Chap 1 Introduction

This chapter elaborates on the introduction to the problem that will be faced in the future by humans in terms of scarcity of resources and how machine learning can be used to address the problem.

1.1 Background

Technology has created several avenues of development in several fields and has made daily lives simpler by solving many problems which arise on an everyday basis. This has become more apparent in the past few years with the advent of machine learning and its widespread use in several everyday applications.

The availability of soil moisture is a determinant cause of plant growth and ultimately affects the productivity of plants. It is seen that a balance of soil moisture in the plant is crucial because very little moisture causes loss of plant or the death of the plant whereas an excess of water in the plant can cause wastage of water and plant diseases. Conservation of water has become the need of the hour for people all over the world and the demand for water-saving processes has increased exponentially. Excess water also affects the soil and can cause it to lose its nutrients which ultimately affect the plant. Water is also a delivery mechanism for the nutrients which are not tightly bound to the soil. The nutrients are provided to the plant through the irrigation system and the movement of the water defines how the nutrients are delivered to the roots of the plant. It is safe to say that efficient and proper management of water is also the proper management of plant nutrients.

The need and importance for growing plants in the backyard has increased because future environmental factors predict that many people will need to grow their own food for sustenance. It extends beyond a hobby and emphasizes the importance of systems that are considered ‘smart’ enough to tackle various aspects of urban farming practices. Several gardeners or residents lack the complete knowledge about the plants they grow and the nutrient and water requirements of the same. Plants can catch air-borne, water-borne, or soil-borne diseases that may go undiagnosed in the absence of professional help. AI, today, can detect the real problem and help maintain the greenery.

Through agricultural robotics, monitoring of soil and the plants, and by making data supported predictions, AI is emerging in the field of farming and agriculture. Soil data and crop yield data is stored and analysed to generate algorithms that decide the water and light exposure to the plant. As a future threat, the world population is expected to increase to 9 billion people by the year 2050 and to assure food security to all; food production must rise by 70% ensuring least amount of environmental impact and by using renewable energy. Making AI technology in farming and gardening accessible and affordable at the grassroots level will be another ground-breaking evolution of humanity.

Growing vegetables, medicinal plants, herbs, fruits, etc. at home using AI-powered machines are characteristics of an urban farmer. The aim is to hyper localise food production irrespective of environmental temperature, negligible farming and gardening expertise and in less time.

A system the waters the plant only when required taking into account various factors ensures that water is conserved. Checking the moisture content of the soil prior to watering the plant opens up avenues of ways to battle the scarcity of water and making the entire process green. In a scenario where water is bound to be depleted, the need for a futuristic system that limits the utilization of water and saves the crucial resource prudently along with money and time has become of increasing importance lately. By predicting the amount of water needed in a simple household plant, it becomes easy to safeguard the water usage of the family and also ensure healthy growth of the plant.

Supervised learning is a type of system in which both input and desired output data are provided. Input and output data are labelled for classification to provide a learning basis for future data processing. Training data for supervised learning includes a set of examples with paired input subjects and desired output (which is also referred to as the supervisory signal). After a sufficient amount of observation, the system should be able to distinguish between and categorize unlabelled images, at which time training can be said to be complete.

Regression analysis is a sub field of supervised learning equations are a crucial part of the statistical output after you fit a model. The coefficients in the equation define the relationship between each independent variable and the dependent variable. However, one can also enter values for the independent variables into the equation to predict the mean value of the dependent variable. Regression models consist just one independent and one or more dependent variables. The variable whose value is to be predicted is known as the dependent variable and the one whose known value is used for prediction is known as the independent variable. The difference is that while correlation measures the strength of an association between two variables, regression quantifies the nature of the relationship.

The use of machine learning algorithms to predict the amount of water required to sustain the plant based on soil moisture readings from the moisture sensor and temperature and humidity readings of the surrounding area, this project presents a novel method of maintaining the water content of household plants and gardens which is useful in the long run. The dataset collected consists of temperature, soil moisture, water level, and humidity readings which are first visualize using data visualization methods and then fed into supervised learning algorithms which predict the amount of water to be added to the system. The readings of the moisture content in the plant are sent to the prediction system wirelessly and the predicted readings are returned to the watering system which automatically waters the plant. The prediction has been computed after rigorous comparison between 4 algorithms and the most robust prediction of the water level required has been selected and sent to the watering system. The prediction models utilized are- simple linear regression model, logistic regression model, SVM regressor model, and KNN regressor model.

1.2 Motivation

The motivation for the project stems from a deep intrigue and interest in the field of machine learning and its utilization to provide solutions to a problem faced in our daily lives. The idea was to work on plants and find a problem that directly affects plant health causes them to wilt indefinitely due to negligence or improper knowledge of water and its effect on plant nutrients and addressing the same through machine learning algorithms. Collection of dataset values practically through the use of sensors and APIs was the first step.

An intermediate step was the wireless communication of dataset values that can aid in making the system automated. Our next step was to collect information about suitable machine learning algorithms that will fit the requirement of the project and fill the scope of the project adequately and perform a rigorous comparison to obtain the best results was necessary. The problem of over-watering of plants and their subsequent disease-ridden state or depreciation of soil quality has plagued several residents and horticulture beginners for a long time. The need for a system that can eliminate the problem of over-watering which maintains plant health and is cost-friendly and reliable was something that caught our eye and made us work in that direction. Our discussion was mainly led by the fact that the project should bridge the gap between theory and practice and should address a daily life problem tactfully. Creating a watering system that was carefully calibrated and automated to prevent the wastage of water and conserve the nutrients of the soil for healthy growth of the plant was our next target. An automated system is perfectly timed and incorporates the careful restrictions that are necessary for a system to work successfully. An implementation of a fully functional module that has every intricate aspect of the system working adequately in real-time with sufficient real-time results was our final goal.

1.3 Objective

This project has been devised to incorporate the discipline of machine learning to the dataset values crucial to plant growth transmitted wirelessly to the system to obtain adequate results which can be further processed fed into an automated watering system. The objectives of the project include-

- To develop a system that is capable of predicting the water level required for the healthy growth of a plant.
- In order to detect this amount of water, the system should be capable of extracting target parameters such as temperature, humidity, soil moisture.
- The proposed system employs machine learning to facilitate the prediction of water level with the inputs being these target features.

- Also, the system should be robust and efficient against different seasons faced by the plant and contribute towards water conservation.
- The proposed system should be automated and consist of a watering system that makes it an independent product which leaves the user free of worry and with healthy plants.

1.5 Scope of the project

- Database collection for the three target parameters over a month.
- Wireless communication between the Node MCU and the machine learning algorithm.
- Prediction of output water level using multiple machine learning algorithms and their comparison.
- Automated plant watering system.

1.6 Brief description of the project

Supervised learning is a type of system in which both input and desired output data are provided. Input and output data are labelled for classification to provide a learning basis for future data processing. Training data for supervised learning includes a set of examples with paired input subjects and desired output (which is also referred to as the supervisory signal). After a sufficient amount of observation, the system should be able to distinguish between and categorize unlabelled images, at which time training can be said to be complete. Regression analysis is a sub field of supervised learning.

Regression equations are a crucial part of the statistical output after you fit a model. The coefficients in the equation define the relationship between each independent variable and the dependent variable. However, one can also enter values for the independent variables into the equation to predict the mean value of the dependent variable. Regression models consist just one independent and one or more dependent variables. The variable whose value is to be predicted is known as the dependent variable and the one whose known value is used for prediction is known as the independent variable. The difference is that while correlation measures the strength of an association between two variables, regression quantifies the nature of the relationship.

- The plant selected for testing and analysis is Periwinkle (scientific name- *Vinca*, family-*Apocynaceae*) plant which is type of periwinkle.
- The dataset collected after months of experimentation consists of the humidity, temperature, soil moisture, and water level readings. The soil moisture values are computed through soil moisture sensors placed inside the plant. These readings are transmitted to the prediction model with the help of Node MCU esp8266 board.
- These features are then visualized through data visualization methods such as box plots, 3D plots,

scatter plots, histograms, and heat maps.

- The dataset is then divided into training, testing, and cross validation datasets which are used to increase the accuracy of the system. This data is then fed into the prediction system which predicts the values of the amount of water to be added to the plant with the help of several distinct prediction models. These models include simple linear regression model, logistic regression model, SVM regressor model, and KNN regressor model.
- The predicted value of the amount of water to be added is then sent to the automated watering system which begins watering the plant based on careful calibration of the water levels and ensures the prevention of over-watering of the plant.
- The use of machine learning algorithms to predict the amount of water required to sustain the plant based on soil moisture readings from the moisture sensor and temperature and humidity readings of the surrounding area, this project presents a novel method of maintaining the water content of household plants and gardens which is useful in the long run.

Chapter 2

Literature Survey

This chapter explores on various works done in the field of Machine Learning Algorithms and relations of environmental conditions on plants.Various international papers have been explored.

The environmental conditions like temperature, humidity, soil moisture have an influence on the amount of water to be given to a plant. Various works have been done in the field of making predictions using a dataset and how environmental conditions affect water needs of plant which has helped us in development of prediction based plant watering system.

1. In the paper, Monitoring plant and soil water status: established and novel methods revisited and their relevance to studies of drought tolerance, Hamlyn.G.Jones says that enhanced inputs from environmental plant physiologists could benefit molecular studies for agriculture purposes, measurement of water status along with environmental stress gives us a better information about the water movement in plant system. [1]

2. In the paper, Effects of salinity and relative humidity on growth and ionic relations of plants, M.Salim has conducted experiments on mung bean, sunflower, red kidney bean, and tomato and spongiosa plants by varying concentration of NaCl and relative humidity. Fresh weight/dry weight ratio of shoots increased when relative humidity increased. In sunflower and tomato plants shoot/root ratio increased with increase in relative humidity. Increased relative humidity increased relative growth in spongiosa, mung bean and sunflower plants. [2]

3. In the paper, Interactive effects of low atmospheric CO2 and elevated temperature on growth, photosynthesis and respiration in Phaseolus vulgaris , S. A. Cowling & R. F. Sage have analysed growth of French bean plant and found that increasing growth temperature under ambient CO2 concentration produced symptoms of stress on plant.[3]

4. In the paper, IoT based smart soil monitoring system for agricultural production Dr.N.Ananthi has tested the soil using various sensors. The obtained sensor values are sent to a field manager and crop suggestions are made. This system is helping farmers increasing agricultural production and also saves money and time of farmers. [4]

5. In this paper, IoT based low cost smart irrigation system , Kiranmai Pernapati has used temperature and humidity sensors to sense water vapour content and temperature around the plant. She has used soil moisture sensor to sense the soil moisture of a plant. Her system checks if the water content is below minimum requirement of the plant and supplies water from water reservoir using relay.[5]

6. The paper, Design and Implementation of an Automatic Irrigation Feedback Control system based on Monitoring of Soil Moisture is dealing with the development in agricultural activities of irrigation, soil temperature and moisture measurement using digital technologies and wireless sensors.[6]

7. The paper, Application of Indoor Temperature Prediction Based on SVM and BPNN is predicting indoor temperature for creating an ambient indoor environment. The support vector machine model and back propagation neural network model of temperature prediction is given here.[7]

8. The paper, SVM with improved grid search and its application to wind power prediction, says that selecting prediction model is very important for wind power prediction. Support Vector Machine model is used with an improved grid search method to optimize the required parameters. It is able to predict real-time wind power. [8]

9. In this paper, On stock prediction based on KNN-ANN algorithm, San-hong Liu has stated that to make accurate stock predictions a systematic predicting model based on KNN an BP Neural Network should be made. KNN-ANN algorithm are smaller than those in KNN algorithm and can do better in stock prediction. [9]

Chapter 3

Proposed System

This chapter provides analysis of the way the project was conceptualised .The problem statement has been tackled and the project has been explained thoroughly. The block diagram provides a detailed explanation for the same

3.1 Introduction

Plants are a vital part of our ecosystem and need nutrients in soil and water to survive. Amount of water they need depend on many atmospheric factors and condition of soil. Soil moisture, atmospheric temperature, atmospheric humidity are few of such factors. Our project tries to predict the amount of water required by the plant based on these factors. We are using temperature, humidity, soil moisture of a vinca plant to predict how much water is to be put in plant for keeping it healthy. We are using various machine learning algorithms for obtaining the required result. We are creating a database which stores soil moisture, atmospheric temperature, atmospheric humidity and water added on a particular day for many number of days. We are using this database to train our machine learning algorithm. We are making use of supervised learning in machine learning. We are using regression based analysis for predicting water value to be added in plant.

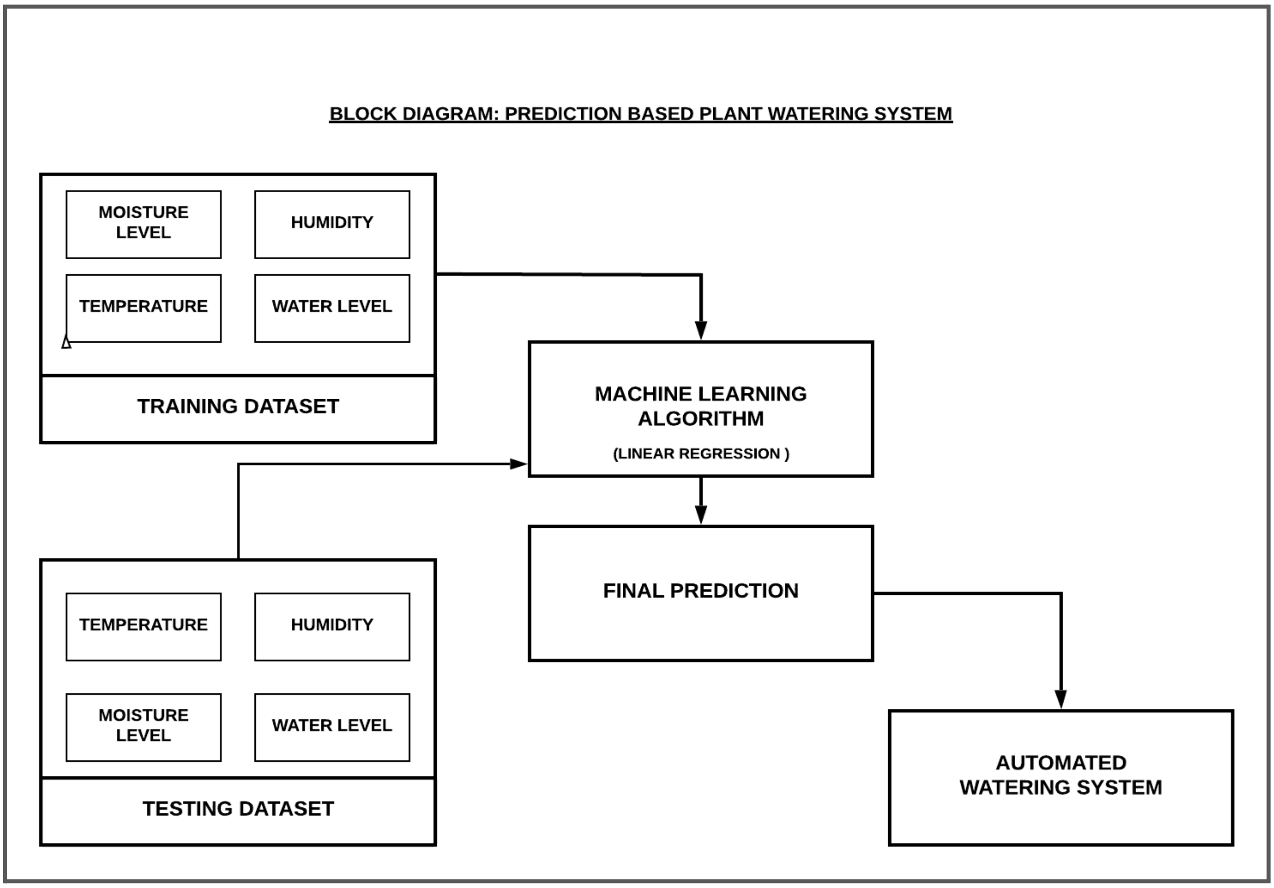


Figure 3.1 :Block Diagram for prediction based plant watering system

3.2 Problem Statement

Plants need water, soil and nutrients to grow. Providing vinca plant less will result in death of the plant. Providing excess water will result in waterlogging and growth of algae which is not healthy for the plant. There is a need to provide the correct amount water for good health of the plant. Amount of water plant needs depends on soil moisture, atmospheric humidity, atmospheric temperature and few other factors. Considering these factors as input to our system, we have to design a system that gives the amount of water to be added in the plant that day. The system trains on a data sets and observes the trends on how these factors affect the water output and based on that it provides the value of water to be added.

3.3 Vinca description

Vinca is a genus of flowering plants in the family Apocynaceae, native to Europe, northwest Africa and southwest Asia. Its common name is Periwinkle and belongs to the genus Catharanthus. In India it is also known as Sadabahar. Two of the species, Vinca major and Vinca minor, are extensively cultivated as a flowering evergreen plant. As the names suggest, all parts, including flowers, of Vinca major are larger than those of Vinca minor. The main difference between Vinca major and Vinca minor is that the leaves of Vinca major are slightly broader, larger, ovate, or heart shaped, while those of Vinca minor are small, elongated, lance-shaped. This can help identify the species. However, if we observe closely, we will see that the leaves of Vinca major have hairy margins, whereas Vinca minor leaves have hairless margins. In this project we have used Vinca minor species for experimentation.

Vinca is most commonly used as a groundcover due to its low, spreading habit and ability to grow in all light conditions. Because stems take root when they touch the soil, it is a useful plant for preventing erosion on hills, slopes or embankments.

It is also well known for its medicinal uses some of them include treatment of Lung cancer, Leukaemia, diabetes, dysentery, diarrhoea, anti-inflammatory and many more.

3.3.1 Physical appearance:

- Vinca can grow up to 7 to 15 cm in height and is 0.5 to 1.5 feet in width.
- The glossy dark green leaves are 2 to 4.5cm long and 1 to 2.5cm broad. They have a leathery texture with hairless margins.
- The flowers are 0.5 to 1 inch wide, five petaled solitary blue/purple flowers that bloom on upright stalks from the leaf axils. The flowers comes in pink, purple, red, white, magenta, and bi-colours. The flowers are also attractive to butterflies.

3.3.2 Conditions essential for good growth of the plant

- Sun: It is better to plant them in partial or full shade. A little morning sun would be enough. The light conditions, in which Vinca grows, however, affect the performance of the plant. With some shade, the leaves tend to be a brighter green than when grown in sun. The sun, however, causes the plant to produce more flowers.
- Soil: - They grow well in fertile, moist, and well-drained soil. However, these species can thrive in poor soils and soils of various pH as well. Vinca performs best in acidic soil with a pH of 5.4 to 5.8. An alkaline soil with a pH above 6.0 may cause the leaves to turn yellow.
- Tolerance: - Vinca minor plants can withstand far more diverse growing conditions as compared to *Vinca major. Well-established plants can tolerate an occasional drought.*
- *Water:* - Vincas are drought-tolerant plants that don't grow well in soggy soil, and wet, humid conditions often promote diseases. Too much water can also contribute to fungal diseases that cause lesions on the leaves, followed by yellowing and leaf drop. They should be watered only when the top 1 to 2 inches of soil feel dry to the touch, and then water deeply enough to saturate the roots. The soil should be allowed to dry before watering again. In this project we are concentrating on the optimal watering of the plant.
- *Nutrients:* - Vincas with yellowing leaves are often reacting to a lack of iron, which is a common problem in alkaline soil. A lack of nitrogen in the soil can also cause yellowing leaves. Use of a fertilizer containing nitrogen, iron and sulphur helps to replace the nutrients while balancing the pH.
- Diseases: - There are no serious insect or disease problems. Potential diseases include root rot, leaf spot, dieback, and blight. Various types of root rot and stem rot may cause yellow leaves, often followed by stunted growth, wilting and eventual plant death.

3.4 Selection of the plant for experimentation

Our main objective is to predict the correct amount of water that is needed for the plant and build a system which can water the plant automatically without human intervention. Vinca is a common flowering plant with low maintenance requirements. These features attract the common man easily who does not have much knowledge about plants and their requirements in detail. Although the plant requires low maintenance, it is easily susceptible to root rot and fungal diseases due to overwatering. Without proper knowledge about the plant, it may be overwatered which would lead to diseases and ultimately death of the plant. Also considerable amount of water maybe wasted. These are the reasons which contributed to the selection Vinca plant for our experimentation.

3.5 Training data collection

The data base made contains soil moisture readings, atmospheric temperature readings, atmospheric humidity readings and water added. These readings are mapped in the following manner:- previous days atmospheric humidity and atmospheric temperature is mapped with current days soil moisture readings. From the literature survey these factors were taken because they show an influence on water requirement of the plant .The water needed by the plant increases or decreases based on the increase or decrease in the value of soil moisture, atmospheric humidity and atmospheric temperature. This data set is used to train the machine learning algorithm.

3.6 Selection of Machine Learning for Regression based Analysis

Machine learning offers promising analytical and computational solutions for the integrative analysis of large, heterogeneous and unstructured datasets on the Big-Data scale, and is gradually gaining popularity in biology. In recent times machine learning has emerged with big data technologies and high-performance computing to create new opportunities for data intensive science in the multi-disciplinary Agri-technologies domain.

The plant science community not only needs to build its own Big-Data-compatible parallel computing and data management infrastructures, but also to seek novel analytical paradigms to extract information from the overwhelming amounts of data. At present, the system we have designed can be used to tackle individual problems, but with further integration it can also be used to enable automated decision-making with an interconnected system, that would help farming practices change into the so-called knowledge-based agriculture that would be able to increase production levels and products quality with extreme precision.

The goal of this project is to predict the amount of water required to sustain the plant based on soil moisture readings from the moisture sensor and temperature and humidity readings of the surrounding area, this can be achieved by the use of supervised learning in machine learning.

3.6.1 Supervised Learning

Supervised learning for predictive analysis can be simply put forth as a technique in which the machine is provided with certain training attributes which contribute to the output value to be predicted and it employs various algorithms to predict the output that is close enough to the real value. Hence, this type of learning is the best fit for our project design as the water output level is dependent on the soil moisture, temperature and humidity and hence the sensor data values of these attributes can be in the training dataset used to compute the predicted water output. The figure below explains the concept of Supervised learning:

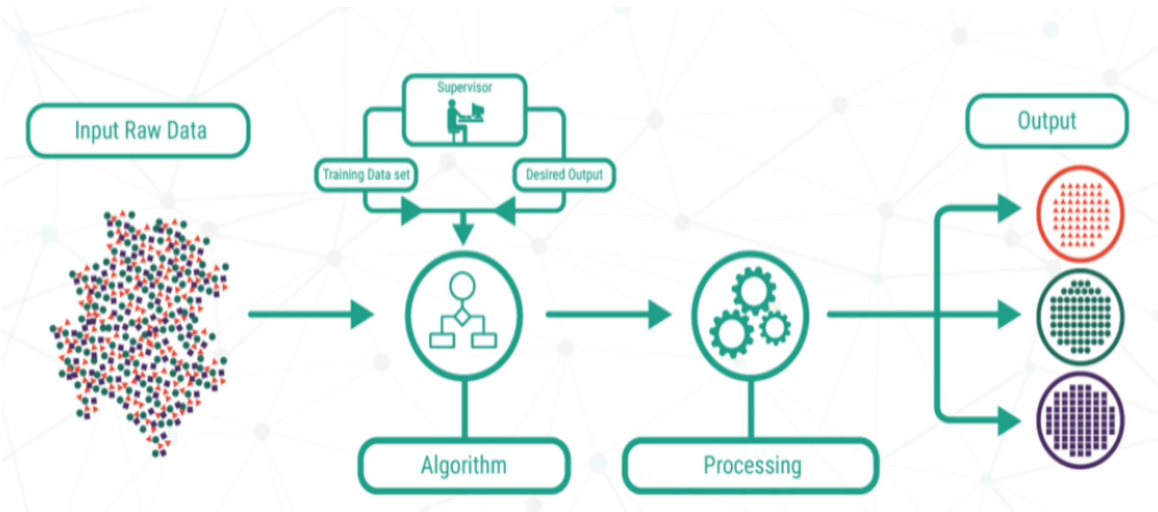


Fig no.:3.6.1

Supervised learning problems can be grouped into **Regression** and **Classification** problems. Both problems have as goal the construction of a succinct model that can predict the value of the dependent attribute from the attribute variables. The difference between the two tasks is the fact that the dependent attribute is numerical for regression and categorical for classification. And since, the training dataset and predicted output required in our project are numerical using a regression model is the best fit for it.

3.6.2 Regression based analysis

A regression problem is when the output variable is a real or continuous value, such as “salary” or “weight”. Many different models can be used; the simplest is the linear regression. It tries to fit data with the best hyper-plane which goes through the points.

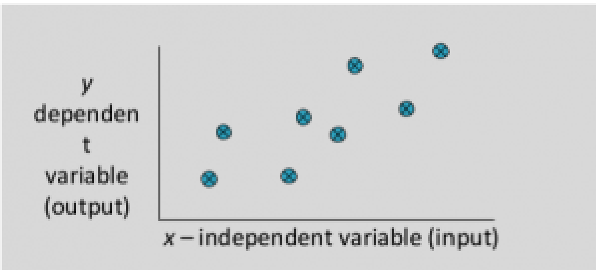


Fig no.:3.6.2

Regression predictive modelling is the task of approximating a mapping function (f) from input variables (X) to a continuous output variable (y). A continuous output variable is a real-value, such as an integer or floating point value. These are often quantities, such as amounts and sizes.

For example, a house may be predicted to sell for a specific dollar value, perhaps in the range of $100,000 to $200,000.

There are multiple benefits of using regression analysis. Such as:

- It indicates the significant relationships between dependent variable and independent variable.
- It indicates the strength of impact of multiple independent variables on a dependent variable.

For the solution of a problem to use regression the following scenarios are important:

- A regression problem requires the prediction of a quantity.
- A regression can have real valued or discrete input variables.
- A problem with multiple input variables is often called a multivariate regression problem.
- A regression problem where input variables are ordered by time is called a time series forecasting problem.

Since our model has most of the above features:

- We are predicting the water output level.

- Our input values have both real and discrete.

- Our problem does have multiple input variables.

which implies that we can use regression to get the prediction.

Now, since there are various kinds of regression techniques available to make predictions such as Linear Regression, Logistic Regression, Polynomial Regression, Stepwise Regression, Ridge Regression, Lasso Regression, Support Vector Regression, KNN based Regression, etc. We wanted to make sure that the model we're using provides the estimated value which is closest to the real value, basically select a model with low error. Therefore we implemented the following algorithms in order to detect which one of them is the best possible solution:

1. Linear Regression
2. Support Vector Regression
3. KNN based Regression

3.6.3 Types of Regression models used

Let us learn in detail about the models that we have used:

3.6.3.1 Linear Regression

It is one of the most widely known modelling techniques. Linear regression is usually among the first few topics which people pick while learning predictive modelling. In this technique, the dependent variable is continuous, independent variable(s) can be continuous or discrete, and nature of regression line is linear.

Linear Regression establishes a relationship between **dependent variable (Y)** and one or more **independent variables (X)** using a **best fit straight line** (also known as regression line).

It is represented by an equation **Y=a +b*X + e**, where a is intercept, b is slope of the line and e is error term. This equation can be used to predict the value of target variable based on given predictor variable(s).

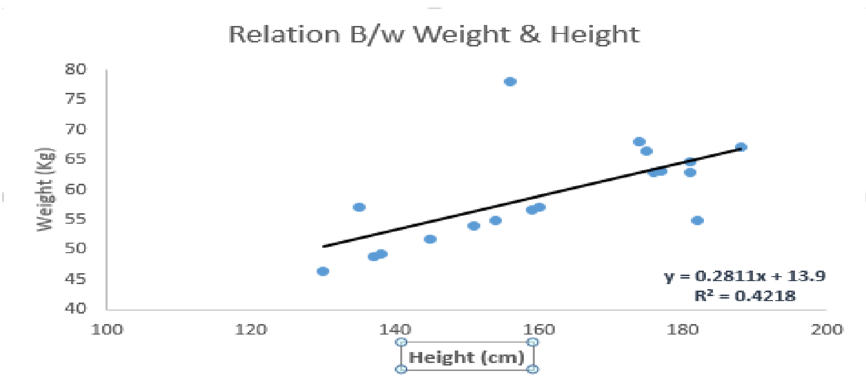


Fig.no: 3.6.3

The difference between simple linear regression and multiple linear regression is that, multiple linear regression has (>1) independent variables, whereas simple linear regression has only 1 independent variable. Since our problem has multiple independent variables we use multiple linear regression.

Equation for multiple linear regression:

$$Y=a+b1*X1+b2*X2+B3*X3+\ldots\ldots+e,$$

The multiple regression model is based on the following assumptions:

- There is a linear relationship between the dependent variables and the independent variables.
- The independent variables are not too highly correlated with each other.
- ‘Y’ observations are selected independently and randomly from the population.
- Residuals should be normally distributed with a mean of 0 and variance σ.

3.6.3.3 Support Vector Regression

The method of Support Vector Classification can be extended to solve regression problems. This method is called Support Vector Regression.

The model produced by support vector classification (as described above) depends only on a subset of the training data, because the cost function for building the model does not care about training points that lie beyond the margin. Analogously, the model produced by Support Vector Regression depends only on a subset of the training data, because the cost function for building the model ignores any training data close to the model prediction.

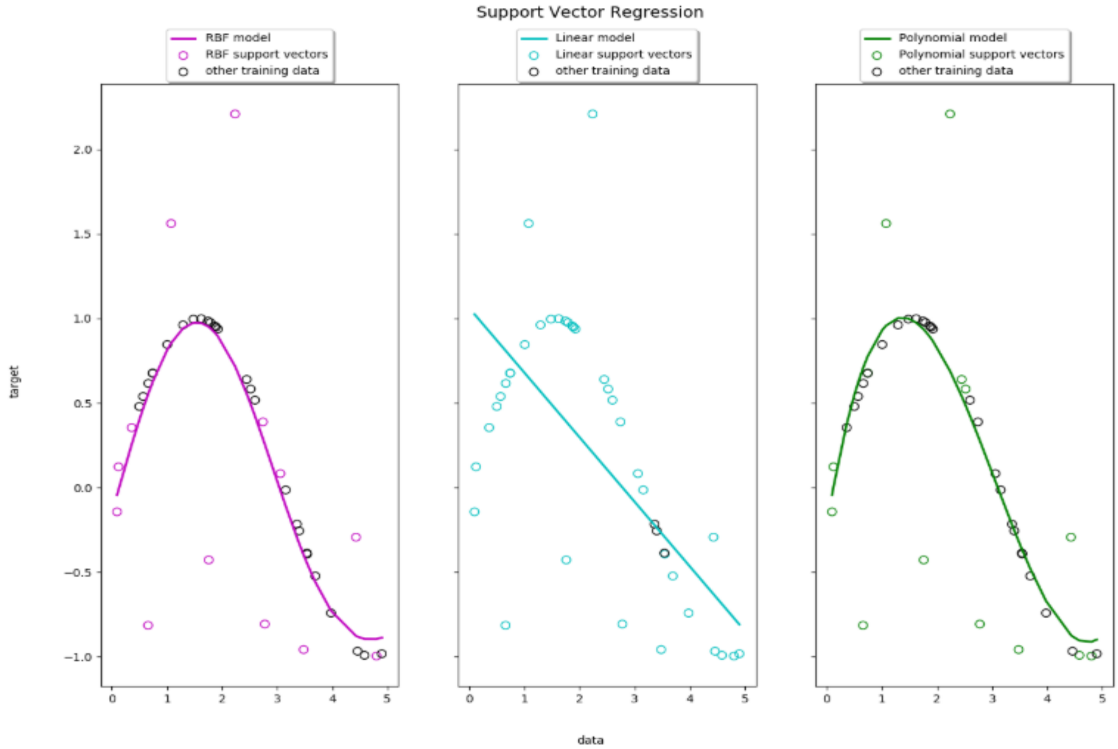


Fig.no: 3.6.4

Support Vector Regression (SVR) using linear and non-linear kernels, there are basically three types of model that we can implement:

1. RBF model 2.Linear model 3.Polynomial model

3.6.3.4 KNN Regression

KNN can be used for both classification and regression problems. The algorithm uses '**feature similarity**' to predict values of any new data points. This means that the new point is assigned a value based on how closely it resembles the points in the training set. K nearest neighbours is a simple algorithm that stores all available cases and predicts the numerical target based on a similarity measure (e.g., distance functions). KNN has been used in statistical estimation and pattern recognition already in the beginning of 1970's as a non-parametric technique.

A simple implementation of KNN regression is to calculate the average of the numerical target of the K nearest neighbours. Another approach uses an inverse distance weighted average of the K nearest neighbours. KNN regression uses the same distance functions as KNN classification.

Distance functions

Euclidean $$\sqrt{\sum_{i=1}^{k}(x_i - y_i)^2}$$

Manhattan $$\sum_{i=1}^{k}|x_i - y_i|$$

Minkowski $$\left(\sum_{i=1}^{k}(|x_i - y_i|)^q\right)^{1/q}$$

The above three distance measures are only valid for continuous variables. In the case of categorical variables you must use the Hamming distance, which is a measure of the number of instances in which corresponding symbols are different in two strings of equal length.

Hamming Distance

$$D_H = \sum_{i=1}^{k} |x_i - y_i|$$

$$x = y \Rightarrow D = 0$$
$$x \neq y \Rightarrow D = 1$$

X	Y	Distance
Male	Male	0
Male	Female	1

Choosing the optimal value for K is best done by first inspecting the data. In general, a large K value is more precise as it reduces the overall noise; however, the compromise is that the distinct boundaries within the feature space are blurred. Cross-validation is another way to retrospectively determine a good K value by using an independent data set to validate your K value. The optimal K for most datasets is 10 or more. That produces much better results than 1-NN.

3.6.4 Cross-Validation and RMSE

Apart from the above algorithms we can also implement the technique of cross-validation since it is simple to understand and because it generally results in a less biased or less optimistic estimate of the model skill than other methods, such as a simple train/test split.

In this section we also discuss the RMSE (Root Mean Square Error) because a regression predictive model predicts a quantity and the skill of the model must be reported as an error in those predictions. Here, we have used RMSE for this purpose.

3.6.4.1 Cross Validation

Cross-validation is a technique to evaluate predictive models by partitioning the original sample into a training set to train the model, and a test set to evaluate it. In k-fold cross-validation, the original sample is randomly partitioned into k equal size subsamples. Of the k subsamples, a single subsample is retained as the validation data for testing the model, and the remaining k-1 subsamples are used as training data. The cross-validation process is then repeated k times (the folds), with each of the k subsamples used exactly once as the validation data. The k results from the folds can then be averaged (or otherwise combined) to produce a single estimation. The advantage of this method is that all observations are used for both training and validation, and each observation is used for validation exactly once.

For classification problems, one typically uses stratified k-fold cross-validation, in which the folds are selected so that each fold contains roughly the same proportions of class labels.

The general procedure is as follows:

1. Shuffle the dataset randomly.
2. Split the dataset into k groups
3. For each unique group:
    - Take the group as a hold out or test data set
    - Take the remaining groups as a training data set
    - Fit a model on the training set and evaluate it on the test set
    - Retain the evaluation score and discard the model
4. Summarize the skill of the model using the sample of model evaluation scores.

3.6.5.1 RMSE

RMSE is a commonly used metric and serves well as a general purpose error metric. There are many ways to estimate the skill of a regression predictive model, but perhaps the most common is to calculate the root mean squared error, abbreviated by the acronym RMSE.

The root-mean-squared error (RMSE) is a measure of how well your model performed. It does this by measuring **difference between predicted values and the actual values**. Let's say you feed a model some input X and your model predicts 10, but the actual value is 5. This difference between your prediction (10) and the actual observation (5) is the **error** term: (y_prediction - y_actual). The error term is important because we usually want to minimize the error. **In other words, our predictions are very close to the actual values**.

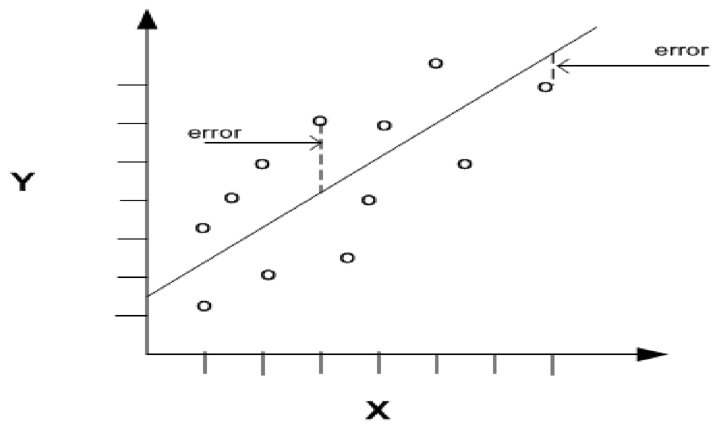


Fig.no: 3.6.5

But there are different ways we could minimize this error term. We could minimize the squared error. Or minimize the absolute value of the error.

In our case, we take the square root of the squared difference (RMSE):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (Predicted_i - Actual_i)^2}{N}}$$

In our example above, we would get: (10−5)2−−−−−−√=5(10−5)2=5

In a good model, the RMSE should be close for both your testing data and your training data. If the RMSE for your testing data is higher than the training data, there is a high chance that your model over fit. In other words, your model performed worse during testing than training.

3.6.5.2 R-Squared

Goodness of fit Linear regression is used for minimizing the distance between the fitted line and the data points, and it is generally said that the model fits the data properly if the difference between the predicted values and the observed values is minimal. Goodness-of-fit is the process of evaluating the observed values against the model's expected (predicted or fitted) values.

R-squared is a statistical measure that is used to find how close the data points are to the fitted regression line. It is also called as the coefficient of multiple determination for multiple regression or simply, the coefficient of determination. It is used in linear models as it evaluates the scatter of data points around the fitted line.

R-squared is the percentage of the dependent variable variation that a linear model explains.

$$R^2 = \frac{\text{Variance explained by the model}}{\text{Total variance}}$$

The value of R-squared will always should be between 0% to 100%.

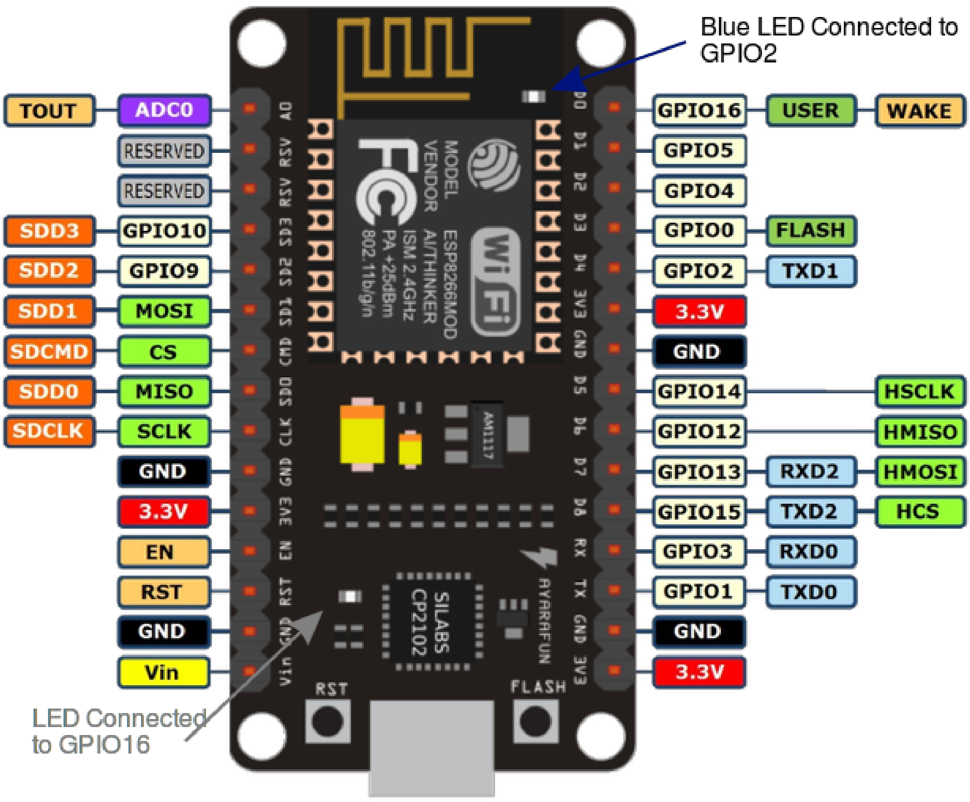
If the value is-

- 0%, the model is not explaining any variation of the response around its mean, because the mean of the dependent variable can be used to predict the dependent variable and the regression model.
- 100%, the model is explaining all of the variation in the response variable around its mean.

3.7 Information about the components

3.7.1 Node MCU Development Board v1.0

Node MCU v3 is a development board which runs on the ESP8266 with the Espressif Non-OS SDK, and hardware based on the ESP-12 module. The device features 4MB of



Flash memory, 80MHz of system clock, around 50k of usable RAM and an on chip Wi-Fi Transceiver. Node MCU Dev Kit has **Arduino like** Analog (i.e. A0) and Digital (D0-D8) pins on its board.
It supports serial communication protocols i.e. UART, SPI, I2C etc.

Using such serial protocols we can connect it with serial devices like I2C enabled LCD display, Magnetometer HMC5883, MPU-6050 Gyro meter + Accelerometer, RTC chips, GPS modules, touch screen displays, SD cards etc.

The Node MCU is a very user friendly and cost effective device which provides internet connectivity to small projects. The module can function as an Access point (can create hotspot) and also as a station (can connect to Wi-Fi), hence it can easily fetch data and upload it to the internet. It can also fetch data from internet using API's using which the project could access any information that is available in the internet, thus making it smarter. Another exciting feature of this module is that it can be programmed using the Arduino IDE which makes it very user friendly.
The physical overview is as follows :

- Wi-Fi Module – ESP-12E module similar to ESP-12 module but with 6 extra GPIOs.

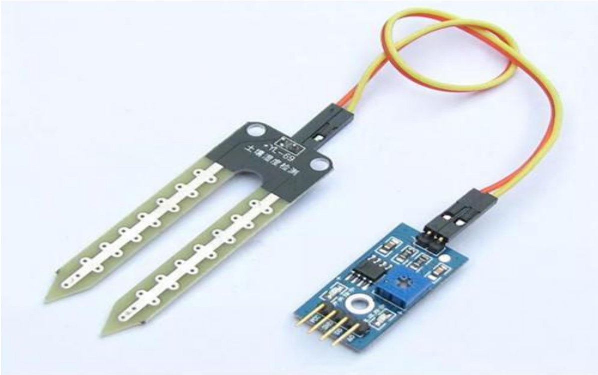
- USB – micro USB port for power, programming and debugging
- Headers – 2x 2.54mm 15-pin header with access to GPIOs, SPI, UART, ADC, and power pins
- Misc – Reset and Flash buttons
- Power – 5V via micro USB port
- Dimensions – 49 x 24.5 x 13mm

The technical specifications are:

- **Developer :** ESP8266 Open Source Community
- **Type :** Single-board microcontroller
- **Operating system :** XTOS
- **CPU :** ESP8266
- **Memory :** 128 kBytes
- **Storage :** 4 MBytes
- **Power :** USB
- **Power Voltage :** 3v ,5v (used with 3.3v Regulator which inbuilt on Board using Pin VIN)
- **Code :** Arduino Cpp
- **IDE Used :** Arduino IDE
- **GPIO :** 10

3.8.2. Soil Moisture Sensor

The Soil Moisture Sensor is used to measure the content of water present in the soil. This makes it ideal for performing experiments in courses such as soil science, agricultural science, environmental science, horticulture, botany, and biology. The Soil Moisture Sensor measures the dielectric permittivity of the soil which is directly related to the water content present in the soil. It The complete sensor probe is inserted into the soil which is to be tested and the volume of water in the soil is returned as percentage.
When there is more water, the soil will conduct more electricity which means that there will be less resistance. Therefore, the sensor reading will be higher. Dry soil conducts electricity poorly, so when there will be less water, then the soil will conduct less electricity which means that resistance is more. Therefore, the sensor reading will be lower.



The technical specifications and physical overview is as follows :

- Accuracy: ±4% typical
- Typical Resolution: 0.1%
- Power: 3 mA at 5VDC
- Operating temperature: –40°C to +60°C
- Dimensions: 8.9 cm × 1.8 cm × 0.7 cm (active sensor length 5 cm)

3.8.3. DC water pump

The basic working principle of a DC water pump is conversion of electric energy to mechanical energy and using it for providing the required pressure for pumping the water.

The technical specifications and physical overview is as follows

- Operating Voltage : 3 ~ 6V
- Operating Current : 130 ~ 220mA

- Flow Rate : 80 ~ 120 L/H
- Maximum Lift : 40 ~ 110 mm
- Continuous Working Life : 500 hours
- Driving Mode : DC, Magnetic Driving
- Material : Engineering Plastic
- Outlet Outside Diameter : 7.5 mm
- Outlet Inside Diameter : 5 mm

Chapter 4 – Implementation and Experimentation

In this chapter, a conclusion based on the result and discussion has been provided. The scope of the project in the future has also been provided.

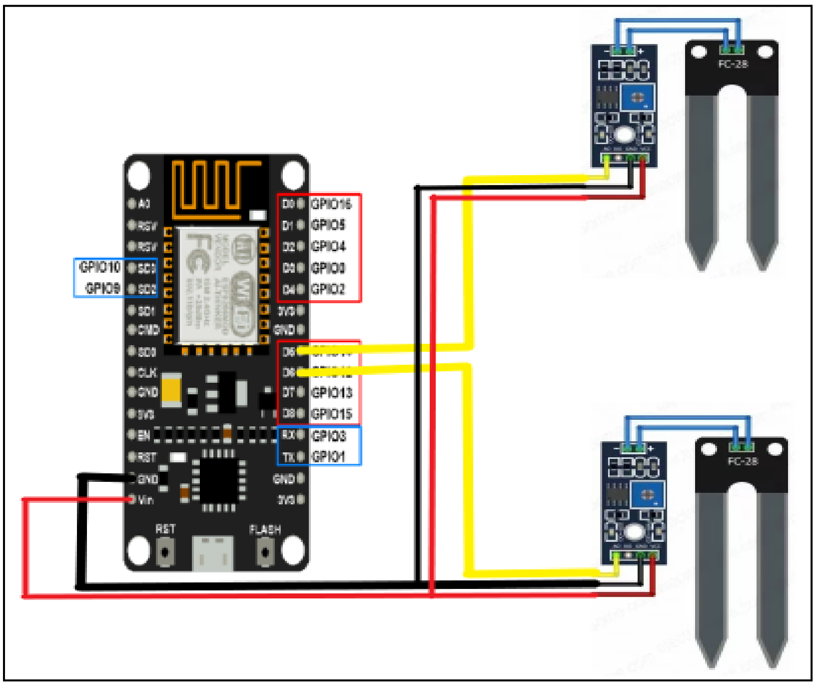
4.1 Steps in collecting sensor data values

The equipment required

- Vinca plant
- ESP8266 module
- Soil moisture sensor
- Measuring cup and water

Steps followed for collecting sensor data values

- The readings were collected everyday between 9 to 10am for more than a month.
- Two soil moisture sensor probes were inserted in the soil at different locations and connected to the ESP8266 module as shown in the circuit diagram.
- Initially while taking the sensor reading, the reading might be higher. The readings decreased and became constant after some time. This reading was collected.
- The two reading collected are sent to ESP8266 module where an averaging function is applied.



- The plant is watered until the sensor reading crosses the value of 70. The amount of water put into the soil was recorded.
- The temperature and humidity readings for that day were collected using the Google weather application.
- Using these readings i.e. the average soil moisture sensor value, temperature, humidity and the water put in the soil, an initial database was created.
- The above steps were performed manually to build a sufficient dataset in order to apply a machine learning algorithm. Since this process was done manually, it might be prone to human error.
- The final watering system completely automates the above process. The average value is directly sent to a database along with the temperature and humidity readings.
- The machine learning algorithm predicts the amount of water to be put next day using this dataset and returns it to the ESP8266 module which pours that amount of water in the soil.

4.2 Sending and Receiving sensor values using the wifi -module

- The library files are added.
- The web server is to port 80. This port is used for Communication.
- The Analog pin A0 is set as the input pin.
- The Serial of the Node MCU module is initialised at 115200 bits/sec frequency. This is the frequency of communication.
- The Node MCU module is provided with ssid and password for it to connect to internet.
- CallerValue() function gets the moisture sensor value from the sensor at connected at A0 pin .
- The Value is mapped to get the value in percentage.
- The handlepath() function returns the soil moisture value in percentage.
- This function calls the callerValue function to get moisture sensor value.
- Then it converts the value to string type data.
- This string is then sent to the data base on the cloud.

4.3 Data Visualization

Data visualization is the process of representing the data in a format is comprehensive to the viewer and helps the viewer's understand the significance of the data by putting it in a visual context. There are several patterns, correlations, and trends in the dataset that can go completely undetected if the data is text based. Through data visualization techniques, this underlying data which could go neglected in some way can be well-exposed, easily absorbed, and also analysed faster and efficiently.

Data visualization is important because it gives insight into data that can be missed by the human eye. It helps the analyst comprehend and discover new trends in the data just by varying some parameters slightly which results in drastic changes.

The data that has been collected after experimentation consists of the values of atmospheric humidity, soil moisture of the previous day, temperature, and water amount added to the plant the next day.

The correlation matrices and graphs obtained are as follows-

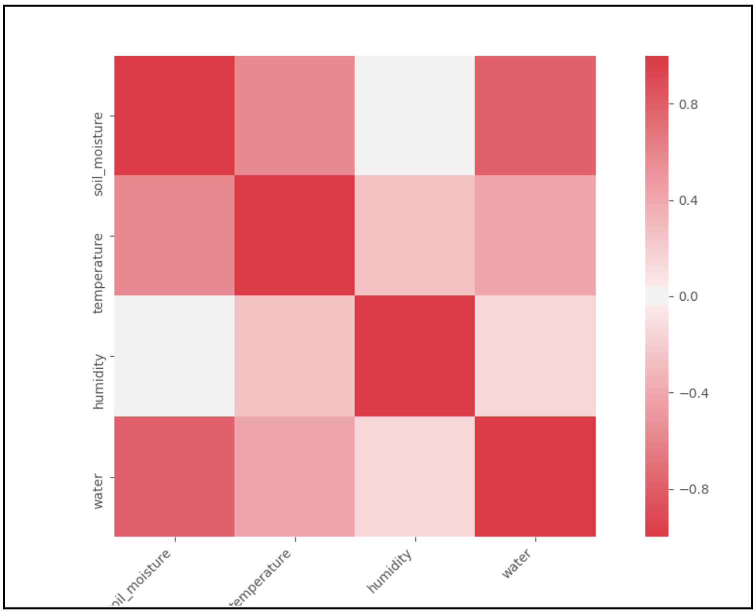


Fig 4.3.1 correlation heat map of dataset values

```
Temperature and soil moisture correlation
[[ 1.         -0.56748147]
[-0.56748147  1.        ]]
humidity and soil moisture correlation
[[1.         0.00740489]
[0.00740489 1.        ]]
Temperature and humidity correlation
[[ 1.        -0.2620878]
[-0.2620878  1.       ]]
```

Fig 4.3.2 correlation matrices of dataset values

The figures above show the following observations-

1. The temperature and soil moisture have a negative correlation, which means that as the temperature increases, the soil moisture decreases.
2. The temperature and atmospheric humidity have a negative correlation, which means that as the temperature increases, the atmospheric humidity decreases.
3. The atmospheric humidity and soil moisture have a positive correlation, which means that as the atmospheric humidity increases, the soil moisture also increases.
4. The water level has the highest negative correlation with the soil moisture, which means that an decrease in the soil moisture will increase the water needed.

The features have been plotted in the form of 3D plots and scatter plots. They are as follows-

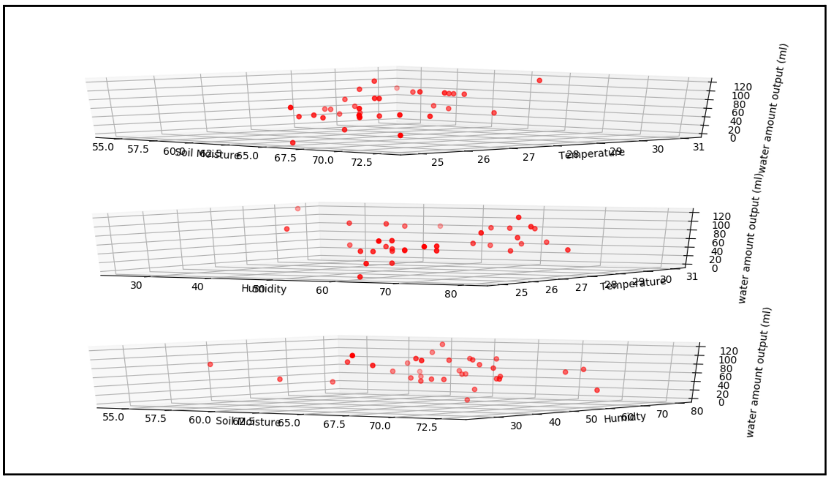


Fig 4.3.3 3D plot of features

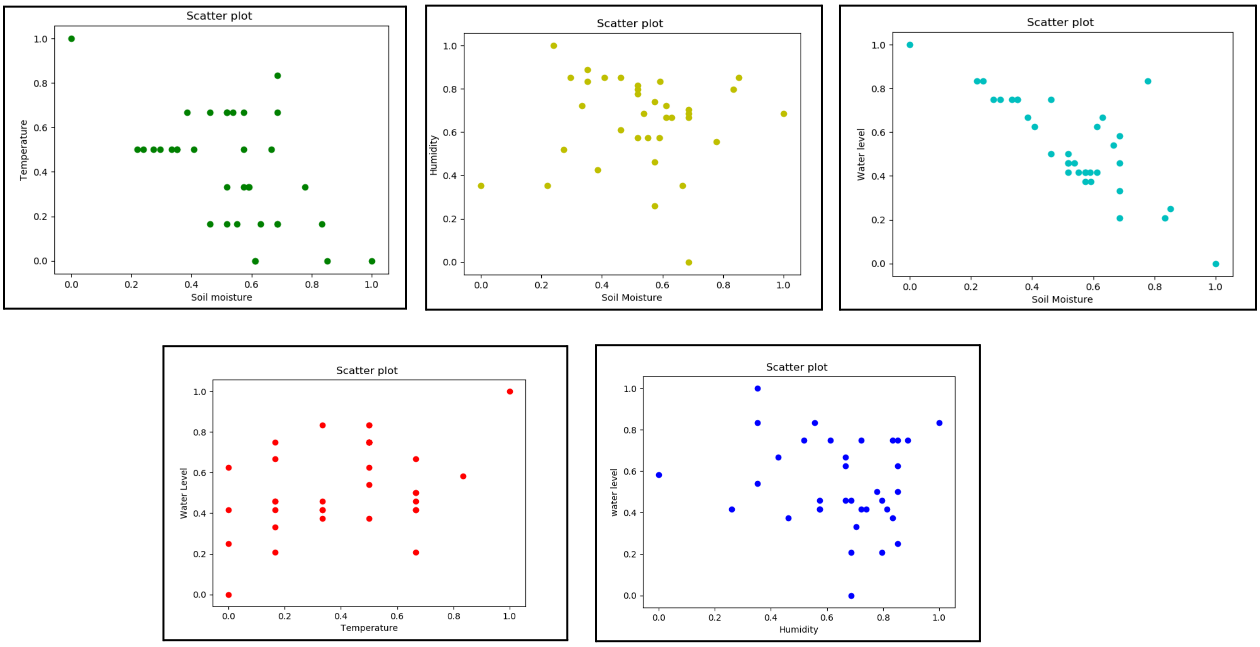


Fig 4.3.4 scatter plots of features

The observations drawn from the above figures includes-

1. The relationship between temperature and soil moisture and humidity and soil moisture is non-linear.
2. There is some linearity between the water level and soil moisture values, where water level has a non-linear relationship with temperature and humidity.

The visualized data points and correlation can help in understanding the data which will be fed to the regression model.

4.4 Results

The results obtained after the process of dataset collection, data visualization and data modelling show the predicted values of the output water amount that can be put in the plant which is adequate for the growth of the plant and also prevents a situation where the water added to the plant is in excess.

The results have been obtained in two methods using 4 algorithms for the prediction of the output amount of water:

1. The dataset obtained has been divided into training and testing data which is then sent as an input in the prediction models. The procedure is known as the train-test-split process.
2. The R-Squared value of the predicted water output is then scrutinized. The R-squared is the square root of the variance of the residuals. It indicates the absolute fit of the model to the data–how close the observed data points are to the model's predicted values. Lower values of R-squared indicate better fit. R-squared is a good measure of how accurately the model predicts the response, and it is the most important criterion for fit if the main purpose of the model is prediction.
3. RMSE is the Root Mean Square Error of the prediction. It tells how much error was computed and it is the difference between the predicted and actual values of the data.

I] **Results of separating dataset into training and testing datasets-**

1. Linear Regression Model

Sr. No	**Splitting of data ratio**	**Case**	**Output water predicted (ml)**	**RMSE**	**R-squared**	**Coefficients**	**Probability values** (corresponding to titles in coefficient column)
1	80% training, 20% testing	Only temperature values with water o/p	50.5	20.07	0.167	tempval 0.3489 const 0.0079	P>|t| 0.015 0.000
		Only humidity values with water o/p	44.01	21.3	0.020	humval -0.1476 const 0.0127	P>|t| 0.413 0.000
		Only soil moisture values with water o/p	101.9	20.9	0.630	soilval -0.8602 const 0.0199	P>|t| 0.000 0.000
		Temperature and humidity combined	59.48	13.41	0.168	tempval 0.3403 humval -0.0398 const 0.0085	P>|t| 0.023 0.819 0.006
		Temperature and soil moisture values combined	51.29	22.62	0.633	tempval -0.0536 soilval -0.8989 const 0.0207	P>|t| 0.633 0.000 0.000
		Humidity and soil moisture values combined	67.54	8.34	0.649	humval -0.1422 soilval -0.8593 const 0.0217	P>|t| 0.198 0.000 0.000

		All three values without constant	62.74	12.03	0.856	tempval 0.5995 soilval -0.0598 humval 0.4563	P>|t| -------- 0.000 0.698 0.003
		All three values with constant	59.664	11.45	0.659	tempval -0.1092 soilval -0.9377 const 0.0239 humval -0.1763	P>|t| -------- 0.348 0.000 0.000 0.133
2	70% training, 30% testing	Only temperature values with water o/p	62.92	18.31	0.167	tempval 0.3489 const 0.0079	P>|t| ---------- 0.015 0.000
		Only humidity values with water o/p	58.064	25.19	0.020	humval -0.1476 const 0.0127	P>|t| ---------- 0.413 0.000
		Only soil moisture values with water o/p	54.426	18.727	0.630	soilval -0.8602 const 0.0199	P>|t| ------- 0.000 0.000
		Temperature and humidity combined	67.28	22.48	0.168	tempval 0.3403 humval -0.0398 const 0.0085	P>|t| --------- 0.023 0.819 0.006
		Temperature and soil moisture values combined	46.94	20.434	0.633	tempval -0.0536 soilval -0.8989 const 0.0207	P>|t| -------- 0.633 0.000 0.000
		Humidity and soil moisture values combined	102.18	11.88	0.649	humval -0.1422 soilval -0.8593 const 0.0217	P>|t| ------- 0.198 0.000 0.000

		All three values without constant	51.898	12.06	0.856	tempval 0.5995 soilval -0.0598 humval 0.4563	P>|t| -------- 0.000 0.698 0.003
		All three values with constant	47.34	9.84	0.659	tempval -0.1092 soilval -0.9377 const 0.0239 humval -0.1763	P>|t| -------- 0.348 0.000 0.000 0.133
3	60% training, 40% testing	Only temperature values with water o/p	59.50	24.55	0.167	tempval 0.3489 const 0.0079	P>|t| ---------- 0.015 0.000
		Only humidity values with water o/p	52.66	20.90	0.020	humval -0.1476 const 0.0127	P>|t| ---------- 0.413 0.000
		Only soil moisture values with water o/p	20.90	13.152	0.630	soilval -0.8602 const 0.0199	P>|t| ------- 0.000 0.000
		Temperature and humidity combined	67.28	22.48	0.168	tempval 0.3403 humval -0.0398 const 0.0085	P>|t| --------- 0.023 0.819 0.006
		Temperature and soil moisture values combined	68.13	33.627	0.633	tempval -0.0536 soilval -0.8989 const 0.0207	P>|t| -------- 0.633 0.000 0.000
		Humidity and soil moisture values combined	45.03	11.832	0.649	humval -0.1422 soilval -0.8593 const 0.0217	P>|t| ------- 0.198 0.000 0.000

		All three values without constant	61.153	12.8	0.856	tempval 0.5995 soilval -0.0598 humval 0.4563	P>|t| -------- 0.000 0.698 0.003
		All three values with constant	73.66	17.63	0.659	tempval -0.1092 soilval -0.9377 const 0.0239 humval -0.1763	P>|t| -------- 0.348 0.000 0.000 0.133

Table 4.8.1

2. Support Vector Regression

Sr. No	**Kernel**	**Splitting of data ratio**	**Output water predicted (ml)**	**RMSE**
1	RBF	80% training, 20% testing	67.23	29.481
		70% training, 30% testing	58.15	24.314
		60% training, 40% testing	65	32.135
2	LINEAR	80% training, 20% testing	53.89	30.977
		70% training, 30% testing	26.32	25.417
		60% training, 40% testing	134.09	15.513
3	POLYNOMIAL Degree=1	80% training, 20% testing	76.28	26.068

		70% training, 30% testing	37.38	22.863
		60% training, 40% testing	107.024	17.336
4	POLYNOMIAL Degree=2	80% training, 20% testing	100.49	31.588
		70% training, 30% testing	52.67	24.861
		60% training, 40% testing	46.72	15.799

Table 4.8.2

4. KNN Regressor

```
The best value of K is=

{'n_neighbors': 9}
The predicted water level is  66.66666666666667 ml
The RMSE value for the KNN regressor is 27.86889816639363
```

Fig 4.8.1 KNN for 80-20

```
The best value of K is=

{'n_neighbors': 2}
The predicted water level is  90.0 ml
The RMSE value for the KNN regressor is 20.532679052936786
```

Fig 4.8.2 KNN for 70-30

```
The best value of K is=

{'n_neighbors': 9}
The predicted water level is  57.22222222222222 ml
The RMSE value for the KNN regressor is 23.875965496604774
```

Fig 4.8.3 KNN for 60-40

Linear Regression

- After extensive case analysis using combinations of the input parameters, it can be observed from the table 4.8.1 that the RMSE value is least for the case when the data is split into 80% and 20% and the inputs considered are humidity and soil moisture values.
- The highest RMSE is seen when the data is split into 70% and 30% and the inputs considered are humidity values only.
- The value of R-squared for cases where temperature and humidity were singly selected is very low. The R-squared for cases where only soil moisture or its combinations were fed to the prediction model is higher compared to the previous scenarios.

- 80-20 scenario:

  1. The RMSE values for the temperature and humidity combined and the humidity and soil moisture values combined is the lowest among the cases considered and analysed.

  2. The RMSE ranges between 0-30 for the entire set of cases.

  3. When the constant is removed from the dataset and all three inputs are considered, the value of R-squared and RMSE change drastically. The RMSE is lower than the case when the constant is added to the inputs, and the R-squared value increases up 85.6%.

- 70-30 scenario:

  1. The RMSE values for only temperature and only soil moisture, as well as the combination of all three inputs with the constant are the lowest among the cases considered and analysed.

  2. The RMSE ranges between 0-30 for the entire set of cases.

  3. When the constant is removed from the dataset and all three inputs are considered, the value of R-squared and RMSE change drastically. The RMSE is higher than the case when the constant is added to the inputs, and the R-squared value increases up 85.6%. The difference in the RMSE is also high, which means that adding the constant is best for the scenario.

- 60-40 scenario:

  1. The RMSE values for the humidity and soil moisture values combined is the lowest among the cases considered and analysed.

  2. The RMSE ranges between 0-30 for the entire set of cases.

  3. When the constant is removed from the dataset and all three inputs are considered, the value of R-squared and RMSE change drastically. The RMSE is higher than the case when the constant is added to the inputs, and the R-squared value increases up 85.6%.

Support Vector Regression

- The data has been split into 3 ratios, i.e. 80-20, 70-30, 60-40 for training and testing respectively. The comparison is done for the three kernels, RBF, Linear, and Polynomial kernels. The polynomial kernel with degree 1 and 2 are considered.
- The RMSE value is least in 60-40 scenario when the Linear kernel is considered. It is the highest for the 60-40 scenario with the RBF kernel.
- The least variation in RMSE values can be observed for the Polynomial kernel with degree 1. However, the values of water output change dramatically with the kernel.

KNN Regressor

- The data has been split into 3 ratios, i.e. 80-20, 70-30, 60-40 for training and testing respectively. The output provides the best number of neighbours for the data and this value of ‘K’ is used for prediction.

- The RMSE is least for the 70-30 scenario.

Chapter 5

Conclusion and Further Work

In this chapter, a conclusion based on the result and discussion has been provided. The scope of the project in the future has also been provided.

Conclusion

The conclusion based on the results obtained can be drawn as follows-

- After extensive experimentation and analysis it is concluded that along with soil moisture, humidity, and temperature values, there are also some more unknown factors that must be taken into consideration for determining the output water amount. The output water amount is dependent on other factors as well. These factors may affect the soil moisture in different ways.
- The values of the output water to be added depend heavily on the soil moisture values. The soil moisture is an important component in predicting the output water to be put in the plant.
- The RMSE values vary for different cases when the data is split differently each time. The RMSE and the coefficients of the model summary present the effect of the different inputs on the output water amount.
- The R-squared value is highest when the soil moisture values and its combinations are put in the prediction model. The data points are best fit when the soil moisture is considered as a feature. The R-squared value does not change when the data splitting ratio is changed.
- The train-test-split method provides unreliable results, due to small value of dataset and lesser parameters. This problem can be combated with cross validation, which provides stable results and better approximations for the output water amount values.

It can also be concluded that most of the prediction models are not well-fitted and undergo ‘under fitting’. This situation owes to the fact that there should an inclusion of more features and the dependency of water level on the features must be thoroughly checked along with an increase in the data collection. Extending the features and increasing the dataset can provide more promising results.

Future Scope

- More number of plant species can be used for experimentation.
- The data collection can be done for multiple seasons. This can increase the accuracy of the predicted output.
- As we saw from the results, there are might be more parameters which have an effect on the water level output. These parameters can be taken into consideration and a better machine learning model can be built.
- Sophisticated equipment can be used in the complete system to increase power efficiency, decrease error and increase accuracy.

Acknowledgment

This undertaking has seen the light of the day because of the priceless commitment of specific individuals whose steady direction, backing and support brought about the acknowledgment of our venture. We genuinely thank our guides Mrs. Jayshree Kundargi for giving their valuable inputs and support which has provided our project with the correct guidance for improvement. They have been our mainstay of help and have constantly given us a valuable analysis and eager consolation. Their eagerness to give their time so liberally has been particularly valued. We are appreciative to our Head of the Department Dr. Ameya Naik for giving us the important help and support at whatever point required. This has brought about the accomplishment and success of our task. We accept this opportunity to express gratitude toward Dr. Shubha Pandit for providing an encouraging and enthusiastic environment in our college, which helped us in focusing on our task. We also want to thank all the staff members from our department, particularly Mr. Santosh Chaudhary, without whose helpful recommendations and significant counsel, the basic thought, which had been borne by us, would not have had the capacity to bloom forward to give such a lovely blossom. To wrap things up, we are thankful to our parents and every one of our companions for their immediate or roundabout encouragement and help over the span of this venture.

Bibliography

1. Jones, H. G. (2006). Monitoring plant and soil water status: Established and novel methods revisited and their relevance to studies of drought tolerance. *Journal of Experimental Botany.*
2. Salim, M. (1989). Effects of salinity and relative humidity on growth and ionic relations of plants. *New Phytologist.*
3. Cowling, S. A., & Sage, R. F. (1998). Interactive effects of low atmospheric CO2 and elevated temperature on growth, photosynthesis and respiration in Phaseolus vulgaris. *Plant, Cell and Environment.*
4. Ananthi, N., Divya, J., Divya, M., & Janani, V. (2017). IoT based smart soil monitoring system for agricultural production. *2017 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR)*.
5. Pernapati, K. (2018). IoT Based Low Cost Smart Irrigation System. *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)*.
6. Gagandeep, Arora, D., & Saini, H. S. (2017). Design and implementation of an automatic irrigation feedback control system based on monitoring of soil moisture. *2017 International Conference on Inventive Computing and Informatics (ICICI)*.
7. Qi, C., Wenbiao, W., & Siyuan, W. (2015). Application of indoor temperature prediction based on SVM and BPNN. *The 27th Chinese Control and Decision Conference (2015 CCDC)*.
8. Meng, L., Shi, J., Wang, H., & Wen, X. (2013). SVM with improved grid search and its application to wind power prediction. *2013 International Conference on Machine Learning and Cybernetics*.
9. Liu, S., & Zhou, F. (2010). On stock prediction based on KNN-ANN algorithm. *2010 IEEE Fifth International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA)*.
10. Tissera, M. D., & Mcdonnell, M. D. (2016). Modular expansion of the hidden layer in Single Layer Feedforward neural Networks. *2016 International Joint Conference on Neural Networks (IJCNN)*.
11. R. Subalakshmi and Anu Amal, (2016). “GSM Based Automated Irrigation using Sensors” presented at Special Issue published in International Journal of Trend in Research and Development (IJTRD),
12. John R. Dela Cruz, Renann G. Baldovino, Argel A. Bandala, Elmer P. Dadios,(2017). “Water Usage Optimization of Smart Farm Automated Irrigation System Using Artificial Neural Network”, (ICoICT).
13. Remco R. Bouckaert,(2013) “Weka Manual for Version 3-7-8”, The University of Waikato.
14. H. B. Lu and J. L. Wang, (2014). Optimizing the exhaust carbon contented of fly ash of power station boiler based on LIBSVM and artificial intelligence algorithm, Journal of Northeast Dianli University.
15. D. K. Sreekantha; Kavya A. M.,” (2017). Agricultural crop monitoring using IOT - a study” 11th International Conference on Intelligent Systems and Control (ISCO)