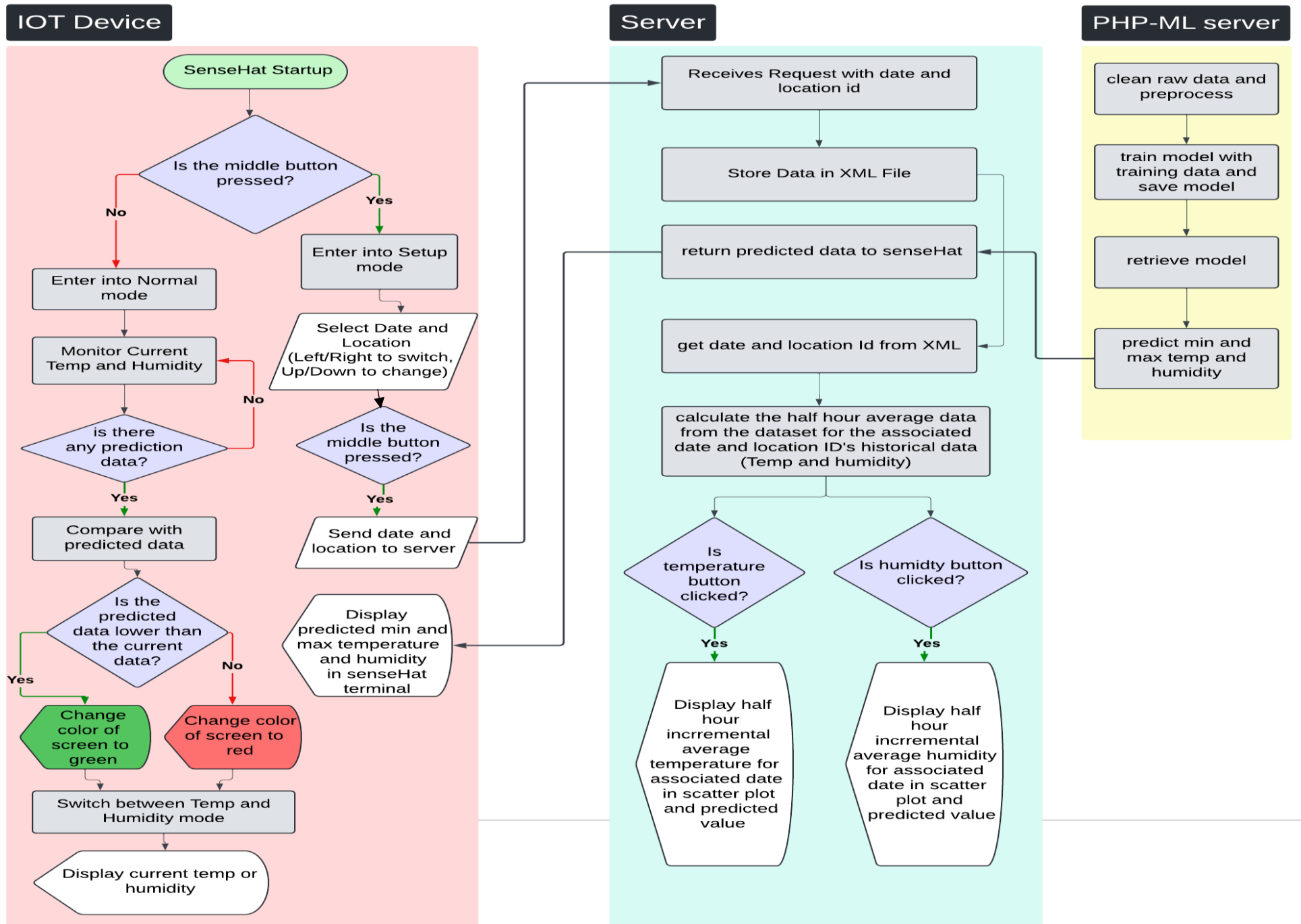
Student ID: 663221

Table of Contents

Flow Chart Diagram:

Diagram explaining data flow in the system

IOT Device

SenseHat Startup

Is the middle button pressed?

No → Enter into Normal mode

Yes → Enter into Setup mode

Monitor Current Temp and Humidity

is there any prediction data?

No → (back to Monitor Current Temp and Humidity)

Yes → Compare with predicted data

Select Date and Location (Left/Right to switch, Up/Down to change)

Is the middle button pressed?

No → (back to Select Date and Location)

Yes → Send date and location to server

Is the predicted data lower than the current data?

Yes → Change color of screen to green

No → Change color of screen to red

Switch between Temp and Humidity mode

Display current temp or humidity

Display predicted min and max temperature and humidity in senseHat terminal

Server

Receives Request with date and location id

Store Data in XML File

return predicted data to senseHat

get date and location Id from XML

calculate the half hour average data from the dataset for the associated date and location ID's historical data (Temp and humidity)

Is temperature button clicked?

Yes → Display half hour incremental average temperature for associated date in scatter plot and predicted value

Is humidty button clicked?

Yes → Display half hour incremental average humidity for associated date in scatter plot and predicted value

PHP-ML server

clean raw data and preprocess

train model with training data and save model

retrieve model

predict min and max temp and humidity

Explanation of Machine Learning Algorithm Choice

For this project, I selected supervised linear regression as the machine learning algorithm, based on various significant factors. The decision was influenced by the advantages it offers in the realm of weather forecasting. Supervised learning establishes a systematic framework where the algorithm learns from labeled data, enabling it to forecast based on past observations. Linear regression, a fundamental technique in supervised learning, is well-suited for this purpose due to its simplicity and ease of interpretation. By establishing a linear relationship between input features (date, location, and time) and target variables (temperature and humidity), linear regression allows for a clear understanding of how each feature impacts the predicted outcome. Additionally, supervised learning promotes adaptability by continuously adjusting the model's parameters to reduce prediction errors, thereby improving its predictive accuracy over time. Utilizing these characteristics, supervised linear regression proves to be a dependable tool for predicting temperature and humidity, providing more accuracy into future patterns and trends.

Training Data and Sampling Steps

To train the supervised linear regression model, the necessary training data was sourced from the provided historical weather dataset spanning almost six years for five different sites. This dataset contained records of temperature, humidity, and wind observations recorded every half hour.

The initial dataset likely contained inconsistencies, missing values, or outliers that could impact the model's performance. Moreover, unnecessary columns were present in the dataset, but they were removed during the data cleaning process. The final cleaned dataset consisted of five essential data points: location ID, date-time, humidity, and temperature. Any other labels were eliminated to streamline the dataset.

In certain instances, data points were not recorded at regular 30-minute intervals. To address this issue, the missing data points were averaged with adjacent observations

and integrated into the dataset. Temperature values of 0 were retained as they can legitimately occur and are not considered abnormal.

Additionally, instances where humidity was recorded as 100% were rectified. To tackle this problem, the preceding and subsequent humidity values were averaged, and the resulting value was substituted for the 100% humidity reading. These measures were implemented to ensure the reliability and accuracy of the training data for the machine learning model.

Data cleaning played a crucial role in enhancing the quality and reliability of the dataset. This involved various tasks such as handling missing values, eliminating duplicates, and rectifying errors to ensure that the dataset was suitable for training the machine learning model.

Accuracy of Predictions

To calculate the accuracy of the predictions, historical dates from the dataset will be used, comparing the actual data with the predicted minimum and maximum humidity values. Specifically, the focus will be on the Hobart location, examining the data for January 1st of each year from 2015 to 2017. This process involves extracting the humidity data for January 1st from 2015 to 2017 for Hobart from the dataset and using the machine learning model to predict the minimum and maximum humidity for January 1st of each selected year. By comparing the actual humidity data with the predicted values, the model's accuracy can be assessed. The results will be tabulated, showing the actual and predicted minimum and maximum humidity values for each year. This approach will provide a clear picture of the model's performance and its ability to accurately predict humidity for specific historical dates.

Location Hobart	Actual Data				Predicted Data				Accuracy in 100%
Date	Min Temp	Max Temp	Min Humidity	Max humidity	Min Temp	Max Temp	Min Humidity	Max humidity	
2015-01-01	7.9	22.6	32	95	7.7	17.2	66.5	92	**44.24%**
2016-01-01	14.9	23.8	56	94	7.7	17.2	66.4	91.9	**34.23%**

The overall accuracy of the prediction is 39.24%.

Scope of Improvement:

Enhancing the accuracy of weather predictions is possible through several key strategies. One approach involves feature engineering - exploring additional relevant factors like wind speed, atmospheric pressure, and cloud cover to gain a more comprehensive understanding of weather patterns. Advanced machine learning algorithms, such as decision trees, random forests, and neural networks, offer another avenue for improvement by capturing complex nonlinear relationships in the data.

Additionally, increasing dataset size, continuous monitoring, and feedback mechanisms can collectively refine the accuracy and reliability of weather prediction models. By adopting these multifaceted strategies, future weather prediction models can deliver more precise predictions of temperature and humidity.