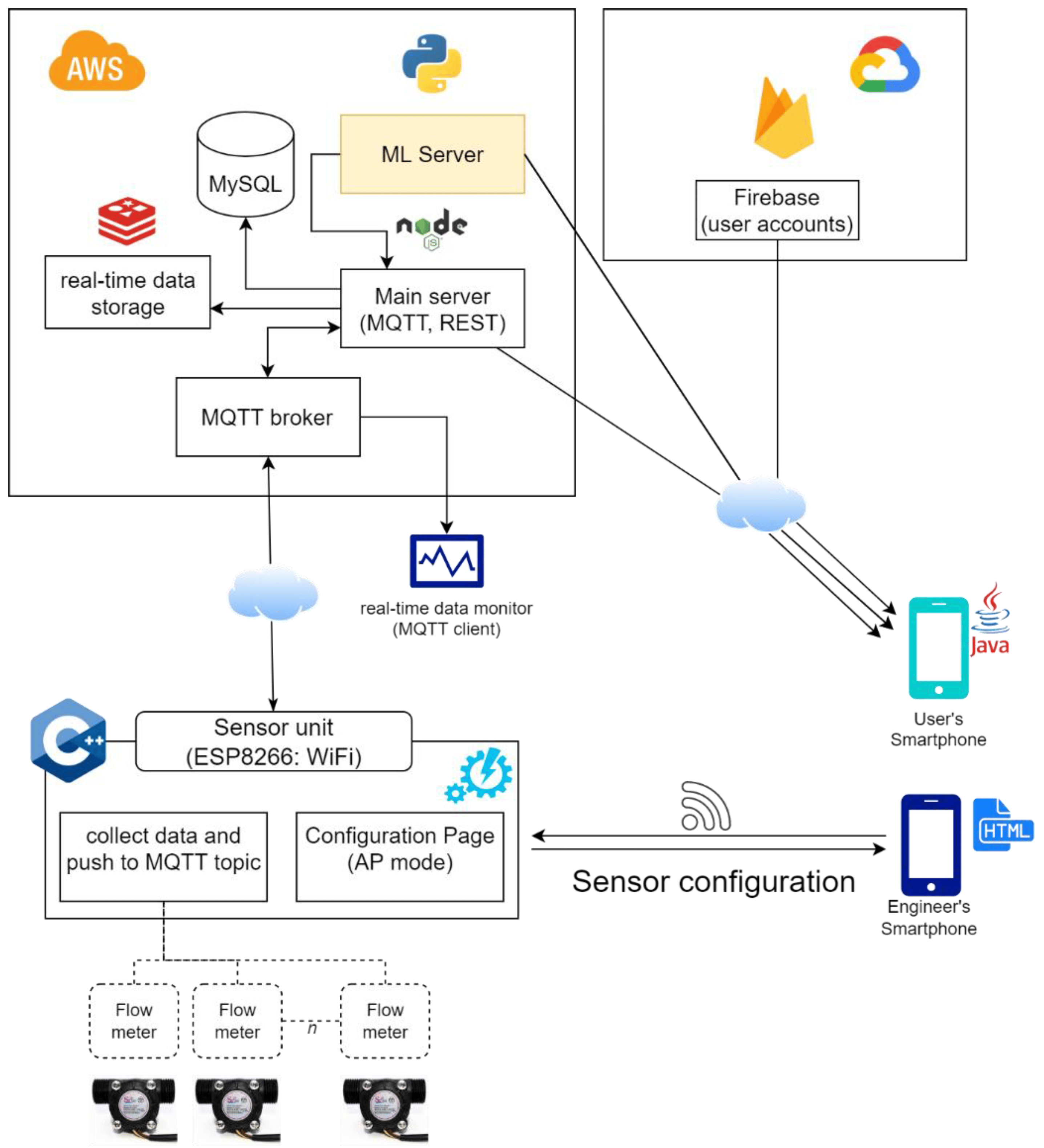
#### System Architecture

An intelligent water network management system is most effective when several factors are considered simultaneously, i.e., water suppliers, decision makers, and the direct involvement of consumers.

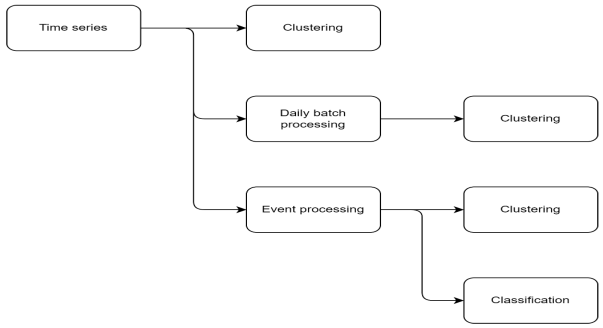
For efficient water consumption data collection, the sensor network is integrated into a cloud-based architecture as shown in [**Figure 1**](https://www.mdpi.com/2073-4441/14/14/2187#fig_body_display_water-14-02187-f001). Data acquisition is performed using a NodeMCU development board based on the ESP8266 microcontroller, with Wi-Fi communication. The platform has multiple GPIO pins (general-purpose input/output) connected to several YF-S201 flow meters and can be programmed using the Arduino environment [[**44**](https://www.mdpi.com/2073-4441/14/14/2187#B44-water-14-02187)] to monitor and collect the water flow through several pipes. The acquisition module is pre-programmed before installation, providing a configurable interface for connecting to the local network and defining a variable number of flow meters attached. To display the required configuration data, the sensor interface queries the server that is hosted directly on the ESP8266 microcontroller, which is configured in both station and Wi-Fi client modes.



**Figure 1.** The water consumption monitoring system architecture

#### Data Processing Pipeline

The processing pipeline is shown in Figure 2. For a better understanding, each step will be described below. The raw data set (time series) was generated by collecting data from sensors installed in multiple households. Four types of water outlets were considered, i.e., sink (cold and hot water measured separately), toilet, and shower. The raw data set were processed to be evaluated and tested by first eliminating nonrelevant data that could negatively impact the results.



**Figure 2.** The processing pipeline.

**Experimental Results**

This section describes the results obtained using the proposed methodology. The consumption profiles are evaluated using clustering methods, showing weekly and daily consumption patterns. Consumption events are processed and evaluated using clustering and classification methods to provide an additional level of detail.

#### Consumption Data Evaluation

The data collected from a multisensor node are shown in [**Figure 3**](https://www.mdpi.com/2073-4441/14/14/2187#fig_body_display_water-14-02187-f003)a, showing water consumption from 4 types of water outlets: sink cold (cold tap water), sink hot (hot tap water), toilet, and shower, from 33 sources, over a time frame of 1 week, with a sampling time of 60 s. The instantaneous flow is measured by the sensors and represented in L/h, while the individual samples are represented on the x-axis.

**Discussion**

As presented in the results section, the results are promising in terms of predicting and identifying consumer outlet types of the four measurement points (hot water sink, cold water sink, toilet, and shower). Data were collected from 33 sources representing various water outlets for one week with a sampling time of 60 s, which allowed for a high level of detail in terms of consumption patterns and events.

In the first stage, the K-means clustering algorithm and evaluation metrics were applied to observe the consumption patterns in terms of variability. Due to a moderate similarity between the extracted clusters and the consumer classes, we extracted daily consumption patterns for a better understanding of the consumers. During this stage, clustering on daily consumption patterns and evaluation metrics was applied and showed better results in terms of similarity, which was higher for the filtered data set.

For the better understanding of the daily consumption patterns, the time series were extracted from the raw data set, while moving average filtering was applied for a better overview. Next, K-means clustering was applied to extract the daily consumption patterns. For a numerical evaluation of the results obtained by filtering and daily consumption, the five metrics were applied again, showing better scores on the filtered data set compared with the raw data set.

Given that the two approaches offer different results in terms of extracting consumption patterns, a graph was made containing the evaluation metrics, where the relative score for each method was calculated based on the deviation from the average score. It turned out that the clustering method applied on the filtered data set offers the best results.

Going in depth with the analysis stages, the individual consumption events were also extracted from the time series. An initial limitation was caused by the separation of hot and cold tap water, and the accuracy was not so encouraging. This is because consumers’ preferences can be very varied in terms of water temperature. As a solution for the sink data to be more consistent, and for the study to be relevant, the hot water tap and cold water tap were combined, making it easier to evaluate the entire sink data set. The results show moderate variability for sink events and high variability for shower events, while the toilet registers small variability; toilet water consumption is mostly constant because the water tank has the same volume each time and requires the same period of time to be refilled.

Next, the clustering of the events extracted from the time series was performed using three clusters corresponding to the three consumer outlet types, i.e., sink, shower, and toilet, to analyze the variability. Three levels of variability could be visualized that matched the characteristics of the three consumption types; we concluded that this scenario is the most suitable for evaluating consumption events on a data set. Furthermore, the evaluation metrics showed a high similarity between the clustering assignments and the known water outlets.

The final evaluation involved testing the machine learning and deep learning algorithms to predict consumption events. For this stage, only two of four events were chosen, i.e., hot water and cold water events taken together and toilet events, these being the most consistent in terms of variability. The machine learning models offer comparably higher accuracy for training and lower accuracy for testing when compared with the deep learning models, which have higher accuracy for prediction.

Therefore, the main objective of the supervised learning in this study was to train the classification models to predict between sink and toilet consumption events from the water consumption data set, which accounted for 9414 consumption events. An additional balancing of the data set resulted in 4785 consumption events, which were used for training the classification models.

**Conclusions**

Monitoring water consumption is essential today for predicting leaks and eliminating water waste that can lead to water scarcity. The current paper proposes the evaluation of several methods for predicting water at consumer outlet types: hot tap water sink, cold tap water sink, toilet, and shower.

The proposed architecture is based on scalable components and requires minimal configuration, which makes it possible to extend the study to multiple households. The proposed configuration using separate sensors for each consumption outlet allowed for evaluating different methods for characterizing consumer behavior in great detail.

Moreover, the same methods can be applied to extract the consumption patterns from combined measurements, i.e., using sensors installed at the mains, and predict the consumption activities based on the overall water consumption data. Possible extensions include location-based clustering and demographic analysis in large-scale deployments.

In contrast with other studies in this field, the four types of household activities were analyzed in this study, applying clustering, classification methods, and evaluation metrics. In the case of clustering methods, high accuracy was obtained by extracting the consumption events from the time series, as confirmed by the evaluation metrics. In the case of classification methods, using both machine learning algorithms and deep learning algorithms, good results were achieved in terms of prediction accuracy.

As in any study, there were some limitations. First, the solution could be extended to more households and the installation of additional sensors to improve the accuracy of the results. Second, the installation process proved to be quite tedious due to the wiring of the flow sensors, which have to be connected to the IoT boards. Another limitation is the high volume of data that is collected regardless of the actual water consumption, which adversely impacts the data access.

As future work, the results can be used to define a decision support system that achieves multiple objectives for improved efficiency of water resource management, i.e., shaping consumption patterns to maintain a constant flow, which can be especially beneficial for central heating systems; showing consumption statistics for household consumers; comparing consumption events between households; and promoting consumer involvement in regulating water flow and overall resource consumption.