



**ÇANKAYA UNIVERSITY**  
**FACULTY OF ENGINEERING**  
**COMPUTER ENGINEERING DEPARTMENT**

**Project Report**

**Version 1**

**CENG 407**

Innovative System Design and Development II

**Team ID: 202005**

**Monitoring System of Water Quality and Efficiency of  
Wastewater Treatment Plants**

Alp Özeren - 201711051

Abdulkерим Güven - 201711033

Oğuzhan Saltık - 201611048

Mustafa Kayhan Arıcan - 201611004

# Abstract

In this project, we aim to design a web-based monitoring system for water quality and efficiency to be used for decision-making involving wastewater treatment plants. The current and present water quality data will be visualized by our web-based system. With our project, the quality of water will be predictable using machine learning algorithms. There are two parts to this project. Analyzing the water quality data for rivers, lakes, seas all around Turkey and analyzing the data for water treatment plants. The data for water treatment plants include samples taken from both the inlets and outlets of these plants.

# Özet

Bu projedeki amacımız su arıtma tesislerinde kullanılmak üzere amaçlanan web tabanlı bir su kalitesi izleme ve tahmin etme sistemi geliştirmek. Önceden girilmiş veriler ve gelecek zamandaki su kalitesi verileri tahmin edilerek web tabanlı sistemimizde çeşitli grafiklerle görselleştirilecek. Projemizde su kalitesi makine öğrenmesi ile tahmin edilebilecek. Bu proje iki kısımdan oluşmaktadır; Türkiye’deki akarsular, göller, denizler ve su arıtma tesislerindeki su kalitesi analizi ve makine öğrenmesi kullanılarak su kalitesinin tahmini. Tesislerden alınan veriler hem tesise giriş hem de tesisten çıkış numunelerini içermektedir.

## Contents

Abstract.....	2
Özet.....	2
Introduction.....	6
Company Background.....	7
Problem Statement.....	7
Literature Search.....	8
Related Work.....	8
Machine Learning Methods for Better Water Quality Prediction .....	8
Study of Short-Term Water Quality Prediction Model Based on Wavelet Neural Network .....	9
Prediction of Water Quality Time Series Data Based on Least Squares Support Vector Machine .....	11
Proposed System.....	12
Dataset .....	12
Water Quality Monitoring System .....	13
Water Quality Prediction .....	14
Software Requirements Specification.....	15
Introduction .....	15
Purpose.....	15
Scope .....	15
Glossary.....	15
Overview.....	16
Overall Description .....	16
Product Perspective.....	16

Product Functions .....	17
User Characteristics.....	18
Constraints .....	18
Risks .....	18
Assumptions and Dependencies .....	18
Requirements Specification .....	19
External Interface Requirements .....	19
Functional Requirements.....	20
Performance Requirements.....	23
Software System Attributes.....	23
Software Design Description.....	25
Introduction .....	25
Purpose.....	25
Scope.....	25
Glossary.....	25
Overview.....	26
Design Considerations .....	26
Approach .....	26
Tools Used.....	27
Constraints .....	27
Assumptions and Dependencies .....	27
Architecture .....	28
Software Architecture.....	28
Hardware Architecture.....	29
System Interfaces.....	29

External System Interfaces.....	29
User Interface Design .....	30
Screen Definitions .....	30
Process Design.....	35
Use Cases .....	35
Sequence Diagrams .....	40
Database Design.....	42
Table Definitions .....	43
ER Diagram.....	45
Conclusions.....	45
References.....	46

# Introduction

Water is the most important source of life. Water covers 71% of the earth and only 3% of water is fresh. Humankind always settled down near freshwater sources. Water has an enormous effect on human life throughout history. Even in ancient times, people found ways to purify water or to keep it clean [1]. Water is the home for the microorganism if there are no toxic chemicals in it. Although most microorganisms are harmless, there can be viruses or bacteria that can cause health damage [2].

The effect of climate change and increasing demand for water by rapidly increasing population, industrialization, agricultural and other sectors is putting serious pressure on quality and quantity of water resources. For those reasons, managing and monitoring water, and detecting potential dangers before they affect water is highly important for protection and cleansing of water resources.

Therefore, The Special Environmental Protection Agency, which is a part of The Ministry of Environment and Urban Planning is monitoring physical, chemical and biological parameters of important rivers, lakes, drainage channels and seas inside the Special Environmental Protection Area (SEPA).

Managers and analysts need operational tools that help understanding the complex information about quality of water. Tools based on statistical approaches are often unable to conduct a detailed analysis due to sparse data and the invisible interactions of analysis results. In this project, samples taken in this area with cooperation of public institutes since 2005, this big and complex data will be visualized. This module will include basic data management functions such as time series management, spatial selection and representation, data availability assessment and data series comparison (visual and statistical).

Improved water quality prediction, accuracy and reduced computational complexity are vital for precise control over water quality. For this purpose, our aim is to develop a machine learning model that effectively predicts water quality and establishes an early warning system for water pollution.

# Company Background

The Special Environmental Protection Agency was established in 1988 by a decree of the Turkish government. They were authorized to take special measures to safeguard Special Protection Areas in various places in Turkey. The agency was placed under direct supervision of the prime minister upon creation, but after a few years, the agency was put under the responsibility of the Ministry of Environment. As of 2011 the agency was reorganized as three different branches, one specialized at town planning, another at water management and another at forestry.

## Problem Statement

The water quality assessment survey is not performed directly by the employees of the department. A firm is hired to take measurements around Turkey and place all the measured data within a database. Information provided by the contractor company is not always checked, often it might be missing more than a few entries in the database.

The decision-makers at the agency primarily use visualizations of readings contained in the database. Visualizations of various measures such as pH value, dissolved oxygen, total coliform, fecal coliform, temperature, and so on are generated by hand, taking valuable time.

These visualizations and statistical approaches are used to understand the complex information about quality of water. Tools based on statistical approaches are often unable to conduct a detailed analysis due to sparse data and the invisible interactions of analysis results.

Rest of this document describes a system that will solve the problems stated above by providing the agency a data-entry system, a visualizations interface and a future water-quality predictor AI, all-in-one.

# Literature Search

## Related Work

### **Machine Learning Methods for Better Water Quality Prediction**

In this research paper [3], the dataset was created with 4 monitoring states on Johor River, a river at Johor State in Malaysia. A comparison is made between the following machine learning algorithms: WDT-ANFIS, ANFIS, RBF-ANN, and MLP-ANN. Due to the presence of noise in the data, it is relatively difficult to make an accurate prediction. Hence, a Neuro-Fuzzy Inference System based augmented wavelet de-noising technique has been recommended that depends on historical data of the water quality parameter.

#### Dataset and Data Processing

Selecting the input variables for a model is very important for Artificial Neural Networks. The following water quality parameters were chosen for ANN modelling: temperature, electrical conductivity, salinity, nitrate( $\text{NO}_3$ ), turbidity, phosphate( $\text{PO}_4$ ), chloride( $\text{Cl}$ ), potassium( $\text{K}$ ), sodium( $\text{Na}$ ), magnesium( $\text{Mg}$ ), iron( $\text{Fe}$ ) and Escherichia coli( $\text{E-coli}$ ). These input parameters were used in many previous studies for ANN models [4,5,6]. Using these parameters the prediction of pH, suspended solids (SS), and ammoniacal nitrogen (AN) is made possible.

#### Model Performance

There are in total 3 models for three primary water quality parameters: AN, SS, and pH. The performance of the models is measured with the Coefficient of Efficiency (CE). Mean Square Error (MSE) is used to see the level of fitness between the network output and the desired output. Performance is better with smaller MSE values. Coefficient of Correlation (CC) is employed to inspect the linear relationship between the measured and predicted dissolved oxygen in the water. Using this methodology, the WDT-ANFIS models outperformed others.

# **Study of Short-Term Water Quality Prediction Model Based on Wavelet Neural Network**

This research paper [7] combines the wavelet transform with a Back Propagation (BP) neural network to build a short-term water prediction model. The trained model is used to predict the water quality on freshwater pearl breeding ponds in Duchang County, Jiangxi province, China. Also, a comparison has been made between Elman Neural Network, Wavelet Neural Network (WNN), and a BP network. The proposed model also features a high learning speed and improved accuracy.

## **Dataset and Data Processing**

The dataset consists of measurements taken from Jishan Lake in Duchang Country, Jiangxi province, China. The research took the ecological environment monitoring data of the mussel aquaculture pond as research samples; each sample includes solar radiation, water temperature, dissolved oxygen, pH, humidity, and wind speed. The sampling period was from July 21 to July 27, 2010. Data were collected every 60 minutes for a total of 168 samples. 144 of the samples were used as training sets and 24 of them were used as test sets. Data was normalized because the dimensions of the sample data were different. This normalization step reduced the influence of the prediction performance. As for model input, the first half-hour of dissolved oxygen, PH, temperature, humidity, wind speed, and solar radiation were used. The subsequent dissolved oxygen predictive values were used as outputs. After properly training the model, the prediction of dissolved oxygen in freshwater pearl aquaculture ponds as possible.

## Model Performance

Model performance was measured by Absolute Percentage Error (APE) and Mean Absolute Percentage Error(MAPE). The Wavelet Neural Network (WNN) outperformed BPNN and Elman NN by significantly lower APE. The model accuracy was greater than 90%. As shown in Figure 1.1, WNN also has higher prediction precision, stronger learning, and generalization ability compared to BPNN and Elman NN [7].

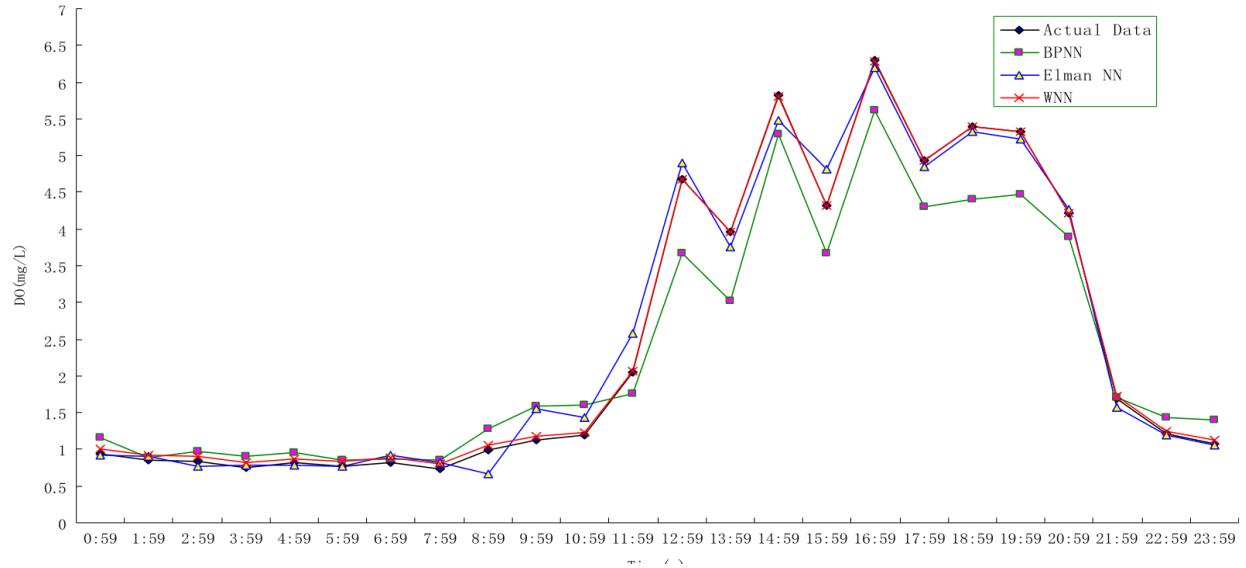


Figure 1.1: WNN compared to BPNN, Elman NN, and Actual Data. WNN fitted the actual data better than other algorithms. (x axis: time in minutes and seconds, y axis: Predicted dissolved oxygen mg/L)[7]

# **Prediction of Water Quality Time Series Data Based on Least Squares Support Vector Machine**

This paper [8] proposes using least squares support vector machine (LS-SVM) algorithm to construct a non-linear time series forecasting model for predicting water quality.

## **Dataset and Data Processing**

They use the small number of samples (actual number not shared) provided by the Beijing Water Authority and after normalizing data points, a variance of 0.01 as random white noise was introduced to training samples.

## **Model Performance**

A comparative study of prediction is performed using the LS-SVM algorithm, Backpropagation (BP), and Radial basis function (RBF) network methods. Predicted values from three models and the true models are compared by examining the percent of deviation. Since the LS-SVM model has the lowest average deviation from the true value, it is considered the best of three. The paper argues that the LS-SVM model can take full advantage of the distribution of the training samples and has a better ability to process small samples. It is concluded that LS-SVM has a lower root mean square error and mean relative error than other methods and has high prediction accuracy, and applicable to real-time water quality data with a small sample.

# Proposed System

The Wastewater Monitoring project will consist of two modules. The first module will be responsible for monitoring the water quality by providing visualizations of physical, chemical and biological factors taken from the dataset. The second module will be responsible for predicting the water quality using machine learning, based on the current data.

In the following sections, the properties of the dataset are given and the two planned modules are explained.

## Dataset

The dataset contains the data gathered from testing important rivers, lakes, drainage channels, and marine areas in 245 Special Environmental Protection Areas in Turkey in terms of physical, inorganic - chemical, and organic parameters since 2005.

In the dataset, we have samples from rivers, seas, lakes, and water treatment plants. Samples taken from water treatment plants started in 2011. The number of samples taken from all of the water sources is increased each year. In 2005, 15 samples were taken from rivers and in 2006 the number of samples was 32. All the samples have columns SAMPLE\_NAME, REGION\_NAME, LOCATION, X\_UTM, Y\_UTM, DATE, TEMPERATURE\_C regarding where and when the sample is taken. Also, the dataset contains the following sample values; pH, SALINITY\_THOUSANDTH, DISSOLVED\_OXYGEN\_MGL, DISSOLVED\_OXYGEN\_PERCENT, for all the samples taken. For water treatments plants we also have ELECTRICAL\_CONDUCTIVITY, BOD\_MG (biochemical oxygen demand), COD\_MG (chemical oxygen demand), TOTAL\_SUSPENDED\_SOLID\_MGL, TOTAL\_NITROGEN\_MGL, TOTAL\_PHOSPHORUS\_MGL, TOTAL\_COLIFORM\_CFU\_100ML, FECAL\_COLIFORM\_CFU\_100ML and FLOW\_RATE. However, the dataset has some null values in alternating years and sometimes there are extra sample parameters for the same sample location.

The dataset is provided in Microsoft Access 2003 (.mdb) format. To be able to visualize the data and train Machine Learning algorithms, we need to access the data inside these ".mdb" files. Since this file format is proprietary and has no public specification, there are only a couple of methods to programmatically obtain the data inside. The first method is to write a Microsoft Access extension using "Visual Basic for Applications" for the users of the software and ask them to use this extension to export the data inside to the XML format that then we can easily read. The second method is to ask the users to install the Access Database Driver that Microsoft provides and then use this driver to access the file contents as if the file itself was a database.

## Water Quality Monitoring System

The governmental agency, The Ministry of Environment and Urban Planning, needs a reporting system that visualizes the observations from important rivers, lakes, and marine special environmental protection areas (SEPA) in Turkey.

The reports are then used for decision making. In the past, the reports and visualizations in these reports were prepared manually. Our goal is to develop a web-based reporting system for SEPA that will automatically read the observation data set and present test results to decision-makers. This system will have detailed filtering features and will be able to perform data visualization. The data collected from the field will be used to produce visualizations for statistical modeling. Data visualization is helpful to decision-makers in identifying features that are not easily noticed by statistical models or humans, such as detection of outlier values of parameters, missing values. Visualizations enable doing correlation analysis, determination of the relationship between dependent variables [9].

In the reports prepared for the years 2005-2009 provided to us by the agency, the tables were often plotted with a bar chart. The graphics were created from samples taken from the inlets and outlets of wastewater treatment plants, as well as from different points of seas, lakes, and rivers. The charts parameters vary by year and region, but generally depend on the parameters of pH, Temperature (OC), Light Transmittance (m), Dissolved Oxygen (mg / L), O<sub>2</sub> (%), Ammonia (mg / L), Total Phenol (mg / L), Total Coliform (CFU / 100mL), Fecal Coliform (CFU / 100mL), Fecal Streptococcus(CFU / 100mL), Oil-Grease (mg / L), Color (Pt-Co), Fragrance (TON).

There are some studies on water quality visualization. One of them used old-fashioned interactive maps and various types of plotting [10]. Besides, tables are used, which contain somewhat similar to the parameters used to determine water quality in the reports provided to us by the agency, sorted by years with parameter values of total phosphorus, total nitrogen, electrical conductivity, pH, dissolved oxygen.

A modern user-friendly interface, more effective and easy-to-understand graphics will be produced by considering the types of graphics and parameters used previously. Also, the locations where the test sample was taken will be displayed on the interactive maps.

### Tools and Frameworks

There are multiple promising options for generating charts and creating user interfaces. The tools that are being considered include (but not limited to): Electron, React.js, Chart.js, ASP.NET, Flask, Qt and more. We are weighing all the options and will be considering the input of the agency since the end product will run on their environment.

## Water Quality Prediction

The second goal of the project is to predict the future properties of a water sample. These predicted properties can then be used to predict the water quality and inform the water treatment plant.

For the problem of forecasting how to treat the water, since the quality of water can be affected by various parameters and such parameters show a complex non-linear relationship with each other and water quality, traditional techniques for data processing are no longer efficient enough.

### Tools and Frameworks

For this project, we decided to use the Python programming language for implementing machine learning algorithms. The machine learning models will be trained locally. As for machine learning framework, we decided to use Tensorflow because, after training, the model can be easily used in Tensorflow.js, therefore making it very easy to run in a browser.

# Software Requirements Specification

## Introduction

### Purpose

The purpose of this document is to provide the technical description of all software requirements of this system. It explains interfaces and the usage of the system. In addition, the document describes in what conditions the system will work.

### Scope

Water Quality Prediction and Monitoring System will be a web based system that is intended to be used by employees of The Ministry of Environment and Urban Planning. Its goal is to visualize water quality parameters from water sample readings entered to the system by data entry operators. The data visualized by the front-end will facilitate the decision making process for the employees of the agency.

WQPMS will also be able to predict the future water quality parameters using previously gathered data using machine learning algorithms. Thereby, it may reveal hidden patterns and maximizing work efficiency.

### Glossary

The Agency	The Ministry of Environment and Urban Planning
WQPMS	Water Quality Prediction and Monitoring System
SEPA	Special Environmental Protection Areas
WQP	Water Quality Prediction
ML	Machine Learning
GUI	Graphical User Interface

## **Overview**

This document is prepared in accordance with the IEEE Std 1016-2009 [11], IEEE Recommended Practice for Software Requirements Specifications [12].

## **Overall Description**

The Water Quality Monitoring system is created for the use of decision-makers at the agency. They will be the main users of the software and therefore the software will be hosted on the agency's servers.

## **Product Perspective**

The system is self-contained and independent from other software.

The previous workflow of decision making related to water quality was a manual workflow, by this we mean, water quality readings were provided to the decision makers at the agency and a report including charts of the readings were produced manually. The system described in this document is the automated alternative to this workflow.

### **System Interfaces**

Since the software is entirely self-contained, there are no external system interfaces needed.

### **User Interfaces**

There will be three different user interfaces for different types of users. The types are the administrator, decision-maker, and the data entry operator.

The administrator will be able to add users, remove users, and update permissions of a user.

The data entry operator will be able to enter new water quality readings to the system, but will not be able to see past readings.

The decision-makers will be able to see the water quality readings in the system, both by examining the raw data provided by the data entry operators and the visualizations automatically generated by the software. They will also be able to access the future water quality predictions provided by the ML subsystem.

## Hardware Interfaces

For users, there are no hardware interfaces required to run the software other than a computer capable of serving and displaying web pages because parameter prediction will be handled by a server.

For the server, a CUDA capable GPU is needed to make predictions as quickly as possible.

## Software Interfaces

The system depends on PostgreSQL 13.1 for persistent data storage. For serving web pages and communicating with database Django Framework version 3.1 is used. The system will include a SQL definition file that describes data tables for the first time setup that will be performed by the agency's database admin.

## Communication Interfaces

An internet connection is required to run the data entry subsystem of the software. The prediction and visualization subsystems can be accessed through a local network connection.

# Product Functions

## Data Entry Subsystem

The water quality measurement survey is not directly conducted by the agency employees. A company is contracted to do the survey travels around Turkey, taking measurements and inserting all the measured data inside a Microsoft Access database. This workflow has some disadvantages.

The first disadvantage is that the data provided by the contractor company is not always validated, sometimes more than a few entries in the database can be missing. The second disadvantage is that the data is sent to the agency only 1 or 2 times a year, not at the time they're taken. By creating a data entry subsystem for the contractors to use, we'll be able to make sure the data entered will be on time and up to standards.

## Prediction of Future Water Quality

The data entered to the system by data entry operators will be used to forecast the future water quality by running a ML model in the background. The decision makers will be able to access these predictions through their visualizations GUI.

## Visualization of Measurements

The decision makers will mainly use the system for visualizing the data obtained by measuring several water sources across Turkey. To be able to decide how to treat water in each source, the users need visualizations of different measurements such as pH value, dissolved oxygen, total coliform, fecal coliform, temperature and so on.

## User Characteristics

The users in decision maker and data entry operator user groups must have basic computer knowledge, such as knowing how to operate a computer and a web browser.

The administrator is expected to have knowledge of database systems and servers, since the setup of the system will require these skills.

## Constraints

Users should open the system using a web browser.

The database should have enough space to hold data and have enough space for future data entries.

Prediction must be done in accordance with The Regulations of Water Quality Management.

## Risks

Due to missing, redundant and noisy data, the dataset should be cleaned before training the model. After cleaning, some information is lost. In machine learning, more data is usually better with a well implemented machine learning algorithm. Therefore, with less data, the trained model's prediction accuracy could be lower than expected and lower prediction accuracy on training sets always results in poor performance in real-world situations.

## Assumptions and Dependencies

For the software to run reliably, servers must be able to run docker containers and PostgreSQL. When these requirements are met, the operating system or other software should not affect the operation of the software.

The users must use the latest version of the Google Chrome web browser to be able to reliably operate the user interface of the software.

# Requirements Specification

- Data entry operators must have connected to the internet.
- The system should be able to be opened by a web-browser.
- Information of users should be stored in a database.
- The database must be able to store data for the past 15 years and should have enough space for future data entry.
- New users should be able to be created by contacting an admin.
- Data entry operators should have different interfaces than User or Admin.
- Data entry operators should not be able to see previously entered data.
- The system should be able to visualize or show raw data to users in a selected time and place.
- The system should be able to predict a selected value in the selected place.
- New resources should be added if needed.
- The system may show a map where the samples taken from.

## External Interface Requirements

### User Interfaces

Users should be able to open the system using a web-browser. The system should be opened by any operating system.

### Hardware Interfaces

There are no external hardware interfaces needed.

### Software Interfaces

There are no external software interfaces needed.

## Communications Interfaces

There are no external communication interfaces needed.

## Functional Requirements

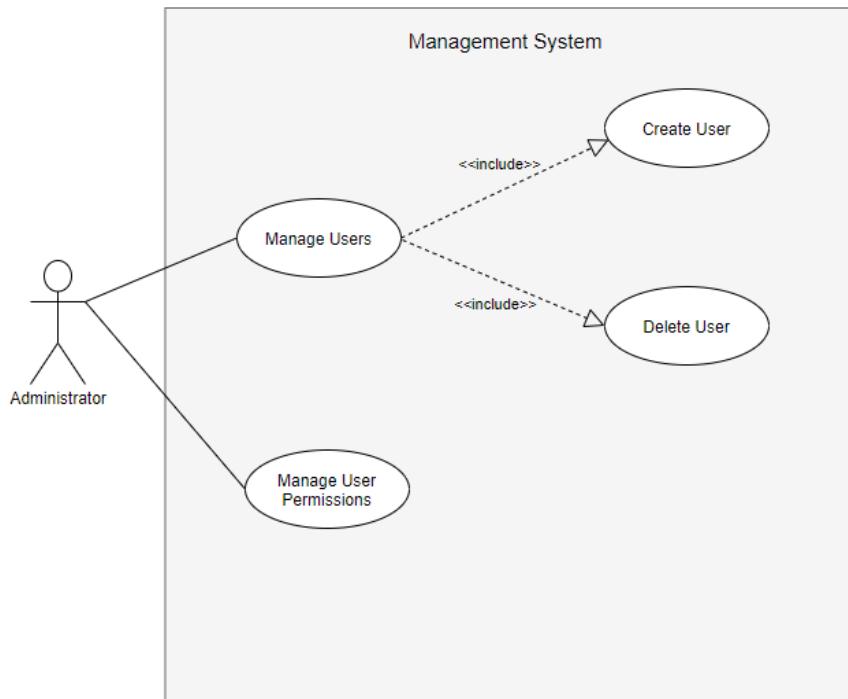
### Administrator Use Case

**Actor :** Administrator

**Use Case:**

- Manage Users
  - Create User
  - Delete User
- Manage User Permissions

**Diagram:**



*Figure 2.1 : Management System Use Case Diagram*

**Brief Description:**

Figure 2.1 shows the management system use case diagram. The administrator has authority to manage the system. Administrators are able to add and delete users and manage user permissions.

### **Initial Step by Step Description:**

- Users can not register.
- Administrators are able to create users.
- Administrators can give users specific permissions and prohibitions.
  - Administrators can assign the user to the data entry operators group or decision makers group.
- Administrators can delete users.

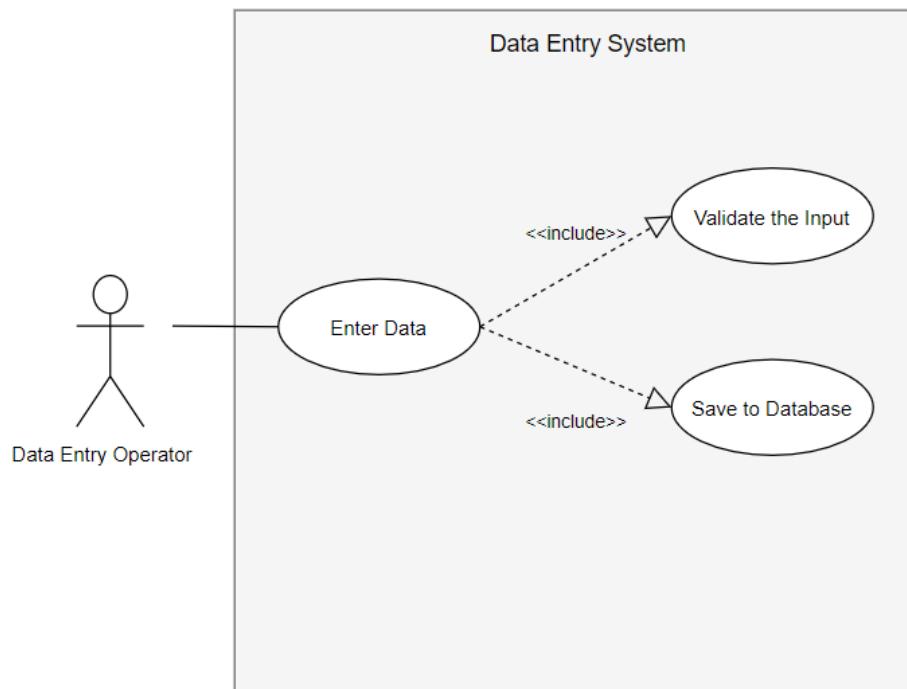
Data Entry Operator Use Case

**Actor:** Data Entry Operator

**Use Case:**

- Enter Data
  - Validation
  - Save to Database

**Diagram:**



*Figure 2.2 : Data Entry System Use Case Diagram*

**Brief Description:**

Figure 2.2 shows the data entry system use case diagram. Data entry operators do not have the same permissions as Decision Makers, therefore they can only enter data in appropriate format to aid decision making.

### **Initial Step by Step Description:**

- Data entry operator can insert data to system
  - If data is not entered in the proper format, a request will be sent to a user to enter the data in proper format.
  - If data is in proper format, data can be inserted to the database.

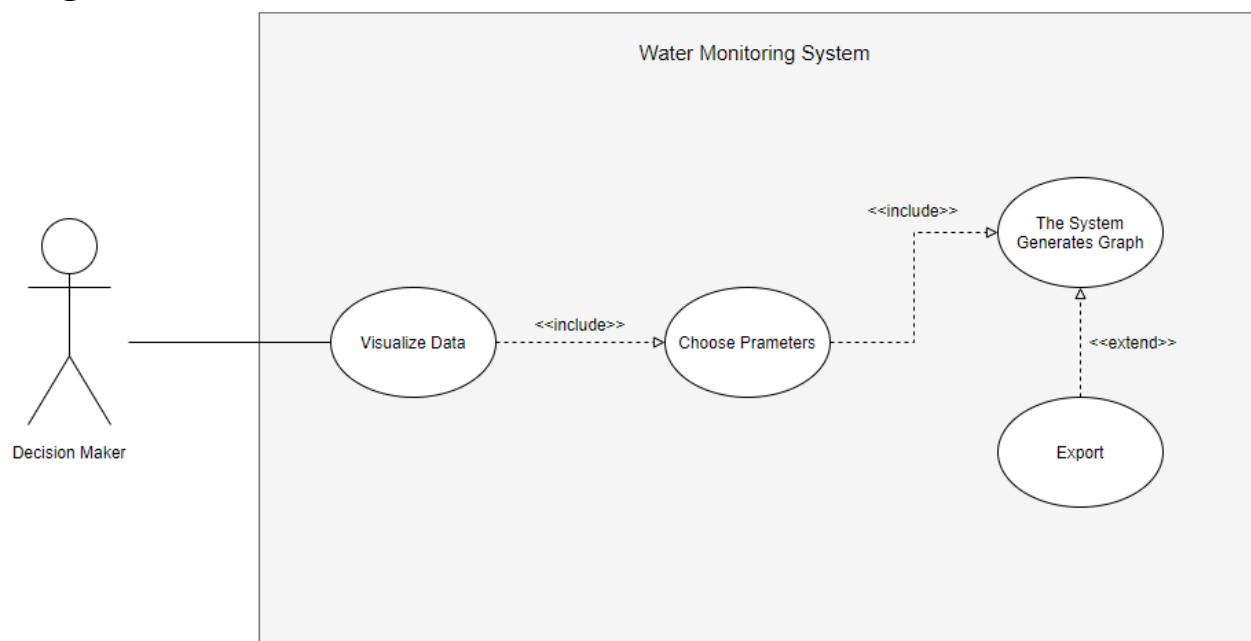
Decision Maker Use Case

**Actor:** Decision Maker

**Use case:**

- Visualize Data
  - Show Graph
    - Export
  - Show Table
    - Export
  - Show Predicted Graph
    - Export

**Diagram:**



*Figure 2.3: Decision Maker Use Case*

**Description:** Decision Makers are allowed to use the system for visualizing selected samples. The graph of a sample includes, time, water quality parameters and prediction results of the selected sample. As you can see in Figure 2.3, Decision Makers select which graph they want to see. The system displays the actual and predicted results of a sample. Also, the reports can be exported.

### **Initial Step by Step Description:**

- Decision Maker log in to the system with username and password.
- The decision maker can select Visualize Data.
  - The decision maker can choose parameters to see the graph.
    - These parameters can be a place, time and value.
    - If Decision Maker selects “Graph”:
      - The system shows a bar-chart value over time.
      - The graph can be exported as an image.
    - If decision Maker selects “Table”:
      - The system shows a table of selected place and time.
      - The graph can be exported as an image.
    - If decision Maker selects “Predicted Graph”:
      - The system shows a bar-chart value over time that ends with predicted value.
      - The graph can be exported as an image.

## **Performance Requirements**

### Response Time

Response time is highly dependent on internet connection speed. Considering an internet connection with at least 8Mbps, the system should respond to requests in less than 3 seconds.

### Workload

The system should support at least 15 concurrent users.

## **Software System Attributes**

### Availability

The software will run and be available as long as the hardware and the Docker environment it runs on works correctly. In a crashing event such as a system crash, if the database is not corrupted, software will run without problems when restarted.

## Security

Most subsystems in the monitoring software will not be connected to the internet, since the users will be able to use the software through the local network of the agency.

The data entry system, however, must be connected to the internet so that the contractors are able to enter the new measurements after they're taken. Since this data entry system can be used as an attack vector, we will validate the user input rigorously using both custom software and Django's validation module.

## Maintainability

Since the database is independent from the server software, administrators are free to update them, as long as they're compatible. The agency will provide a development environment for the testing of WQPMS, this will allow us to make sure the software will work for many years into the future with as little maintenance as possible.

## Usability

The GUI will be designed to be similar to the previous workflow for both the Data Entry Operators and the Decision Makers. For the decision makers, charts will be designed to be similar to the charts contained in previous years' reports. For the data entry operators, since the previous workflow involved Microsoft Access, a similar Table-like data entry UI will be produced.

# Software Design Description

## Introduction

### Purpose

The Water Quality Prediction and Monitoring System aims to visualize and predict water quality parameters from a given water sample. Collected data from various different rivers, seas, lakes and underwater sources are uploaded to the database by data entry operators. Using the collected data, future values of water quality parameters will be predicted and visualized along with the present parameter readings for use by the decision makers.

### Scope

The scope of this document is to elucidate the essential components of the system. Existing data structure of administrator, data entry operator and decision maker defined on SRS document will be used. This document comprises WQPMS design principles with its specifications, functionalities and meanings.

### Glossary

The Agency	The Ministry of Environment and Urban Planning
WQPMS	Water Quality Prediction and Monitoring System
WQP	Water Quality Prediction
ML	Machine Learning
GUI	Graphical User Interface

## Overview

- In the “Design Considerations” section, tools used while designing the system and why we decided to use them are explained.
- “Architecture” section includes software and hardware architecture diagrams.
- In the “System Interfaces” section, information about the database and application framework is given.
- “User Interface Design” section includes figures and their explanations about GUI design.
- “Process Design” section has use case and sequence diagrams for each type of user.
- “Database Design” section includes an ER diagram of the database and explanations of fields in tables.

## Design Considerations

## Approach

- Data will be broken into several different categories depending on water quality parameters, location and name of the sample that was taken from a specific water source.
- Attributes listed above that belong to each data sample will be stored in the database.
- Decision Makers will be able to add new parameters to be entered by the Data Entry Operators.
- The specific ML algorithm that will be used is going to be decided upon reviewing similar projects and research papers.
- The ML algorithm that achieves the highest accuracy, precision, recall and F1 score will be used for the prediction.
- User selected water quality parameters of a sample will be visualized as bar charts.

## **Tools Used**

- Bootstrap is a and CSS will be used for front-end and design of the webpage. Bootstrap saves a lot of time in terms of designing the site.
- Chart.js will be used to generate graphs.
- Django web framework will be used for back-end programming. Django provides useful features such as easy communication with the database, generating web pages from template files and so on.
- PostgreSQL will be used for persistent data storage. The database connection is handled by Django.
- Docker will be used for distributing the WQPMS to the Agency's servers as an image that needs a minimal amount of work to set up.
- Python and Tensorflow will be used for training and deploying Machine Learning and Deep Learning models to predict future values of water quality parameters.

## **Constraints**

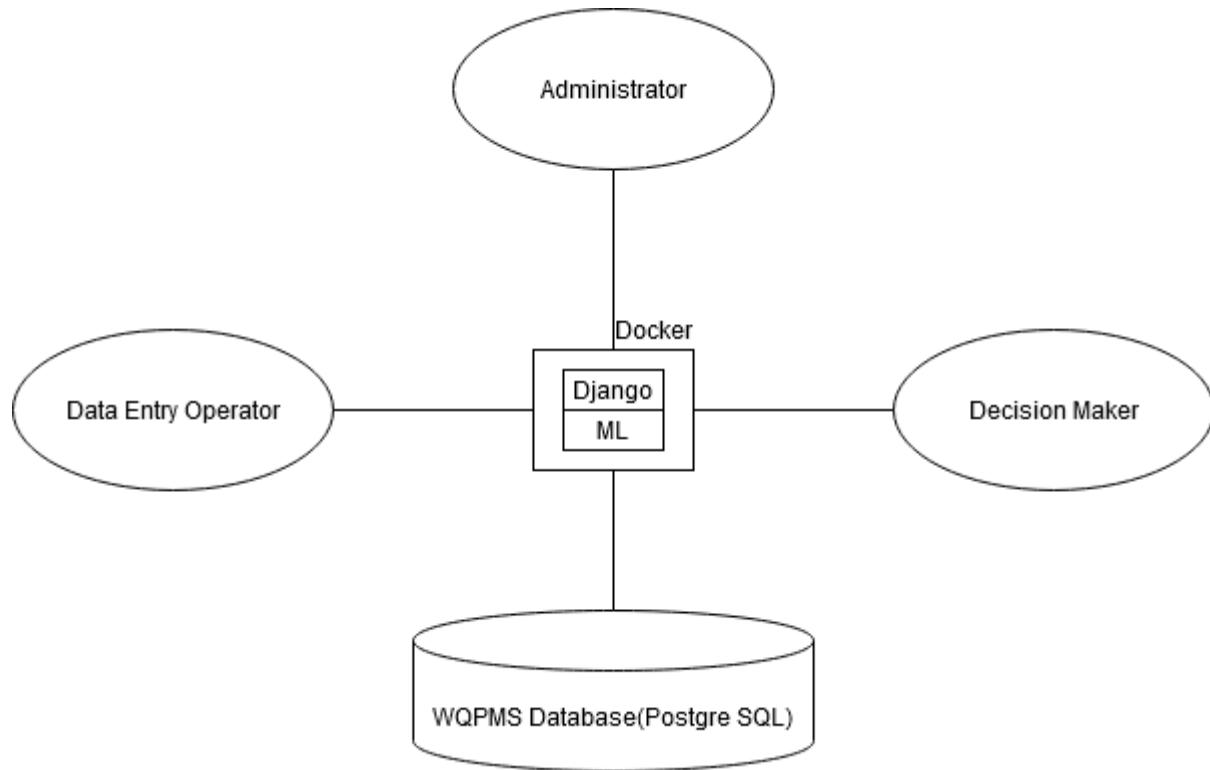
Data Entry Operator should not be able to see previously entered data.

## **Assumptions and Dependencies**

There are no assumptions nor dependencies for the design of WQPMS.

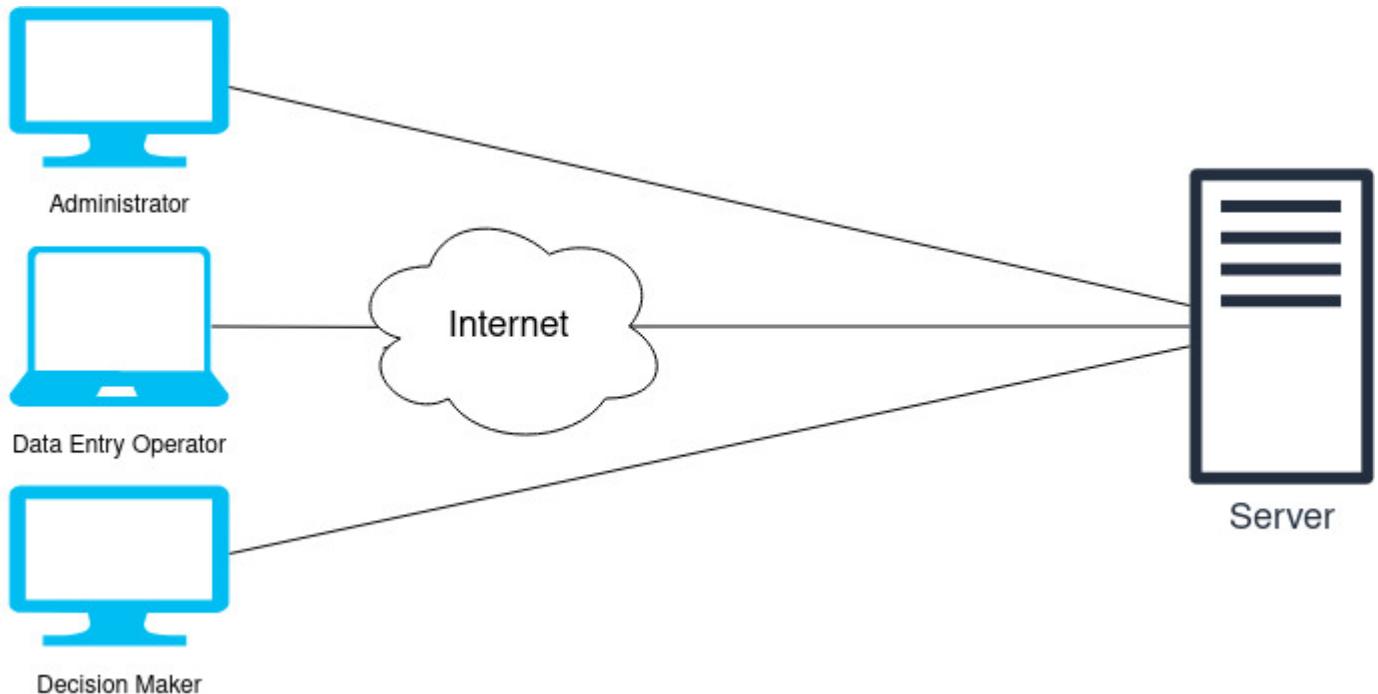
# Architecture

## Software Architecture



*Figure 3.1: Software Architecture*

## Hardware Architecture



*Figure 3.2: Hardware Architecture*

## System Interfaces

For persistent data storage, the system uses PostgreSQL 13.1. Django framework version 3.1 is used to serve web pages and communicate with the database. The system will be distributed with a SQL definition file detailing the first configuration of data tables that will be used to configure the persistent data storage by the administrator at the agency.

## External System Interfaces

There are no external system interfaces needed.

# User Interface Design

Water Quality Monitoring and Prediction System is a web-based recommendation system that has 3 different types of user. Decision Maker, Data Entry Operator and Admin. Therefore, system will include 3 different interfaces:

- An interface of visualizing and predicting data for only Decision Makers to see.
- An interface of entering data for only Data Entry Operators to see.
- An interface of managing users for Admin.

## Screen Definitions

### Login

The WQPMS is only reachable with a registered account. Accounts contain a username and a password. There are two input areas for that purpose and a button to log in. Accounts can only be created by an administrator. Therefore, there is no “Create a User” button. The logo and the name of the system which is “Water Quality Prediction and Monitoring System” is shown in the login page top of the input boxes.

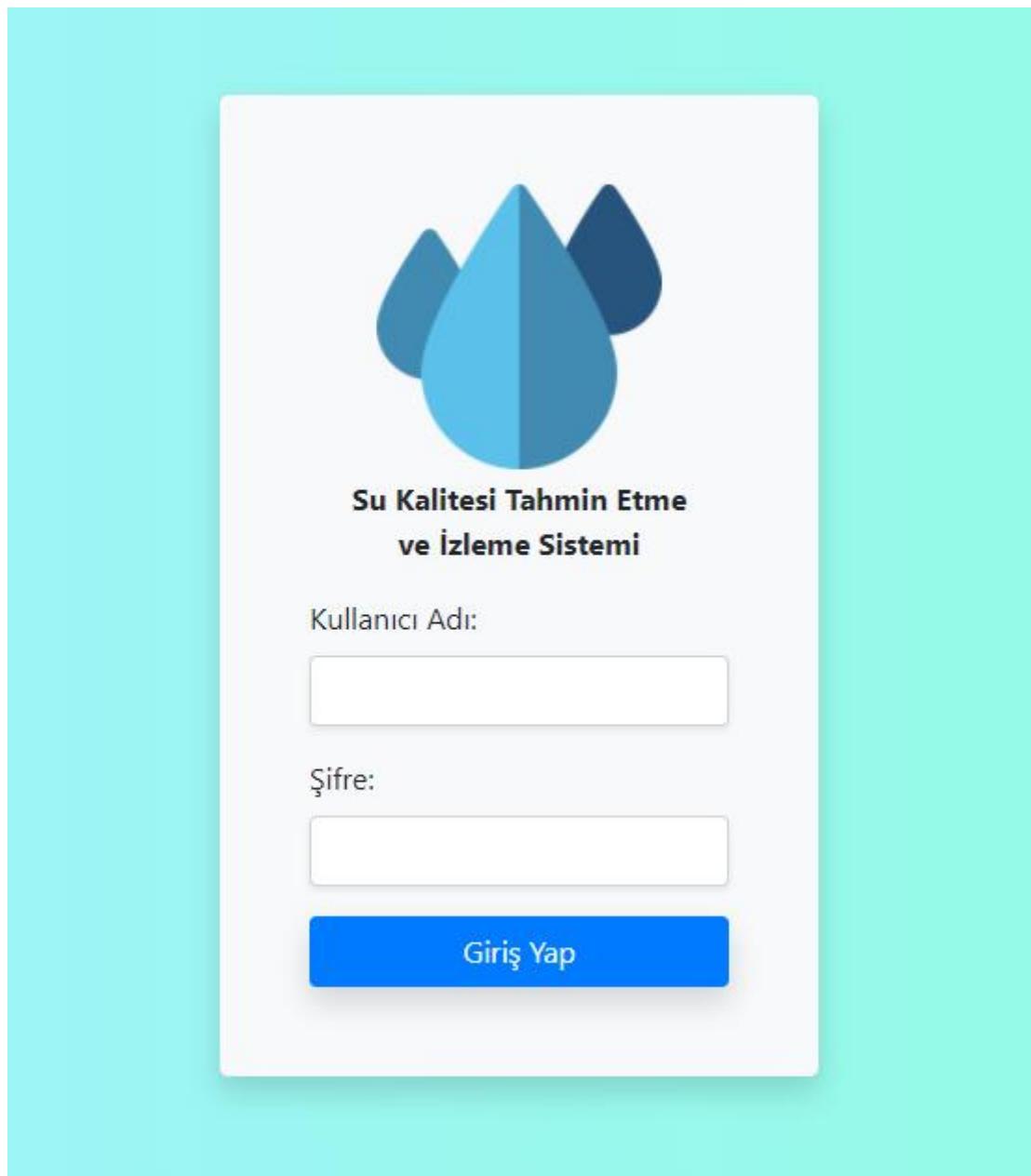


Figure 3.3: Login Page

## Decision Maker

The screenshot shows the 'Decision Maker' page. At the top left is the WQMP logo. To its right is a 'Çıkış Yap' (Logout) button. Below the logo are four dropdown menus: 'Su Kaynağı Seçiniz', 'Yer Seciniz', 'Deger Seciniz', and 'Yıl Seciniz'. A checked checkbox labeled 'Tahmin' (Prediction) is positioned between the third and fourth dropdowns. Below these controls are two blue links: 'Tabloyu Getir' and 'Grafik Ciz'. The main area of the page is a large, empty rectangular box.

Figure 3.4: Decision Maker Page After Successful Login

Decision Maker page is a single page that has 3 parts. Nav-bar, Search Box, Graph box.

First part is the nav-bar at the top of the page. There are the logo and “WQMP” on the left of the nav-bar and when clicked, it should clear the search box and graphs. On the right of the nav-bar there is a button “Log out”. When clicked the system should return to the login page and the user should be logged out.

The second part is the search box between the nav-bar and the graph box. In order to see the graph, table or the predicted value, decision makers must select the water source, region, location, value and the year. At first there is only one box for selecting a water source. Water source options are “River”, “Wastewater Treatment Plant”, “Sea”, “Lake”.

After selecting a water source, new selection boxes will appear. One of the boxes is for selecting the region that contains locations where samples are taken from selected water sources. The options for this box will be dynamically generated by getting relevant info from the database. Also, Decision Makers should be able to search a location by using the search box.

The other box is for selecting the water quality parameter. Decision Makers should be able to select more than one parameter. The parameters for different water sources may vary.

There is also a checkbox for prediction. When the checkbox is checked, predicted values are visualised in the graph. Therefore, there will be only one button for the drawing graph in the last box.

There is also a box to select the time interval that will be visualized. This box should not be shown if there is no entered data. In the last box there should be 2 buttons for bringing graphs or tables of selected options.

The third part is the graph box. System should draw the graph or show a table in this box. There should also be a button for exporting the graph or table as an image.

### Data Entry Operator

The screenshot shows a web-based data entry interface. At the top left is a logo with the letters 'WQMP'. To its right is a dropdown menu set to 'Akarsu'. On the far right is a dark grey button labeled 'Çikis Yap' (Logout). Below this header is a large form area with several groups of input fields:

- Location:** 'Bölge Adı:' (Region Name) and 'pH:' (pH) input fields.
- Sample:** 'Numune Adı:' (Sample Name), 'Sıcaklık (°C):' (Temperature (°C)), 'Askida Kati Madde (mg/L):' (Solid Substances in Water (mg/L)), and 'Fekal Koliform (CFU/100mL):' (Faecal Coliform (CFU/100mL)) input fields.
- Time:** 'Tarih:' (Date) with a date picker, 'Yer:' (Location), and 'E:' (Address) input fields.
- Chemical Parameters:** 'Çözünmüş Oksijen (mg/L):' (Dissolved Oxygen (mg/L)), 'Amonyum Azotu (mg/L):' (Ammonium Nitrogen (mg/L)), 'Nitrit Azotu (mg/L)' (Nitrite Nitrogen (mg/L)), 'Nitrat Azotu (mg/L):' (Nitrate Nitrogen (mg/L)), 'Toplam Kjeldahl Azotu (mg/L):' (Total Kjeldahl Nitrogen (mg/L)), 'Elektriksel İletkenlik (µS):' (Electrical Conductivity (µS)), and 'Toplam Fosfor (mg/L):' (Total Phosphorus (mg/L)) input fields.
- Other:** 'Toplam Koliform (CFU/100mL):' (Total Coliform (CFU/100mL)), 'Kimyasal Oksijen İhtiyacı (KOJ) (mg/L):' (Chemical Oxygen Demand (mg/L)), 'Orto Fosfat (mg/L):' (Orthophosphate (mg/L)), 'Toplam Pestisit:' (Total Pesticides), and 'Debi (m<sup>3</sup>/sn):' (Discharge (m<sup>3</sup>/s)) input fields.

At the bottom center of the form is a blue button labeled 'Veriyi Gir' (Enter Data).

Figure 3.5: Data Entry Operator Page With Selected “River” to Enter Data

The Data Entry Operator page is also a single page with 3 parts. Nav-bar, Water Source Selection, Data Entry Box. After the Data Entry Operator logged in to the

system, there should be the nav-bar at the top of the page and the water source selection box.

The Nav-bar should be at the top of the page with similar features in Decision Maker Page. There are the logo and name of the system at the left of the nav-bar and “Logout” button at the right of the nav-bar.

The Water Source Selection part only contains a selection box. Data Entry Operators should select between the options “River”, “Wastewater”, “Sea”, “Lake”. After selection of the water source, the system brings the data entry box.

The Data Entry Box part will change according to the selected water source. The input may change between different water sources. Therefore, the system should bring predetermined data entry boxes. Also, in order to prevent wrong input all the inputs should be checked to see the entered value is between the determined interval. If it is not in the given internal system should give warning about it that shows the acceptable interval. After every data has been entered correctly, the Data Entry Operator clicks the “Enter Data” button in the bottom of the data entry part.

## Admin

Admin interface is automatically generated by Django. Since it is automatically generated, it is subject to constant change.

# Process Design

## Use Cases

Administrator Use Case

**Actor :** Administrator

**Use Case:**

- Manage Users
  - Create User
  - Delete User
- Manage User Permissions

**Diagram:**

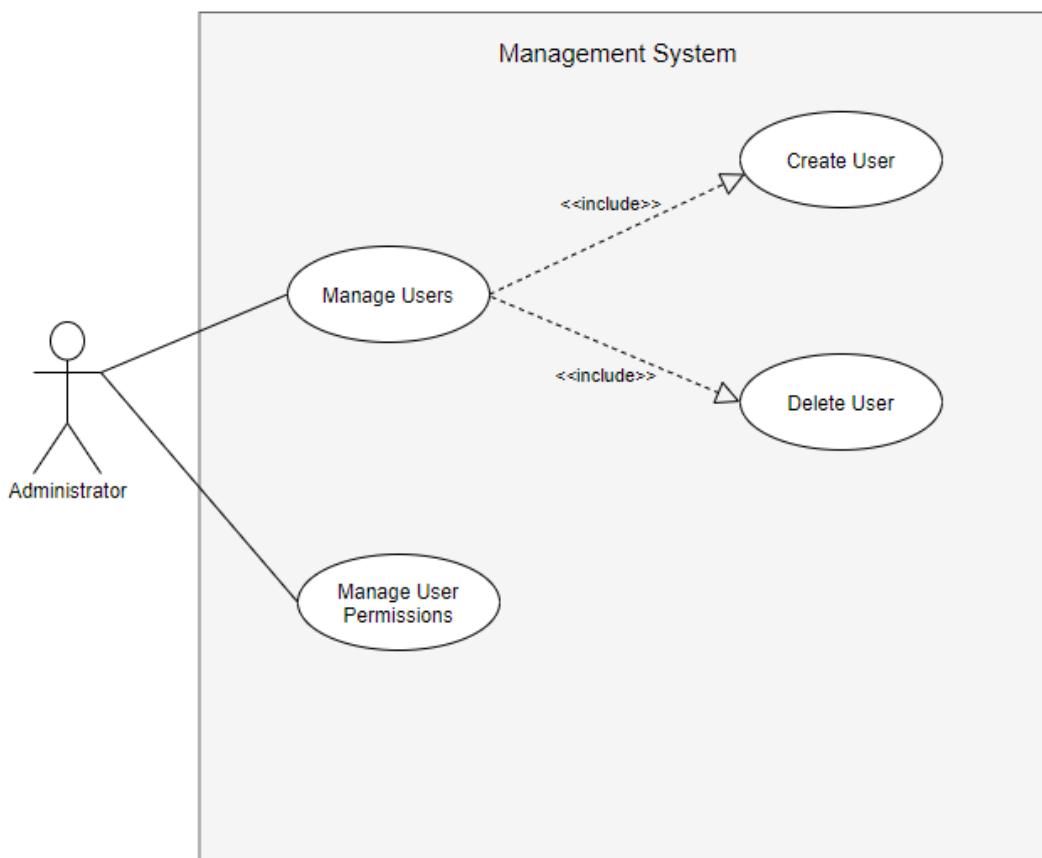


Figure 3.6 : Management System Use Case Diagram

**Brief Description:**

Figure 3.6 shows the management system use case diagram. The administrator has authority to manage the system. Administrators are able to add and delete users and manage user permissions.

**Initial Step by Step Description:**

- Users can not register.
- Administrators are able to create users.
- Administrators can give users specific permissions and prohibitions.
  - Administrators can assign the user to the data entry operators group or decision makers group.
- Administrators can delete users.

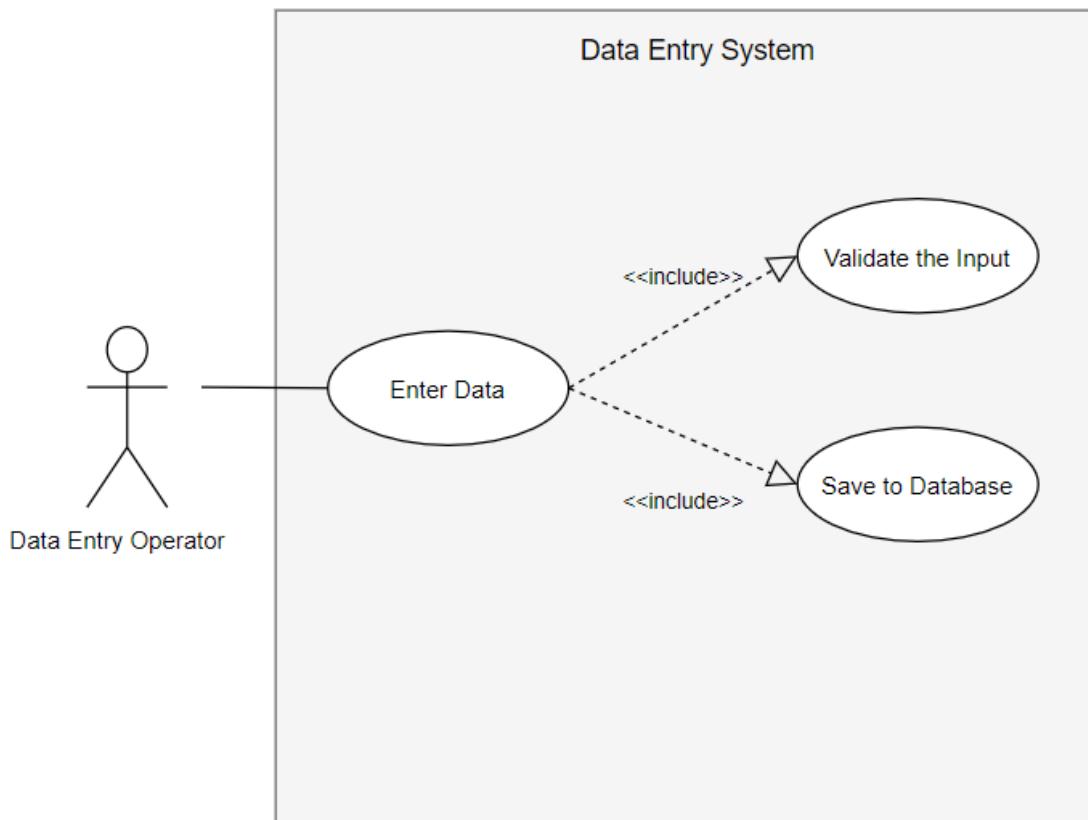
## Data Entry Operator Use Case

**Actor:** Data Entry Operator

**Use Case:**

- Enter Data
  - Validation
  - Save to Database

**Diagram:**



*Figure 3.7 : Data Entry System Use Case Diagram*

**Brief Description:**

Figure 3.7 shows the data entry system use case diagram. Data entry operators do not have the same permissions as Decision Makers, therefore they can only enter data in appropriate format to aid decision making.

**Initial Step by Step Description:**

- Data entry operator can insert data to system
  - If data is not entered in the proper format, a request will be sent to a user to enter the data in proper format.
  - If data is in proper format, data can be inserted to the database.

## Decision Maker Use Case

**Actor:** Decision Maker

**Use case:**

- Visualize Data
  - Show Graph
    - Export
  - Show Table
    - Export
  - Show Predicted Graph
    - Export

**Diagram:**

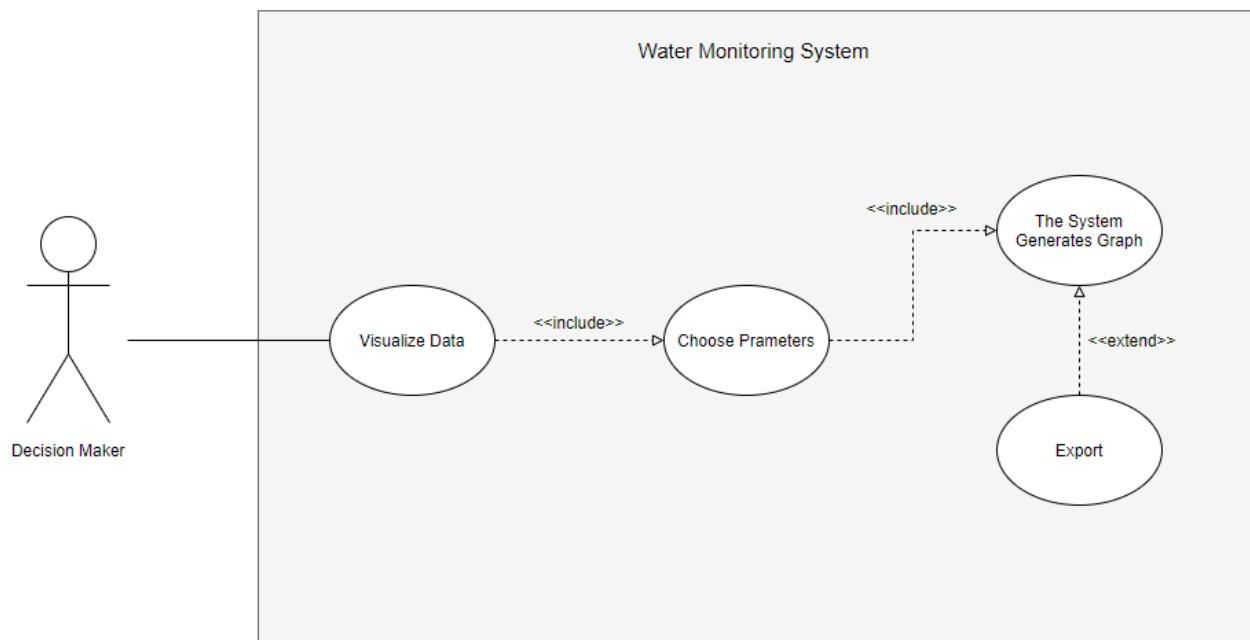


Figure 3.8: Decision Maker Use Case

**Description:** Decision Makers are allowed to use the system for visualizing selected samples. The graph of a sample includes, time, water quality parameters and prediction results of the selected sample. As you can see in Figure 3.8, Decision Makers select which graph they want to see. The system displays the actual and predicted results of a sample. Also, the reports can be exported.

### **Initial Step by Step Description:**

- Decision Maker log in to the system with username and password.
- The decision maker can select Visualize Data.
  - The decision maker can choose parameters to see the graph.
    - These parameters can be a place, time and value.
    - If Decision Maker selects “Graph”:
      - The system shows a bar-chart value over time.
      - The graph can be exported as an image.
    - If decision Maker selects “Table”:
      - The system shows a table of selected place and time.
      - The graph can be exported as an image.
    - If decision Maker selects “Predicted Graph”:
      - The system shows a bar-chart value over time that ends with predicted value.
      - The graph can be exported as an image.

## Sequence Diagrams

Administrator Scenario Sequence Diagram

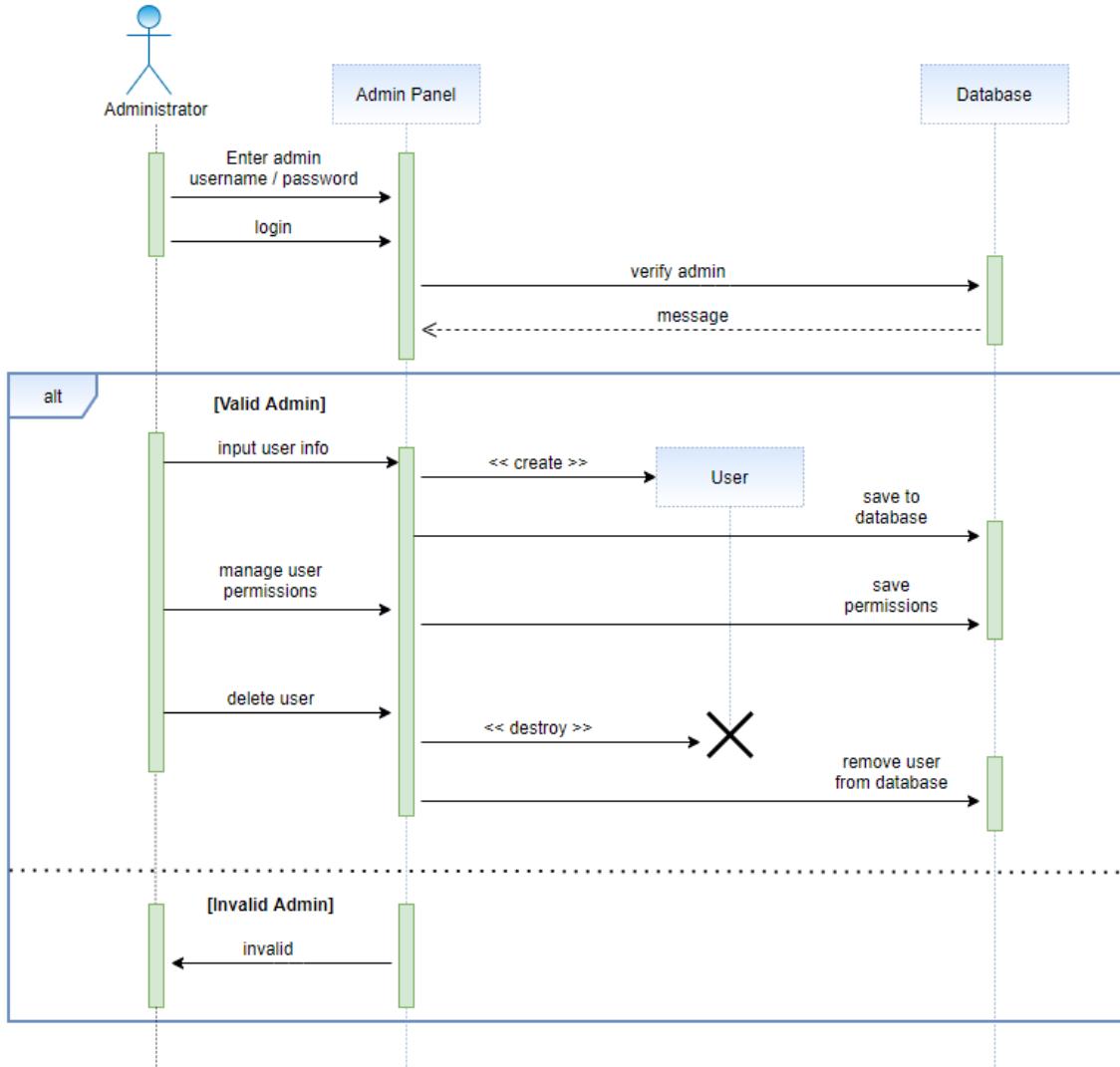
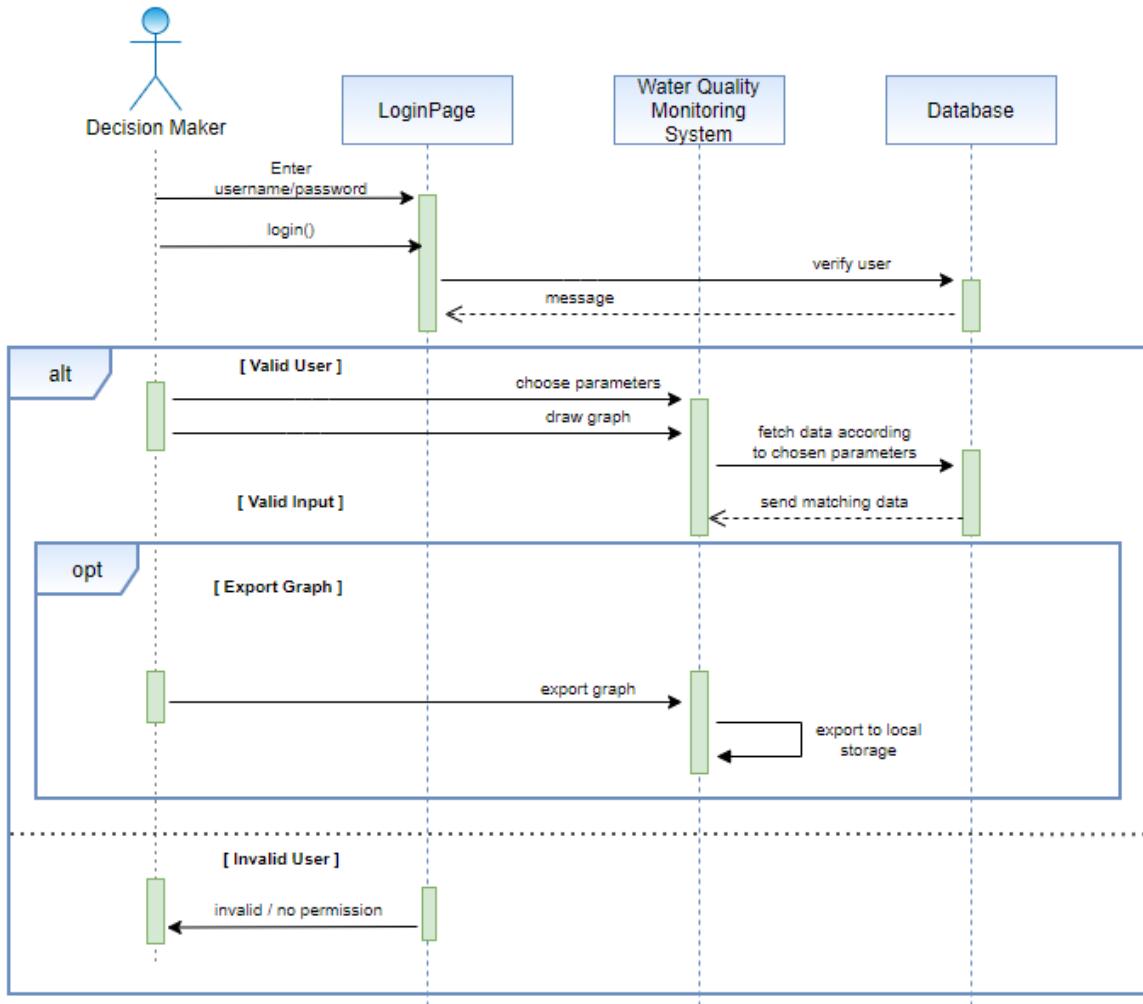


Figure 3.9: Administrator Scenario Sequence Diagram

## Monitoring Scenario Sequence Diagram



*Figure 3.10: Monitoring Scenario Sequence Diagram*

## Data Entry Scenario Sequence Diagram

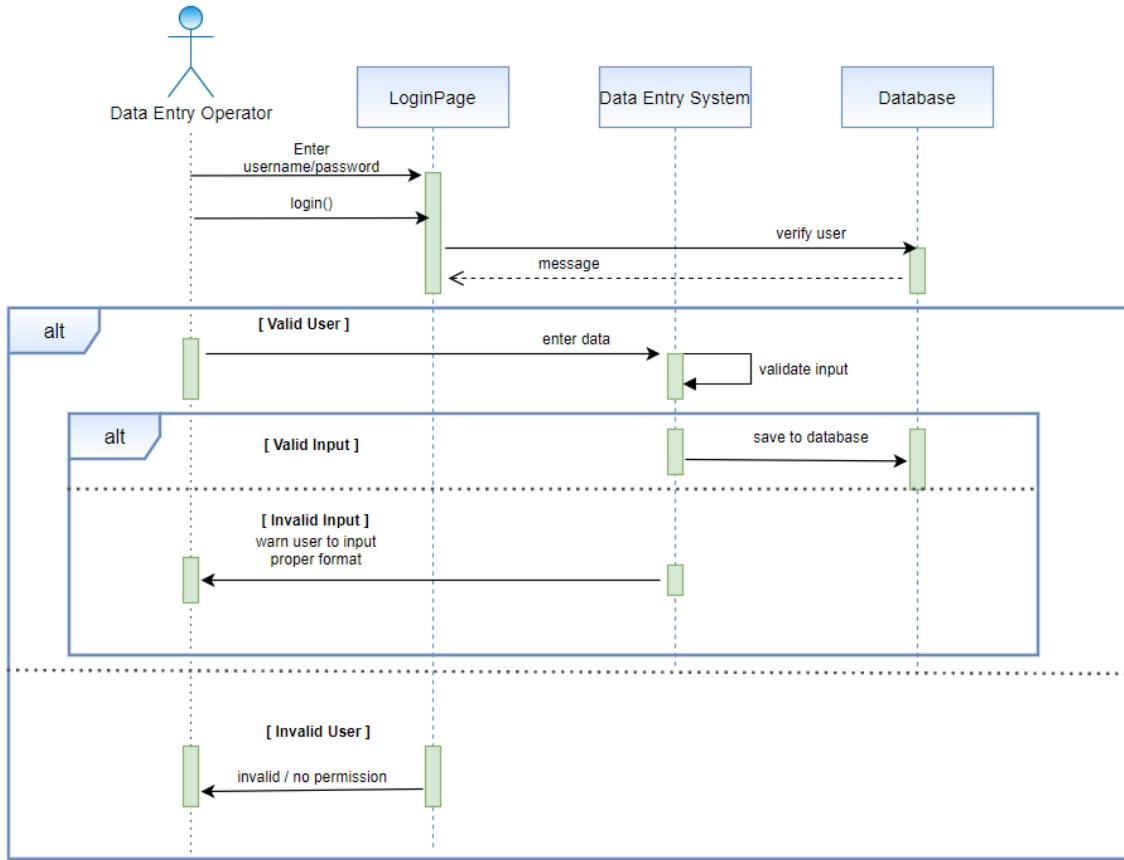


Figure 3.11: Data Entry Scenario Sequence Diagram

## Database Design

There are 4 tables in the planned database of the system. The tables are User, Reading, Location and Reading Type. Their content is explained in their respective table definition sections.

## Table Definitions

User

Field	Type	Description
id	AutoField	Auto generated positive integer id for table entry
email	EmailField	Email address for the user
first_name	CharField	First name of the user
last_name	CharField	Last name of the user
username	CharField	Username of the user
password	CharField	Password of the user
last_login	DateTimeField	Last login time
is_dmaker	BooleanField	True if user is a decision maker
is_dataent	BooleanField	True if user is a data entry operator
is_superuser	BooleanField	True if user is an administrator

Reading

Field	Type	Description
id	AutoField	Auto generated positive integer id for table entry
added_by	ForeignKey(id)	FK to User table
location	ForeignKey(id)	FK to Location table
reading_type	ForeignKey(id)	FK to ReadingType table
date	DateField	The date that reading has been taken
reading_string_value	TextField	The reading in string format
reading_value	FloatField	The reading in floating point format, if possible

## Location

<b>Field</b>	<b>Type</b>	<b>Description</b>
id	AutoField	Auto generated positive integer id for table entry
bolge_adi	CharField	The general location name
numune_adi	CharField	The code for a reading taken at this location
yer	CharField	Specific location
utm_x	PositiveIntegerField	Universal Transverse Mercator x coordinate
utm_y	PositiveIntegerField	Universal Transverse Mercator y coordinate

## Reading Type

<b>Field</b>	<b>Type</b>	<b>Description</b>
id	AutoField	Auto generated positive integer id for table entry
name	CharField	Can be “pH”, “Toplam Koliform”, etc.
max_value	FloatField	Maximum value for a reading of this type
min_value	FloatField	Minimum value for a reading of this type

## ER Diagram

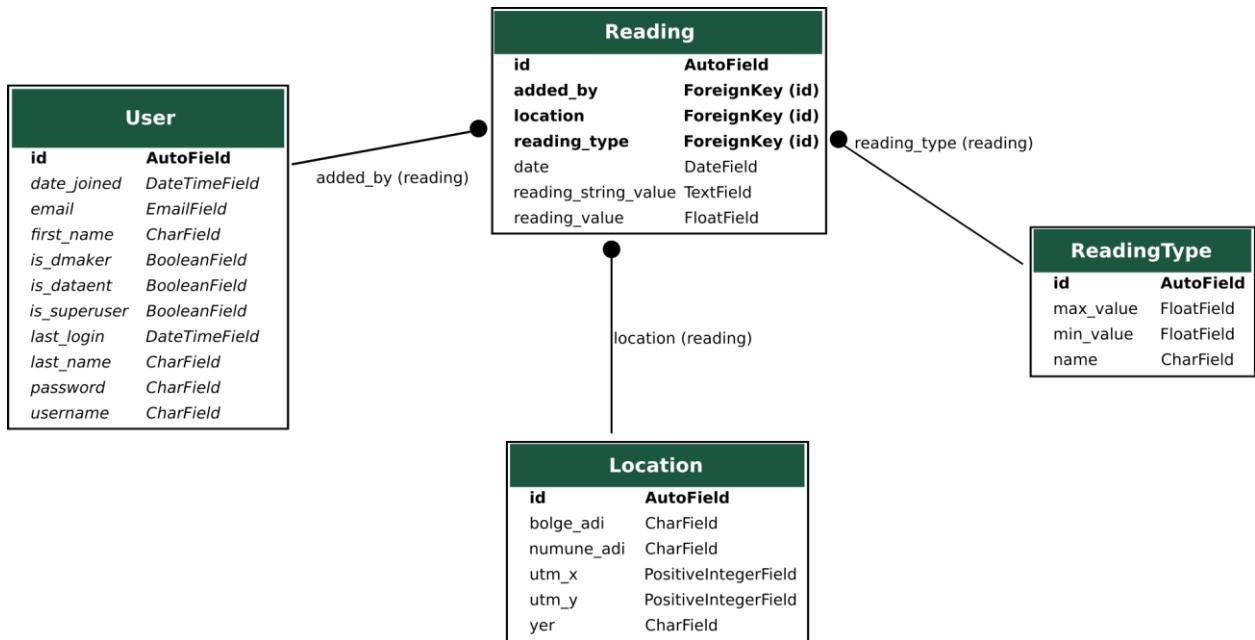


Figure 3.12: ER Diagram

## Conclusions

Our project will provide a user-friendly GUI and overall better visualizations than previously used workflow at the Ministry of Environment and Urban Planning. WQPMS will also predict future water quality parameters for a given sample using machine learning algorithms. With better visualizations and an accurate water quality prediction, our system will allow users to easily understand the data, and make accurate decisions concerning the water quality in Turkey in the future.

# References

- [1] APEC. 'The History of Clean Drinking Water', 2018. [Online]. Available: <https://www.freerdrinkingwater.com/resource-history-of-clean-drinking-water.htm> [Accessed: 2020/11/01]
- [2] Minnesota Department of Health, 'Bacteria, Viruses, and Parasites in Drinking Water', 2019. [Online PDF]. Available: <https://www.health.state.mn.us/communities/environment/water/docs/contaminants/parasitesfactsht.pdf> [Accessed: 2020/11/01]
- [3] A. N. Ahmed, F. B. Othman, H. A. Afan, R. K. Ibrahim, C. M. Fai, M. S. Hossain, M. Ehteram, and A. Elshafie, "Machine learning methods for better water quality prediction," *Journal of Hydrology*, vol. 578, p. 124084, Aug. 2019.
- [4] J.-T. Kuo, M.-H. Hsieh, W.-S. Lung, and N. She, "Using artificial neural networks for reservoir eutrophication prediction," *Ecological Modelling*, vol. 200, no. 1-2, pp. 171–177, 2007. Retrieved from: <https://www.sciencedirect.com/science/article/abs/pii/S0304380006002985?via%3Dhub>
- [5] A. Zaqoot, A. K. Ansari, M. A. Unar, and S. H. Khan, "Prediction of dissolved oxygen in the Mediterranean Sea along Gaza, Palestine – an artificial neural network approach," *Water Science and Technology*, vol. 60, no. 12, pp. 3051–3059, 2009. Retrieved from: <https://iwaponline.com/wst/article-abstract/60/12/3051/13774/Prediction-of-dissolved-oxygen-in-the?redirectedFrom=fulltext>
- [6] Sengorur, B , Dogan, E , Koklu, R , Samandar, A . "Dissolved Oxygen Estimation using Artificial Neural Network for Water Quality Control", *Electronic Letters on Science and Engineering* 1 pp. 13-16, 2005. Retrieved from: <https://dergipark.org.tr/en/pub/else/issue/29326/313793>
- [7] L. Xu and S. Liu, "Study of short-term water quality prediction model based on wavelet neural network," *Mathematical and Computer Modelling*, 22-Dec-2012. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0895717712003676>. [Accessed: 01-Nov-2020].
- [8] Tan, G., Yan, J., Gao, C. and Yang, S. "Prediction of water quality time series data based on least squares support vector machine", *Procedia Engineering*, 31, pp.1194-1199. 2012.
- [9] Unwin, A. (2020). Why is Data Visualization Important? What is Important in Data Visualization? *Harvard Data Science Review*, 2(1). Retrieved from: <https://doi.org/10.1162/99608f92.8ae4d525>

- [10] Ramsay, Ian & Shen, S. & Tennakoon, S.. (2009). Water Quality Visualisation and Tracking - Generic Decision Support Tool. Retrieved from:  
[https://www.researchgate.net/publication/237627349\\_Water\\_Quality\\_Visualisation\\_and\\_Tracking\\_-\\_Generic\\_Decision\\_Support\\_Tool](https://www.researchgate.net/publication/237627349_Water_Quality_Visualisation_and_Tracking_-_Generic_Decision_Support_Tool)
- [11] "IEEE 1016-2009 - IEEE Standard for Information Technology--Systems Design--Software Design Descriptions", Standards.ieee.org, 2020. [Online]. Available: <https://standards.ieee.org/standard/1016-2009.html>. [Accessed: 01-Dec- 2020].
- [12] "IEEE 830-1998 - IEEE Recommended Practice for Software Requirements Specifications", Standards.ieee.org, 2020. [Online]. Available: <https://standards.ieee.org/standard/830-1998.html>. [Accessed: 01- Dec- 2020].