EFFICACY OF MACHINE LEARNING TECHNIQUES IN PREDICTING GROUNDWATER FLUCTUATIONS IN CHENGALPATTU DISTRICT

INTERNALLY FUNDED PROJECT REPORT

2022 – 2023

SUBMITTED BY

SAM DEVAVARAM JEBARAJ (193002090)

SANTHOSH SRINIVAS L (193002094)

KARAN K (201001013)

DEVADHARSHINI B (201001006)

SUPERVISORS

Dr. P. Kaythry

Associate Professor, Dept. of ECE, SSNCE

Dr. P. Sangeetha

Associate Professor, Dept. of Civil Engineering, SSNCE

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

SSN COLLEGE OF ENGINEERING KALAVAKKAM-603110

IFP-REPORT

CERTIFICATE OF COMPLETION

This is to certify that the project titled “Efficacy of Machine Learning Techniques in predicting Groundwater fluctuations in Chengalpattu District ” undertaken by Sam Devavaram Jebaraj (193002090) and Santhosh Srinivas L (193002094) of final year B.E. Electronics and Communication Engineering and Karan K (201001013) and Devadharshini (201001006) of third year B. E. Civil Engineering has been completed as per the proposed aim and objectives.

Faculties In-Charge

Dr. P. Kaythry,
Associate Professor, ECE

Dr. P . Sangeetha,
Associate Professor, Civil

Head of Department, ECE

Dr. P.Vijayalakshmi,
Professor, ECE

TABLE OF CONTENTS

CHAPTER NO	TITLE	PAGE NO
	TABLE OF CONTENTS	**iii**
1	**INTRODUCTION**	**4**
	1.1 OBJECTIVE	4
	1.2 MOTIVATION	4
	1.3 TARGET BENEFICIARIES	5
	1.4 GROUND WATER LEVEL METER	5
2	**LITERATURE SURVEY**	**7**
3	**METHODOLOGY**	**11**
	3.1 THE DATA COLLECTION STAGE	11
	3.2 ACQUIRING THE COMPONENTS	13
	3.3 DEEP LEARNING MODEL	13
4	**RESULTS**	**17**
5	**CONCLUSION AND FUTURE WORK**	**20**

CHAPTER 1

INTRODUCTION

1.1 OBJECTIVES

The objective of our project is to Predict the ground water fluctuation of open wells located at different places of the Chengalpattu district using various ML models and compare their results based on certain metrics.

The following are the major objectives to be kept in mind for this project:

- Identifying the location of open well in and around Chengalpattu district
- Estimate the ground water level of the selected open wells for the period of six months and analyse the level difference
- Predict the ground water fluctuation using various Machine Learning

1.2 MOTIVATION

Chengalpattu district came into existence recently on 29 November 2019 when it was carved out of a much larger Kanchipuram district. This study demonstrates a pragmatic framework for predicting seasonal groundwater levels at this new district of Tamil Nadu.

Groundwater is one of the key natural resources that supplies a large portion of the water used by a nation.

- Rural households and public water supplies
- Farmers
- Commercial businesses and industries

For administrative reasons, the district has been divided into 3 Revenue divisions comprising of 8 taluks with 636 revenue villages. On the development side, it is divided into 8 development blocks with 359 Village Panchayats.

1.3 TARGET BENEFICIARIES

The results obtained from the study gives confident to use ground water resources for the stipulated period. The data-driven modelling approaches can perform sufficiently well in predicting future groundwater level changes. Different evaluation metrics confirm and highlight the capability of these models to catch the trend of groundwater level fluctuations.

1.4 GROUND WATER LEVEL METER

Monitoring of groundwater levels is necessary to obtain data on the water column head under the earth's surface. It is useful to keep track of changes in the amount of groundwater level in accordance with compliance. Continuous groundwater monitoring gives detailed data and makes it simple to make observations for development and future projections. Piezometers, commonly referred to as level sensors, are crucial in providing all the data underneath the earth's surface. It keeps track of the starting point of the water table. To determine the amount of water present in sand pores and aquifers, it can be installed in wells, bore wells, and tube wells at the desired depth. The electromagnetic flow metre and telemetry system are connected to the level sensor with appropriate wires up to the surface level, and the data are then transmitted to the server for real-time access on data management software regardless of location. With the aid of various level sensors, various changes in the groundwater can be seen.

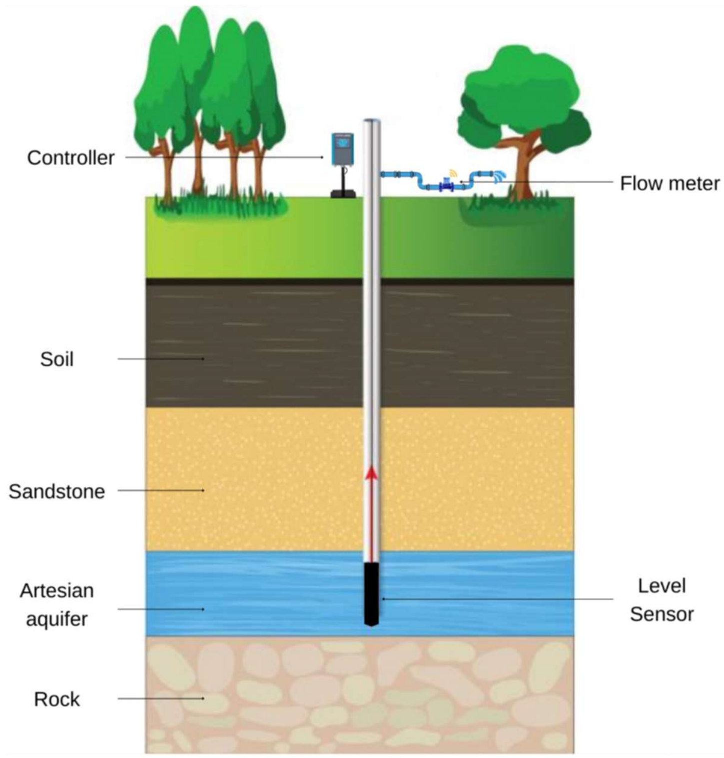

Fig. 1.1 Working of a Ground Water Level Sensor

Fig 1.2 Ground Water Level Senor Used

Accurate long-term groundwater level monitoring data collection can help with correct planning for development and management of groundwater level and prevent future water shortages.

CHAPTER 2

LITERATURE SURVEY

In this section, the available literature in the field of ground water level detection and time series detection is discussed and analyzed to arrive at a conclusion on which techniques we have to implement.

P. D. Sreekanth, N. Geethanjali, P. D. Sreedevi, Shakeel Ahmed, N. Ravi Kumar and P. D. Kamala Jayanthi in the year 2009 conducted a study on forecasting ground water level using Artificial Neural Networks. The elixir of life, WATER, is essential for long-term growth. It was once thought to be an endless or at least completely renewable natural resource, but over the past 20 years or so, there has been a great deal of pressure placed on this priceless resource, primarily as a result of rapid industrialization and population growth, as an increase in the human population will only increase the demand for irrigation to meet food production requirements. Although agricultural technology has advanced significantly, in many areas poor irrigation management has led to significant groundwater table depletion, damaged soils, and deterioration in water quality, making the availability of water in the future highly uncertain. Water scientists and planners must now quantify the available water resources for their judicial usage due to the limited supply of water resources in the near future and the looming risks to it. As a result, it is imperative to have a ready reckoner to track changes in groundwater levels far in advance of when it is necessary to develop sustainable water management strategies. In this direction, a number of research have been conducted to predict groundwater levels using conceptual/physical models. These studies are arduous and have practical limits because there are so many interconnected variables involved. Artificial neural networks (ANNs), a type of soft computing tool, have recently been employed more and more for prediction in a variety of science and technology disciplines. ANNs have been proven to be very helpful for groundwater modelling.

As an all-purpose model with a constrained set of variables, the ANN is employed as a functional approximator. Compared to traditional simulation approaches, it can forecast a wide variety of nonlinear time series events. Essentially, ANNs are intelligent systems that share some characteristics with a biologically skewed model of the human brain. They are made up of numerous small, parallel operational units known as neurons that are coupled to one another in the forward direction by certain multipliers known as connection weights. ANNs are typically taught by changing the connection weights between the network members. These networks have applications in a variety of domains, including forecasting, pattern identification, classification, speech recognition, image processing, etc. They also have the potential to learn on their own and are fault-tolerant as well as noise-immune. In order to have a precise forecasting with increased accuracy over the present approaches being used, a dependable forecasting model for predicting the groundwater level utilising weather parameters through ANNs has been built in this study.

Purna C.Nayak, Y.R.Satyaji Rao & K P Sudheer in the year 2006, tried to study the level of ground water in shallow aquifers and they provided a basis of prediction using similar Artificial Neural Network models. Planning conjunctive use in each basin necessitates an accurate forecast of changes in ground water levels. In this publication, a research study that looked into the feasibility of using artificial neural networks to forecast changes in groundwater levels in an open coastal aquifer in India is reported. The best set of input variables for the model are chosen using a combination of statistical analysis of the available data series and domain expertise. Two observation wells' water levels are predicted using a number of ANN models. The findings indicate that, when measured by several statistical indices, the model predictions are fairly accurate.

The absence of antecedent values from the water level time series may hinder the model's ability to accurately represent the aquifer's recharge period and may worsen the model's performance, according to an input sensitivity analysis. The results generally imply that the ANN models are capable of reasonably accurate water level forecasting up to 4 months in advance.Such projections could be helpful in the coordinated use planning of groundwater and surface water in coastal locations that help preserve the natural gradient of the water table to prevent seawater intrusion or water logging conditions.

Sujay Raghavendra Naganna, Beste Hamiye Beyaztas, Neeraj Dhanraj Bokde, and Asaad M. Armanuos suggested a different approach of high effectiveness and high precision for the prediction of ground water level using Gradient Tree Boosting.Although groundwater can be replenished, overuse has led to a growing concern of its depletion. As a result, a key responsibility of governmental water boards and agencies for sustainable water management is the monitoring and forecasting of groundwater levels. In order to anticipate groundwater levels in two coastal aquifers, the current study compared the performance of the Gradient Tree Boosting (GTB) model with that of the more traditional Adaptive Neuro-Fuzzy Inference System (ANFIS) and Group Method of Data Handling (GMDH) models. In the current study, data from two groundwater level monitoring wells that penetrate unconfined aquifers and are situated in Shirtadi and Rayee close to Mangalore city in Karnataka state, India, were taken into consideration. For model simulation, monthly groundwater level data from the years 2000 to 2013 were used, with 70% of the data used for model training and the remaining 30% used for model testing. A comparative analysis of the results reveals that the proposed GTB approach for forecasting groundwater levels one month out was substantially more accurate than the other models for the same time period and collection of data.

The error statistic, RRMSE, of the GTB, GMDH, and ANFIS models derived during the test phase were 0.473, 0.517, and 0.7522, respectively, for Rayee monitoring well. With the use of several performance measurements, the contrast is further explored. For water scientists and hydro-geologists around the world, groundwater level or water table is a significant indication for monitoring or estimating groundwater reserves in order to forecast future droughts and famine conditions. Here, using the antecedent time-series, the implemented algorithms (GTB, GMDH, and ANFIS) serve as forecasting models to estimate groundwater levels for the following month. Different machine learning and hybrid techniques have been put forth by numerous academics to predict multi-scale groundwater levels. However, in the current situation, a supervised machine learning method called Gradient TreeBoosting (or GTB), which uses ensembles of so-called decision trees, is utilised for predicting to obtain high precision results.

CHAPTER 3

METHODOLOGY

3.1 THE DATA COLLECTION STAGE

As discussed earlier, Groundwater is one of the key natural resources that supplies a large portion of the water used by a nation. This project is aimed at collecting data across various places in the Chengalpattu district. The Chengalpattu town is the District headquarters. For administrative reasons, the district has been divided into 3 Revenue divisions comprising of 8 taluks with 636 revenue villages. On the development side, it is divided into 8 development blocks with 359 Village Panchayats. The water resources of the Chengalpattu district that were taken into consideration are shown here.

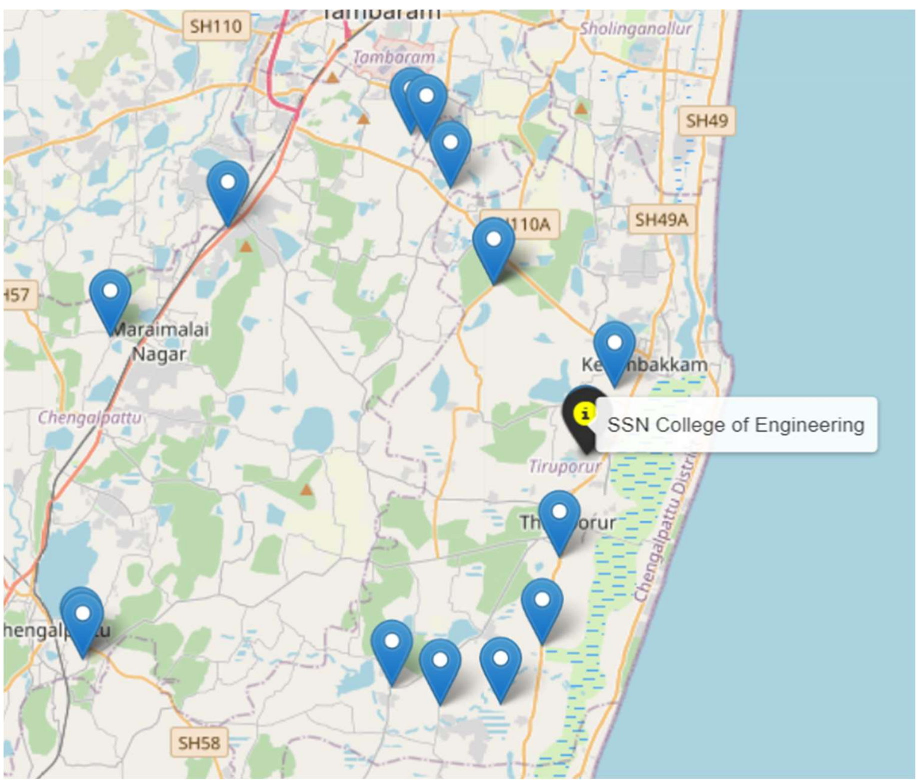

Fig 3.1 Various locations from which data is to be collected.

For this stage, as it is one of the most important stages in any prediction projects, data is collected from various locations in the distract which shows different climate conditions, soil types, rainfall rates and so on. This variation helps in better prediction of the Deep Learning Model. As a result, Permission to take readings from 15 wells consisting of Major and minor irrigation tanks as well as private wells were approved. Readings were taken every fortnight (period of 15 days) by the members of the project with the first reading taken at 11/09/2022 over a span of two days until 05/04/2023. The ground water level meter was used for this purpose of measuring the water level from various pumps and wells. The list of places where the readings were taken are listed below.

- Thandalam
- Rajiv Gandhi IT Expy
- Karunguzhipallam
- Sirudavoor
- Acharavakkam
- Alapakkam
- Thenur
- Kattankolathur
- Guduvancheri
- Rathinamangalam
- Kolapakkam
- Kandigai
- Kolathur – I
- Kolathur – II
- Senganmal

Apart from these, various wells locate within the Sri Sivasubramaniya Nadar College campus was also used for taking the readings every fortnight.

The proposed methodology for the data collection is given below.

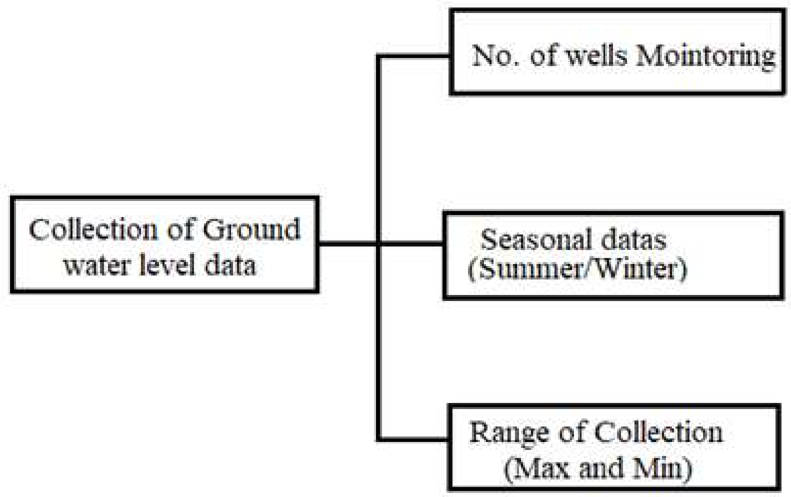

Fig 3.2 – Methodology for the collection of ground water level

3.2 ACQUIRING THE COMPONENTS

The next process in our project is getting the working parts. It is necessary to get them initially so that there is adequate time to study them and design our model later based on our variation in the data obtained. Sources can be easily found on the internet and other physical shops. A complete list of components is available in the budget section of this document.

3.3 DEEP LEARNING MODEL

Once the data has been collected, timely verification needs to be done to check if the data has been collected in the correct format and the data can be used for the date and time. This data can then be used for developing the deep learning model.

This stage is very important as data collection and verification plays a key role in any project and in monitoring rainfall and water levels, it is absolutely necessary to obtain accurate reading of the level of water as an inaccuracy may lead to a poor outcome.

Fig 3.3 Certain wells that were used for timely recording of data

Fig 3.4 A member of this study at a few wells recording the data using the instrument mentioned earlier

The proposed methodology of the project is given below

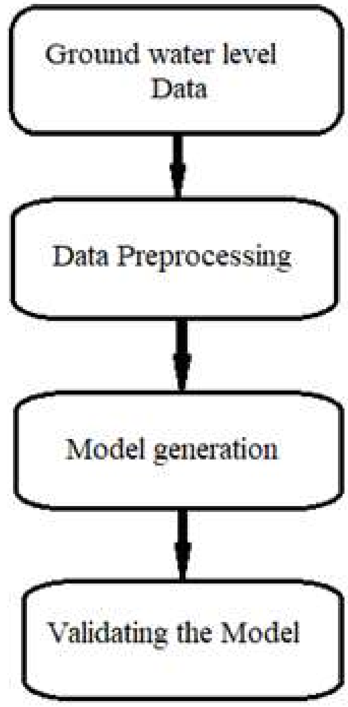

Fig 3.5 - Proposed methodology for the collection and evaluation of ground water level

We have used a library called Autogluon which consists of a collection of Time Series models. By automating machine learning operations, AutoGluon makes it simple to develop apps with robust prediction performance. One can train and use highly accurate machine learning and deep learning models on picture, text, time series, and tabular data with just a few lines of code. Only hurdle was to reshape and redesign the tabular data to be an acceptable input to the library, which was done by Transpose and grouping and index functions. A 2 hour training was assigned, it automatically stopped earlier due to no significant improvement. Numerous models such as Temporal Fusion Transformer (Temporal Fusion Transformer (TFT) is a Transformer-based model that leverages self-attention to capture the complex temporal dynamics of multiple time sequences.) , ARIMA (ARIMA models provide another approach to time series forecasting. Exponential smoothing and ARIMA models are the two most widely used approaches to time series forecasting, and provide complementary approaches to the problem. While exponential smoothing models are based on a description of the trend and seasonality in the data, ARIMA models aim to describe the autocorrelations in the data.), Seasonal Naïve (in the case of seasonal data, there is a simple forecasting method that can be considered as a good benchmark in many situations. Similar to Naïve, Seasonal Naïve relies only on one observation, but instead of taking the most recent value, it uses the value from the same period a season ago.) were present under one package.

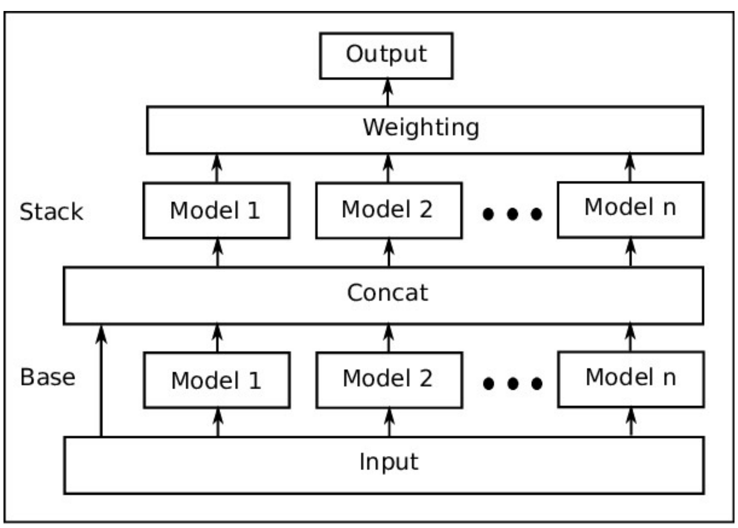

Fig 3.6 – The working of an Autogluon model

CHAPTER 4

RESULTS

An deep learning model was developed which was used to predict the ground water level from the data collected using the Ground water level meter varying across different regions and different times and different climatic conditions. A small sample of the data collected is shown below.

Location/Date	11/9/2022	22/09/2022	15/10/2022	26/10/2022
Thandalam	1.8	1.4	1.6	1.4
Rajiv Gandhi IT Exp	3.2	2.9	3.1	3.1
Karunguzhipallam	2.7	2.2	2.3	2.6
Sirudavoor	2.9	2.5	2.6	3.3
Acharavakkam	1.3	1.6	1.4	0.8
Alapakkam	1	0.7	0.5	0.6
Thenur	4.8	5.2	5	4.6
Kattankolathur	4.4	4.4	4.8	4.1
Guduvancheri	1.65	1.75	2.05	1.85
Rathinamangalam	1.7	1.9	2.2	1.7
Kolapakkam	1.2	0.9	0.7	1.2
Kandigai	0.4	0.1	0.2	0.1
Kolathur-I	0.3	0.6	0.4	0
Kolathur-II	1.5	1.4	1.8	1
Senganmal	1.35	1.45	1.55	0.95

Fig 4.1 – Small sample of the data collected.

Some samples from Acharavakkam, Thennur and Thandalam have been forecasted in this duration and plot. A 10% - 90% confidence interval is also plotted to show the range of values within which an unknown population parameter is estimated. These places for output estimation are chosen at random and the possible ground water levels are predicted using the most suitable model. The results of the same are shown in the corresponding figures.

The parametric results obtained are shown below:

	model	score_test	score_val	pred_time_test	pred_time_val	fit_time_marginal	fit_order
0	WeightedEnsemble	-0.102377	-0.058583	0.236483	102.471823	3.986929	13
1	ARIMA	-0.105073	-0.061705	0.681781	0.598753	0.001002	6
2	SimpleFeedForward\T1	-0.107704	-0.076466	0.171731	0.127995	52.954464	12
3	AutoETS	-0.109718	-0.062338	62.129611	60.631308	0.001002	7
4	TemporalFusionTransformer	-0.111110	-0.074443	0.181594	0.072997	58.173907	11
5	ETS	-0.125063	-0.068257	0.166349	0.163598	0.003000	3
6	SeasonalNaive	-0.131607	-0.095140	0.033231	0.053071	0.000982	2
7	Naive	-0.131607	-0.095140	8.470385	8.398296	0.001002	1
8	DynamicOptimizedTheta	-0.132089	-0.099735	64.797131	63.646260	0.001000	8
9	Theta\T2	-0.135328	-0.103295	0.764066	0.794916	0.000000	5
10	Theta\T1	-0.135328	-0.103295	0.798166	0.776831	0.000000	4
11	AutoARIMA	-0.141788	-0.105893	44.336594	40.950169	0.001999	10
12	DeepAR\T1	-0.174232	-0.144771	0.455149	0.384281	55.522120	9

Fig 4. Parameters of various models obtained using the data collected.

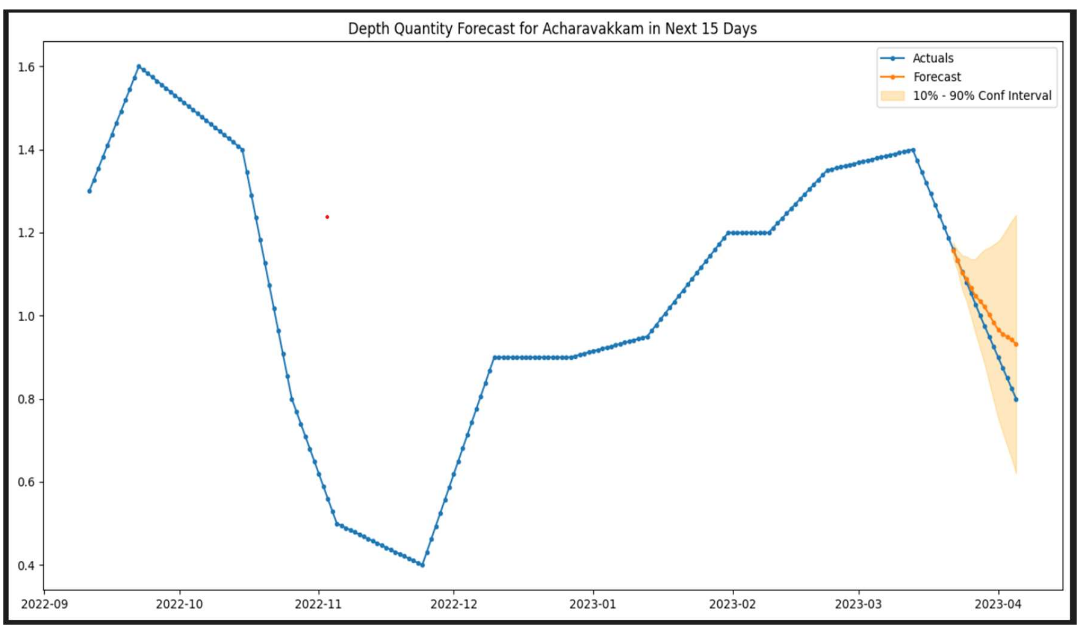

Fig 4.3 Depth Quantity forecast for Acharavakkam

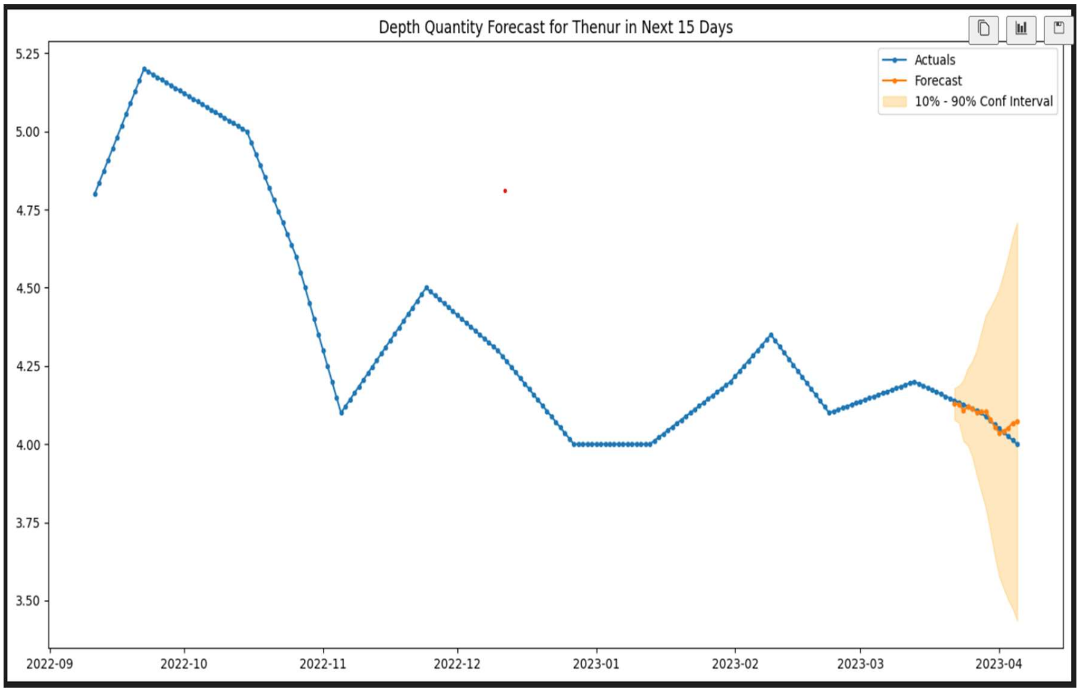

Fig 4.4 Depth Quantity forecast for Thenur

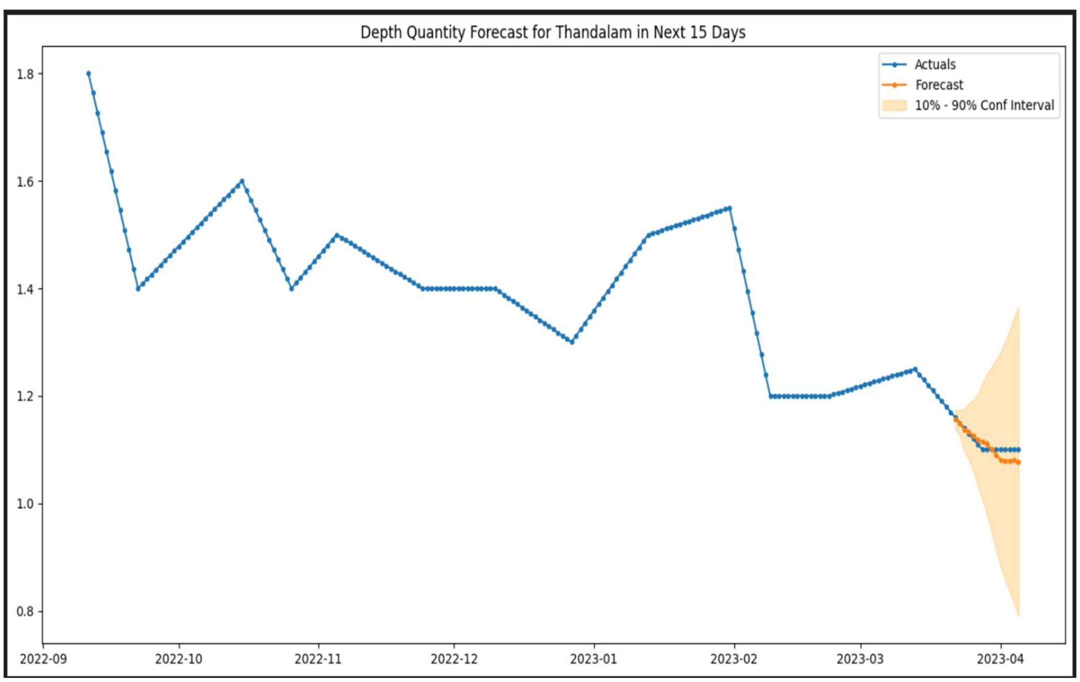

Fig 4.5 Depth Quantity forecast for Thandalam

CHAPTER 5

CONCLUSION AND FUTURE WORK

Looking at the problems that we wish to solve using the product at the end of the project is extremely vital to lay a proper direction to the project and fully satisfy objectives.
For nearly 10 years, the expanding Chengalpattu town and its residents situated on the outskirts of Chennai have been waging a lopsided battle against water scarcity.

Even though river Palar, which is one of the drinking water sources for local bodies in Chennai, flows through here, Chengalpattu residents have not been fortunate enough to reap the benefits of it, which was again not helped by the river going dry two decades earlier. Then, the monsoon of 2015 happened and Palar was brim for once, after a generation's gap.

But, in under three months of that happening, water scarcity has reared its ugly head again in this rapidly urbanising town located off GST Road.

Residents in areas like Natham, Shastri Nagar, Gandhi Salai, Mettu street, Jeevanandam street in wards one, seven, eight, nine, 10 and 21 complained that piped water supply is restricted to once every 10 days

Monitoring this issue on a regular basis and taking timely readings is extremely important. It is also to be noted that the data should be widespread and regular in order to predict the results with higher degree of accuracy. Moreover a larger amount of data is also required to enable in improving the efficacy of the above model/s. It should also be noted that the data collecting individual should be highly skilled in the process.

This project can thus be extended to various places in India where water scarcity is a very important issue. Due to the global warming crisis and depletion in the Ozone layer, this could be deemed to be helpful in estimating the water shortages and plan for the future. Hence more data would be required for the better operation of the model and this can enable in fighting the major water crisis which is due to occur in the coming future.

EXPENDITURE SUMMARY

S. NO	DETAILS	QUANTITY	PRICE
1	Ground Water Level Meter	1	26,082
2	Shipping Charges	-	4,999

X