Forecasting Water Consumption in Residences of Catmon, Malolos City: A Predictive Modeling Approach

ABSTRACT

This study addressed the water shortage crisis in Catmon, Malolos City. The researchers presented a water consumption forecasting system designed for PrimeWater Infrastructure Corp. in Barangay Catmon, Malolos City. Leveraging historical data from January 2018 to December 2022, the research utilizes data mining and specifically employed the PROPHET algorithm for predicting water consumption trends in residential and commercial areas. The resulting dashboard system offers actionable insights for effective water resource management, highlighting the crucial role of accurate forecasting in addressing the expected surge in water demand. The study emphasizes the reliability of the PROPHET algorithm in capturing seasonal trends, thereby supporting well-informed decision-making as it favored the PROPHET algorithm for water consumption forecasting due to its exceptional performance metrics, including a low Mean Absolute Percentage Error (MAPE), minimized Mean Absolute Error (MAE), and normalized performance metrics. The study's integration of the Knowledge Discovery in Databases (KDD) methodology and the Agile framework ensures a systematic and flexible approach throughout the research process. Recommendations include providing uniform file formats, broadening study scopes, implementing error detection mechanisms, incorporating additional attributes, and maintaining continuous data monitoring for improved forecasting accuracy and sustainable water management practices.

CCS Concepts

• Computing methodologies → Machine learning → Time series analysis

Keywords:

Algorithms; Visual analytics; Forecasting; Prophet; Dashboard; Water Consumption

1. INTRODUCTION

With nearly 2.3 billion people experiencing water-related stress and millions experiencing acute water shortages, the global water crisis is an urgent problem that impacts billions of people globally (UN-Water, 2021; FAO, 2020). With the wide-ranging effects this problem is having on businesses, communities, and ecosystems, comprehensive solutions are needed to guarantee a secure and sustainable water supply. In this environment, regional water-related issues have gained prominence. President Ferdinand Marcos Jr. recently raised the alarm in the Philippines, where countless families are coping with an impending drought as the dry season draws near (Cepeda, 2023). This is an illustration of a global issue manifesting as a regional emergency. It demonstrates how critical it is to address water scarcity issues in a variety of methods as soon as possible.

Water-related concerns are particularly serious in the Philippines. Bulacan is one such locality experiencing water supply challenges, with dropping water levels at Angat Dam raising worries about Metro Manila's and adjacent provinces' water supplies, including Bulacan and Pampanga. (From the Philippine Daily Inquirer, 2022). According to the report, Malolos, Bulacan, too had water scarcity issues that affected several regions of the city. Understanding and solving these

water concerns on several levels is critical to establishing effective solutions and guaranteeing water security.
The progress of technology is crucial in reducing this situation. Predictive analytics is one such technical area that has gained traction. Predictive analytics forecasts future occurrences using past information, machine learning methods, and statistical approaches. This technology has proven beneficial in a variety of industries, promising to improve resource management, optimize operations, and solve crucial concerns such as water scarcity.

Predictive analytics, a type of data analytics, shows enormous promise for addressing complicated situations like water scarcity. Its primary duty is to create precise forecasts about future occurrences based on previous data. Predictive analytics is being seen as a game changer in a variety of industries, from banking to healthcare (Halton, 2023). In healthcare, for example, it assists in anticipating disease outbreaks and patient cases, allowing for early treatments and resource allocation. Predictive analytics may be used in water management to estimate water use trends, improve resource distribution, and avert shortages.

The PrimeWater Infrastructure Corp., located in Malolos City, is the client for this case study. PrimeWater is responsible for a variety of tasks, including the design of water distribution systems, the development of water sources, the construction of transmission and distribution lines, the administration of water supply, and the supervision of facilities for septage management. The primary objective of PrimeWater, which is to supply communities in the Philippines with water that is drinkable, dependable, and sustainable, is becoming more apparent as the company embarks on increasing its reach and initiatives.

According to the data that PrimeWater has given, it has been observed that Barangay Catmon represents one of the locations that is greatly impacted by water shortage. This adds urgency to the need for an effective solution, especially increasing the amount of water that customers use in order to fulfill their demand. The reason for this is that they have been obtaining a large number of inquiries from their clients, all of which concerned the water pressure issues that they have been experiencing. The implementation of an integrated approach is necessary in order to solve these difficulties and make progress toward averting water shortages. The purpose of this study was to construct a forecasting model for the amount of water that is consumed by residents of Catmon, which is located in Malolos City.

According to PrimeWater, Barangay Catmon is severely water-short, which supports this proposal. Water consumption forecasting helps Prime Water efficiently plan water production, optimize resource allocation, and avoid shortages or excess production, which can waste resources and increase operational costs. Estimating water use helps find leaks and system issues, decreasing water loss and preserving resources. This project developed a water usage forecasting system. Not only can Catmon, Malolos City, use predictive analytics and visualization technologies to estimate water demand, but other water suppliers and towns experiencing similar difficulties may too.

1.1 PROJECT OBJECTIVES

The main objective of this study was to develop an analytics model that would forecast the water consumption of the residences in Catmon, Malolos City.

Specifically, the researchers aim to consider the following objectives: (1) To collect the historical water consumption data from residences in Catmon, Malolos City, spanning five inclusive years, and apply pre-processing techniques to optimize the dataset for building the forecasting model using an algorithm. (2) To apply the most appropriate data mining algorithm to forecast the water consumption in the residential area of Catmon, Malolos. (3) To determine the accuracy and validity of the developed analytics model using the comparison of actual data and forecasted data. (4) To develop a dashboard system that presents the water consumption forecasts and provides insights on water usage patterns to aid water resource management in Catmon, Malolos.

2. PROJECT METHODOLOGY

2.1 Methodology for Analytics Modeling

For the purpose of conducting their data analysis, the researchers that participated in this study utilized the Knowledge Discovery in Databases (KDD) approach. The KDD methodology is a methodical technique that includes the selection of data, preparation of that data, transformations, and the detection of patterns. It acts as the basis for the use of data mining methods, which are used to bring to light previously concealed information within the dataset. The major objective of knowledge discovery and development (KDD) is data mining, which includes the analysis of data, the building of models, and the development of algorithms to uncover hidden patterns. Data cleaning and integration, which removes noise from the data; data selection and transformation, which prepares the data for analysis; data mining, which seeks out patterns and associations; pattern evaluation, which evaluates the value of created patterns; and knowledge representation through visualization tools, which communicates the results to stakeholders. These are the five key steps that make up the process.

Figure 1 shows the KDD process employed in the study. More specifically, these are the steps involved in the data mining process in line with the study used by the researchers:

1. Selection - After collecting data from the customer, the researchers used Excel and Python to clean up and remove noise from the data.
2. Preprocessing - The researchers cleaned the dataset and transformed it into a Comma-separated values (CSV) file, which was then opened with Python. The data was then fitted for the following stage.
3. Transformation - For the data mining process, the converted dataset was examined using Python methods.

4. Data Mining - The researchers assessed the quality of each model generated by the data mining method in order to determine which prediction model was the best or the most reliable.
5. Interpretation / Evaluation - Appropriate graphs and charts were used to convey the findings from the assessment of the top prediction models to all parties involved.

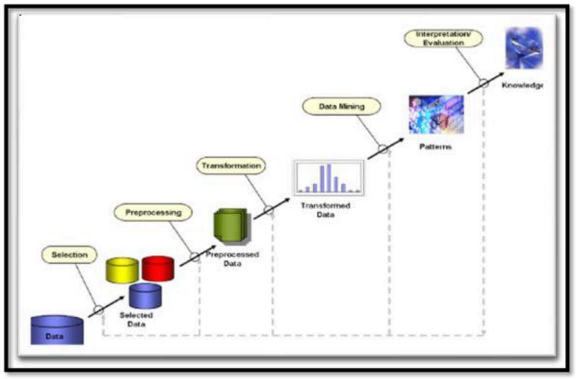


Figure 1. Process of Knowledge Discovery in Databases.

2.2 Methodology for System Development

The field of agile software development has received a lot of attention nowadays. On the other hand, agile software development is a lengthy and iterative process as opposed to a single, large project. Agile development approaches are best suited to adaptable and adaptable software development projects because of their emphasis on cooperation, communication, and flexibility.

Open communication and teamwork are key components of agile development while developing software. Everyone is pursuing the same objectives as a consequence, and any problems may be resolved right away. Agile development prioritizes the potential benefits for the client. Software is built iteratively, thus with each new release, the user should benefit. This ensures that the project continues on track and that the client is satisfied with the final product.

Effective agile software development necessitates collaboration, candid communication, and creativity. Its use in many industries and growing popularity both continue to expand. Learning software programming as part of a capstone project would be extremely beneficial for students who want to work in the industry.The Agile framework is a versatile and incremental method for managing software development projects. It places a strong emphasis on teamwork, flexibility, and incremental advancement via brief development cycles called sprints. The handling of water resources optimization is a primary focus of this research since agile emphasizes customer-centricity and constant communication.

The study's application of Agile Software Development is depicted in figure 2. These are the six primary stages in particular:

1. Plan: - Project developers will create plans for the next endeavor during the planning stage, sometimes referred to as the feasibility stage. Establishing the objectives for their new processes as well as the nature of the issue and the scope of any existing systems are beneficial. (Preston, 2023) Agile planning is an iterative process that facilitates adaptability and quick response to shifting priorities or objectives. In this study, the project is divided into smaller jobs, or user stories, by the researchers, who will then rank the stories according to how valuable and important they are to the end user. This enabled the creation and delivery of the most crucial features first, adding value for the customer and enabling the integration of feedback into the process of development.
2. Design - Prior to the primary development stage, the design stage is a crucial step that has to be taken. (Preston, 2023) The researchers concentrated on design throughout the early stages of the software development process, which included the important responsibilities of conception and planning. This stage was essential to setting the groundwork for the project as a whole. The researcher described the characteristics and capabilities of the program as well as its functionality. In order to improve the user experience, focus was also paid to creating an intuitive user interface. With careful planning and selection of suitable technologies and frameworks, the entire architecture served as a blueprint for further growth.
3. Develop – After a clear design was established, the development stage got underway. Developing a forecasting system specifically for Catmon homes was the main objective. Coding the features that were specified during the design stage was required for this. The development team set out to turn the abstract concepts into a workable, operational system that could forecast trends in water usage.
4. Test - The researchers provided a thorough testing technique to guarantee the forecasting system's resilience and dependability. This technique emphasized the automatic, ongoing assessment of newly written code. The focus was on finding and fixing any problems as soon as they occurred throughout the development phase.

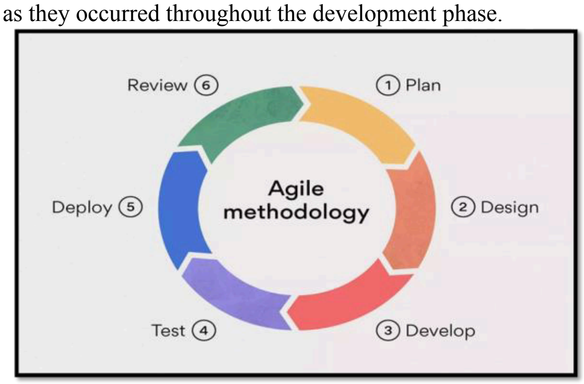


Figure 2. Agile Software Development

3. RESULTS AND DISCUSSION

Prior to applying the four selected algorithms such as SARIMA, ARIMA, Prophet, and Exponential Smoothing, Prophet yielded the best value based on the evaluation metric. The examination emphasized that Prophet showcased superior predictive accuracy, evidenced by lower RMSE, NRMSE, and BIC values, as well as a lower MAE and MAPE. Consequently, Prophet emerged as the most effective forecasting technique. Therefore, the researcher forecasted the water consumption of residences for Catmon, Malolos for the next few years using the Prophet algorithm.

Figure 3 displays the comparison between the actual testing data and the forecasted data for the total water consumption of residential water consumption from 2018 to 2024, utilizing the Prophet model.

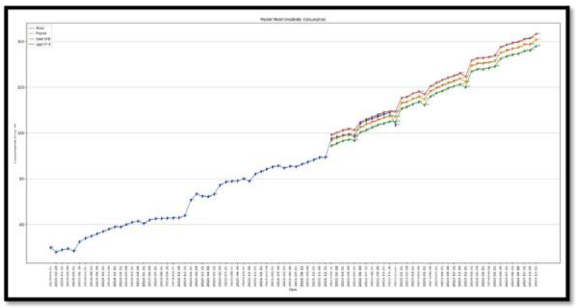


Figure 3. Prophet Forecast, Upper & Lower Limit for Residential Consumption

With an exceptionally low Mean Absolute Percentage Error (MAPE) of 1.19%, which indicates an extremely small average percentage variance between projected and actual values, the Prophet model demonstrates great forecasting ability for household water usage. With an emphasis on absolute differences, the model's excellent accuracy in forecasting water use is demonstrated by its stated Mean Absolute Error (MAE) of 12.49. When contrasted with the other models, the Root Mean Squared Error (RMSE) is significantly lower at 15.38, indicating higher forecasting precision. Additionally, the Normalized RMSE is significantly less at 0.0149, suggesting a better match. At 0.0125, the Normalized BIC is consistent with the other models, suggesting a fair trade-off within model fit and complexity.

Figure 4 displays the comparison between the actual testing data and the forecasted data for the total water consumption of commercial water consumption from 2018 to 2024, utilizing the Prophet model.

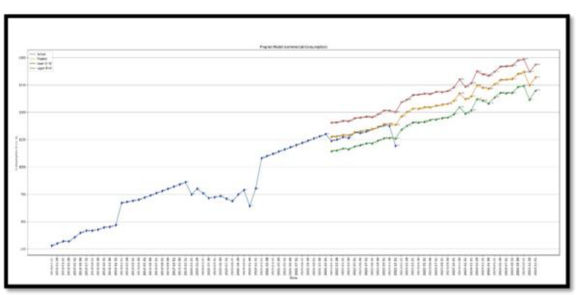


Figure 4. Prophet Forecast, Upper & Lower Limit for Commercial Consumption

The Prophet model applied to commercial water consumption forecasting demonstrates moderate predictive performance, as indicated by the metrics. The Mean Absolute Percentage Error (MAPE) is reported at 2.41%, reflecting the average percentage deviation between predicted and actual values. The Mean Absolute Error (MAE) is 29.68, indicating the average absolute difference between predicted and actual commercial water consumption. The Root Mean Squared Error (RMSE) is 57.89, suggesting a moderate spread of errors.

Overall, among the four commercial water consumption forecasting models with low BIC and MAPE values, the Prophet model stands out. Although its Normalized BIC is 0.0155, its MAPE is competitive (2.41%). ARIMA, SARIMA, and ETS also have reasonable predictive performance, but the Prophet model seems to balance model complexity with accuracy. Prophet, with its lower BIC and equivalent MAPE, may be a good model for commercial water usage predictions, depending on goals.

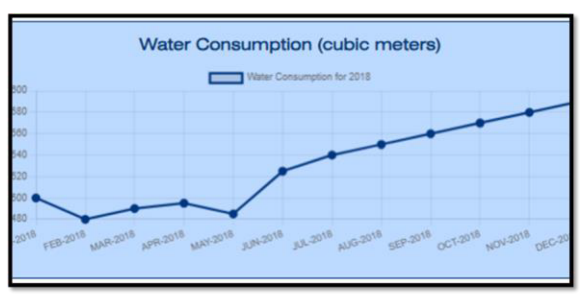
The Lower and Upper Limits are numerical figures that estimate commercial water consumption. These projections were made using a prophet algorithm for Prime Water, Malolos water usage. The algorithm's Lower Limit estimates lowest future water use at 1144.56 to 1699.28 units. This lower level indicates the algorithm's projection of the most cautious or restrained company water consumption. The algorithm's "Upper Limit" is 1404.69 to 1984.04 units for the greatest estimated water use. This high limit represents the algorithm's predictions for excessive commercial water demand.

Visualization Insights and Interpretations

The researchers conducted a thorough analysis of company datasets, employing various chart types. The generated graphs were integrated into the system's dashboard feature. An insight for Water Consumption Data from 2018-2021 could be seen in the graph. Figure 5 shows an example of insight from the year 2018.

Figure 5: Residential Water Consumption Trends

Figure 5 shows Catmon, Malolos' 2018 monthly household water usage climbing steadily. The year starts with 500



units in January and drops to 480 in February. March marks the start of a steady increase in water use. Data suggests that residential water usage may grow in the coming months owing to weather, seasons, or family activities. The use rises significantly from May to August, maybe due to warmer weather or more water demand for horticulture and air conditioning. December saw 590 units as the second half of the year continued to rise. This may be related to holiday water consumption or other special activities. Understanding these trends reveals 2018 Catmon, Malolos residential water usage practices. Researching the reasons for these trends may improve water management or residential infrastructure.

Additionally, an insight for Water Consumption Data from 2018-2021 could be seen in the graph. Figure 6 shows an example of insight from the year 2018 for commercial use.

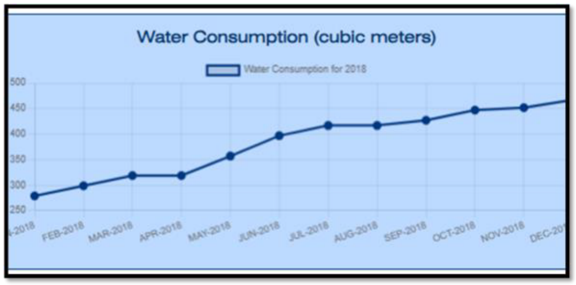
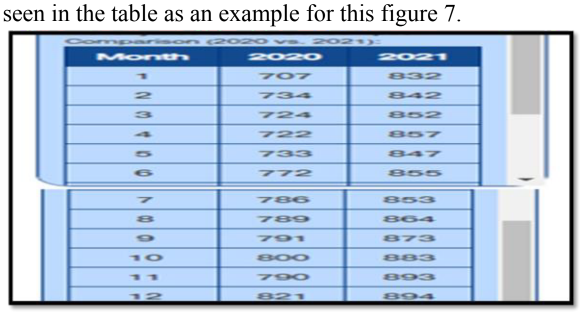


Figure 6: Commercial Water Consumption Trends

The 2018 Catmon, Malolos business water consumption figures in Figure 6 exhibit many findings. Water usage rises steadily from January to December, indicating demand will climb all year. May to July had higher consumption, perhaps due to greater business activities. Seasonal or commercial factors may explain why water usage was stable in August and September and increased in October and November. The year concludes with a high in December, suggesting higher business activity during the holidays. These patterns may assist explain business water use trends and lead to further investigation into certain events or financial periods.

Next is the comparison of years in terms of water consumption obtained by the customers. Two years can be seen in the table as an example for this figure 7.



Comparison (2020 vs. 2021):

Month	2020	2021
1	707	832
2	734	842
3	724	852
4	722	857
5	733	847
6	772	855
7	786	853
8	789	864
9	791	873
10	800	883
11	790	893
12	821	894

Figure 7: Comparison of 2020 and 2021 Water Consumption for Residential

Figure 7 shows Catmon, Malolos, monthly home water use increasing steadily in 2020 and 2021. The two years differ greatly, with 2021 frequently showing higher consumption data. Seasonal water usage may be higher in some months. For instance, January 2021 had a substantial increase over 2020. These disparities suggest population growth, lifestyle changes, and external events may affect residential water usage. Local governments must understand these patterns to create effective water management strategies, allocate resources, and sustain long-term water use habits to meet community needs.

Likewise, a figure shows the comparison of years in terms of water consumption obtained by the customers. Two years can be seen in the table as an example for this figure 8 for commercial use.

Comparison (2020 vs. 2021):

Month	2020	2021
1	707	832
2	734	842
3	724	852
4	722	857
5	733	847
6	772	855
7	786	853
8	789	864
9	791	873
10	800	883
11	790	893
12	821	894

Figure 8: Comparison of 2020 and 2021 Water Consumption for Commercial

Figure 8 illustrates significant monthly commercial water usage fluctuations and trends for Catmon, Malolos, in 2020 and 2021. Note that business water use has grown each month in both years. Consumption is up significantly from 2020 to 2021, reflecting rising demand. The month-by- month comparison shows that consumption increased significantly in January 2021. Disparities may suggest a growing commercial sector, economic expansion, or regional business activity. Analyzing these trends can assist local governments and enterprises adjust water management strategies to meet increased demand, promote sustainable use, and build infrastructure to suit commercial needs.

In the next feature, 6 sitios of Catmon shows the total number of consumption usage and number of customers who avail the water of Prime Water for years.

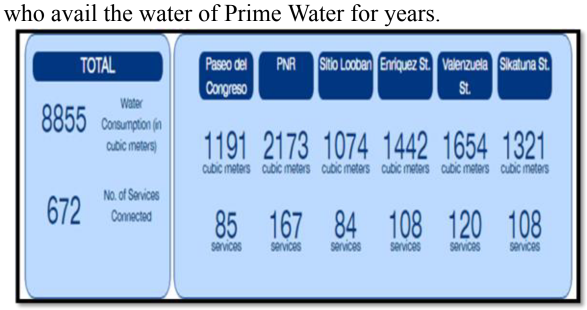


Figure 9: Commercial Total Water Use and Number of Services Connected in 2020

Figure 9 illustrates Catmon, Malolos' 2020 commercial water use of 8,855 cubic meters. Paseo del Congreso accounts for 1,191 cubic meters, PNR for 2,173, Sitio Looban for 1,074, Enriquez St. for 1,442, Valenzuela St. for 1,654, and Sikatuna St. for 1,321. In the same year, 672 business services were connected. This data shows commercial water demand and service distribution among locations, helping develop water policy and infrastructure in Catmon, Malolos in 2020.

Same with the previous feature, 6 sitios of Catmon shows the total number of consumption usage and number of customers who avail the water of Prime Water for years in this figure 10.

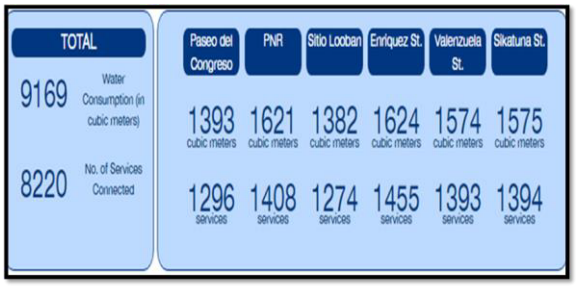


Figure 10: Residential Total Water Use and Number of Services Connected in 2020

In Catmon, Malolos, 9,169 cubic meters of residential water were used, with 8,220 services connected (Figure 10). Paseo del Congreso (1,393), PNR (1,621), Sitio Looban (1,382), Enriquez St. (1,624), Valenzuela St. (1,574), and Sikatuna St. (1,575) have the largest water intake. Paseo del Congreso (1,296 cubic meters), PNR (1,408), Sitio Looban (1,274), Enriquez St. (1,455), Valenzuela St. (1,393), and Sikatuna St. (1,394) also consume water. This study details home water use and service distribution in Catmon, Malolos over time.

For the map tab, it presents the map of Barangay Catmon Malolos, City including their Sitios. The figure 11 presents an example insight for Barangay Catmon.

Figure 11: Heatmap

This visual representation allows quick interpretation of patterns and trends within the data, making it a valuable tool for conveying complex information in an easily understandable format. In this map, figure 41 shows that Valenzuela Valenzuela St., and Paseo St. among the 6 Sitios got the green color implying that sufficient water is needed. Lastly, the administrators can choose what type of consumption they want to present.

The forecast page displays visualizations for the projected consumption results, featuring three-line graphs depicting forecasts for residential consumption, commercial consumption, and combined residential and commercial consumption. Figures 12 to 13 showcase the forecasting features and insight of the system.

Figure 12: Forecast in Residential Consumption

Catmon residential water use will start at 967 cubic meters in January 2022 and rise to 1071 cubic meters in December 2022. Starting at 1130 cubic meters in January 2023, it will rise to 1182 cubic meters in June 2023 before stabilizing again. The estimated water usage would grow abruptly to 1294 cubic meters in January 2024, remain constant, climb to 1349 cubic meters in June 2024, then slowly rise to 1407 cubic meters by December 2024.

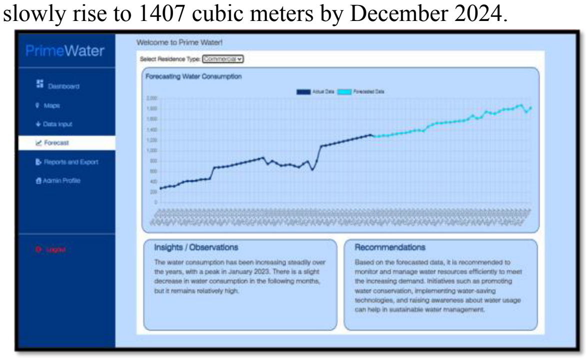


Figure 13: Forecast in Commercial Consumption

The estimated water usage for Catmon's commercial districts would start at 1274 cubic meters in January 2022, climb moderately to 1388 in October 2022, then dramatically to 1460 in January 2023 and 1500 in February 2023. Water consumption gradually rises to 1604 cu m in October 2023, then sharply rises to 1670 cu m in November 2023, returns to above 1619 cu m in December 2023 and 1645 cu m in January 2024, and finally rises to 1744 cu m in February 2024. The estimated water usage then lowers to 1714 cu m in April 2024 before progressively climbing to 1792 cu m by June 2024. It rises to 1868 cubic meters in October 2024, drops to 1745 cubic meters in November, and rises to 1818 cubic meters in December 2024.

System Features

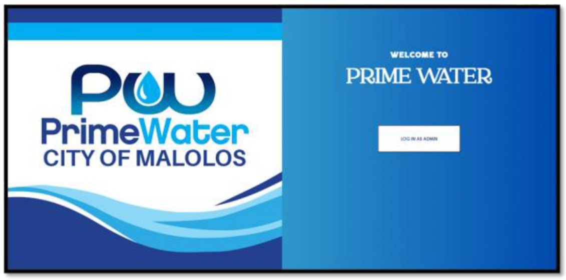


Figure 14: Home Page

Figure 15: Administrators Sign In Page

Figure 16: Incorrect Credentials

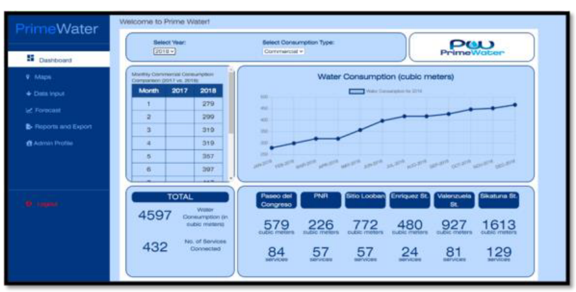


Figure 17: Dashboard

Figure 18: Heatmap

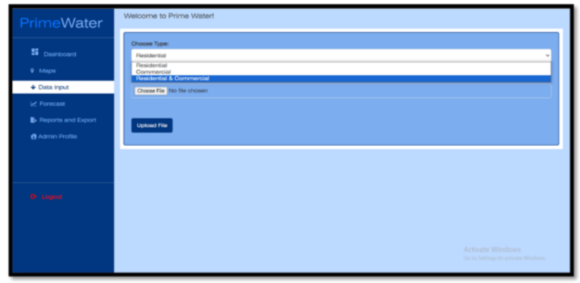


Figure 19: Data Input

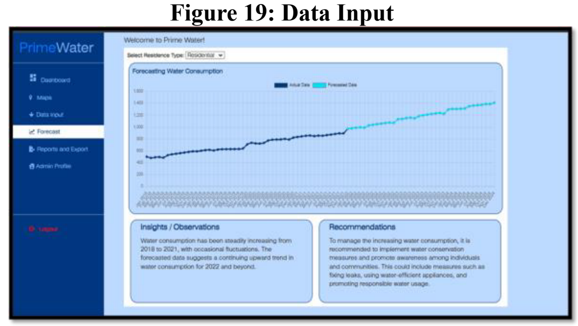


Figure 20: Forecasting Tab (Residential)

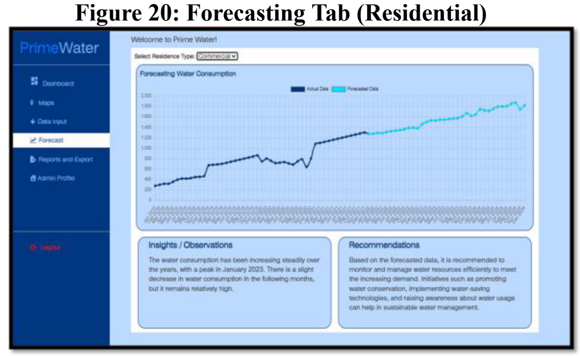


Figure 21: Forecasting Tab (Commercial)

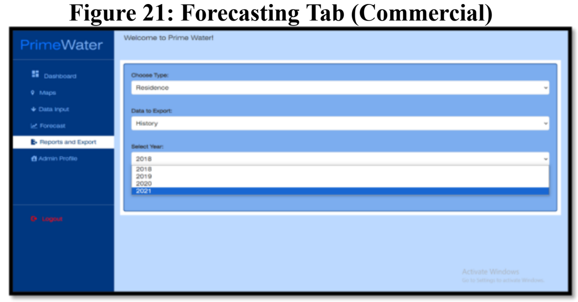


Figure 22: Reports and Exports

Figure 23: Admin Profile

Figure 24: Logout Feature

4. CONCLUSION

This research examined Catmon, Malolos City water use statistics from January 2018 to December 2022. This project aimed to create an analytics model to predict Catmon, Malolos City residents' water use. The study used ARIMA, SARIMA, ETS, and Prophet to anticipate Catmon water consumption across all demand categories until December 2022. The data demonstrate these models' water consumption prediction capabilities. The PROPHET model could anticipate water consumption across all demand categories and was the most resilient and dependable. This study also showed an expected rise in water demand, emphasizing the need of good forecasting for water resource management.

This study shown that short-term water demand forecasting aids operational planning and management. It also prepares water demand analysis benchmarks. A broader dataset and other forecasting methods might improve the findings. If water demand and supply data is available in the region, future researchers may consider population, climatic factors, and spatial attributes in long-term forecasting methods.

5. RECOMMENDATIONS

The following recommendations were made after taking into account the design and execution of the study, and they can be applied by upcoming Information Technology researchers who are interested in conducting studies of the same nature:

1. The file format submitted into the system should be standard to avoid conflicts and mistakes. Although the system may automatically preprocess files, file formats can affect speed.
2. Researchers suggest expanding future studies to include more of Malolos, beyond Catmon, or pursuing other study topics. For a better understanding, avoid focusing on one barrio.
3. Add a function to automatically detect errors and duplicate Excel files in submitted files. Errors in preprocessing can severely impair system performance.
4. To enhance forecasting models, try adding factors beyond year, month, consumption, and location. Demographics, climatic data, and geographical features can enhance water consumption projections.
5. Establish quality control and monitoring methods. Updating and verifying the dataset and incorporating real- time data sources can improve water usage projections by making the system more adaptable and flexible.

6. REFERENCES

Alipala, J. B. (2022). In Bulacan, Angat Dam level drops due to lack of rainfall. Retrieved From https://newsinfo.inquirer.net/1584544/in-bulacan-angat-dam-level-drops- due-to-lack-of-rainfall

Cepeda, M. (2023). 11 million families in Philippines facing water crisis. Retrieved from https://www.straitstimes.com/asia/se-asia/11-million-families-facing-water-crisis-in-philippines

Halton, C. (2023, January 30). Predictive Analytics: Definition, model types, and uses. Investopedia. https://www.investopedia.com/terms/p/predictive-analytics.asp

Preston, M. (2023, October 18). 7 Phases of the System Development Life Cycle Guide. CloudDefense.AI. https://www.clouddefense.ai/system-development-life-cycle/

Un-Water. (n.d.). Water Scarcity | UN-Water. UN-Water. https://www.unwater.org/water-facts/water-scarcity