NEAR EAST UNIVERSITY

INSTITUTE OF GRADUATE STUDIES

DEPARTMENT OF CIVIL AND ENVIRONEMTAL ENGINEERING

STATISTICAL AND MACHINE LEARNING TECHNIQUES APPLIED TO THE PREDICTION OF TOTAL RAINFALL IN URBAN CITIES, NORTHERN PART OF IRAQ

M.Sc. THESIS

TAHIR SHAMSALDDIN ABDALSAMAD

Nicosia

June, 2022

Tahir Shamsalddin Abdalsamad

STATISTICAL AND MACHINE LEARNING TECHNIQUES APPLIED TO THE PREDICTION OF TOTAL RAINFALL IN URBAN CITIES, NORTHERN PART OF IRAQ

MASTER THESIS

2022

NEAR EAST UNIVERSITY

INSTITUTE OF GRADUATE STUDIES

DEPARTMENT OF CIVIL AND ENVIRONEMTAL ENGINEERING

STATISTICAL AND MACHINE LEARNING TECHNIQUES APPLIED TO THE PREDICTION OF TOTAL RAINFALL IN URBAN CITIES, NORTHERN PART OF IRAQ

M.Sc.THESIS

TAHIR SHAMSALDDIN ABDALSAMAD

Supervisor

Prof. Dr. Hüseyin Gökçekuş

Co Supervisor

Assoc. Prof. Dr. Youssef Kassem

Nicosia

June, 2022

Approval

We certify that we have read the thesis submitted by Tahir Shamsalddin Abdalsamad titled **"Statistical And Machine Learning Techniques Applied To The Prediction Of Total Rainfall In Urban Cities, Northern Part Of Iraq"** and that in our combined opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Educational Sciences.

Examining Committee	Name-Surname	Signature
Committee Member:	Prof. Dr. Aşkın kıraz	…………
Committee Member:	Assoc. Prof. Dr. Anoosheh Iravanıan	…………
Committee Member:	Assıst. Prof. Dr. Mustafa Alas	…………
Supervisor:	Prof. Dr. Hüseyin Gökçekuş	…………
Co Supervisor:	Assoc. Prof. Dr. Youssef Kassem	…Youssef…

Approved by the Head of the Department

.09./.09./2022

…………………………

Prof. Dr. Kabir Sadeghi
Head of Department

Approved by the Institute of Graduate Studies

…../…../2022

Prof. Dr. Kemal Hüsnü Can Başer
Head of the Institute

Declaration

I hereby declare that all information, documents, analysis and results in this thesis have been collected and presented according to the academic rules and ethical guidelines of Institute of Graduate Studies, Near East University. I also declare that as required by these rules and conduct, I have fully cited and referenced information and data that are not original to this study.

Tahir Shamsalddin Abdalsamad

29/06/2022

Acknowledgements

First and foremost, I will not forget to express my heartfelt appreciation to Allah Almighty who has given me strength, protection, zeal, energy, bravery, above all and has made this effort possible. This thesis would not have been possible without the help, support, and patience of my principal supervisor; my deepest gratitude goes to Prof. Dr. Hüseyin Gökçekuş and Assoc. Prof. Dr.Youssef Kassem for his constant encouragement and guidance. He has walked me through all the stages of the writing of my thesis. Without his consistent and illuminating instruction, this thesis could not have reached its present form.

Above all, my unlimited thanks and heartfelt love would be dedicated to my dearest family for their loyalty and their great confidence in me..

Tahir Shamsalddin Abdalsamad

Abstract

Statistical And Machine Learning Techniques Applied To The Prediction Of Total Rainfall In Urban Cities, Northern Part Of Iraq

Tahir Shamsalddin Abdalsamad

MA, Department of Civil and Environmental Engineering

June, 2022, 125 pages

Rainfall is considered the main source for groundwater in Iraq. The objective of this thesis is to design a model to predict rainfall in urban cities, of northern Iraq, and give a predictable outcome. To achive this, dataset have been collected from Ministry of planning and statistics office, during the period of 2012-2020. To this aim, eight empirical models (Support Vector Machine (SVM), K-Nearest Neighborsi (KNN), Linear Regression (LR), Decision Tree (DT), Neural network (Cascade-forward backprop, Elman backprop, Feed-forward backprop,and Layer recurrent) are utilized to predict the monthly rainfall in four urban cities in Northern Part of Iraq. The accuracy of the prediction models are KNN 0.96%, SVM 0.93%, NN(Layer recurrent) 0.90%, DT 0.97% and LR 0.89%. From the experimentations carried out it showed that the DT had the best accuracy with 0.97% while the least is the LR with 0.89%. The accuracy is one of the main criteria used to show how accurate the prediction model is. The results shows that the DT will be insightful when utilized to predict the rainfall in the urban region of Iraq. It is believed that this work will develop an insightful model with high and strong accuracy to help people and specialists whenever thought about.

Key Words: support vector machine, machine learning, statistics, decision tree, k-nearest neighbors

Table of contents

CHAPTER III

CHAPTER IV

CHAPTER V

CHAPTER VI

List of Figures

List of Tables

List of Abbreviations

SVM:	Support Vector Machine
CNN:	Convolutional Neural Network
KNN:	Convolutional Neural Network
ML:	Machine Learning
ANN:	Artificial Neural Network
DL:	Deep Learning
SL:	Supervised Learning
DT:	Decision Tree
NN:	Neural Network
AI:	Artificial Intelligence
FIS:	Fuzzy Inference System
GDP:	Gross Domestic Product
IMD:	Iraqi's Meteorological Department
EDA:	Exploratory Data Analysis

CHAPTER I
Introduction

Subcontinental rainfall expectation is so significant for Iraq's economy and everyday life is profoundly founded on horticulture. Around 61% of Iraq's geological region is utilized for farming (Scaife, et. al., 2019; Martin Stendel, 2021). Around 70% of the populace lives in urban areas and the larger part of them are subject to rain to care for their horticulture. Iraq horticulture, which represents 18-21% of gross domestic product (GDP), is generally subject to rain (Nikolaos and Peter, 2021). The GDP of the Iraqis economy is displayed in Figure 1. The changeability of rainfall in space, time, and sum influences horticulture, which hampers the economy. Around 12-17% of Iraq's yearly commodities are agricultural produce.

Around 81% of Iraq's yearly rainfall is gotten between June to September. 60% of the cultivated region has substantial rainfall, as just 41% of the region is under a water system. The water system is additionally subject to a rainstorm. This is the primary season for downpour for agribusiness (Gokcekus, 2020). Significant issues identified with rainfall prediction are, variety in predicting and indicator relationship, entomb connection between the predicators, and evolving consistency (Reder, et. al., 2018). The arrival of storm rainfall impedes harvest efficiency, horticulture venture, and movement of individuals from region to region, poor frameworks, and administrations. Subsequently, the use of processing instruments/strategies to culture this issue is the fundamental inspiration of this work. Albeit, numerous specialists (Hussain, and Dimililer, 2021; Frame, et. al., 2021; Hussain, and Al-Turjman, 2020; Vogel, et. al. 2018; Alotaibi, et. al. 2018) have dealt with expectation utilizing machine learning for various predicting problems and also for forecasting subcontinental rainfall, yet it stays a subject of exploration to further develop the expected precision.

The unseemly or poor rainfall forecast is likewise one major problem for management in water reserves. For this reason, it is vital to plan and work on a framework that would aid in the exact forecast and can be user-friendly (Alizadeh, et. al. 2017; Hussain, et. al. 2021a). The exact and right rainfall forecast cannot just add to the compelling and productive usage of this resource nature provides however, it can likewise help in dealing with the study and plans for generating power. Machine Learning (ML) has developed in wide prevalence as of late. The development guide of ML is displayed in Figure 2. Machine Learning (ML) can get input data and cycle it for a helpful result (Gleixner, et. al. 2017). Machine Learning (ML) methods for rainfall forecast is one of the most reasonable and dependable frameworks for predicting rainfall that has as of now helped administrators for rainfall forecast (Hussain, and Al-Turjman, 2021; Parida, et. al. 2017).

Hence, having a suitable methodology for forecasting rainfall makes it conceivable in making preventive and moderation lengths for these natural peculiarities (Bahrawi, et. al. 2021). This paper expects to give start to finish AI life cycle right from Data preprocessing to carrying out models to assessing them. To settle this vulnerability, we utilized different ML strategies and models to make precise and opportune expectations. We executed models, for example, Logistic Regression, Decision Tree, K Nearest Neighbor, Rule-based, and Ensembles. Information Preprocessing steps incorporate crediting missing qualities, highlighting change, encoding absolute elements, including scaling and element choice.

ML strategies needn't bother with past information on the handling of data that gives it a benefit over different information handling frameworks (Wu, et. al. 2018; Hussain, et. al. 2021b). To investigate the exhibition of these calculations; the co-connection coefficient will be a vital pointer in this review. The rainfall expectation will likewise coordinate versatile Neuro-ML methods for an expanded exactness and improved nature of the anticipated result. The ML method is likewise one of the viable calculations utilized for information investigation for the arrangement (Kassem, et, al., 2021). Regulated learning is the ablest and successful apparatus for the expectation of rainfall that really adds to the most dependable determining (Wadoux, et. al. 2017; Hussain, et. al. 2021c). Thus, each time information is dissected; it allows that information to the most appropriate or most comparative class it has a place with. It relegates classes and apportions cases to comparable gatherings/classifications. This aids in making the relapse and permits the client to make a forecast for the comparable arrangements of information or data got each time (Khan, et. al. 2017).

Figure 1

Iraq GDP from agriculture (Roger Guiu. 2015)

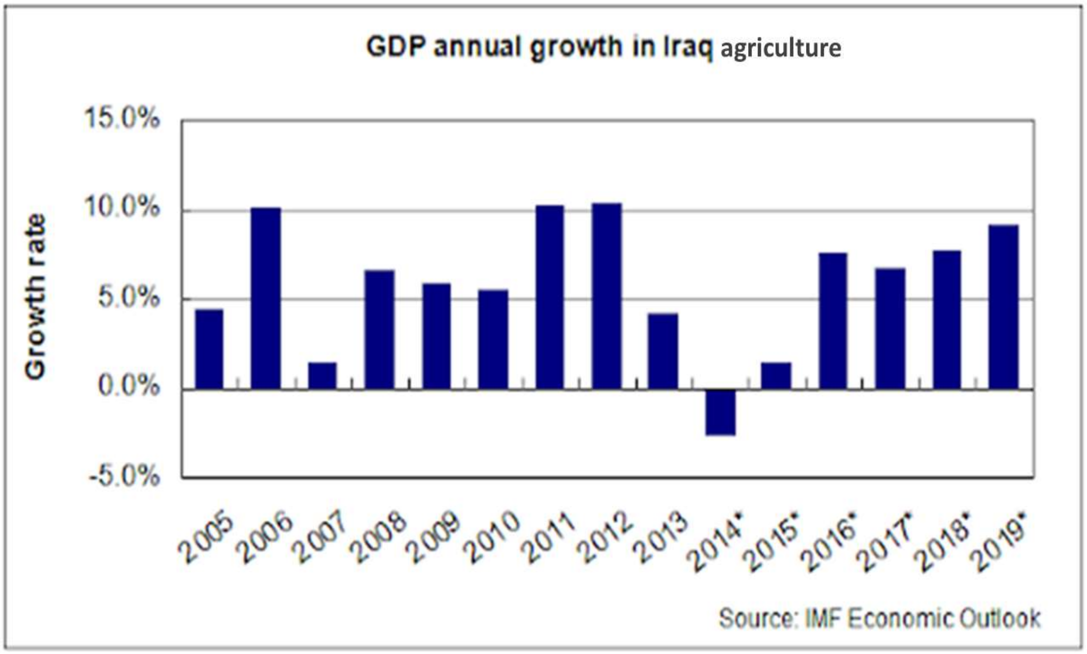
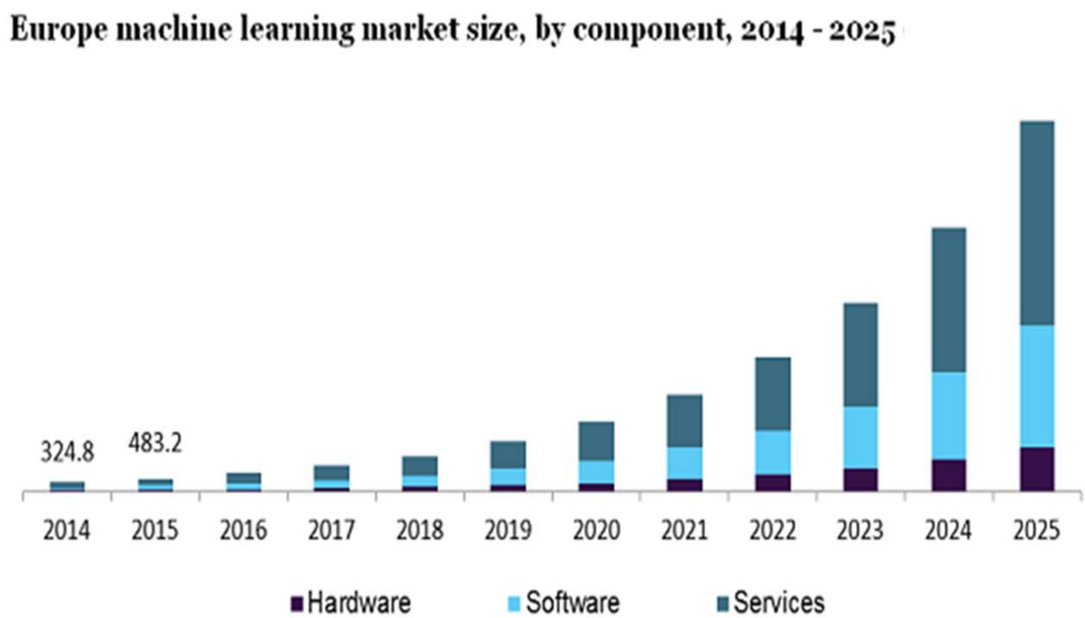


Figure 2

Machine learning growth in recent years (Million insights, 2020)



Problem Statement

The proposed research project work expects to foster the Iraq rainfall forecast model. The exact and exact rainfall expectation is as yet deficient with regards to which could aid assorted fields like agribusiness, flood forecast, and water reservation. Rainfall in Iraq is unequivocally reliant upon the geological area of the Iraq subcontinent; consequently, it is proposed to foster a rainfall forecast model over homogeneous areas of Urban Cities, in the Northern Part of Iraq. The issue is to plan the computations for the rainfall forecast that would be founded on the past discoveries and similitudes and will give the result expectations that are solid and suitable. Iraq Meteorological Department has a few climate stations where everyday rainfall, stickiness, temperatures, and wind speed are estimated. The uncertain and mistaken forecasts are the exercise in futility as well as the deficiency of assets and lead to wasteful administration of emergencies like helpless farming, helpless water stores, and helpless administration of floods. Thusly, the need isn't to detail just the rainfall anticipating framework yet, in addition, a framework that is more exact and exact when contrasted with the current rainfall indicators. Information has been gathered from Iraq Meteorological Department.

Objectives of the Study

The goal of the research is to predict rainfall from Urban Cities, in the Northern Part of Iraq using ML and statistical approaches and picking which approach gives the best extent for prediction. This thesis targets building up a dynamic, smart, and exact framework for the determination and early prediction of rainfall. The objective of the prognosticating philosophy is to style a model that can deduce characteristics of foretelling from the blend of other information.

This research work depicts a framework for early identification and prediction for nations affected with limited rainfall like Iraq for example. This research work set forth depends on using present-day advancements like ML to deduce and predict the rainfall cycle in Iraq. This work in this thesis will offer a plausible guide to specialists, government, and individuals in understanding and predicting the rainfall of a nation.

Significance of the Study

This research work plays a critical significance at fathoming and upgrading the capacity to give a quality framework to individuals, and nations affected with limited rainfall and giving them managerial control and furthermore encourage a way to increase their productivity effectively.

Other significance includes

- Accurately predicting and identifying the arrival of rainfall.
- This will be one of the first research that concentrates on building up another dynamical framework like ML and statistics for predicting rainfall.
- This will be one of the first research that concentrates on offering a plausible guide to professionals, individuals, and government bodies in the creation of momentary goals during the analysis of this rainfall.
- This will be one of the first research to use diverse experimental and ML models for the forecast of conclusion breast cancer.
- This will be one of the first pieces of research that concentrate on the creation of a new programming bundle for statistics for predicting rainfall.
- This will be the principal study that concentrates on delivering a report and adequate findings for assessments.

Limitations of the Study

Building a classifier and a regress utilizing machine learning can be a troublesome task if the dataset utilized isn't organized properly or on the off chance that it isn't effectively deciphered. To set up the dataset, we should discover proper channels and set up the training set before it can be classified (Cramer, et. al., 2017). Accordingly, a significant aspect of this work will be spent on planning and fathoming the dataset to maintain a strategic and efficient analysis. With the headway of ML applications, a few challenges still arise, e.g., low entrance pace of advanced climatology, in spite of quick mechanical development and continued eagerness for computerized climatology among climatologists and scientists (Hussain, et. al., 2020).

The different proposition has been put forth to address these issues, anyway, they still contain impediments with different issues like robustness, security, interoperability, usability, etc. With the ongoing advances in machine vision and ML, the arrangement and analysis of rainfall utilizing historical data pulls in much interest, and creating regressive pattern recognition-based indicative frameworks will assist specialists with improving their analytic quality. In this research, primary methodology depends on utilizing a statistical approach to analyze the data and utilizing the ML to predict those data to achieve a substantial outcome.

Thesis Overview

The remaining portion of the thesis is coordinated in the following manner:

- Chapter 1 gives an introduction to this thesis. It provides an overview of the thesis, aims, problem, limitations, and significance of the study.
- Chapter 2 provides an overview of rainfall, its explanation, progress in agriculture, flood prediction, and water reservation. In addition, enough information on the approaches and methods for calculating rainfall in depth.
- Chapter 3 presents related exploration radiating a clarification of each assessed research paper utilizing ML algorithms and models for the prediction of rainfall.
- Chapter 4 describes the methodology and the framework configuration related to the thesis.
- Chapter 5 presents the experimental outcome and its analysis. Comparison of the experimental result between the implored strategies.
- Finally, the general conclusions just as future work and recommendations are distinctively introduced in chapter 6 where it is expressed that repetition of the experiment would be made utilizing other techniques, to discover a more ideal outcome.

CHAPTER II

Background of Rainfall

Rainfall happens as stratiform or convective downpour; the high scope regions experience stratiform downpour which is a significant predominant type of downpour. It is essential to gauge the conveyance of the rainfall on the worldwide level and for that, as of now the remote satellite detecting methods are helping with estimating the appropriation of the downpour on the worldwide level. These regions incorporate the tropical and subtropical and they experience half to 80% of stratiform downpour rainfall (Mishra, et. al., 2018; Hussain, et. al., 2021d). Unique Sensor Microwave Imager (SSM/I) installed with the US Defense Metrological Satellite Program (DMSP) are utilized to get-together the data about the rainfall with other space-borne instruments like microwave instrument, flying onboard the US – Japanese Tropical Rainfall Measuring Mission (TRMM), and rainfall radar (PR) that work on various frequencies and are aiding the information assortment and in getting the impressions precisely (Bagirov, et. al., 2017; Schumacher, 2017).

Rainfall is one of the main barometrical events that aren't just valuable for the actual climate however for every one of the living creatures on the earth (Gökçekuş, and Nourani, 2018). The rainfall essentially affects the widespread check of climatic dissemination and it influences the nearby climate conditions also. It influences everything straightforwardly or by implication and on the grounds that it is quite possibly the main regular phenomenon; the human beings really must contemplate the rainfall changes with the adjustment of the environment (Alhamshry, et. al., 2019; Htike, 2018). The rainfall helps in adjusting the expanding temperature and in the endurance of the people (Raval, et. al., 2021). Rainfall is likewise remuneration to this large number of stores and it is fundamental for the agribusiness and its creation also. The expanding temperature of the world is related to the unnatural weather change and the water is one of the scant and most valuable assets which in the aftereffect of this expanding temperature are dissipating from the stores. The rainfall peculiarity likewise contrasts with the distinction of districts, planes, bumpy, and levels (Johny, et. al., 2020). The peculiarity of rainfall contrasts with the distinction in scope and longitude.

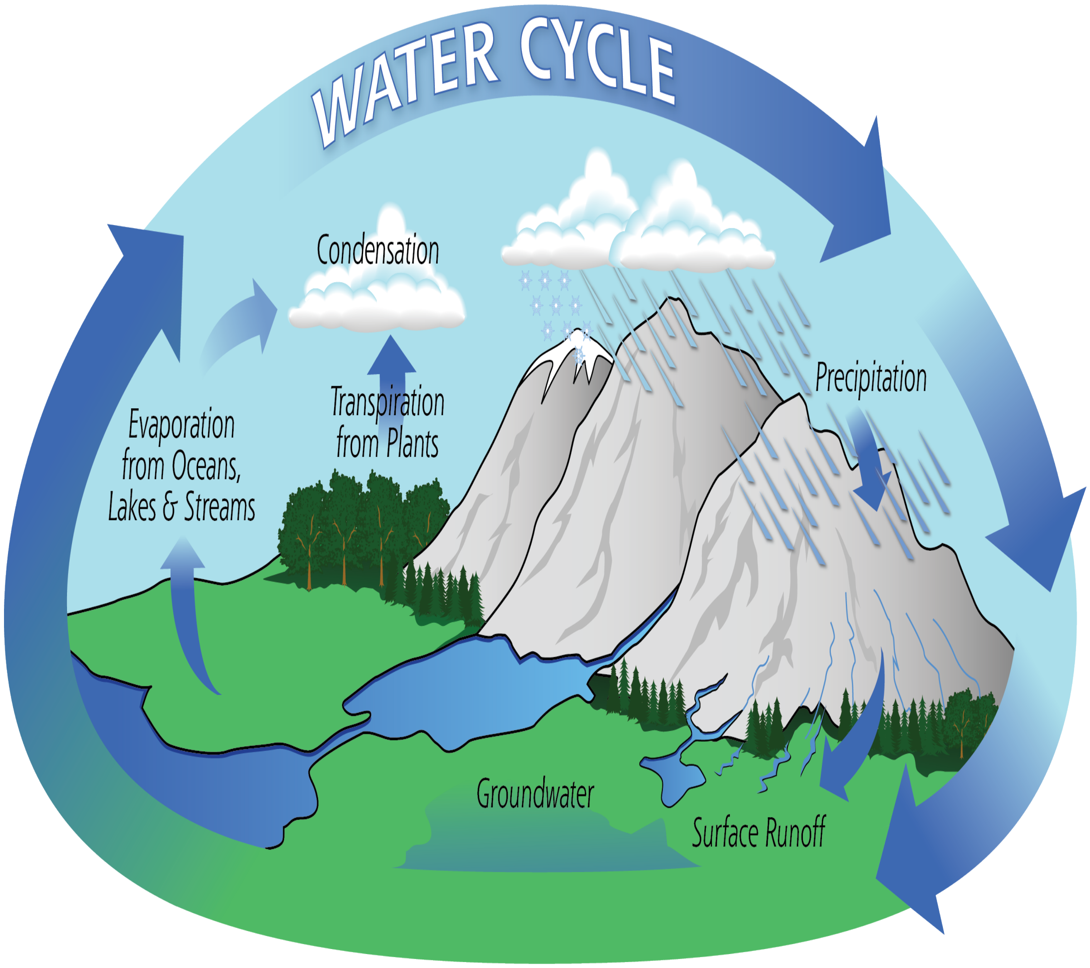
Iraq is a tropical country, and its rainfall changes extensively in recurrence, force, and length. Other than these two significant rainfall periods, a minor amount of downpour falls either as a post rainstorm downpour or as a pre-storm downpour. The biggest measure of rainfall (around 75%) falls as a downpour during the storm season, though, a moderate measure of rainfall falls as snow (around 25%) throughout the colder time of year season. Canli, et. al., (2018) proposed that storm causes around 75% of yearly rainfall covering practically the whole country. These breezes get isolated into two branches, the Arabian Sea branch, and the Bay of Bengal branch. During the summer season, south-westerly rainstorm twists blow at speeds as high as 30 km/hr. On the Earth's surface, there are deviations of land and Ocean (Kassem, and Gökçekuş, 2020). The rainstorm winds of these two branches travel through and strike the mountains or slopes which cause the downpour over the country. The southwesterly wind stream happening over most pieces of Iraq and the Iraq Sea brings about southwest rainstorms over Iraq from June to September. The differential warming of land and Ocean causes varieties in the power of the yearly wavering of the warm equator and henceforth territorial varieties in the force of rainstorm.

Water cycle

The water cycle shows the persistent development of water inside the Earth and the environment. Fluid water dissipates into water fume, gathers to shape mists, and accelerates back to earth as rain and snow (Beusch, et. al., 2018; Davolio, et. al., 2017; Nguyen, and Han, 2017). It is a mind-boggling framework that incorporates a wide range of cycles. Water in various stages travels through the air (transportation). Groundwater moves into plants (plant take-up) and dissipates from plants into the environment (happening). Fluid water streams across the land (overflow), into the ground (invasion and permeation), and through the ground (groundwater). The inverse can likewise occur when water fume becomes strong (statement). Strong ice and snow can transform straightforwardly into a gas (sublimation). The water cycle illustration is given in Figure 3.

Figure 3

Depiction of the water cycle (Precipitation education, 2020)

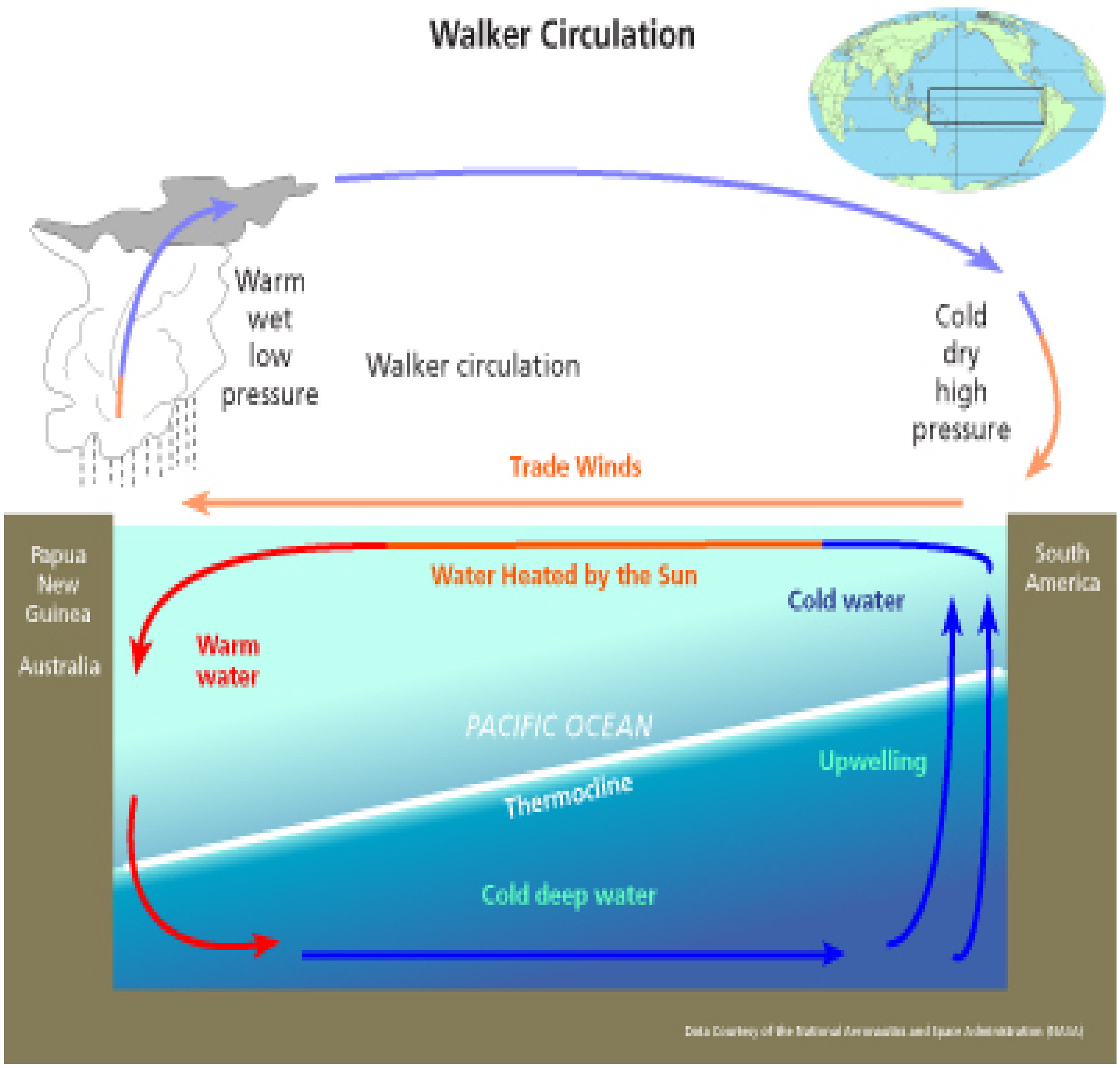


Walker circulation

Sir Gilbert Walker was the primary meteorologist who methodically inspected the connection between subcontinental storm rainfall and worldwide course factors. The walker cycle is depicted in Figure 4. The Walker Circulation or the Southern Oscillation is a significant method of activity of the tropical environment by and large portrayed by the trading of air between the eastern and western halves of the globe (Walker, et al., 2019; Xiang, et al., 2020). The connection between the environment and the ocean at the air-ocean interface brings about the coupling of the flow frameworks of the climate at the sea. Since 70% of the world's surface is covered with water and changes in the ocean surface temperature are a lot more slowly than air deviations, it is obvious that the between yearly inconstancy of the ocean surface temperature might be answerable for the fluctuation of the climatic course and might be all around reflected in the rainfall framework (Khosla, et. al., 2020).

Figure 4

Depiction of walker circulation (Thomas, 2020)

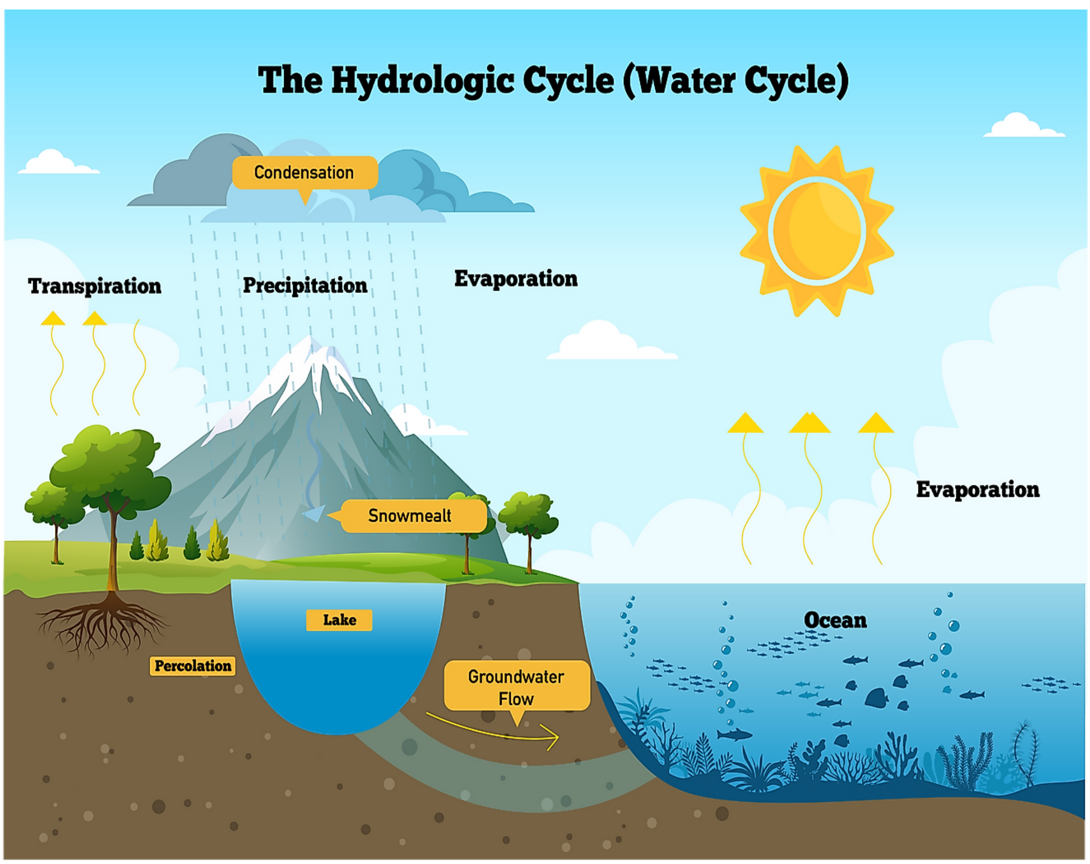


Hydrological cycle

A large portion of the world's water sources, for example, streams, lakes, seas, and underground sources get their provisions from the downpours. Water ascends to the climate as fume from the earth, which is then hastened back as a downpour. This vanishing and rainfall proceeds perpetually and along these lines, an equilibrium is kept up with between the two. This interaction is known as the hydrologic cycle (Kim, et al., 2020; Yaseen, et al., 2018). It is graphically addressed as displayed in the Figure 5. Water dissipates from uncovered water surfaces like streams, waterways, and seas as water fume. As vanishing proceeds, how much barometrical fume continues expanding. Yet, since the environment can hold just a specific fixed measure of water fume, a state is arrived at when any further expansion of fume will get dense on the surfaces (Kassem, et al., 2020). The fume might get consolidated in various structures, for example, fog, downpour, snow, hail, and so forth The water which returns to the outer layer of the earth in its different structures like a downpour, snow, and so on is known as rainfall. The planned rainfall in the space is essentially reliant upon the climate states of the space like barometrical strain, temperature, dampness and so on The communication between these factors prompts the development of winds and typhoons, which cause rainfall. The fluids are continually changed into fume through the course of vanishing (Choudhury, et al., 2019; Lu, et al., 2017). The dissipation misfortunes from water surface rely on elements, for example, space of the water surface, the profundity of the water in the water body, moistness, wind speed, temperature, climatic strain, nature of water.

Figure 5

Depiction of the water cycle (Carly, 2021)



Rainwater measurement

The rainfall is a characteristic peculiarity that is estimated in mm. This is a pipe that assembles the rainfall into a chamber and has the limit of estimating u to 25mm of rainfall (Li, et al., 2017). The estimating instrument is 203mm in distance across. The methods for estimating rainfall as depicted underneath:

Rain gauge (Self-recording)

The conventional technique for recording rainfall is wasteful and gives mistaken outcomes as it's done physically so it can include human blunders (Golding, et al., 2019). The regular self-recording rainfall check is more proficient and more viable in estimating the downpour than that of the standard measure for the estimation of a downpour. Oneself recording check comprises of a tipping pail and a switch balance that gauges the downpour 16 in like manner. Oneself recording rainfall check is seen to involve basic methods and instruments to deliver better estimations to have better likelihood and precision. It has a checking pen that marls each specific level recorded with the development of the container and the rainfall for every hour is recorded by the measure naturally without the help of people (Wang, et al., 2018).

Zonal distribution of rain

The nations along the red ocean are the most important to encounter this sort of rainfall conveyance and the Mediterranean nations are more unsteady to these conditions. The Middle Eastern area encounters a serious fall rainstorm. The North African district is likewise included inside this space and experiences a similar climate. They experience extraordinary rainstorms yet the climate for summer is a greater amount of blistering and dry. These nations additionally experience hailing and may at some point experience flooding. Both the climates are outrageous; if there should arise an occurrence of summers and winter being on the virus front (Graham, and Mishra, 2017; Akiner, 2021; Brown, et al., 2018).

The examples of the rainfall are not steady; it fluctuates from season to prepare and area to area. The rainfall of the rainfall as mean worldwide dissemination is contemplated to be impacted by the latitudinal zones, land and ocean surfaces, and rainfall. There are sure various zones that get more rainfall than not many of them getting less rainfall. It closes by the early August Peninsula and for Japan; it keeps going longer from mid-September to the end of October (Hancock, et. al., 2017). The East Asian district rainfall including China, Korea, and Japan assesses that rainstorm begins from mid-finish of May to the furthest limit of July for China and September in the event that if Korea.

Rain gauge (Ordinary)

It has been seen that the normal check is the non-programmed perception and utilization of glass to gauge the downpour at standard stretches. The customary downpour check estimation is a less viable and less precise strategy of estimating the rainfall. It isn't successful for the heavier and significant rainfall like the cyclonic/front-facing rainfall (Moreno-Rodenas, et al., 2017; Takahashi, et. al., 2019). It has a shell, a capacity bottle with a capacity vessel, and a glass for estimating the downpour. It is less exact and the information gathered may not be exact (Ghamariadyan, and Imteaz, (2021). It utilizes a rainfall record book to look at and measure the rainfall for a specific period. The perceptions acted in the common rainfall are manual so the mistakes are not limited. The normal technique for the rainfall estimation is helping in the neighborhood level estimations and those that are less exact and less exact however this strategy for rainfall assortment is fitting for the estimation record for a bigger level.

CHAPTER III

Literature Review

Long-range gauging can be extensively isolated into four classifications, expanded succinct methodology, Correlation approach, dynamical methodology, and periodicity approach (Chen, et al., 2019). Rainfall forecast is anything but a simple occupation particularly while expecting the exact and exact digits for foreseeing the downpour. Iraqi's Meteorological Department (IMD) is an essential government office in Iraqi for climate forecast. The rainfall expectation is usually used to secure the farming and creation of occasional leafy foods and to support their creation and quality corresponding to how much rain is needed by them (Hadi, et al., 2020). The first figure was given by IMD in quite a while, in view of the connection between Himalayan snow cover and Iraqi's late spring storm rainfall (Panagos, et al., 2017). The rainfall forecast utilizes a few organizations and calculations and gets the information to be given to the agribusiness and creation offices. Sir Gilbert presented connection and relapse methods. In 1988 IMD presented power relapse and parametric model. The rainfall forecast is essential and required particularly in the spaces where there is weighty rainfall and it's all the more frequently anticipated (Farajzadeh, and Alizadeh, 2018). IMD utilized a 16 boundaries power relapse model for the time of 1988 to 2002 with a model mistake of 4%.

There are colossal economies like thosei of Asia like India and China that procure an enormous extent of their income from horticulture and for these economies; rainfall expectation is in reality vital (Staley, et al., 2017; Kim, et al., 2020; Sheen, et al., 2017). These 16-boundaries depend on temperature, strain, wind, and snow-cover. The rainfall gauging is overarching as famous examination in the logical regions in the advanced universe of innovation and development; as it immensely affects simply the human existence however the economies and the living creatures in general. The relationship of these boundaries with rainfall is declining step by step. Rainfall expectation with a few Neural Networks has been breaking down already and the scientists are as yet making a decent attempt to accomplish the more great and exact outcomes in the field of rainfall forecast (Bermúdez, et al., 2017). On account of this declining relationship, 10 boundaries were dropped in 2003. The expectation of occasional rainfall on a month-to-month premise by utilizing the surface information to frame yearly forecast is additionally fundamental for the rural exercises and in this way the creation and management of the horticulture and yields. In 2003 IMD utilized a two-stage determining technique dependent on the past 6 boundaries and some of the recently added boundaries. It very well may be finished by perceiving the varieties in the stock of dampness in the air. The first stage figure was given in April and the subsequent stage was given in June. The instance of the African district delineates that how this succeeded and how West Africa advantaged from the rainfall forecast in dealing with their farming exercises (Sofiati, and Nurlatifah, 2019). From 2003-2006 anticipating was done utilizing 8-boundary and 10-boundary power relapse models (Kim, and Hong, 2021). In 2009, the second stage (August – September) rainfall conjecture was given over the whole country. In 2007 BMD presented a measurable figure framework dependent on outfit strategy.

From 2010, the functional gauge for the rainfall during the final part of the rainstorm season (August-September) and that during the September over the nation is begun. Essentially, the present moment streamflow estimating for the rainfall is additionally solid and inclination-free. BMD has presented many changes in the methodology and extent of rainfall expectation techniques (Tian, et al., 2017; Samanta, et al., 2019; Blum, et. al., 2019). However, they are very little powerful in foreseeing the flood and post-handling of rainfall forecast. Despite the fact that the IMD rainfall expectation model exactness is great, the model is subject to a set of indicators. A methodology called crude mathematical climate forecast (NWP) was presented in 2013, where the methodology zeroed in on the Bayesian joint likelihood model to plan expectation information. 4 Most of the investigations of Bangladeshi's storm anticipating depend on observational and factual methods. The methodology framed conjecture probability disseminations for every area and it possessed expectation energy for 7 it; cooperative figures connected with existence were created in the Southern piece of Australia (Karunakaran, et al., 2019; Ouyang, and Lu, 2018). These factual methods are from basic connection examination to cutting-edge techniques like sanctioned relationship investigation and neural systems administration. This methodology zeroed in on the Schake mix to create the estimate by the conjecture plausibility appropriations (Fereidoon, et. al., 2019). Practically every one of the indicators distinguished up to this point has been founded on relationship examination, however, it is exceptionally touchy to slack period over which it is determined.

Moreover, the present moment streamflow estimating could likewise be utilized through the fake neural organizations as investigated by Zealand, Burn, and Simonovic in 1999. This forces specific restrictions on the dependability of the indicators. The review directed illustrated that ANNs capacity to estimate for transient stream and laid out a portion of the issues that the methodology experienced with ANNs (Navid, and Niloy, 2018). The most regularly involved procedure for storm determining is direct relapse examination. Despite the fact that ANNs with transient streams can compute and introduce perplexing and nonlinear connections among info and result with a capacity to layout the point of interaction impact too yet has issues in handling some information with specific sort and number. Doss-Gollin, et al., (2018) involved Artificial Neural Network to anticipate the rainfall overflow relationship in a catchment space of Japan. The ANNs additionally experienced trouble with aspects of the secret layers. They proposed a model with the utilization of the Feed-forward back engendering with exaggerated digression neurons in the secret layer and direct neurons in the result layer. This examination result was addressed by the information of Winnipeg River framework in Ontario, Canada utilizing the quarter month to month information. Model execution is assessed by other factual records like connection coefficient and mean square blunder. The results of the review were empowering with AANs performing very well for the four expectation lead times. The proposed model is viewed as more precise. Pham, et al., (2020) recommended a dry season hazard file by applying a fake neural organization classifier to bioclimatic time series. The RMSE for the test information of 8 years laid out variety from 5cms to 12.1cms in a gauge from four-time step to double-cross stride ahead separately (Kratzert, et al., 2019). The functional worldly units are created by window moving along time series and window is characterized by quick Fourier change (FFT). The ANN for functional worldly units computes a DRI esteem running between - 1 and 1. There is the degree for examination of the methods utilized for rainfall forecast.

CHAPTER IV
Methodology

Study Area

The study areas in this work are gotten from Urban Cities, in the Northern Part of Iraq. Below more emphasis will be made on each location area. The location has been studied profoundly to be used in the work. Each location has a part to play in the everyday life of the people of Iraq.

- **Erbil:**

Erbil is situated in the north of Iraq. It is a rocky locale. Snowfall occurs in winter the temperature fluctuates from 6 ^{o}C to 10 ^{o}C. Its land span is up to 438 320 km2 with a populace of 28.8 million. Its scope is 36.191113 and its longitude is 44.009167. There is weighty precipitation that occurs over time and other months have lesser precipitation. There are four seasons experienced in the north of Iraq and the winters are very serious. The climate of north Iraq has varieties and the metrological offices measure the boundaries like moistness, gaseous tension, wind course, and wind speed for anticipating the precipitation every day, month to month, or yearly information. The regional depiction is shown in Figure 6.

The water system framework in north Iraq additionally needs to zero in on the booking of water plays a major role in the rural economy; it is likewise huge for this locale to use water effectively. In addition, the area of northern Iraq is wealthy in farming exercises and creation of horticultural items related to the need of precipitation expectation for successful and effective inventory and creation of agrarian items. This likewise makes it critical for the review to examine the weather conditions estimated for this area.

Figure 6

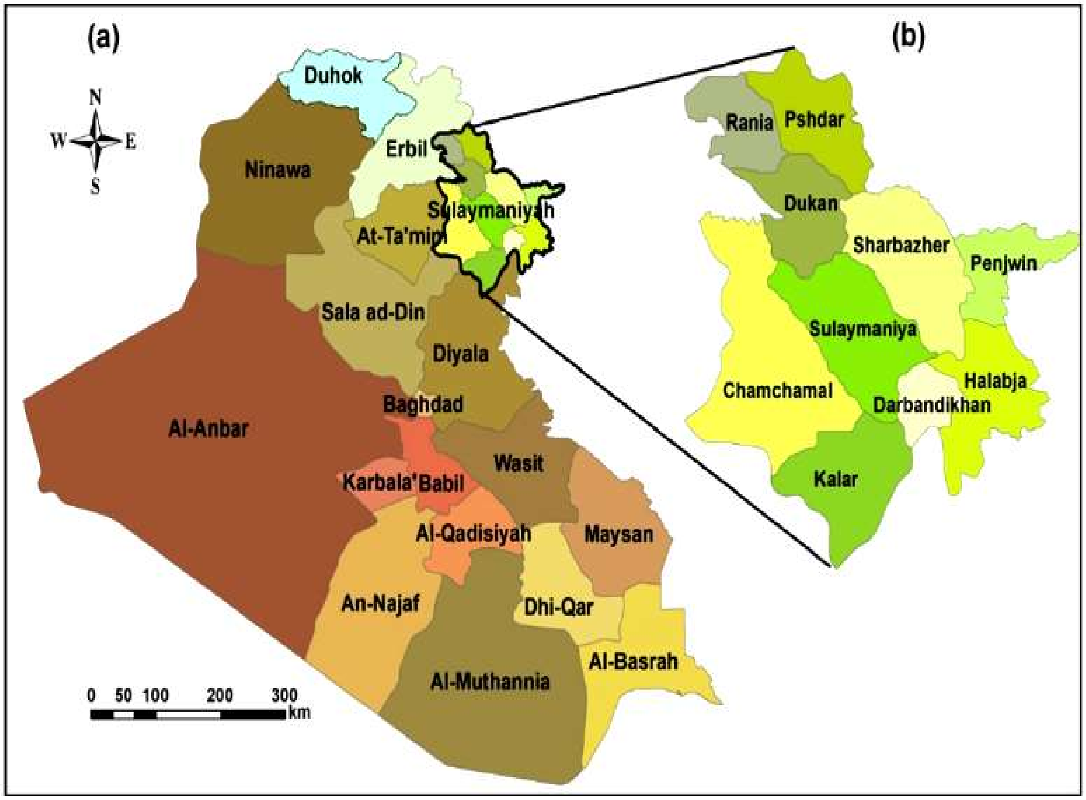
Regional depiction of the erbil

- **Sulamaniah:**

Sulaymaniyah is 831m above ocean level and situated at 35.53° N and 45.45° E. Sulaymaniyah, regional image is displayed in Figure 7. It is a rocky city in the Kurdistan area, Iraq. The climate of Sulaymaniyah is like the Middle-East climate. Sulaymaniyah is encircled by Goizha (1525m), Azmar mountain from Northern-East (1700m), and Qaiwan mountain, and encompassed from the East, Shahrazur plain began with 45km length and 15km width and Baranan mountain from South (1373m). The normal high temperature is 7.9 ^{0}C to 38,9 ^{0}C. The temperature in summer arrives at around 45 ^{0}C and in Winter it arrives at 3.5 ^{0}C or under 0 ^{0}C. The normal low temperature is -0.2 ^{0}C to 24.1 ^{0}C. The regional depiction is shown in Figure 7.

Figure 7

Regional depiction of the sulamaniah

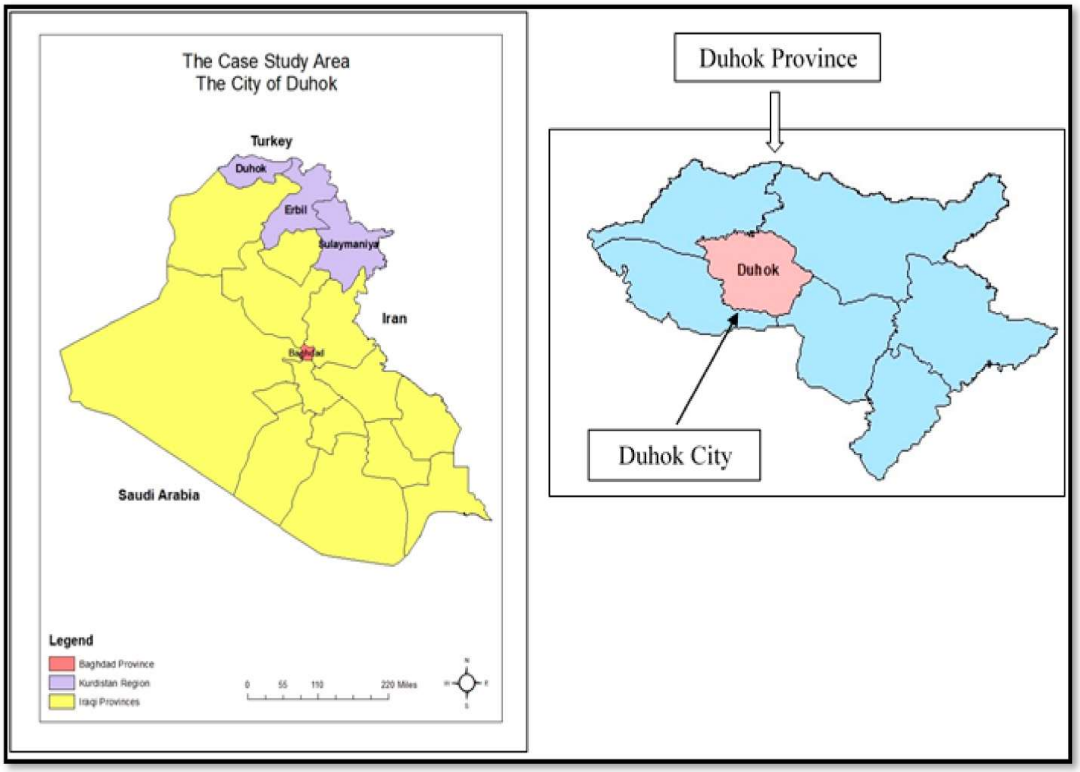


- **Duhok:**

The environment is warm and mild in Dohuk. It is profoundly troublesome, if certainly feasible, to estimate what the weather conditions will resemble at a specific time in an exceptionally exact spot. The normal temperature in Dohuk is 18.5 °C. Around 810mm of precipitation falls every year. In winter, there is considerably more precipitation in Dohuk than in summer. But, all voyagers might want to know ahead of time the environmental conditions to put together their future excursion. The regional depiction is shown in Figure 8. The temperatures referenced henceforth are communicated in degrees Celsius and address the month-to-month midpoints seen over an extraordinary number of years. Normal temperatures or precipitation can assist individuals with finding out about the issue.

Figure 8

Regional depiction of the duhok

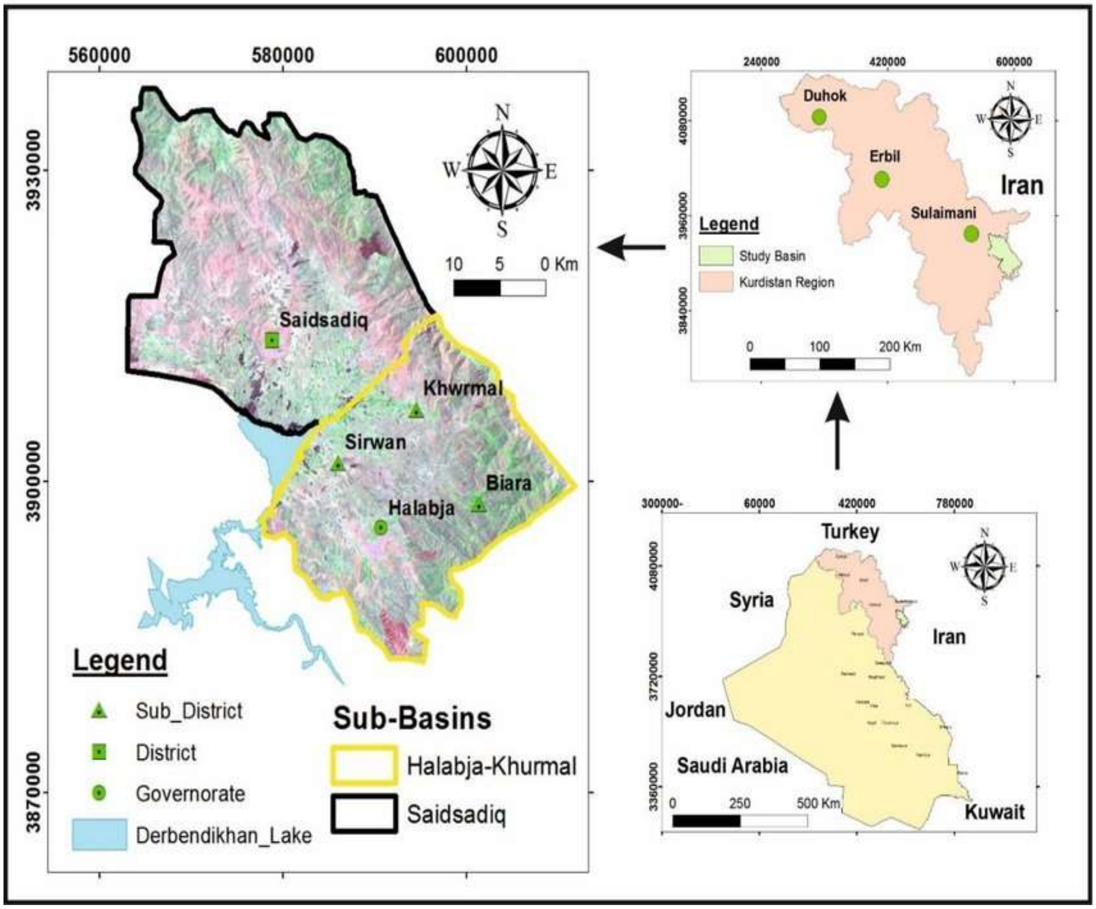


- **Halabja:**

Halabja is encircled by Hawraman and Shnrwe range in the upper east, Balambo range in the south, and Sirwan stream in the west. The city lies at the foundation of what is frequently alluded to as the more noteworthy Hewraman locale extending across the Iran-Iraq line. Rainfall in the city is predominantly needful, however, a few inhabitants of the encompassing towns use the rainfall for farming and basic amenities. Halabja is encircled by Hawraman and Shnrwe range in the upper east, Balambo range in the south, and Sirwan waterway in the west. The regional depiction is shown in Figure 9. The city lies at the foundation of what is frequently alluded to as the more noteworthy Hewraman locale extending across the Iran-Iraq line.

Figure 9

Regional Depiction of the Halabja



The System Model

This is a conventional depiction and it depicts the proposed system overall, fused with the goal and moreover different parts for planning the structure. This section incorporates the depiction and technique, and a nearer perspective on the framework.

Machine Learning (ML)

A subset of computing that enables PC-aided gadgets with the ability to learn without it being emphatically adjusted is referred to as ML. The examination of computational learning theory in ML and pattern recognition explores the algorithmic development and exploration that can gain from, and make propositions on the information. It possesses an extraordinary execution rate that is irksome or impossible; an instance of such applications fuse PC vision, email filtering, sorting out some way to rank, and distinguishing between intruders or pernicious insiders trying to steal data and optical character recognition (OCR). In the scope of computing assignments, Machine learning is utilized were laying out and programming fast calculations (Purnomo, et. al., 2017).

With computational experiences, ML routinely focuses on prediction making utilizing PCs. Information mining is once in a while conflicted with ML, where the previous subsection focuses on exploratory information examination and is called unsupervised learning. ML has strong ties to scientific upgrade, which passes on speculation, techniques, and application zones to the segment. In like manner, ML can be used to learn and construe a measure for various substances and afterward used to find critical peculiarities and subsequently termed unsupervised.

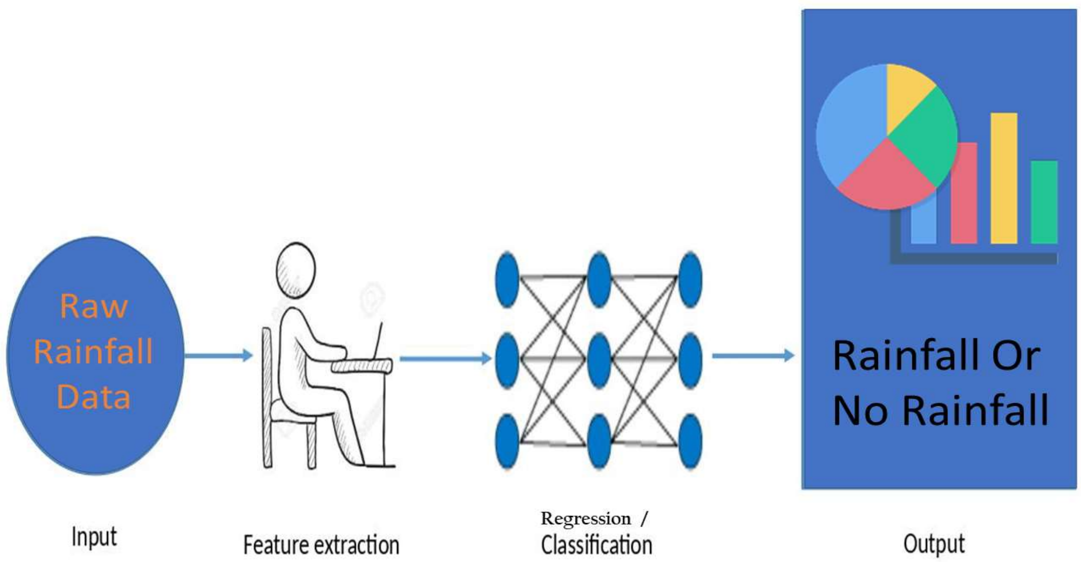
ML in the sector of information investigation is a procedure used to deduce complicated models and computations that are tedious to figure out normally. It is an analytical model that grants information analysts, researchers, and investigators to make repeatable decisions, bringing about reliable and uncovering hidden pieces of information through learning from designs in the information and furthermore irrefutable accurate result. This is referred to as prescient analytics for business use in agriculture (Xiang, and Demir, 2020).

To diminish data dimensionality and hasten imperative inferences and covariance of fast ML approaches will help to speak to data quickly and effectively. ML incorporates characterizing a learning approach to handle an issue contingent upon the data sources. These methods figure extraction of features inside the goal overwhelming various standards on the projection region. While trying to diminish information dimensionality or to learn in an information-driven arrangement element, Dominant and Autonomous Component Analyses have been utilized and besides fairly related to the KNN approach. Then again, unpredictable in reverse systems utilize an outfit of choice trees, where each tree is set up on a substitute subset of the arranging cases, working on the liberality of the general classifier. Every application by then required a fundamental exertion in recognizing the best highlights, which would then be dealt with into a learnable decision computation either for regression or classification. Though, the classification procedure is inclined by probabilistic improvement trees, which form a twofold tree of strong classifiers using a boosting approach to prepare each center point by uniting a ton of weak classifiers (Kajewska-Szkudlarek, 2020).

In this work, we utilized various ML approaches to test and validate the prediction capability. A lot of computations have been suggested, thus, an ordinary choice being the SVM, on account of the straightforwardness of utilization and the notable bits that are nonlinear. A useful representation of the proposed ML approach is depicted in Figure 10. The accompanying segment underneath will portray the ML strategy used for rainfall prediction in this project. In the subsequent subsection, more emphasis on the utilized ML techniques is discussed.

Figure 10

Illustration of machine learning description for rainfall prediction



Application of the Used Techniques

Here, machine learning techniques were utilized to predict rainfall. The utilized techniques are Linear Regression (LR), K-Nearest Neighbors (KNN), Support Vector Machine (SVM), nueral network (NN) and Decision Tree. These techniques were analyzed thoroughly. The input features are moderately and strongly related to environmental variables for predicting rainfall. The proposed techniques are discussed below and further contrasting arguments are presented in the resulting section. Additionally, the ensemble technique is used to deduce the best model after experimenting. The best techniques were identified and recorded based on their computational matrixes.

Support Vector Machine (SVM)

This strategy is effective especially in high layered spaces and useful spaces since they use a subset of focuses on support vectors. it is known as a supervised learning algorithm; it is used for characterization and relapse arrangements. It involves theoretical and numeric abilities to deal with the relapse issue. It gives the most imperative accuracy rate while doing an estimate of the immense dataset. It is a solid ML strategy that depends on 3D and 2D demonstration. SVM groups basic features from all categories, these features are called support vectors. SVM (Support Vector Machine) is one of the directed grouping models that is generally applied in the area of forecast. The Linear classifier of SVM intends to augment the separation between the hyperplane and closest information point, which is known as the restricted distance. SVM has a great precision of up to 98% on information analysis and 78.35% of precision utilizing the polynomial portion.

The highlight of SVM functionality is that it limits the upper bound of speculation mistakes through expanding the space between isolating hyperplane and dataset. SVM is a machine learning technique utilized as a device for information grouping, function estimate, and so on, because of its speculation capacity and has discovered accomplishment in numerous applications. The exhibition of SVM generally relies upon the kernel. SVM has an additional advantage of programmed model determination as both the ideal number and areas of the premise functions are naturally acquired during preparation. Its flowchart application is portrayed in Figure 11. It is a sequence classifier, which utilizes a hyperplane to isolate two classes of sequence-dependent on a given model {x (i), y (i)} $i=1_n$. Where (i) is a vector in the information space $I=R_k$ and y (i) indicates the class file accepting worth 0 or 1. The training information along the hyperplanes close to the limit of the class are known as support vectors, and the space is the demarcation between these vectors and the hyperplanes. A support vector machine is an ML strategy that arranges parallel classes by finding and utilizing a class limit hyperplane boosting the space in the given training data.

The SVM depends on the idea of choice planes that characterize decision limits. A classification process typically includes preparing and testing information which comprises of some instances. SVM is a helpful procedure for information classification. Each occurrence in the preparation set contains one "target label" (class marks) and a few "attributes" (highlights).

Given a preparation set (xi, yi), i=1,… ,l where xi ∈ Run and y ∈ (1,- 1)l, the SVM needs the solution for optimization is calculated using Equation 1 and 2.

$$Min_{w, b, £}\ 1/2w^{T}\ w + c\ \Sigma^{1}{}_{i=1\varepsilon I} \quad (1)$$

$$\text{Subject to } y_i\ (w^{T}\ Ø(x_i)\ +b)>1_{-\varepsilon I},\ \varepsilon_i>=0. \quad (2)$$

Afterwards, SVM looks for a linear approach isolating hyper plane with the maximal space in a higher-dimensional space. Here, the vectors xi is processed into a higher dimensional space by the variable Ø. Moreover, k (xi, xj) = Ø (xi) Ø (xj) is known as the kernel variables.

These involve a linear polynomial, RBF, and sigmoid. Several parts can be utilized in SVM models is calculated using Equation 3.

$$Ø= \{x_i{*}x_j \text{ linear } (\gamma x_i x_j+coeff)^{d} \text{ polynomial Exp } (-\gamma|x_i-x_j|^{2})$$

$$RBF= Tanh\ (\gamma x_i x_j+coeff) \text{ sigmoid}\} \quad (3)$$

There is a similar connection between the Radial Basis Function (RBF) and SVMs classifiers. The RBF is by far the most mainstream decision of part types utilized in SVM. In the area of this thesis imaging, the important utilization of SVMs is in rain discovery. The main SVM is a linear classifier. Similarly, nonlinear SVMs can be made. The component space is a nonlinear guide from the main information space. They are developed by finding a bunch of planes that differentiate at least two classes of information. By developing these planes, SVM finds the limits between the information classes; the components of the information that characterize these limits are called support vectors.

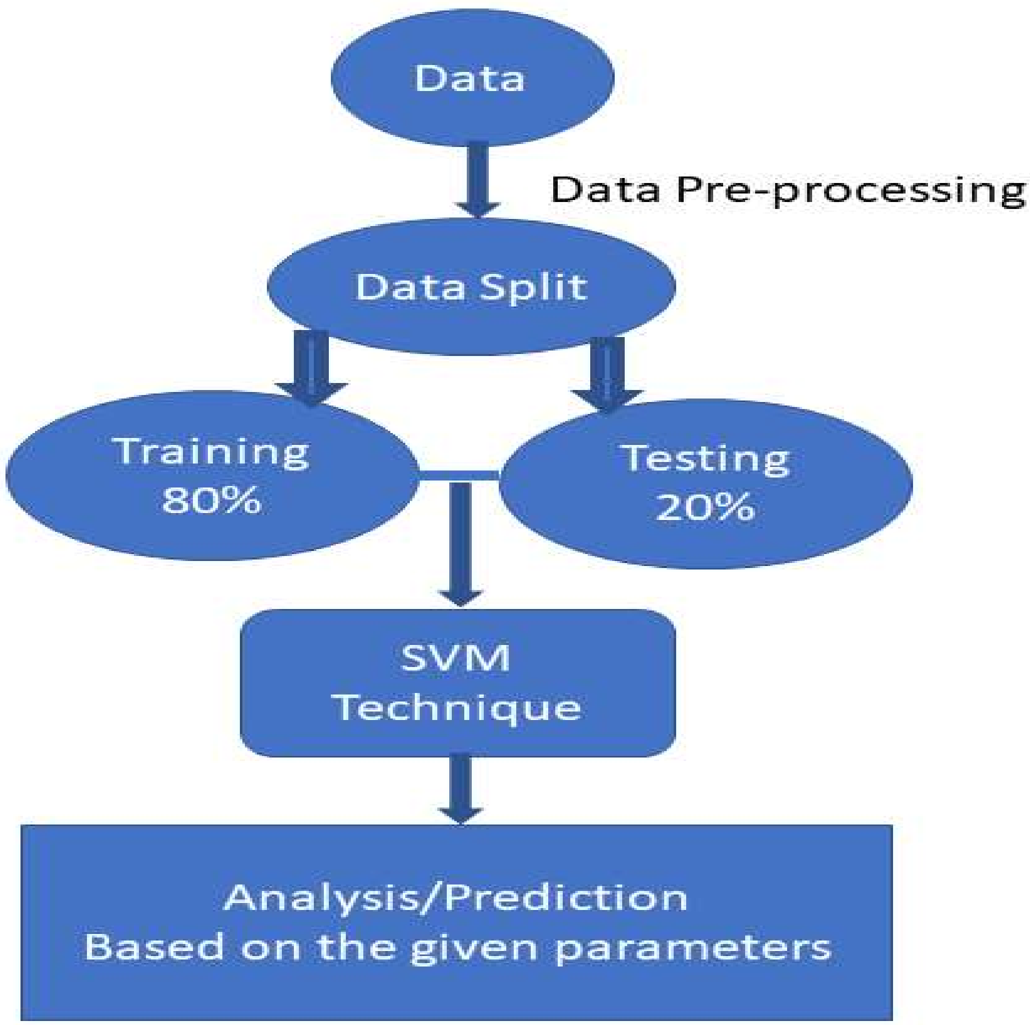
For radial Gaussian basis function is calculated using Equation 4:

$$K\ (x, x') =exp\ (-|x-x'|^{2} / (2\sigma^{2}\ )). \quad (4)$$

The altered kernel is utilized to get the last classifier. The piece is then adjusted in the information path by utilizing the gotten support vectors.

Figure 11

Illustration of SVM application



K-Nearest Neighbors (KNN)

The results of this calculation could vary contingent on how to gauge the space between the information, thus the two techniques, Manhattan distance, and Euclidian distance were studied this time. K-Nearest Neighbor calculations portray the data by finding its closest neighbors in a multi-layered part space involved by known models from a planning informational assortment. The better mixes of dimensions (information highlights) for closest neighbors will in general lead to a more grounded prescient outcome. As per past examination on analyzing a dataset, closest neighbor utilizing Manhattan distance and Euclidian distance brought about 93.567252% and 94.736844% of preciseness, separately.

In KNNs, k is utilized to allude to the quantity of closest neighbors that are to be involved in the voting cycles. KNN keeps all the training information and arranges the inquiry information dependent on a comparability measure. KNNs use highlight similitude. To improve execution, KNN boundary tuning is done by picking a proper estimation of k. The likeness between two points is determined utilizing, for instance, the Euclidean distance. A useful flowchart representation showing the working of the technique is given in Figure 12.

To pick a value for the variable k:

Input: Give an example of N models and their classes.

The class of example x is c(x):

Give another example y:

Decide the k-nearest neighbors of y by estimating the distances.

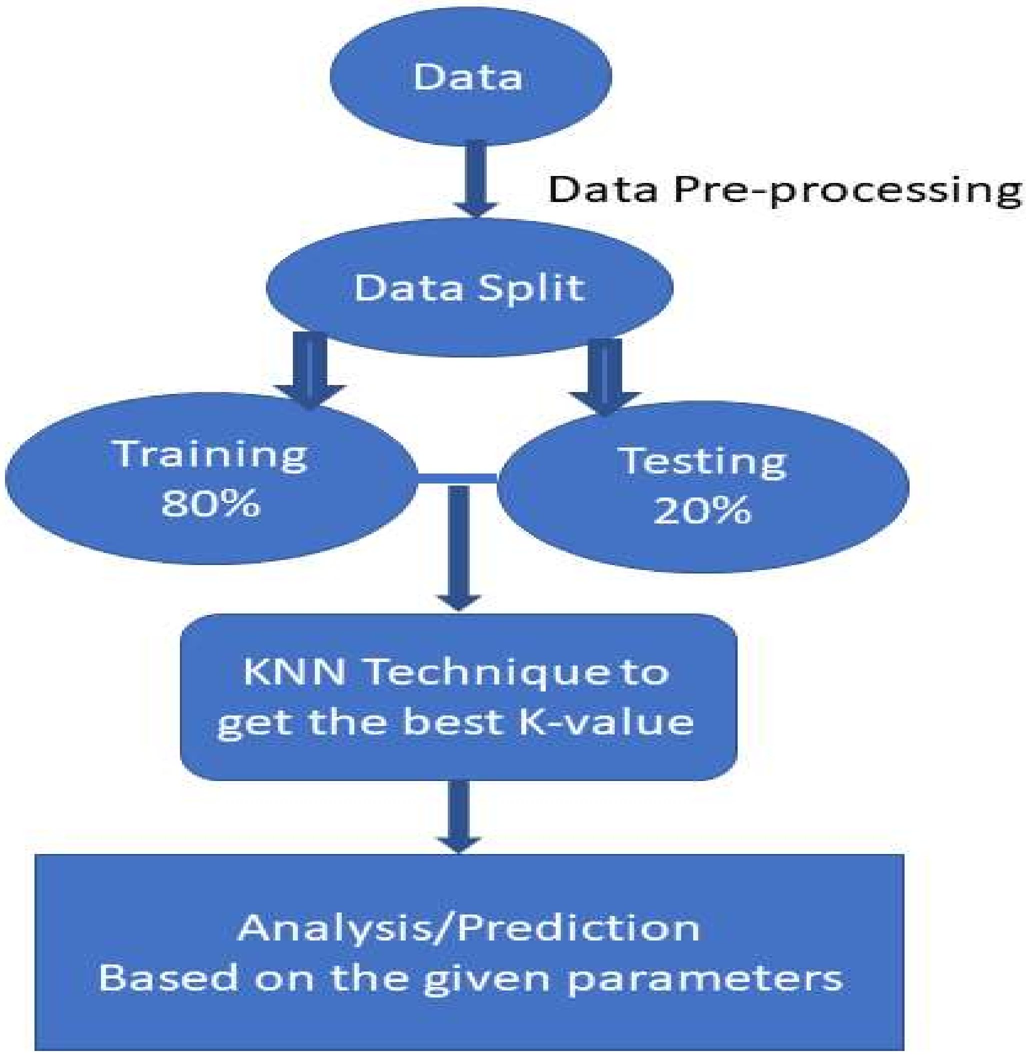
Merge classes of these y models in a single class c

Yield: The class of y is c(Y) = c

The impact of noise on the arrangement is decreased when the noises picked for k is more noteworthy, however, this makes the limits between classes less unmistakable. The decision of the variable k (k 2 N) is controlled by the user, this decision relies upon the information. In this research, the value of k that limits the expected mistake was picked. A decent decision of the value of k can be chosen by various heuristic methods, for example, cross-validation. In the event of balance, I can expand the value of k from 1. On account of a paired order, it is additionally welcoming to pick an odd value for k, it dodges the equivalent votes.

Figure 12

Illustration of KNN application



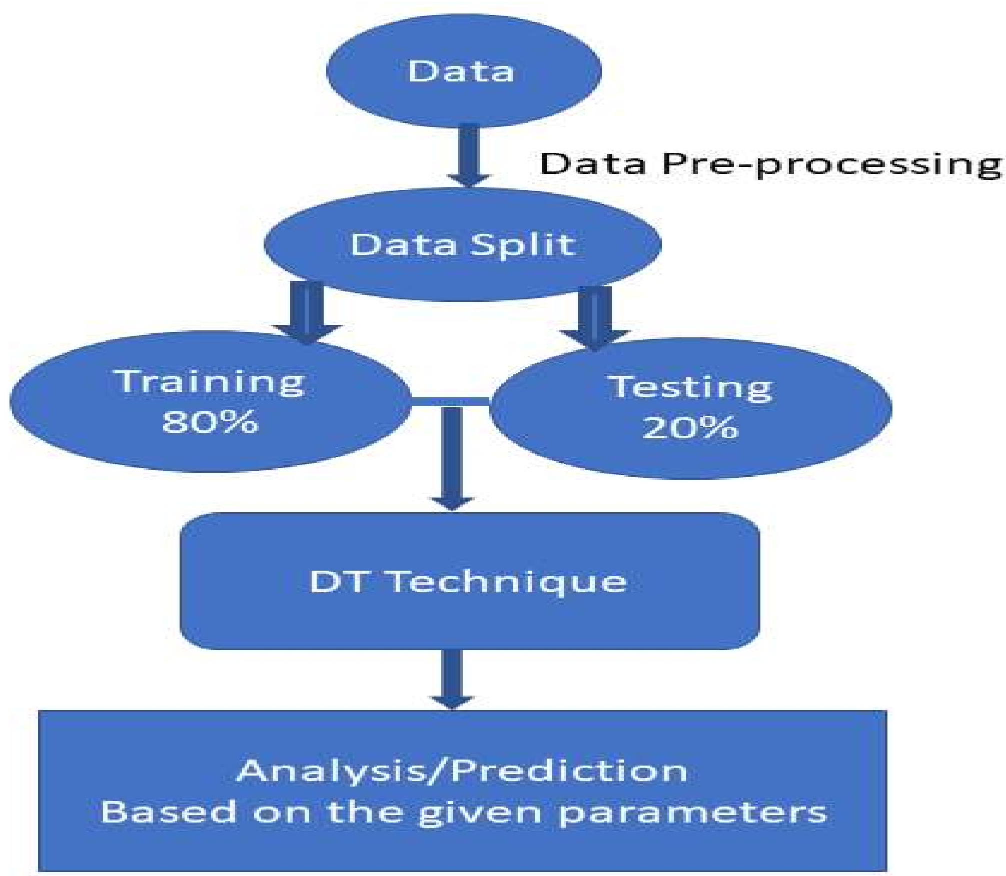
Decision Tree (DT)

A decision tree involves centre points that have definitively one moving toward the edge, aside from the root centre that has no oncoming edges. A Decision Tree is quite possibly the most comprehensively used arrangement strategy. They are extraordinarily simple to appreciate, interpret and moreover work on human significance. A centre point with dynamic edges is an inside centre, while different centres are called leaves or terminal centres, or choice centre points. I applied the best mix of boundaries on cross-validation, and with the aim of the framework should achieve a better accuracy on the revelation dataset. They are minimal and can deal with missing information. A practical flowchart showing the working of the technique is depicted in Figure 13. Unpredictable decision trees can interpret and manage irrelevant qualities in a clear manner.

Decision trees foresee reactions to information. The decision tree is a traditional and efficient factual learning algorithm and has been generally used to tackle classification and regression issues. The leaf hub contains the reaction. Classification trees give reactions that are ostensible, while regression trees give numeric reactions. To foresee a reaction, inquire about the new information by the choices from the root hub to a leaf hub in the tree. In this research, a classification strategy is utilized.

Figure 13

Illustration of DT application



Linear Regression (LR)

This technique is quite possibly the most surely known and notable algorithm in the real and AI world. It has been created in the field of insights, which has been studied as a model for understanding the connection between info and the result of mathematical factors. Also, now it has been acquired by AI. It showers a measurable calculation like an AI calculation. It is an exceptionally straightforward methodology for administered learning. However it might appear, there are a few other best practices and most utilized calculation contrasted with linear regression, yet is a helpful and broadly utilized measurable learning strategy. It is utilized to anticipate a quantitative reaction Y from the indicator variable X.

For linear relapse, it is expected that there is a direct connection between X and Y.

The numerical relationship or condition resemble this Equation 5.

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n \quad (5)$$

Here:

• β is known as model coefficients. It is “learned” during the model training/fitting phase.

• y is the response

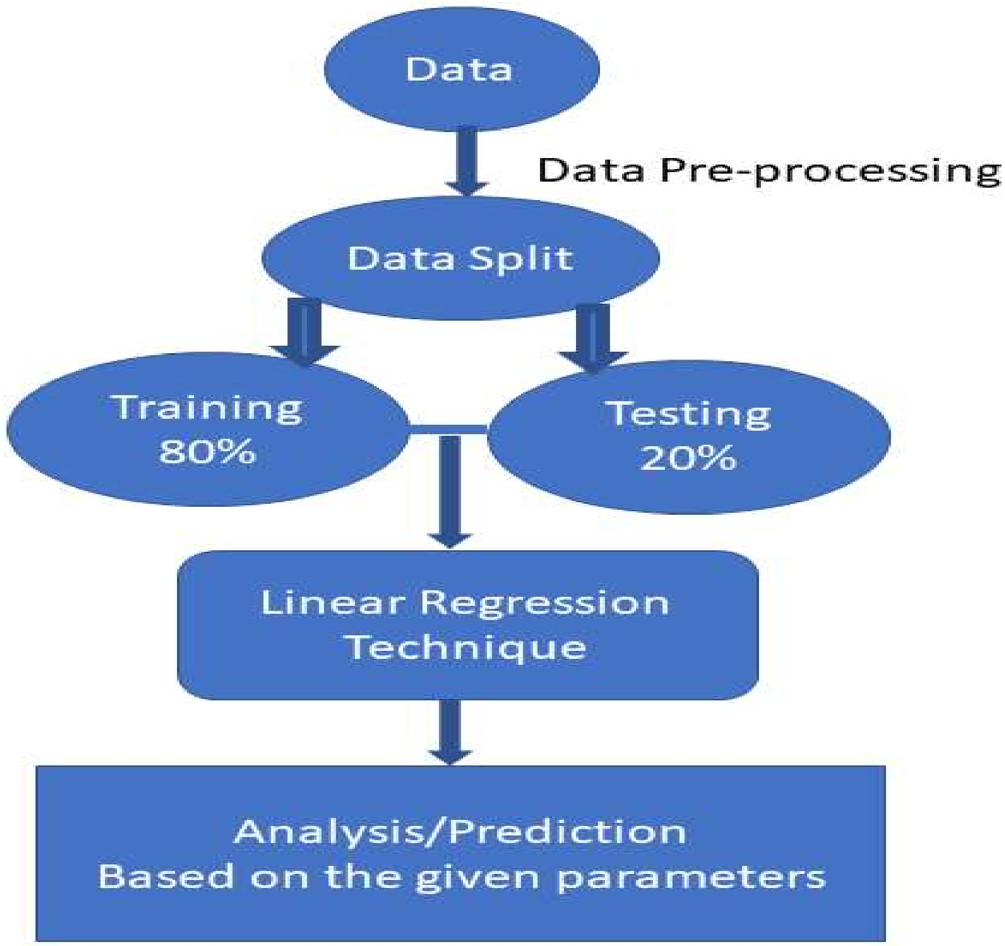
• β1 known as coefficient for X1 (the first feature)

• β0 known as intercept

The data used in this technique was split to 80:20 percentage testing and training after data set processing. The technique discussed above was applied to it and the result analysis will be discussed further in the resulting section. The technique flowchart is given in Figure 14.

Figure 14

Illustration of LR application



Neural Network for Analysis

Neural Networks (NN) are separated from the single shrouded layer neural network as far as their profundity. The further deep into the network, the more unpredictable the highlights the hubs can identify and learn. In profound learning networks, every layer trains on an unmistakable arrangement of highlights dependent on the preceding layer's result. The structure of the NN model is made up of hidden layers. Neural network can arrive at 95% of prediction on analysis information. Note that a few layers consist of boundaries and others do not. Thus, it change the original picture tier by tier from the original pixel contents to the last class scores. The boundaries in the NN tiers will be prepared with slope drop so the class scores that the NN processes are predictable with the marks in the preparation set for each picture. Then again, the RELU tiers will actualize a fixed capacity.

Let f be a NN with N tiers sequential structure (f1; f2; … ; fN). A NN comprises of multi-sub-testing, sub-inspecting, nonlinear, and altogether interconnected tiers. Mappings can be carried out between input (w) and yield (u) as demonstrated in Equation (6) below:

$$u = f(w; X_1, X_2, X_3, \ldots, X_N) = f_1(w; X_1)\ of_2(\mathrm{w}; X_2) \ldots of_{N\text{-}1}(\mathrm{w}; X_{N\text{-}1})\ of_N(\mathrm{w}; X_N) \quad (6)$$

Where XN indicates inclination and weight vector for the nth tier fN. fN has been assigned to do spatial convolution or non-linear enactment or classification. As indicated by the scope of η preparing information {(w(i), u(i))}_(i=1)^η, vectors (X1, X2, X3, … , XN) can be decided as follows Equation (7):

$$\underset{X_1, X_2, X_3 \ldots X_N}{\arg min} \frac{1}{\eta} \sum_{i=1}^{\eta} f_{Loss}(f(w^{(i)}; X_1, X_2, X_3, X_N), u^{(i)} \quad (7)$$

Stochastic decrement and reverse engendering methodologies can achieve Equation (3.6). Where f_{Loss} infers loss work. The element map FM_m^h the equation at m level is depicted in equation (8). In the calculation of a component map, a convolutional tier commonly uses convolutional channels.

$$FM_m^h = f\ (\propto_m^h + \sum_j FM_j^{h-1} * G_{jm}^h) \quad (8)$$

Predispositions and bits individually are G_{jm}^h and α_m^h. FM_{in}^{h-1} and FM_{out}^h are a few qualities of input and yield. Two segments make highlight maps for every convolution tier. The advantages of this technique are the capacity to gauge the picture size of the information and to make a positive distinction in neighbourhood districts. The nearby responsive territory is the main component, and shared weights are the next part. The subsequent equation is utilized to decide the capacity as indicated in Equation (9)

$$\psi_j = \max(\psi_i^{n*n} z(n, n)) \quad (9)$$

At the point where $x \in X$, $\sum_p \mu_p (X \mid \Theta) = 1$. In this situation, π_{py} and $\pi = (\pi_p)p \in P$ is a probability that the example would enter leaf p on class y and recognize by $\mu_p (X \mid \Theta)$. Choice hubs depend on the stochastic schedule and have been portrayed as Equation (10):

$$f_d(x; \Theta) = \sigma(f_r(x; \Theta)) \quad (10)$$

The sigmoid capacity σ (*x*) for this situation is set to σ (X) = $\frac{1}{(1+e^{-x})}$ the choice timberland is known as a gathering of dynamic trees and is characterized by the subsequent Equation (11)

$$F = \{T_1, T_2, \ldots, T_z\} \quad (11)$$

Let *I* be a = $I_1, I_2, I_3, \ldots, I_Q$, where *Q* shows a bunch of pixels and I_Q implies the size of the dark degree of a pixel L, $K = (K_1, K_2, K_3, \ldots, K_Q$ where $K_Q \in LL = \{0, 1\}$ can be summed up to a set with positive names as described in Equation (12).

$$K^* = \arg\min_{k}\{Y(I \mid K, \quad)Y(K)\} \quad (12)$$

$Y(K)$ is a dissemination of Gibbs. Equation (13) and can be composed as in the Expect-Maximization computation.

$$K^* = \arg\min_{K \in k}\{U(I \mid K, \quad)U(K)\} \quad (13)$$

Here *U* relates to urinary potential or energy of possibility and is shown by Equation (14)

$$U(I \mid K, \quad) = \sum_Q [\frac{(I_Q - \mu_{KQ})^2}{2\sigma_K^2} + In\ \sigma_K] \quad (14)$$

Data Description

Performing data analysis exploration is valuable when it comes to problem-solving in machine learning since it brings one closer to the certainty that the future predicted results will be valid. Also, several advantages like, correctly interpreting data that can be applicable in a business desired context. With this, archiving certainty can be guaranteed only after checking and validating raw data for anomalies and ensuring no errors during the collection of the dataset. It also helps locate various insights not worth investigating or made evident to researchers or business stakeholders. The data here consist of four stations, data within the last 9 years, 12 months in a calendar year, and the sum total of the rainfall per year in (mm). Also, the data gotten is recorded in CSV extension format. The data is kept in a record that lays out a complex and vectors networks binary arrangement. Some questionable data like imbalanced data, for instance, spam/coercion area and N/A were taken care of impeccably. The factors of the data are depicted in Table 1. The datasets in our task are gotten from Urban Cities, in the Northern Part of Iraq. The dataset is outstandingly skewed and it processes close to 600 non-repeated outcomes and 47 repeated outcomes. Some data cleaning operations were performed and will be discussed below.

Data pre-handling

Pre-dealing with and Exploratory Data Analysis (EDA) are the principle worries in data science. This pre-dealing with process serves to significantly appreciate the datasets and close how to control the elements that reliably influence the outcome. This approach can also be referred to as data preprocessing. It is a mining approach that transforms or processes raw data into a format that is understandable. Raw data are most times inconsistent, incomplete, and lacks certain characteristics, trends, or behaviors, and they mostly contain huge errors. A certain step was carried out, it will be listed below.

Feature Expansion: The features gotten from the date are expanded to year, month and day. The new features created from these will be used further for other pre-processing steps. Missing Values: With respect to the pre-processing step, it is learned that there will be few null values. Hence, this step becomes very important. In order to input the set of missing values, it will be grouped based on the date and location. Hence the null values will be replaced by their mean value respectively.

Handling Class Imbalance: The dataset has to be balanced to avoid imbalance. Imbalanced data produces a biased output. With this, the model wouldn't learn much and prediction might be void.

Feature Selection: This is a process where features can be manually or automatically selected from those which contribute tremendously to the prediction model from the dataset. Here, all features are important to perform this experiment.

Data normalization: For guaranteeing that the dataset is solid; that is, every information has the right highlights and information esteem. Data standardization was executed trailed by data assessment and insight.

Table 1.

Dataset attributes and meaning

Attributes	Meaning
Station	This shows the station location of where the data was collected
Year	This is the year in which the data was recorded
Average humidity	This denotes the average humidty for the given region.
Average temperature	This is the average temperature for the given region
Month	This is the month in which the data was recorded and shows the rainfall of each month
Total	This shows the total rainfall for the year

Statistical Approach

This approach helps provide data collection, exploration, and presentation to make adequate understanding and assist in discovering hidden trends and patterns. This approach is utilized for daily prediction schemes, in government organizations and industries to generate adequate information and provide decision making. The utilized statistical approach is mean, standard deviation (S.D), median, range, maximum, and minimum of the average rainfall with the yearly period. The results and the computational parameters utilized in this approach will be presented and discussed in the resulting section.

Visualization

A vital inspiration driving portrayal is depicting the dataset and graphically conversing with it. This is with the creative mind that the outcome is a plan that contains the assistance of using visual instruments, so the test results are depicted. Here, the outcome, and pattern gotten from the results are investigated with the use of visual representation. The resulting outcome will be analyzed and discussed in the result section.

Computational Environment

VScode is a standard environment for programming that contains a solid set-up of instruments for information evaluation and techniques. Python is one of the most notable programming paradigms, and it offers programming chances on various libraries that can manage data science, for instance, data examination, import ¡datasets, data pre-taking, and others. The examinations done in this paper were executed using the VScode, which is an open-source condition that ¡drives the utilization of ML systems. It is used on various platforms like Windows, macOS, or Linux, and with these current features can be incorporated. Tensorflow is a library that puts forth various exhaustive properties with regard to ML. It is likewise the most natural and effective library and it is utilized in this examination. Pandas, Keras, SciPy, TensorFlow, NumPy, Matplotlib, and Scikit-Learn, were also utilised. Also, Matlab and MS-excel are used for the statistical approach. The experiment was assessed on a pc with, intel center i7 Processors: 2.3Ghz, GPU: EFORCE, RAM: 12GB, Disk: 1TB.

Experimental Procedure

To take care of the prediction issue, the dataset is gathered and it’s followed by information processing strategies, then the novel ML technique is being played out, which is the critical process. From standard criticism and understanding of the thesis, the thesis has been on target so far. Since this is a prediction process ML supervised learning techniques were applied. The process underneath shows a rundown of our machine learning technique. Afterward, it was analyzed using different attributes and evaluations matrices. After the execution of the ML technique, at that point, a split ratio was given using a cross-validation approach followed by the testing system to close which model fit best utilizing the heat map and confusion matrix. Furthermore, an ensemble technique was used to deduce which approach fits perfectly. The illustration of the experimental working process is depicted in Figure 15. The results will be discussed in the accompanying resulting section.

The means of our ML implementation of rainfall prediction can be summed up as follows:

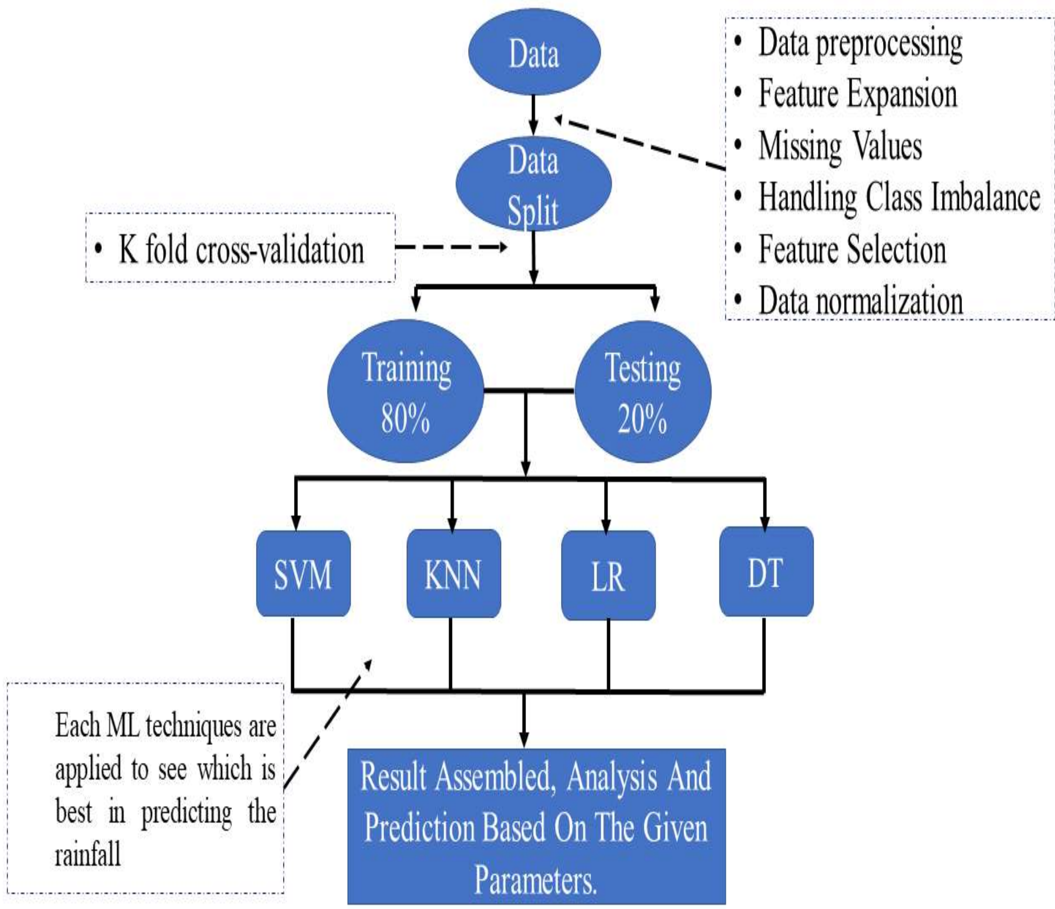
- Data selection: Choosing datasets that are from the Urban Cities, in the Northern Part of Iraq.
- Data pre-processing: Cleaning up the datasets, using various pre-processing steps were applied. Various useful features were selected and the used dataset was picked then stacked and arranged.
- Visualization: Various plots were carried out to view the data and several statistical approaches were made to view the data. This approach can help our expectation of determining what the prediction of rainfall will look like. Following that data, the exploratory assessment was done to conclude whether there are any special cases or anything of interest that might influence the outcomes.
- Several statistical methods were applied efficiently each data efficiently.
- Next was the preparation cycle to start the experiment with various ML techniques.
- Validation: The data was split into train /test split (80% to 20%) based on the K-fold validation. The models were trained for the ML technique to understand what operation it is carrying out. The "test" was utilized on the grounds that most writing surveys didn't make reference to their split proportions by any means. Testing outcomes were predominantly recorded to contrast and to confirm if the results were substantial. K-fold validation was used in order to validate how efficient the models are.
- From the k-fold where k=4 it was observed that we can apply one more procedure for differentiating models. Next appraisal measures and computational boundaries were applied to approve every ML model. The results were recorded and the techniques were audited, using a confusion matrix and heat map.
- Ensemble method: Based on some of the proposed ML techniques, parameters were tweaked and upgraded in order for the results to be better. For example: Getting the best value of K in the KNN technique. Averaging ensemble technique is used to pick which model is the best fit for prediction. The averaging technique is like the max voting procedure, based on individual data points in averaging multiple predictions are

made. The average prediction for each ML technique is used to get the final prediction result.

- Results were all assembled and similar investigations were applied for each process. At last, the best technique for predicting was determined and discussed further.

Figure 15

Illustration of the ML experimental process



Experimental Parameters

- The F1 score depicts the harmony between the recall and precision. It is calculated using Equation 6:

$$F_{1=}2 \times \frac{precision \times recall}{precision+recall} = \frac{TP}{TP+\frac{1}{2}(FP+FN)} \tag{6}$$

FP= False positives

TP= True positives

FN= False negatives

- Precision

This is a decent measure to decide when the expenses of False Positive are high. It works out how exactly our model is out of those anticipated positive, the number of them are really positive. It is calculated using Equation 7:

$$\frac{True\ Positive}{True\ Positive+Fal\quad positive} \tag{7}$$

- Recall

This ascertains the number of the Actual Positives the proposed model catch by naming it as the rue Positive. It is calculated using Equation 8:

$$\frac{True\ Positive}{True\ Positive+False\ Negative} \tag{8}$$

- Accuracy

This approach ascertains how accurate the model is. It is calculated using Equation 9:

$$\frac{True\ Positive+Tru\quad Negative}{True\ Positive+True\ Negative+\ False\ positive+\ False\ Negative} \tag{9}$$

- SD

Statistically, the standard deviation is a proportion of how much variety or scattering of a bunch of values. It is is calculated using Equation 10:

$$\sigma = \sqrt{\frac{\sum(x_i - \mu)^2}{N}} \tag{10}$$

N = population size

σ = standard deviation of population

x_i = every population value

μ = mean of the population

- Mean

The mean is the value of the average number and is utilized to quantify the focal inclination of the information. It is calculated using Equation 11:

$$m = \frac{Sum\ of\ term}{Number\ of\ term} \tag{11}$$

m = mean

- Median

Statistically, this is worth isolating the higher half from the lower half of an information test.

- *R*-Square (R^2)

The R-square or coefficient of assurance is a straightforward measure and is frequently used to decide the presentation of a regressive model. It gives an outline of the independent variable in foreseeing the dependent variable. The computation of the R-square is as per the Equation 12:

$$R^2 = \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y}_i)^2} \times 100\% \tag{12}$$

y_i is depicts the value of observation vduring the period i, $\hat{y}_i$ depicts the forecast value in period i, and the observations number is given as N.

- Root Mean Squared Error (RMSE)

The RMSE is a well known proportion of assessing the decency of a predicting models. RMSE is the root worth of the mean squared predicting bias. The estimating inclination can be deciphered as the contrast between the actual value and the predicted value. The RMSE equation can be communicated as Equation (13):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N}} \tag{13}$$

The component of RMSE is given in equation (12).

CHAPTER V
Result

To assess the reasonability of our methodology, we attempted the dataset on various ML models. We have done a ton of different assessments with the most consoling arrangement of hyperparameters' classifiers. Each model and component anticipate that a basic limit should get the higher evaluation model. The models have been endeavored with various settings to accomplish the main accuracy. We have utilized four models KNN, SVM, DT, and LR to foresee the data with different game plans for preparing tests. A statistical analogy was also utilized in the experimentation to speak directly to the data. Right when we recognize the best model for prediction, we lessen the arrangement set in size to see what the limit is to best guess this data. The utilized data is part into two areas the preparation and testing sets. Likewise, a gigantic instructive assortment of 2592 instances and 6 characteristics are utilized in this assessment. VScode was involved which includes various libraries for this errand. Each model has the respect to examine the rainfall in each local. As follows the consequence of the models is clarified in the system area. Then contrast the best-recognized classifier and evaluation metric communicated close to the start of the errand. In Table 2 we portray a section of the input rainfall data. Each model utilized a comparable area of enlightening assortments.

Table 2.

Rainfall data depiction

Year	Station	Months	Rainfall	Station Index	Avg humidity	Avg temperature
2012	Erbil	1	56.1	1	74.4	8.2
2012	Erbil	2	31.5	1	63.6	9.55
2012	Erbil	3	72.1	1	61.7	11.2
2012	Erbil	4	14.5	1	49	21.95
2012	Erbil	5	26.8	1	41.5	27.1

The Statical Performance Analysis

The consequences of the ML models will be found in the figures and tables beneath. To guarantee model consistency, all models executed in this attempt utilized a 80% by 20% train and test split. The table includes the result of our used models for the determination of rainfall prediction. From the models, performing proficiently, the best model for rainfall prediction is the DT obtaining a precision of 97% then SVM at 93%, and KNN at 96%. Moreover, in view of the qualification in pre-taking care of and train/test split the results were useful for each model. From the beginning of the experimentation, default limits were applied from the beginning. Likewise, the models were analyzed with default limits, we similarly cultivated the DT is best accuracy model followed by KNN and SVM.

Every component and model accept a key part to getting the higher evaluation model. To survey the sufficiency of our strategy, we tried the dataset on different arranging models. Figure 16 begins the depiction of the statistical approach and the ML, it shows the heatmap of the utilized dataset so it can be visualized and spoken to graphically. While in Figure 17 shows the relationship in seclusion. It shows the correlation between each attribute of the dataset. This concept shows each dataset's individual similarities. It shows that the relationship between the mean range and the edge is solid, additionally the mean border and region, and also shows the connection between the highlights.

Figure 16

Heat map showing the mean area target output

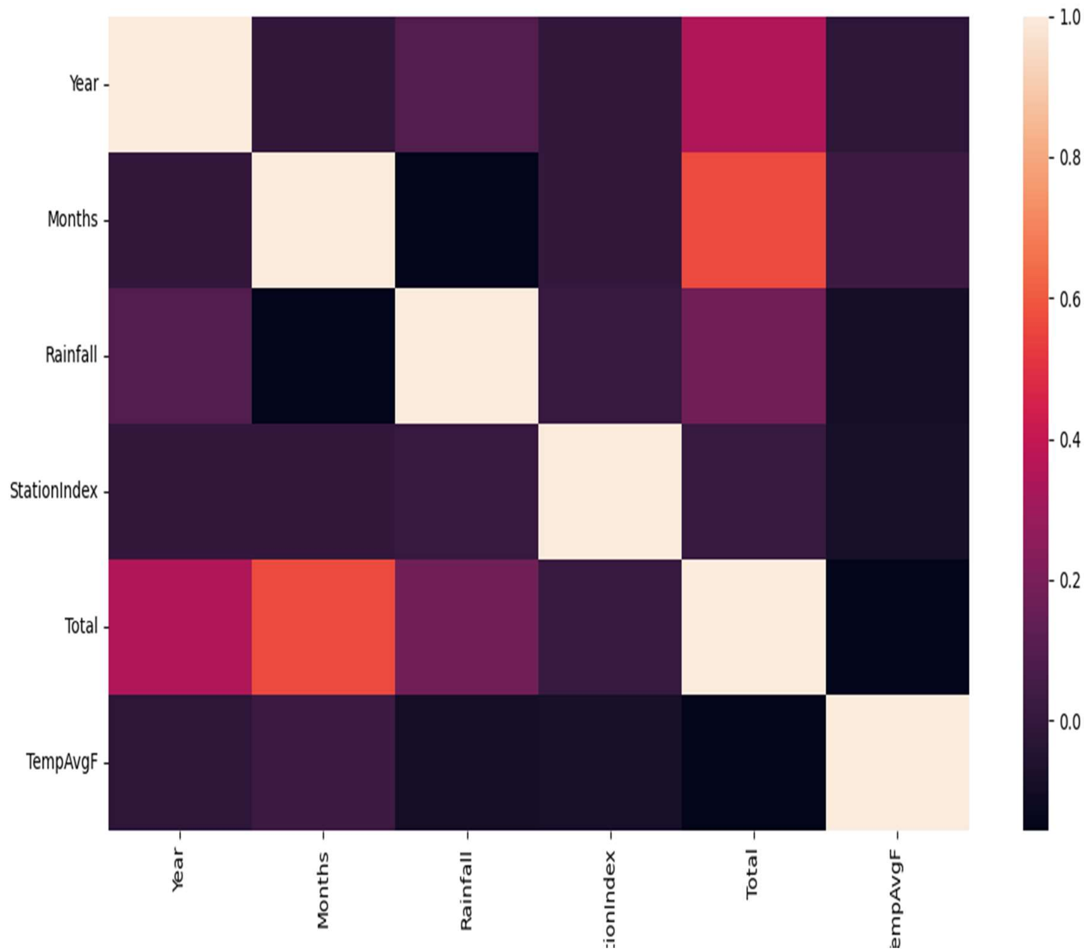


Figure 17

Depicting the correlation of the target output

Table 3 shows the statistical analysis of the rainfall. This factual examination shows logical intrusion that depicts, gathers, and investigates the information of the rainfall as well as the patterns to recognize any significant data change in the rainfall. The table depicts the mean, median, and standard deviation of the rainfall which can be utilized as unmistakable measurable strategies for the ML model to change crude perceptions into what can be shared and comprehended. Figure 18 shows the year vs the rainfall. This significes the amount of rainfall individually in each year for all the stations. This significes the total amount of rainfall in each year for all the stations. In basic words, this result measures the examination that will help in investigating the rainfall to determine how much rainfall will fall that particular year or based on cumulative years for all station indexes. It helps make significant inferences from the years and total rainfall information. Therefore, this inferential measurable techniques can be utilized to reason from each station and its rainfall for the year based on the information depicted in the entire space. Figure 19 shows the months vs the rainfall. This significes the amount of rainfall individually in each month for all the stations. This significes the total amount of rainfall in each month for all the stations. This tends to characterize the study by gathering and breaking down information to recognize patterns gotten from depicting the rainfall against the month. With this, individuals can plan ahead in expectation of what the rainfall will be for that particular month. The ends are drawn utilizing measurable examination working with navigation and assisting with making awareness of future forecasts based on the depicted patterns. Figure 20 shows the stations vs the rainfall. This significes the amount of rainfall individually in each stations. This significes the total amount of rainfall for each stations The image shows there was more rainfall in station 2. This factual examination includes working with numbers for each weather station, organization and different establishments can consider the rainfall to be expected in each region and this is a significant pattern to be considered.

Table 3.

Statistical report of the dataset

	Year	Months	Rainfall	Station Index	Total
count	432	432	432	432	432
Mean	2016	6.5	49.654606	2.5	356.125394
std	2.584983	3.456055	63.397233	1.11933	196.817108
Min	2012	1	0	1	36.2
25%	2014	3.750000	0.0	1.75	210.8
50%	2016	6	22	2	341.4
75%	2018	9.25	83.85	3.25	460.4
max	2020	12	400	4	1235.2

Figure 18

The year vs the individual rainfall

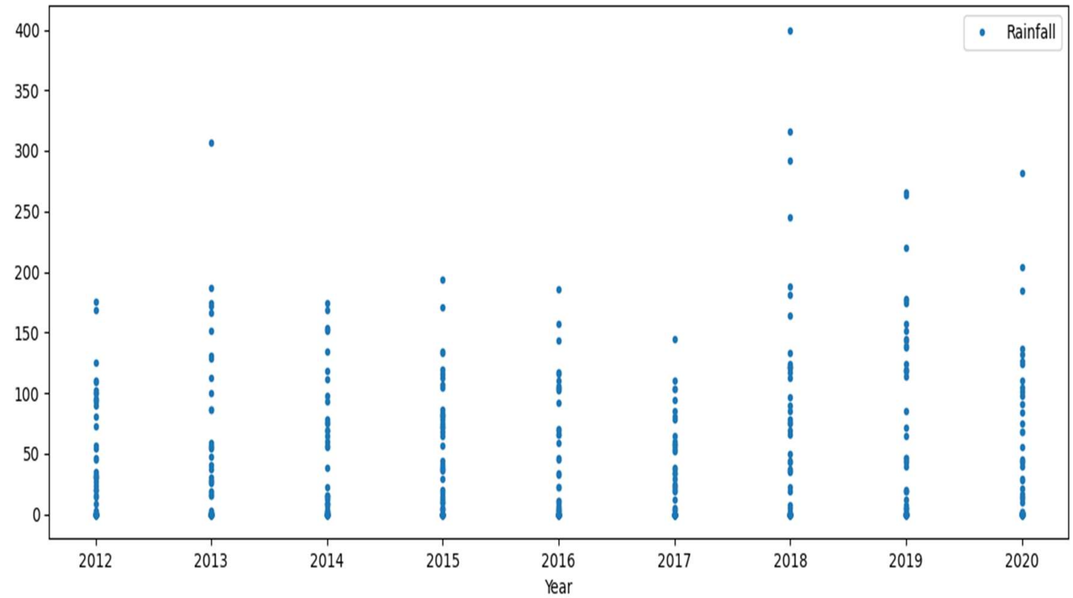


Figure 19

The months vs the individual rainfall

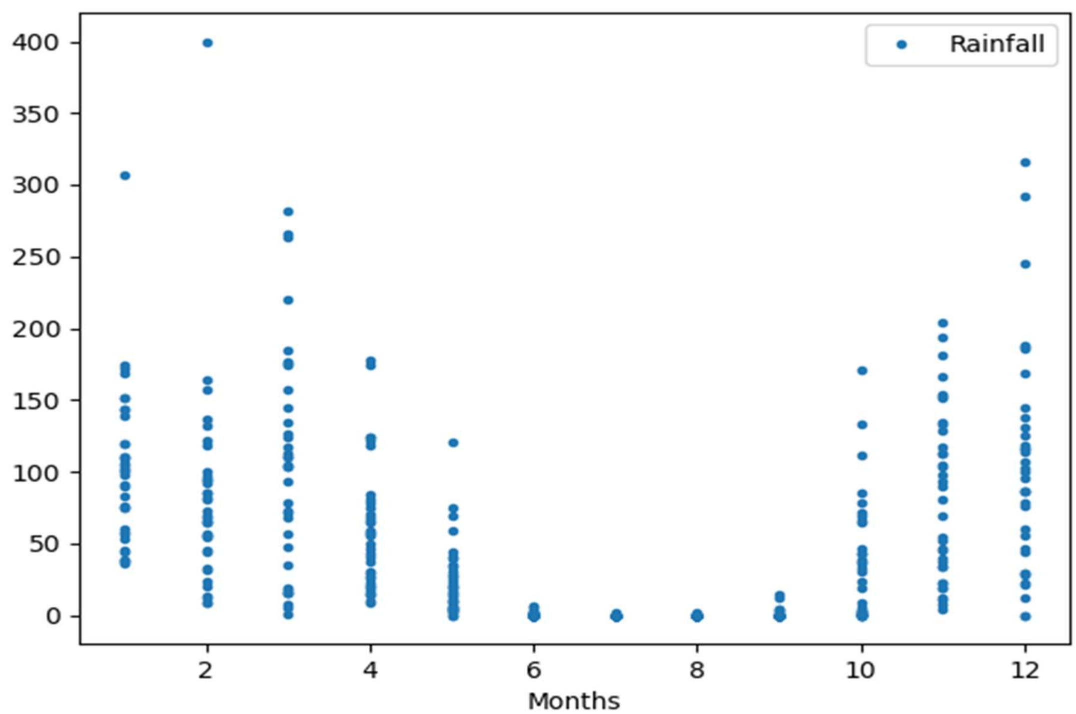
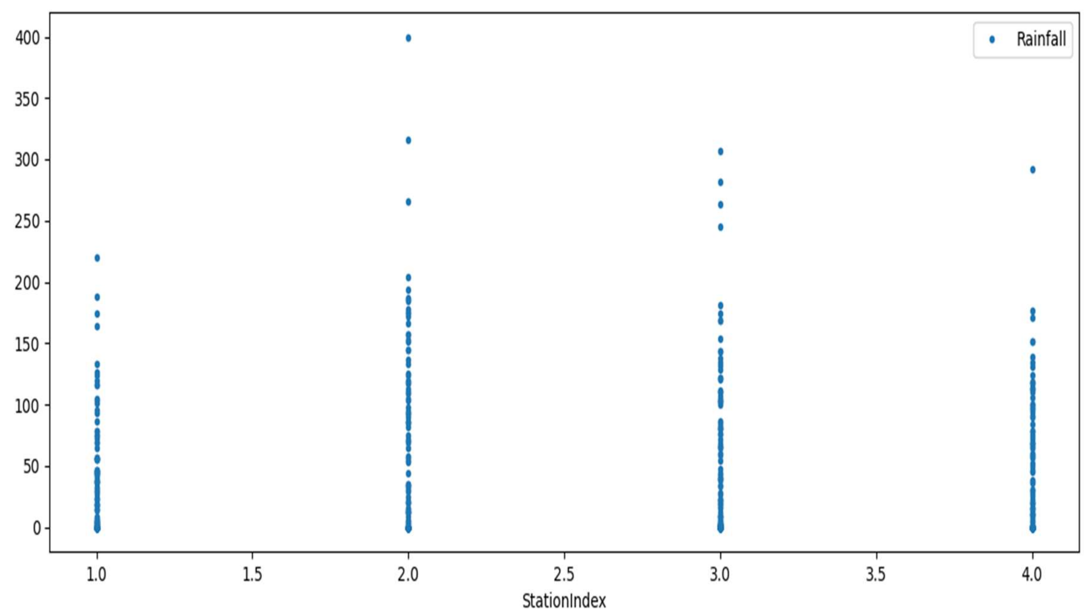


Figure 20

The stations vs the individual rainfall



The Performance of the Neural Networks for Analysis

Table 4 shows the result of the model after the training and testing split. This result will make us understand the proficiency of the utilized NN models. After the data was gathered, data normalization was applied for each data too be between 1 and 0. The Cascade-forward backprop and Elman backprop and Feed-forward backprop and Layer recurrent approach was applied after the data normalization. The result experiment was carried out 10 times for each model in order to determine how proficient the result is. The data was also splited into 80% to 20% training and testing. The table 4 shows the results after the split. The table shows the application of different NN models. The neural network is sufficient when it comes to analysis. The best model from the result is picked and it shows that Erbil Rsq 0.021924949 and RMSE 0.147071117, Sulamaniah Rsq 0.25258569 and RMSE 0.02548661, Duhok Rsq 0.49551079 and RMSE 0.148538344, Halabja Rsq 0.049418163, and RMSE 0.167684707 after testing. Table 5, 6, 7, 8 shows the predicted result and the actual result for each models of NNs to Erbil. This give a clear understandinmg on how each model was able to predict the rainfall agaist the the provided data. The result is sufficient were it shows that the predicted value are very similar to thwe actual value.

This will be a good approach for the weather station in Iraq were predicting of rainfall will be of great significance at the point of utilizing this approach. Figure 21, 22, 23, 24 shows the visualization of the predicted rainfall result of each NN models for Erbil stration. The stations shows are Erbil, Sulamaniah, Duhok and Halabja respectively. It shows the predicted result of two years for each station. This is sufficient enough when it comes to statistical analysis. The resulting graph shows the rain against the month. The months total twenty four because it is of two years. In the figure 25 shows Erbil station actual vs the predicted result for (cascade-forward backprop , elman backprop, feed-forward backprop, and layer recurrent) models.

Table 4.

Experimental result for each stations and model after split using the NN

Stations	models of neural network	Training periods		Testing periods	
		R sq (max)	RMES(min)	R sq (max)	RMES(min)
Erbil	cascade-forward backprop	0.6233793	0.1545735	0.0285908	0.1496602
	elman backprop	0.6588813	0.1350089	0.0873398	0.1739043
	feed-forward backprop	0.6397080	0.1451501	0.0689060	0.1495228
	layer recurrent	0.6254990	0.1526363	0.0219249	0.1470711
Sulamaniah	cascade-forward backprop	0.6365879	0.1636275	0.2082960	0.1645815
	elman backprop	0.7168301	0.1603569	0.2337896	0.1367277
	feed-forward backprop	0.7079764	0.1423053	0.2525856	0.0254866
	layer recurrent	0.6410395	0.1632020	0.2324822	0.1232401
Duhok	cascade-forward backprop	0.5965440	0.1699731	0.5780664	0.1542442
	elman backprop	0.7818381	0.1657365	0.5052257	0.1602129
	feed-forward backprop	0.6162230	0.1533035	0.4955107	0.1485383
	layer recurrent	0.5522402	0.1687917	0.5742940	0.1532137
Halabja	cascade-forward backprop	0.5419531	0.1392161	0.0268224	0.2054100
	elman backprop	0.6795395	0.1356162	0.0368582	0.2031419
	feed-forward backprop	0.6573702	0.1267927	0.0494181	0.1676847
	layer recurrent	0.50286674	0.1249531	0.1530782	0.2040373

Table 5.

Cascade-forward backprop model for Erbil

Actual rainfall	Predicted rainfall
0.29775	0.000645
0.1605	0.080816
0.55025	0.00242
0.30875	0.019306
0.0135	7.68E-05
0.00175	0.01308
0	0.074751
0	0.083933
0	0.070977
0.04825	0.025621
0.00975	0.00031
0.1145	0.017386
0.2535	0.009082
0.13925	0.206356
0.31625	0.518851
0.108	0.294168
0.03575	0.233158
0.00375	0.204696
0.00025	0.001377
0	0.001477
0	0.045848
0.0005	0.000346
0.1135	0.083437
0.07225	0.09005

Figure 21

Erbil station actual vs the predicted result

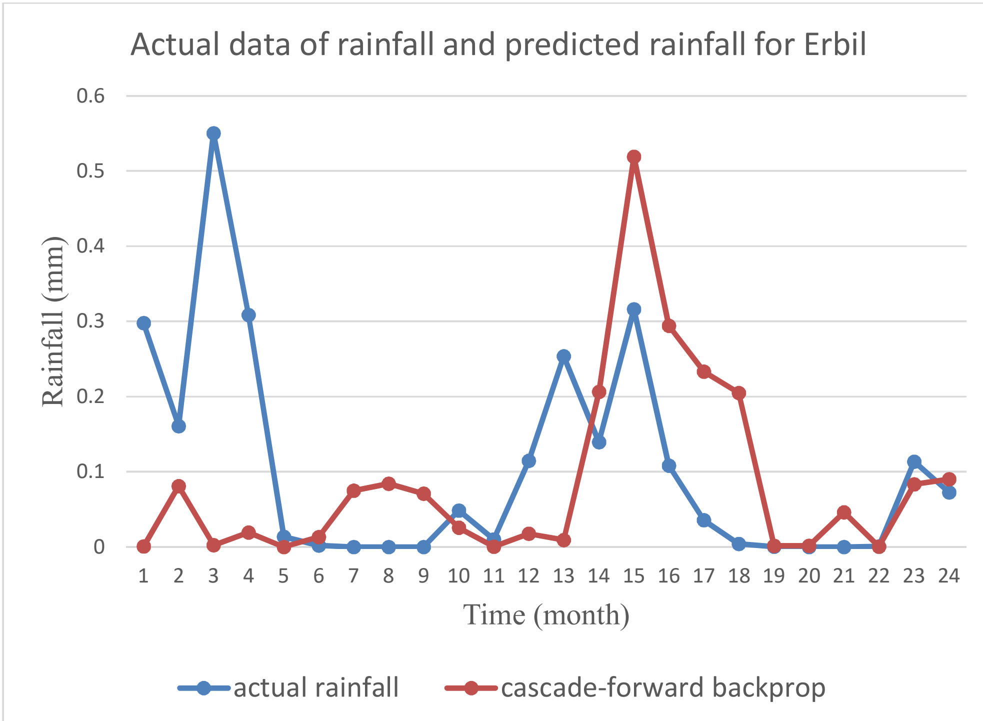


Table 6.

Elman backprop model for Erbil

Actual rainfall	Predicted rainfall
0.29775	0.146841
0.1605	0.063132
0.55025	0.057624
0.30875	0.060216
0.0135	0.036559
0.00175	0.059139
0	0.076611
0	0.13178
0	0.092855
0.04825	0.12235
0.00975	0.108915
0.1145	0.241534
0.2535	0.141833
0.13925	0.233826
0.31625	0.360045
0.108	0.336283
0.03575	0.306063
0.00375	0.343254
0.00025	0.125499
0	0.097602
0	0.086577
0.0005	0.042945
0.1135	0.066611
0.07225	0.058385

Figure 22

Erbil station actual vs the predicted result

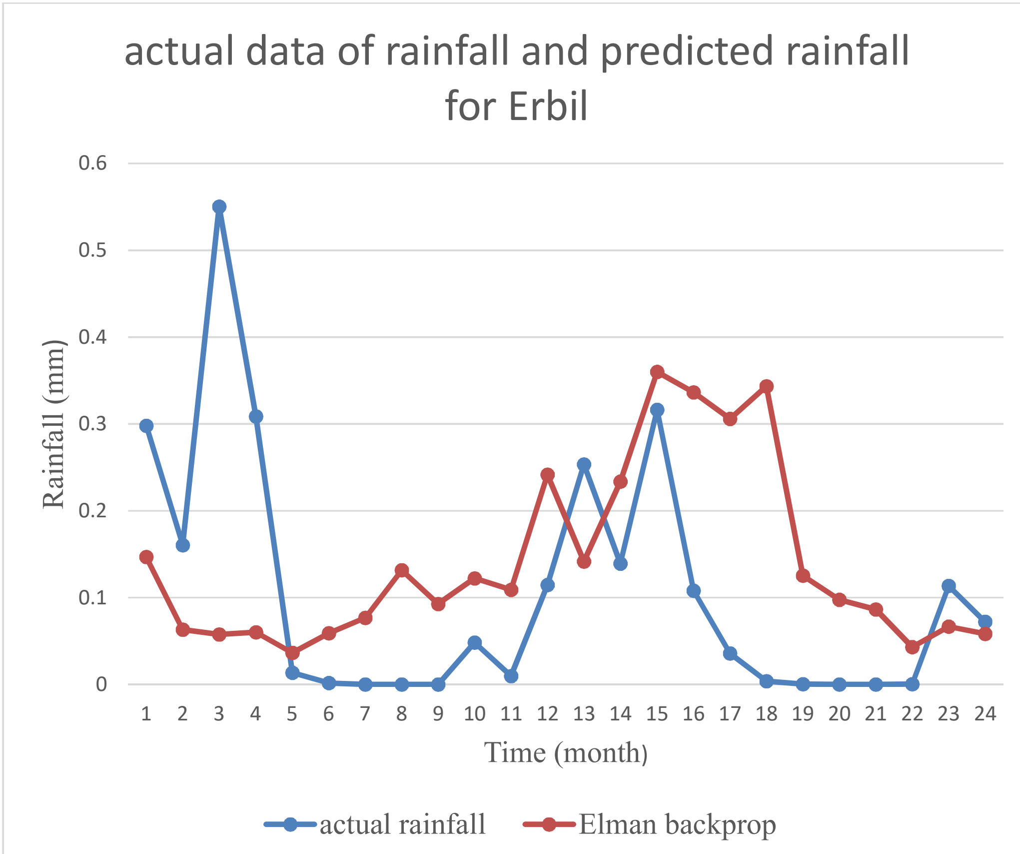


Table 7.

Feed-forward backprop model for Erbil

Actual rainfall	Predicted rainfall
0.29775	0.173613167
0.1605	0.052866095
0.55025	0.138217233
0.30875	0.114526165
0.0135	0.130154235
0.00175	0.118024045
0	7.20E-02
0	0.133834855
0	0.118757796
0.04825	0.110248918
0.00975	0.132610326
0.1145	0.183014163
0.2535	0.199307531
0.13925	0.201911589
0.31625	0.239647555
0.108	0.238303489
0.03575	0.201582852
0.00375	0.192922881
0.00025	0.130039558
0	0.097395899
0	0.083463128
0.0005	0.153645304
0.1135	0.082577122
0.07225	0.061328401

Figure 23

Erbil station actual vs the predicted result

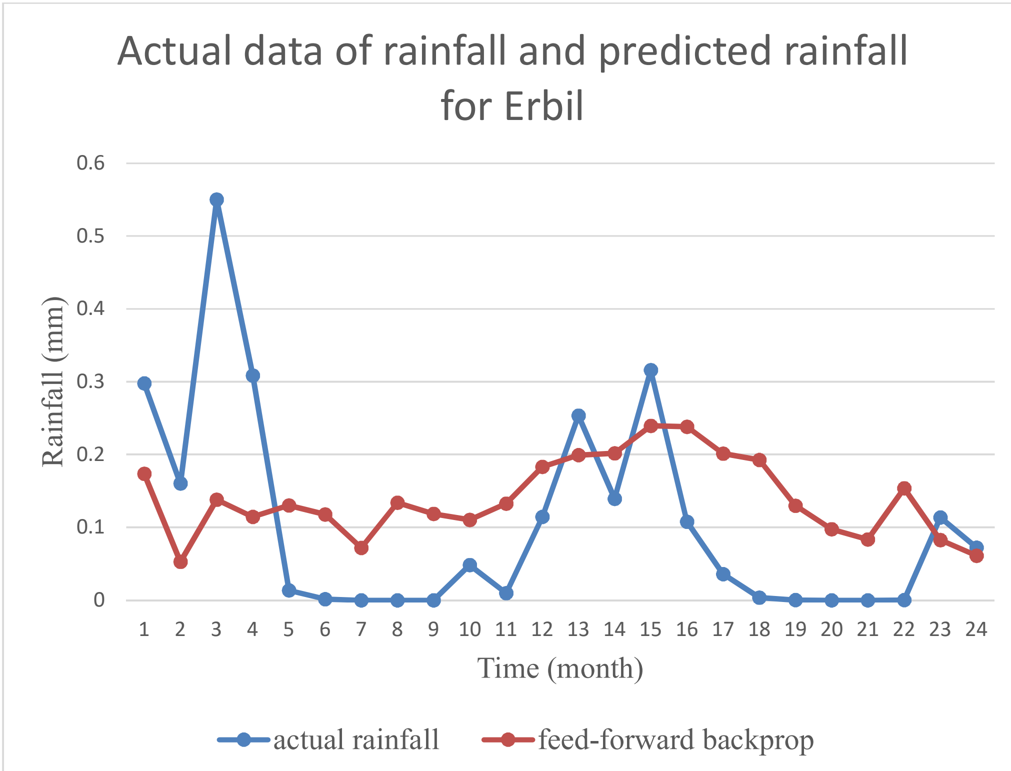


Table 8.

Layer recurrent model for Erbil

Actual rainfall	Predicted rainfall
0.29775	0.173613167
0.1605	0.052866095
0.55025	0.138217233
0.30875	0.114526165
0.0135	0.130154235
0.00175	0.118024045
0	7.20E-02
0	0.133834855
0	0.118757796
0.04825	0.110248918
0.00975	0.132610326
0.1145	0.183014163
0.2535	0.199307531
0.13925	0.201911589
0.31625	0.239647555
0.108	0.238303489
0.03575	0.201582852
0.00375	0.192922881
0.00025	0.130039558
0	0.097395899
0	0.083463128
0.0005	0.153645304
0.1135	0.082577122
0.07225	0.061328401

Figure 24

Erbil station actual vs the predicted result

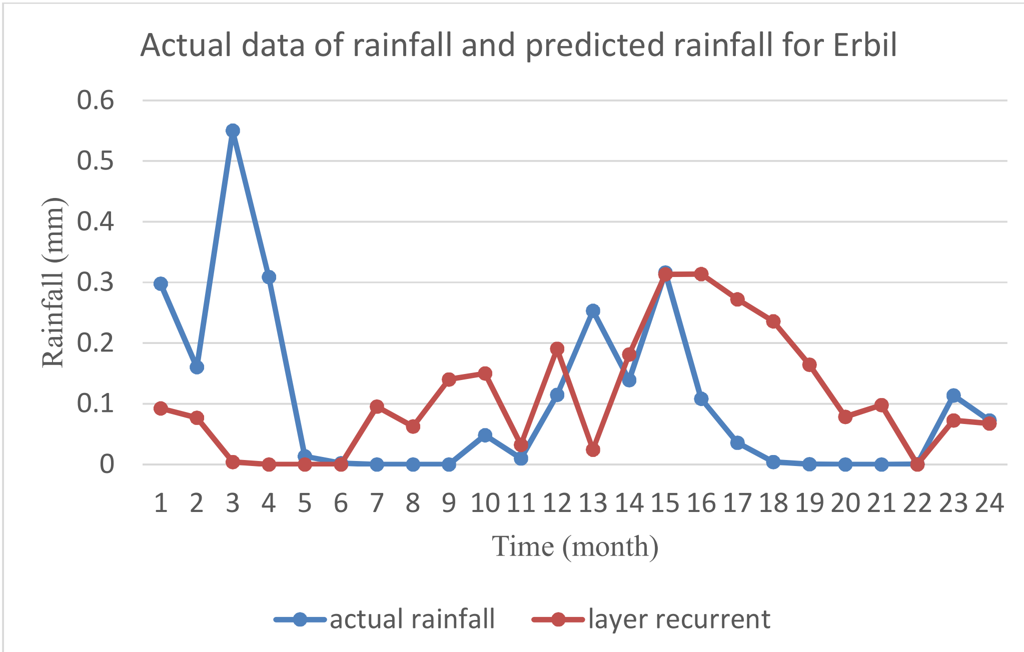
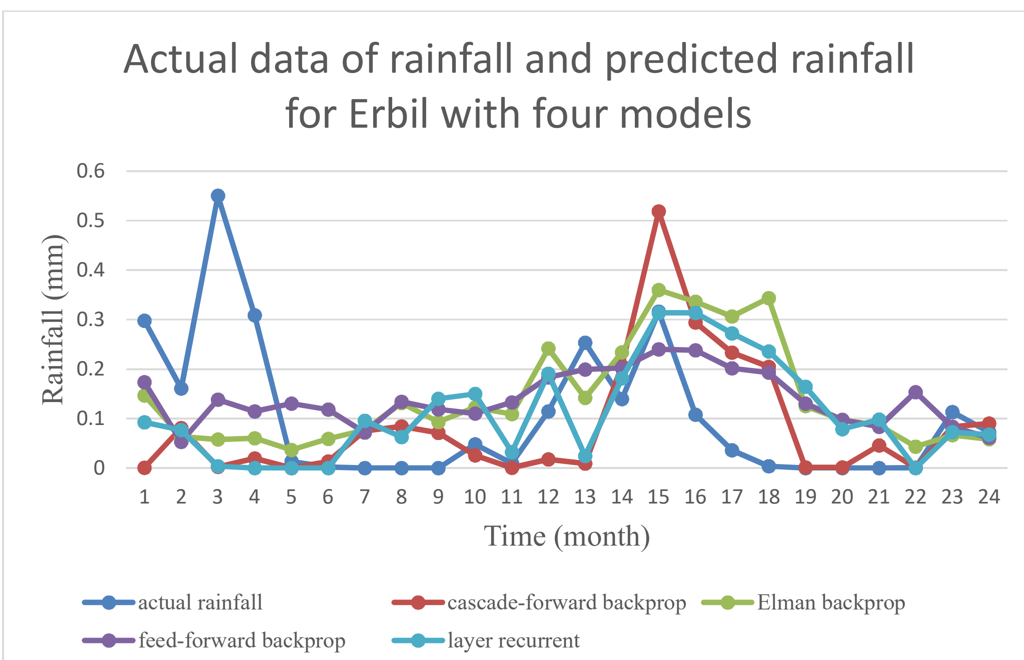


Figure 25

Erbil station actual vs the predicted result for four models



The Performance of the the KNN, SVM, DT and LR

Most times confusion matrices are used to know how accurate the models are. Since the thesis is considering applying a bit of a statistical approach, the work will be looking toward residual errors and how accurate the models are based on the given parameters in the method section. In the KKN, the certifiable class is set apart with the best value of K. After performing several experiments to consider the best value of K, Figure 26, shows the residual error of the KNN. This works to see how valuable the proposed model is. Residuals in ML and statistics are the distinctions among noticed and anticipated results of information. They are asymptomatic measures utilized while evaluating the nature of a model. They derived the name errors because of this. Considering this several cross-validations were also applied to achieve how the model can be trained efficiently. Table 9 shows that Erbil has Rsq 0.77 and RMSE 0.13, Sulamaniah Rsq 0.74 and RMSE 0.12, Duhok Rsq 0.75 and RMSE 0.17, Halabja Rsq 0.72 and RMSE 0.12 using the KNN for analysis. The result is sufficient when it comes to model analysis. The parameters are used to justify how proficient the model is. The result was splited into 80% training and 20% testing after normalization as discussed in the previous resulting section. Table 10 demonstrates the precision of the analysis. It shows the actual rainfall in mm vs the predicted result to show the validity of the proposed model. A while later, we needed to derive the best worth of K for us to upgrade the precision and productivity of our module, then the above-discussed table was deduced. It can be reasoned that after the planning the best worth of considering a K value can yield a better result. Figure 27 shows the visualization of the actual rainfall and the predicted rainfall as explained in Table 10.

Figure 26

The residual error of the KNN

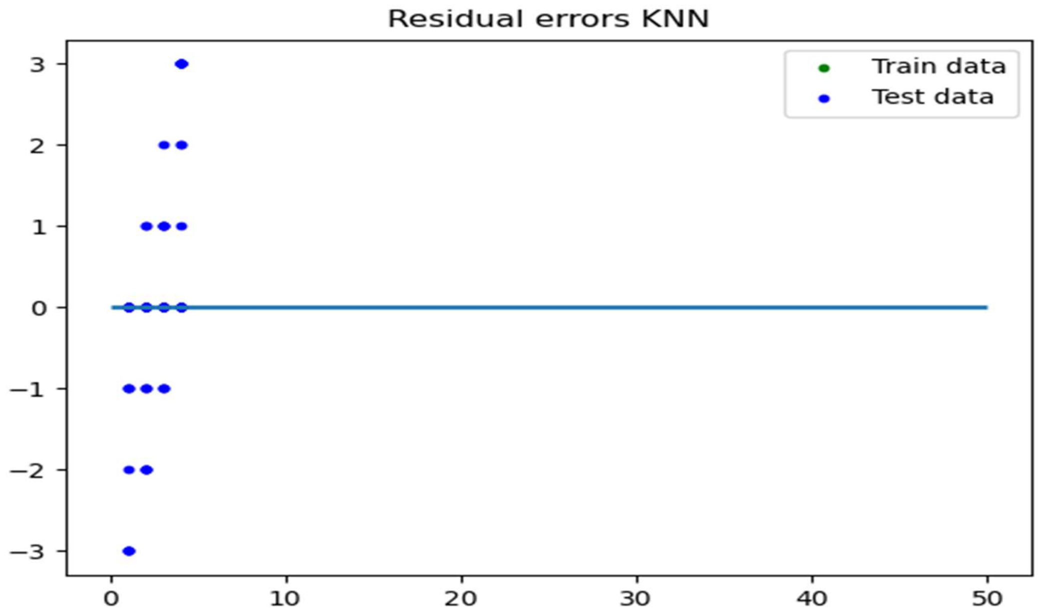


Table 9.

Experimental result for each stations and model after split using the KNN

Stations	Model	Training period		Testing period	
		R square	RMSE	R square	RMSE
Erbil	KNN	0.76	0.12	0.77	0.13
Sulamaniah		0.75	0.14	0.74	0.12
Duhok		0.77	0.13	0.75	0.17
Halabja		0.78	0.15	0.72	0.12

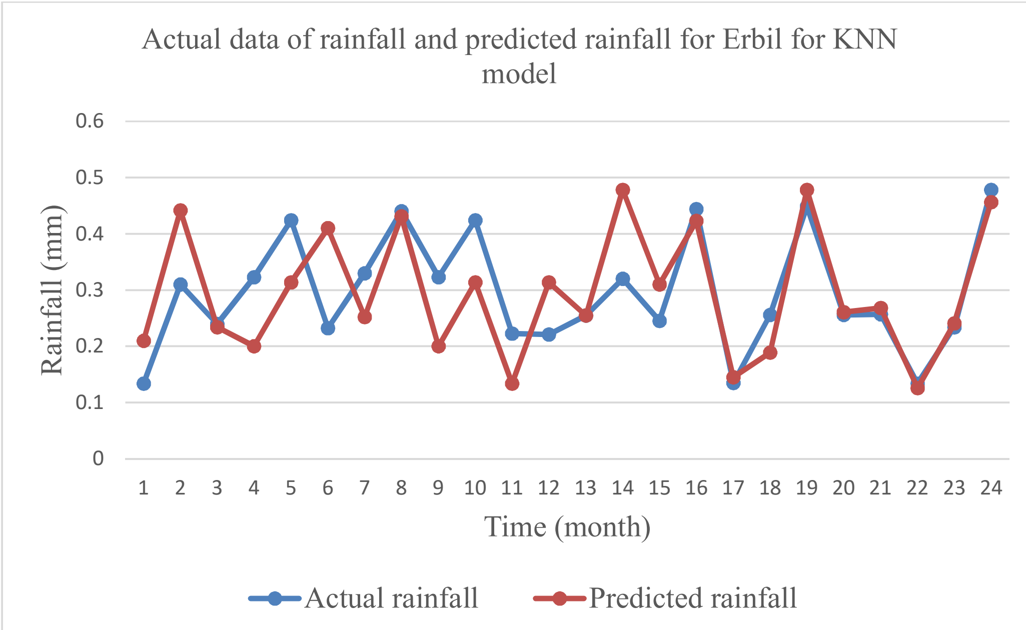
Table 10.

KNN model for Erbil

Actual rainfall	Predicted rainfall
0.1340	0.2100
0.3100	0.4420
0.2400	0.2340
0.3230	0.2
0.4240	0.3140
0.23240	0.4100
0.3300	0.2520
0.4400	0.4310
0.3230	0.2000
0.4240	0.3140
0.2230	0.1340
0.2214	0.3140
0.2550	0.2550
0.3200	0.4780
0.2450	0.3100
0.4440	0.4230
0.1350	0.1450
0.2560	0.1890
0.4490	0.4780
0.2560	0.2610
0.2570	0.2680
0.1340	0.1260
0.2340	0.2410
0.4780	0.4560

Figure 27

Erbil station actual vs the predicted result for KNN model



Linear regression is one of the main models in statistics because of its mode of prediction linearly. Table 11 shows that Erbil has Rsq 0.73 and RMSE 0.14, Sulamaniah Rsq 0.75 and RMSE 0.17, Duhok Rsq 0.78 and RMSE 0.18, Halabja Rsq 0.71 and RMSE 0.19 using the LR for analysis. The result is sufficient when it comes to model analysis. The parameters are used to justify how proficient the model is. The result was splited into 80% training and 20% testing after normalization as discussed in the previous resulting section. While Table 12 shows the consequence of the actual and predicted output. This table demonstrates the exactness of the finding of the rainfall for a particular time. We can conclude after the planning that the LR is fair in execution. Subsequently, this, shows in after improving its efficiency is still fair. Figure 28 shows the visualization of the actual rainfall and the predicted rainfall as explained in Table 12. Figure 29 shows the residual error of the LR. Analyzing the picture carefully shows the error visually noticing the mark from the test and the predicted values. The asymptomatic measure shows the nature of a model's practicality.

Table 11.

Experimental result for each stations and model after split using the LR

Stations	Model	Training period		Testing period	
		R square	RMSE	R square	RMSE
Erbil	LR	0.72	0.16	0.73	0.14
Sulamaniah		0.74	0.18	0.75	0.17
Duhok		0.72	0.13	0.78	0.18
Halabja		0.73	0.15	0.71	0.19

Table 12.
LR model for Erbil

Actual rainfall	Predicted rainfall
0.3010	0.2400
0.2210	0.2400
0.2130	0.2000
0.1390	0.4230
0.4310	0.4210
0.3210	0.2630
0.2370	0.2410
0.4370	0.4830
0.3280	0.3620
0.3710	0.3210
0.3210	0.3700
0.1230	0.4300
0.1760	0.2130
0.4740	0.3710
0.3140	0.2930
0.4370	0.3710
0.3270	0.4210
0.3680	0.3430
0.2410	0.3760
0.2140	0.1680
0.3420	0.2850
0.4310	0.3420
0.3120	0.3420
0.3620	0.4410

Figure 28

Erbil station actual vs the predicted result

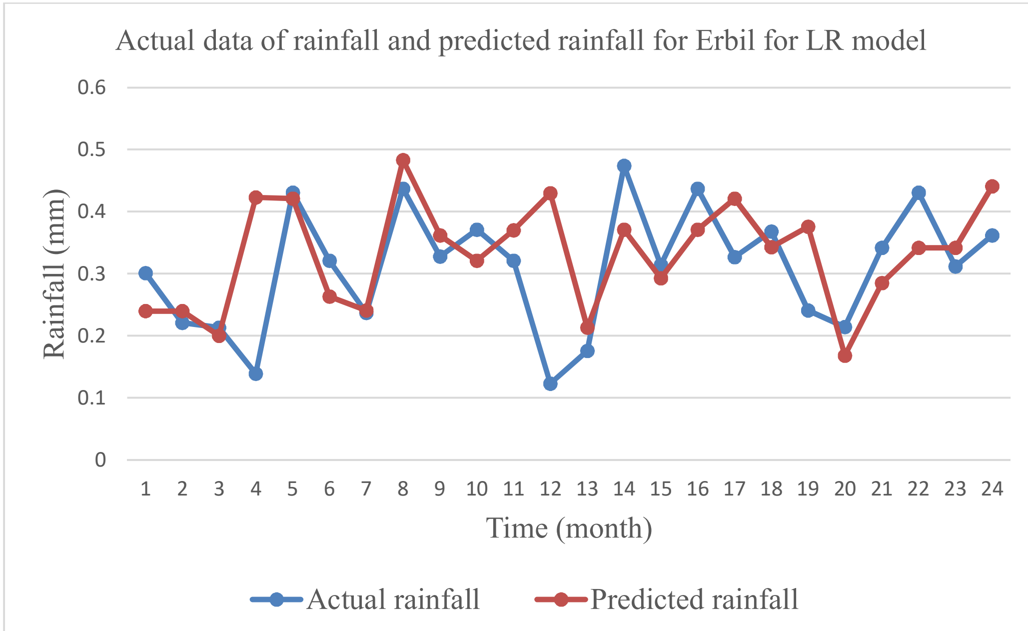
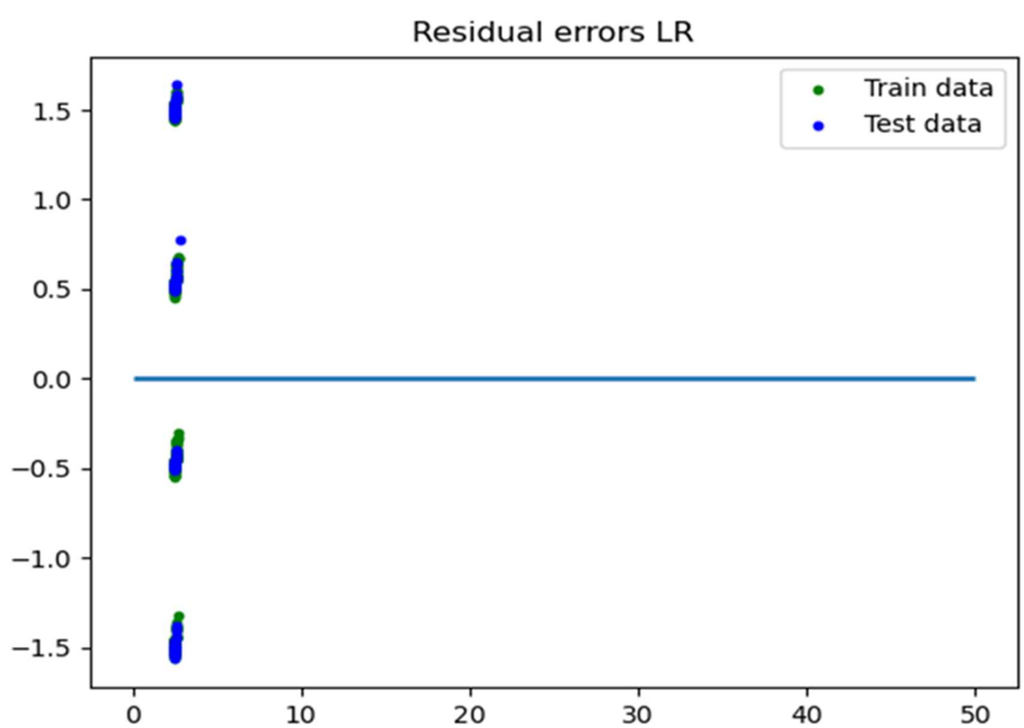


Figure 29

The residual error of the LR



The analysis below shows the sturdiness this model has for analysis. In the wake of executing a similar technique as above, we could see the SVM model performance was not fair enough. Figure 30 shows the residual error of the SVM. Analysing the picture carefully shows the error visually noticing the mark from the test and the predicted values. The asymptomatic measure shows the nature of a model's practicality, with this it can be noticed that there is an adjustment to the consistency of the model. Table 13 shows that Erbil has Rsq 0.73 and RMSE 0.15, Sulamaniah Rsq 0.75 and RMSE 0.12, Duhok Rsq 0.73 and RMSE 0.16, Halabja Rsq 0.72 and RMSE 0.18 using the SVM for analysis. The result is sufficient when it comes to model analysis. The parameters are used to justify how proficient the model is. The result was splited into 80% training and 20% testing after normalization as discussed in the previous resulting section. Contrasting our outcome and the previously discussed model as displayed in Table 14 is the equivalent even in the wake of making do for the LR. We can find in the table that it had a shallow precision without making do. The table shows that in the wake of making do normalize and upgrading some utilized parameters the results were linearly separable even after improving the Gamma and C Parameters. This shows the less superior consequence of the SVM model, we can likewise see after adjustment a better output is far from it than anyone might have expected. Figure 31 shows the visualization of the actual rainfall and the predicted rainfall as explained in Table 14.

Figure 30

The residual error of the SVM

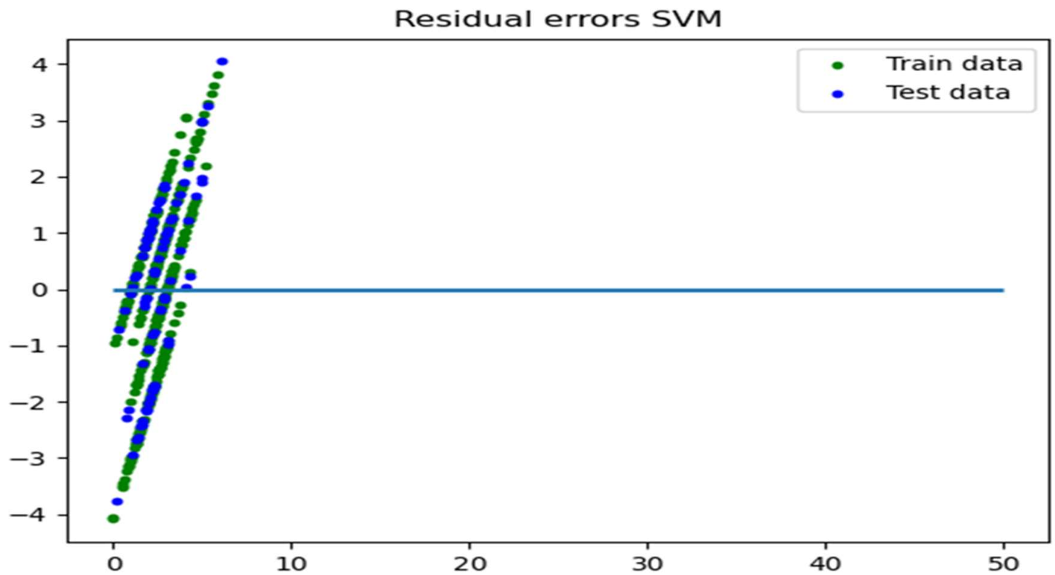


Table 13.

Experimental result for each stations and model after split using the SVM

Stations	Model	Training period		Testing period	
		R square	RMSE	R square	RMSE
Erbil	SVM	0.72	0.12	0.73	0.15
Sulamaniah		0.71	0.14	0.75	0.12
Duhok		0.76	0.17	0.73	0.16
Halabja		0.73	0.15	0.72	0.18

Table 14.

SVM model for Erbil

Actual rainfall	Predicted rainfall
0.3210	0.2680
0.2520	0.2430
0.2300	0.2510
0.1290	0.1380
0.2310	0.2420
0.3320	0.3630
0.3670	0.2410
0.3610	0.4460
0.4360	0.4640
0.3450	0.2780
0.2430	0.3140
0.1510	0.1630
0.2540	0.2710
0.4210	0.3570
0.3250	0.2780
0.4150	0.4230
0.4220	0.3810
0.4390	0.4770
0.2240	0.2310
0.2220	0.1850
0.3310	0.2550
0.3430	0.3850
0.4140	0.4320
0.2550	0.3420

Figure 31

Erbil station actual vs the predicted result

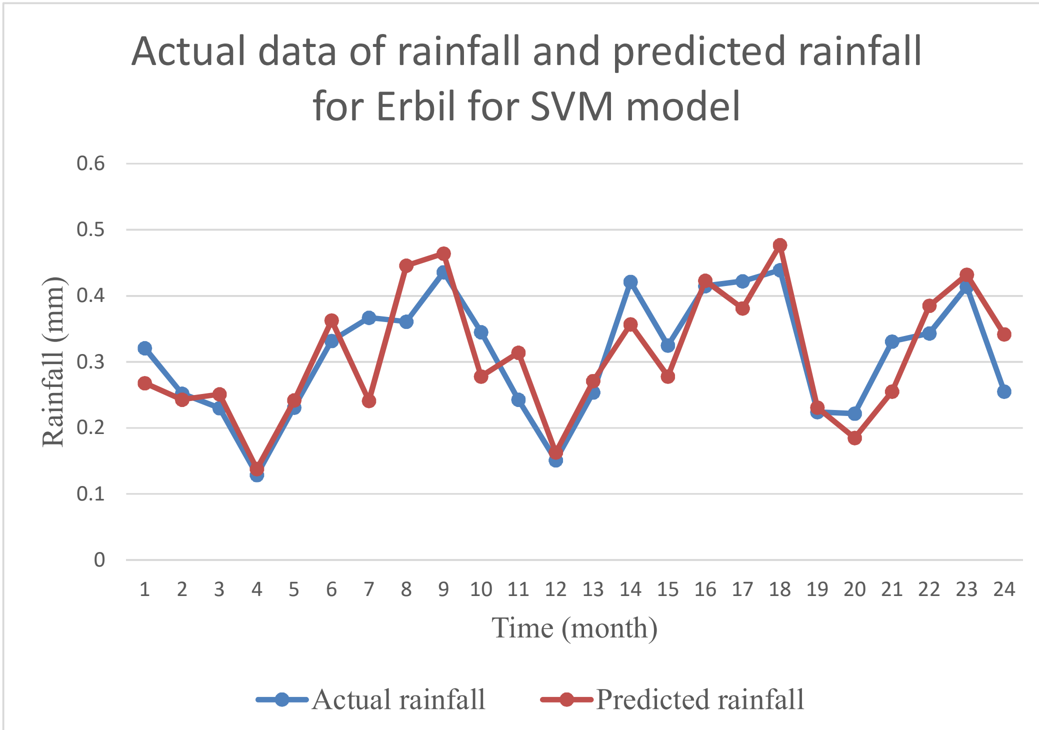


Table 15 shows that Erbil has Rsq 0.75 and RMSE 0.13, Sulamaniah Rsq 0.71 and RMSE 0.16, Duhok Rsq 0.74 and RMSE 0.14, Halabja Rsq 0.76 and RMSE 0.23 using the LR for analysis. The result is sufficient when it comes to model analysis. The parameters are used to justify how proficient the model is. The result was splited into 80% training and 20% testing after normalization as discussed in the previous resulting section. Table 16 and Figure 32 depicts the DT as the best ordering technique. It provides the most elevated exactness against other experimented model precision scores individually. The image shows the residual error while the table shows the actual against the predicted value. The outcome referenced was gotten after the improvement of the models after standardization. We additionally contrast our outcomes as it was done to the other following ML model as referred to as the DT as we display the resulting image and table. The figure shows the residual error outline of the consistency of our model. Albeit the model demonstrated to be viable as that of the past models. The DT demonstrated to be compelling in the wake of hitting prediction. Besides, it is still better contrasted with the other learning model which we will talk about in the following segment. The result shows that LR at the least performs as other model outperforms it. Figure 33 shows the visualization of the actual rainfall and the predicted rainfall as explained in Table 16. From the results it can be deduced that LR wont be proficient when it comes to rainfall prediction.

Figure 32

The residual error of the DT

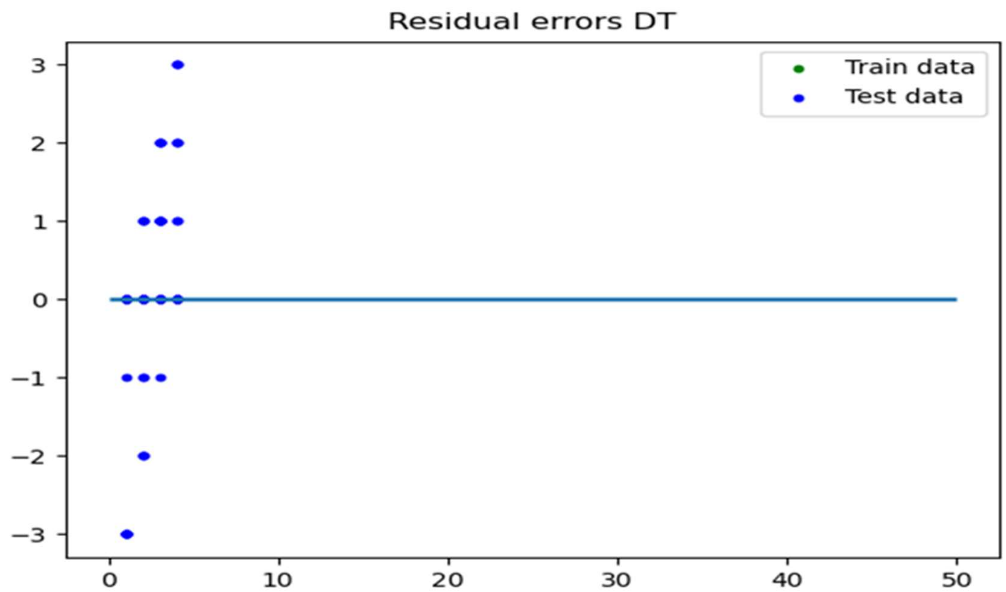


Table 15.

Experimental result for each stations and model after split using the DT

Stations	Model	Training period		Testing period	
		R square	RMSE	R square	RMSE
Erbil	DT	0.73	0.14	0.75	0.13
Sulamaniah		0.71	0.16	0.71	0.16
Duhok		0.73	0.13	0.74	0.14
Halabja		0.77	0.17	0.76	0.23

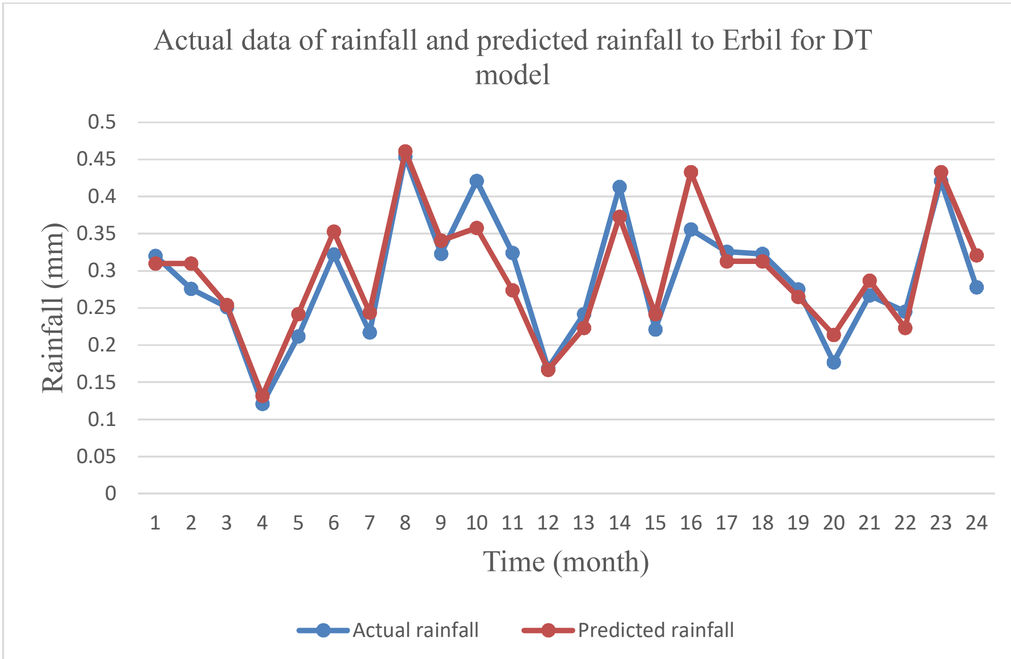
Table 16.

DT model for Erbil

Actual rainfall	Predicted rainfall
0.3200	0.3100
0.2760	0.3100
0.2510	0.2540
0.1210	0.1320
0.2120	0.2420
0.3220	0.3530
0.2170	0.2440
0.4540	0.4610
0.3230	0.3410
0.4210	0.3580
0.3240	0.2740
0.1690	0.1670
0.2420	0.2230
0.4130	0.3730
0.2210	0.2420
0.3560	0.4330
0.3260	0.3130
0.3230	0.3130
0.2750	0.2650
0.1770	0.2140
0.2670	0.2870
0.2450	0.2230
0.4210	0.4330
0.2780	0.3210

Figure 33.

Erbil station actual vs the predicted result



Experimental Result Discussion and Comparison

To deduce the best model in prediction we have touse certain parameters as discussed in the the methodology section. This section is set to justify each accuracy of the utilised models. To guarantee consistency of each used model, they were all executed in this experiment using a train/test parting of 80% to 20% extent. We have utilized four models KNN, SVM, DT, NN and LR were applied to a relatively immense informational indexes a brief time later was part into two segments the train and test sets to show the presence of precision on the rainfall data for analysis and forecast. Accuracy relationship for the AI model of the rainfall analysis after average ensemble method is utilized is displayed in Table 17. From the models, performing effectively, the best model for disease analysis is the ML model known as the DT getting a precision of 97%. It is followed by KNN 96%, SVM 93%, NN 90% and LR 89%. Also, Table 18 shows the parameter comparison of the proposed ML models. It shows the DT outperforms other models. It also shows comparison with other research and the proposed model outperforms other model. This gives furthure validation to the proposed research model.

The used ML approach in this study provided a promising presentation as depicted in this area. The made request structure for SVM is shown to be good for gathering rainfall analysis capably. Similarly, due to the differentiation in preprocessing the train/test split the outcome was viable for every used approach. Additionally, when we experimented with the model ourselves with default values, DT at this point is better separable. In any case, the residual errors were efficient in analyzing each model. The least performed model is the LR. This checks that LR can get on rapidly and definitively in the arrangement stage, in any case, it will in general be weak in summarizing and seeing concealed data. Consequently, the result derived from this test shows that experts can use all ML proficiently in anticipating and analyzing rainfall in Iraq effectively at whatever point used. Note that the distinction in precision was somewhat similar for both LR and SVM. This is a reminder to all researchers and practitioners to use the proposed strategy.

Table 17.

ML models experimental result (%)

	KNN	SVM	DT	LR	NN
Initial accuracy	0.94	0.92	0.96	0.85	0.86
Accuracy	0.96	0.93	0.97	0.89	0.90

Table 18

Experimental result comparison with other models (%)

Authors	Recall	Precision	F1 score	Accuracy	Root square	RMSE
Oswal, (2019)	-	85%	-	0.88		-
Vijayan et al. (2020)	-	80%	-	0.82		-
Liyew, and Melese, (2021)	-	-	-	0.79		-
This thesis (KNN)	0.96	0.95	0.96	0.96	0.77	0.13
This thesis (SVM)	0.92	0.89	0.91	0.93	0.76	0.18
This thesis (DT)	0.98	0.97	0.97	0.97	0.78	0.12
This thesis (NN)	0.89	0.88	0.89	0.90	0.57	0.15
This thesis (LR)	0.88	0.87	0.88	0.89	0.75	0.23

CHAPTER VI

Conclusion and Recommendation

Machine learning has indicated impressive improvement for the prediction of rainfall. Rainfall proffers an exceptional setting for countries and climatologists by also contemplating the atmospheric condition and their response towards finding a way to make it beneficial. This thesis presents an examination in a smart way to forecast and analyze rainfall data. Regardless, there are several issues with the ideal forecast of precipitation paying little mind to trend-setting innovations. It is expected to join the climate and demographic string of information to work on the improvement of the farsighted modules. KNN, DT, SVM, NN and LR were selected for the ML technique for this project and the statistical approach for its analysis. This technique was compared dependent on various parameters set during the study. Likewise, there was a comparison between the techniques to survey the accuracy of each in order to find the techniques that perform best for this project. Also, the NN was used for analysis applying the Root square and the RSME. The result here shows that Erbil Rsq 0.005237097 and RMSE 0.147071117, Sulamaniah Rsq 0.211838663 and RMSE 0.02548661, Duhok Rsq 0.273075886 and RMSE 0.148538344, Halabja Rsq 0.049418163, and RMSE 0.167684707.

In this thesis a predictive model was built to empower specialists to get up to speed in their working environment, endeavoring to give delegates a dynamically secure working environment. This module licenses nations and pioneers to screen the weather status of their environment and to follow any event of rainfall respectively. Once the module is being implemented, it channels information based on existing rainfall data and provides reliable information on predicting later rainfall. This enables nations, climatologists, and individuals to get conversant on the weather condition and plan appropriately based on this factor.

Usage of the unique dataset from Urban Cities, in the Northern Part of Iraq, was utilized in this paper. ML techniques have since been efficiently utilized in various sectors and have functioned as an important analytic thread that assists practitioners in analyzing the accessible information and designing the various expert framework. The outcomes of these classifiers were studied and analyzed utilizing distinctive performance measures such as precision, accuracy, median, F1 score, recall, root square and the RSME and standard deviation. Their approaches and primary highlights were described and discussed. Hence, the model with the best accuracy score for prediction was selected.

Insights from this preliminary outcome showed that it is in no way, shape, or form prudent to use just a single ML methodology on the datasets with fewer elements since a diminishing of the component vectors would bring about misclassification and possibly terrible execution accuracy. The utilized DT method has a better performance than the other ML utilized models, where we arrived at a performance of 97%, and the least being the LR with 89%.

Recommendation and Future Works

I suggest that this research work should be taken with utmost seriousness and executed with quick effect. Because of the availability in analyzing, creating, and actualizing this thesis, there are significant areas for additional exploration in the future that might be done, which are mentioned in this section. For information development, the framework designed is an interactive framework that educates individuals about the present condition of the weather. This methodology has incredible potential for speculations, and it must be additionally upgraded with more open informational indexes. The prediction could help the specialist to get ready for a better understanding of the weather and make provisions for later rainfall.

Dependent on the measures while preparing this research, listed below are the future study areas which I recommended:

- Different highlight algorithms for classification that will assist with deciding the least subset of highlights that will aid in precisely distinguishing the types of rainfall are recommended.
- Researchers can also include various analysis techniques to fine-tune the rainfall data.
- It is useful to use various information for precipitation to survey and lift the proficiency of the suggested calculation. These revelations in the current exploration might be a respectable beginning for further exploration.
- The combination of IoT to this research can also increase the smartness in the prediction accuracy.
- The system implemented gives careful investigation and expectation, executing it with other techniques like fuzzy logic is emphatically recommended.

References

Akiner, M. E. (2021). Long-Term Rainfall Information Forecast by Utilizing Constrained Amount of Observation through Artificial Neural Network Approach. Advances in Meteorology.

Alhamshry, A., Fenta, A. A., Yasuda, H., Shimizu, K., & Kawai, T. (2019). Prediction of summer rainfall over the source region of the Blue Nile by using teleconnections based on sea surface temperatures. *Theoretical and Applied Climatology*, 137(3), 3077-3087.

Alizadeh, M. J., Kavianpour, M. R., Kisi, O., & Nourani, V. (2017). A new approach for simulating and forecasting the rainfall-runoff process within the next two months. *Journal of hydrology*, 548, 588-597.

Alotaibi, K., Ghumman, A. R., Haider, H., Ghazaw, Y. M., & Shafiquzzaman, M. (2018). Future predictions of rainfall and temperature using GCM and ANN for arid regions: a case study for the Qassim Region, Saudi Arabia. Water, 10(9), 1260.

Bagirov, A. M., Mahmood, A., & Barton, A. (2017). Prediction of monthly rainfall in Victoria, Australia: Clusterwise linear regression approach. *Atmospheric research*, 188, 20-29.

Bahrawi, J., Alqarawy, A., Chabaani, A., Elfeki, A., & Elhag, M. (2021). Spatiotemporal analysis of the annual rainfall in the Kingdom of Saudi Arabia: predictions to 2030 with different confidence levels. *Theoretical and Applied Climatology*, 146(3), 1479 1499.

Bermúdez, M., Neal, J. C., Bates, P. D., Coxon, G., Freer, J. E., Cea, L., & Puertas, J. (2017). Quantifying local rainfall dynamics and uncertain boundary conditions into a nested regional-local flood modeling system. *Water Resources Research*, 53(4), 2770-2785.

Beusch, L., Foresti, L., Gabella, M., & Hamann, U. (2018). Satellite-based rainfall retrieval: From generalized linear models to artificial neural networks. Remote Sensing, 10(6), 939.

Blum, A. G., Zaitchik, B., Alexander, S., Wu, S., Zhang, Y., Shukla, S., ... & Block, P. (2019). A Grand Prediction: Communicating and Evaluating 2018 summertime Upper Blue Nile rainfall and streamflow forecasts in preparation for Ethiopia's new dam. Frontiers in Water, 1, 3.

Brown, J. N., Hochman, Z., Holzworth, D., & Horan, H. (2018). Seasonal climate forecasts provide more definitive and accurate crop yield predictions. *Agricultural and forest meteorology*, 260, 247-254.

Canli, E., Loigge, B., & Glade, T. (2018). Spatially distributed rainfall information and its potential for regional landslide early warning systems. *Natural hazards*, 91(1), 103 127.

Carly Dodd, (2021). The Water Cycle. *In Environment*, [Online] Available: https://www.worldatlas.com/the-water-cycle.html

Chen, X., He, G., Chen, Y., Zhang, S., Chen, J., Qian, J., & Yu, H. (2019). Notice of Retraction: Short-term and local rainfall probability prediction based on a dislocation support vector machine model using satellite and in-situ observational data. IEEE Access.

Choudhury, D., Mehrotra, R., Sharma, A., Sen Gupta, A., & Sivakumar, B. (2019). Effectiveness of CMIP5 decadal experiments for interannual rainfall prediction over Australia. *Water Resources Research*, 55(8), 7400-7418.

Cramer, S., Kampouridis, M., Freitas, A. A., & Alexandridis, A. K. (2017). An extensive evaluation of seven machine learning methods for rainfall prediction in weather derivatives. *Expert Systems with Applications*, 85, 169-181.

Davolio, S., Silvestro, F., & Gastaldo, T. (2017). Impact of rainfall assimilation on high resolution hydrometeorological forecasts over Liguria, Italy. *Journal of Hydrometeorology*, 18(10), 2659-2680.

Depiction of the water cycle (Precipitation education, (2020). The Water Cycle, [Online] Available: https://gpm.nasa.gov/education/water-cycle.

Doss-Gollin, J., Muñoz, Á. G., Mason, S. J., & Pastén, M. (2018). Heavy rainfall in Paraguay during the 2015/16 austral summer: Causes and subseasonal-to seasonal predictive skill. *Journal of Climate*, 31(17), 6669-6685.

Farajzadeh, J., & Alizadeh, F. (2018). A hybrid linear–nonlinear approach to predict the monthly rainfall over the Urmia Lake watershed using wavelet-SARIMAX LSSVM conjugated model. *Journal of Hydro informatics*, 20(1), 246-262.

Fencl, M., Dohnal, M., Rieckermann, J., & Bareš, V. (2017). Gauge-adjusted rainfall estimates from commercial microwave links. *Hydrology and Earth System Sciences,* 21(1), 617-634.

Fereidoon, M., Koch, M., & Brocca, L. (2019). Predicting rainfall and runoff through satellite soil moisture data and SWAT modelling for a poorly gauged basin in Iran. Water, 11(3), 594.

Frame, J., Kratzert, F., Klotz, D., Gauch, M., Shelev, G., Gilon, O., Qualls, L. M., Gupta, H. V., and Nearing, G. S. (2021) Deep learning rainfall-runoff predictions of extreme events. *Hydrology Earth System Science Discussion*. [preprint], https://doi.org/10.5194/hess2021-423, in review.

Ghamariadyan, M., & Imteaz, M. A. (2021). Prediction of Seasonal Rainfall with One year Lead Time Using Climate Indices: A Wavelet Neural Network Scheme. *Water Resources Management*, 1-19.

Gleixner, S., Keenlyside, N. S., Demissie, T. D., Counillon, F., Wang, Y., & Viste, E. (2017). Seasonal predictability of Kiremt rainfall in coupled general circulation models. *Environmental Research Letters*, 12(11), 114016.

Gokcekus, H., Ozsahin, D., & Mustapha, M. (2020). Simulation and evaluation of water sterilization devices. *Desalination Water Treat*, *177*, 431-436.

Gökçekuş, Hüseyin & Nourani, Vahid. (2018). Lessons from Groundwater Quantity and Quality Problems In the Güzelyurt Coastal Aquifer, North Cyprus. *Nova Science Publishers, Inc*. ISBN: 978-1-53614-010-1.

Golding, N., Hewitt, C., Zhang, P., Liu, M., Zhang, J., & Bett, P. (2019). Co development of a seasonal rainfall forecast service: Supporting flood risk management for the Yangtze River basin. *Climate Risk Management*, 23, 43 49.

Graham, A., & Mishra, E. P. (2017). Time series analysis model to forecast rainfall for Allahabad region. *Journal of Pharmacognosy and Phytochemistry*, 6(5), 1418 1421.

Hadi, M. P., Suprayogi, S., & Herumurti, S. (2020). Impact of rainfall intensity, monsoon and MJO to rainfall merging in the Indonesian maritime continent. *Journal of Earth System Science,* 129(1), 1-20.

Hancock, G. R., Verdon-Kidd, D., & Lowry, J. B. C. (2017). Soil erosion predictions from a landscape evolution model–An assessment of a post-mining landform using spatial climate change analogues. *Science of the Total Environment*,` 601, 109-121.

Htike, K. K. (2018). Predicting rainfall using neural nets. *International Journal of Computational Science and Engineering,* 17(4), 353-364.

Hussain, A. A., Bouachir, O. Al-Turjman F. and Aloqaily, M. (2020). AI Techniques for COVID-19. *In IEEE Access*, vol. 8, pp. 128776-128795, doi: 10.1109/ACCESS.2020.3007939.

Hussain, Adedoyin Ahmed & Al-Turjman, Fadi & Gemikonaklı, Eser & Kirsal Ever, Yoney. (2021)d. Design of a Navigation System for the Blind/Visually Impaired. 10.1007/978-3-030-69431-9_3.

Hussain, Adedoyin Ahmed & Al-Turjman, Fadi & Sah, Melike. (2021)b. Semantic Web and Business Intelligence in Big-Data and Cloud Computing Era. 10.1007/978 3-030 66840-2_107.

Hussain, Adedoyin Ahmed & Al-Turjman, Fadi. (2021). Artificial intelligence and blockchain: A review. *Transactions on Emerging Telecommunications Technologies*. 10.1002/ett.4268.

Hussain, Adedoyin Ahmed & Dawood, Barakat & Al-Turjman, Fadi. (2021)a. IoT and AI for COVID-19 in Scalable Smart Cities. 10.1007/978-3-030-76063-2_1.

Hussain, Adedoyin Ahmed & Dawood, Barakat & Al-Turjman, Fadi. (2021)c. Application of AI Techniques for COVID-19 in IoT and Big Data Era: A Survey. 10.1007/9783 030-60188-1_9.

Hussain, Adedoyin Ahmed & Dimililer, Kamil. (2021). Student Grade Prediction Using Machine Learning in Iot Era. 10.1007/978-3-030-69431-9_6.

Hussain, Adedoyin Ahmed and Al-Turjman, Fadi. (2020). Resource Allocation in Volunteered Cloud Computing and Battling COVID-19. 10.1201/9781003098881 2.

Johny, K., Pai, M. L., & Adarsh, S. (2020). Adaptive EEMD-ANN hybrid model for Indian summer monsoon rainfall forecasting. *Theoretical & Applied Climatology*, 141.

Kajewska-Szkudlarek, J. (2020). Clustering approach to urban rainfall time series prediction with support vector regression model. *Urban Water Journal*, 17(3), 235-246.

Karunakaran, V., Joseph, S. I., Teja, R., Suganthi, M., & Rajasekar, V. (2019). A wrapper based feature selection approach using bees algorithm for extreme rainfall prediction via weather pattern recognition through svm classifier. *International Journal of Civil Engineering and Technology (IJCIET*), 10(1).

Kassem, Youssef & Gökçekuş, Hüseyin & Aljamal, J. (2020). Surface water resource and effect of weather parameters in estimating the annual rainfall: A case study in Lebanon. *IOP Conference Series: Materials Science and Engineering*. 800. 012028. 10.1088/1757-899X/800/1/012028.

Kassem, Youssef & Gökçekuş, Hüseyin & Çamur, Hüseyin & Esenel, Engin. (2021). Statistical analysis and determination of best-fit probability distribution for monthly rainfall in Northern Cyprus. *Desalination and Water Treatment*. 215. 347-379. 10.5004/dwt.2021.26556.

Kassem, Youssef & Gökçekuş, Hüseyin. (2020). Water resources and rainfall distribution function: a case study in **.Lebanon***Desalination and Water Treatment*. 177. 306 321. 10.5004/dwt.2020.24811.

Khan, M. Z. K., Sharma, A., & Mehrotra, R. (2017). Global seasonal rainfall forecasts using improved sea surface temperature predictions. *Journal of Geophysical Research: Atmospheres*, 122(9), 4773-4785.

Khosla, E., Dharavath, R., & Priya, R. (2020). Crop yield prediction using aggregated rainfall-based modular artificial neural networks and support vector regression. Environment, *Development and Sustainability*, 22(6), 5687-5708.

Kim, H. J., Moon, I. J., & Kim, M. (2020). Statistical prediction of typhoon-induced accumulated rainfall over the Korean Peninsula based on storm and rainfall data. *Meteorological Applications*, 27(1), e1853.

Kim, J. S., Chen, A., Lee, J., Moon, I. J., & Moon, Y. I. (2020). Statistical Prediction of Typhoon-Induced Rainfall over China Using Historical Rainfall, Tracks, and Intensity of Typhoon in the Western North Pacific. *Remote Sensing,* 12(24), 4133.

Kim, Y., & Hong, S. (2021). Very Short-Term Rainfall Prediction Using Ground Radar Observations and Conditional Generative Adversarial Networks. *IEEE Transactions on Geoscience and Remote Sensing*.

Kratzert, F., Klotz, D., Herrnegger, M., Sampson, A. K., Hochreiter, S., & Nearing, G. S. (2019). Toward improved predictions in ungauged basins: Exploiting the power of machine learning. *Water Resources Research*, 55(12), 11344-11354.

Kulkarni, S., Mandal, S. N., Sharma, G. S., & Mundada, M. R. (2018). Predictive analysis to improve crop yield using a neural network model. *In 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)* (pp. 74-79). IEEE.

Kumar, P., & Varma, A. K. (2017). Assimilation of INSAT-3D hydro-estimator methodretrieved rainfall for short-range weather prediction. *Quarterly Journal of the Royal Meteorological Society*, 143(702), 384-394.

Li, J., Wang, B., & Yang, Y. M. (2017). Retrospective seasonal prediction of summer monsoon rainfall over West Central and Peninsular India in the past 142 years. *Climate Dynamics*, 48(7-8), 2581-2596.

Liyew, C.M., & Melese, H.A. (2021). Machine learning techniques to predict daily rainfall amount. J Big Data 8, 153 (2021). https://doi.org/10.1186/s40537-021 00545-4

Lu, B., Scaife, A. A., Dunstone, N., Smith, D., Ren, H. L., Liu, Y., & Eade, R. (2017). Skillful seasonal predictions of winter rainfall over southern China. *EnvironmentalResearch Letters*, 12(7), 074021.

Martin Stendel, Jennifer Francis, Rachel White, Paul D. Williams, Tim Woollings, The jet stream and climate change, *Climate Change*, 10.1016/B978-0-12 821575-3.00015 3, (327-357), (2021).

Million insights. Machine Learning Market Analysis Report By Component, By Enterprise Size, By End Use And Segment Forecasts From 2019 To 2025. [Online] Available: https://www.millioninsights.com/industry-reports/global machine-learning-market. August 2020.

Mishra, N., Soni, H. K., Sharma, S., & Upadhyay, A. K. (2018). Development and Analysis of Artificial Neural Network Models for Rainfall Prediction by Using Time-Series Data. *International Journal of Intelligent Systems & Applications*, 10(1).

Moreno-Rodenas, A. M., Cecinati, F., Langeveld, J., & Clemens, F. H. (2017). Impact of spatiotemporal characteristics of rainfall inputs on integrated catchment dissolved oxygen simulations. *Water*, 9(12), 926.

Navid, M. A. I., & Niloy, N. H. (2018). Multiple linear regressions for predicting rainfall for Bangladesh. *Communications*, 6(1), 1.

Nguyen, D. C., & Han, M. Y. (2017). Proposal of simple and reasonable method for design of rainwater harvesting system from limited rainfall data. Resources, *Conservation and Recycling*, 126, 219-227.

Nikolaos Christidis, & Peter A. Stott, The influence of anthropogenic climate change on wet and dry summers in Europe, *Science Bulletin*, 10.1016/j.scib.2021.01.020, (2021).

Oswal, N. (2019). Predicting rainfall using machine learning techniques. arXiv preprint arXiv:1910.13827.

Ouyang, Q., & Lu, W. (2018). Monthly rainfall forecasting using echo state networks coupled with data preprocessing methods. *Water resources management*, 32(2), 659 674.

Panagos, P., Ballabio, C., Meusburger, K., Spinoni, J., Alewell, C., & Borrelli, P. (2017). Towards estimates of future rainfall erosivity in Europe based on REDES and WorldClim datasets. *Journal of Hydrology*, 548, 251-262.

Parida, B. R., Behera, S. N., Bakimchandra, O., Pandey, A. C., & Singh, N. (2017). Evaluation of satellite-derived rainfall estimates for an extreme rainfall event over Uttarakhand, Western Himalayas. *Hydrology*, 4(2), 22.

Pham, B. T., Le, L. M., Le, T. T., Bui, K. T. T., Le, V. M., Ly, H. B., & Prakash, I. (2020). Development of advanced artificial intelligence models for daily rainfall prediction. *Atmospheric Research*, 237, 104845.

Purnomo, H. D., Hartomo, K. D., & Prasetyo, S. Y. J. (2017). Artificial neural network for monthly rainfall rate prediction. *In IOP conference series: materials science and engineering* (Vol. 180, No. 1, p. 012057). IOP Publishing.

Raval, M., Sivashanmugam, P., Pham, V., Gohel, H., Kaushik, A., & Wan, Y. (2021). Automated predictive analytics tool for rainfall forecasting. *Scientific Reports*, 11(1), 1-13.

Reder, A., Rianna, G., and Pagano, L. (2018) Physically based approaches incorporating evaporation for early warning predictions of rainfall-induced landslides, *Nat. Hazards Earth Syst. Sci.*, 18, 613–631, https://doi.org/10.5194/nhess-18-613-2018.

Roger Guiu. (2015). Farewell to growth in Iraq, until when?. [Online] Available: http://www.meri-k.org/publication/farewell-to-growth-in-iraq-until-when/.

Samanta, D., Karnauskas, K. B., & Goodkin, N. F. (2019). Tropical Pacific SST and ITCZ biases in climate models: Double trouble for future rainfallprojections?. Geophysical Research Letters, 46(4), 2242-2252.

Scaife, A. A., Ferranti, L., Alves, O., Athanasiadis, P., Baehr, J., Dequé, M.,. & Yang, X. (2019). Tropical rainfall predictions from multiple seasonal forecast systems. *International Journal of Climatology*, 39(2), 974-988.

Schumacher, R. S. (2017). Heavy rainfall and flash flooding. *In Oxford Research Encyclopedia of Natural Hazard Science*.

Sheen, K. L., Smith, D. M., Dunstone, N. J., Eade, R., Rowell, D. P., & Vellinga, M. (2017). Skilful prediction of Sahel summer rainfall on inter-annual and multi year timescales. *Nature communications,* 8(1), 1-12.

Sofiati, I., & Nurlatifah, A. (2019). The prediction of rainfall events using WRF (weather research and forecasting) model with ensemble technique. *In IOP Conference Series: Earth and Environmental Science* (Vol. 374, No. 1, p. 012036). IOP Publishing.

Staley, D. M., Negri, J. A., Kean, J. W., Laber, J. L., Tillery, A. C., & Youberg, A. M. (2017). Prediction of spatially explicit rainfall intensity–duration thresholds for post-fire debris-flow generation in the western United States. *Geomorphology*, 278, 149-162.

Takahashi, N., Ushio, T., Nakagawa, K., Mizutani, F., Iwanami, K., Yamaji, A., & Kawasaki, M. (2019). Development of multi-parameter phased array weather radar (Mp-pawr) and early detection of torrential rainfall and tornado risk. *Journal of Disaster Research*, 14(2), 235-247.

Thomas J. Eley, (2020). Walker Circulation. *In: Multimedia Atlas of Global Warming and Climatology*. [Online] Available: https://gpm.nasa.gov/education/water-cycle Carly Dodd, (2021). The Water Cycle. *In Environment*, [Online] Available: https://www.worldatlas.com/the water-cycle.html

Tian, J., Liu, J., Wang, J., Li, C., Yu, F., & Chu, Z. (2017). A spatio-temporal evaluation of the WRF physical parameterisations for numerical rainfall simulation in semi humid and semi-arid catchments of Northern China. *Atmospheric Research*, 191, 141-155.

Vijayan, R., Mareeswari, V., Kumar, P. M., Gunasekaran, G., & Srikar, K. (2020). Estimating Rainfall prediction using machine learning techniques on a dataset. International Journal of Scientific and Technology Research (IJSTR), 1(06), 440-445.

Vogel, Peter, et al. (2018). Skill of global raw and postprocessed ensemble predictions of rainfall over northern tropical Africa. *Weather and Forecasting* 33.2: 369 388.

Wadoux, A. M. C., Brus, D. J., Rico-Ramirez, M. A., & Heuvelink, G. B. (2017). Sampling design optimisation for rainfall prediction using a non stationary geostatistical model. *Advances in Water Resources*, 107, 126 138.

Walker, D. P., Birch, C. E., Marsham, J. H., Scaife, A. A., Graham, R. J., & Segele, Z. T. (2019). Skill of dynamical and GHACOF consensus seasonal forecasts of East African rainfall. *Climate Dynamics*, 53(7), 4911-4935.

Wang, B., Li, J., Cane, M. A., Liu, J., Webster, P. J., Xiang, B., & Ha, K. J. (2018). Toward predicting changes in the land monsoon rainfall a decade in advance. *Journal of Climate*, 31(7), 2699-2714.

Wu, M. C., Yang, S. C., Yang, T. H., & Kao, H. M. (2018). Typhoon rainfall forecasting by means of ensemble numerical weather predictions with a GA-Based integration strategy. *Atmosphere*, 9(11), 425.

Xiang, Z., & Demir, I. (2020). Distributed long-term hourly streamflow predictions using deep learning–A case study for State of Iowa. *Environmental Modelling & Software*, 131, 104761.

Xiang, Z., Yan, J., & Demir, I. (2020). A rainfall-runoff model with LSTM-based sequence to-sequence learning. *Water resources research*, 56(1), e2019WR025326.

Yaseen, Z. M., Ghareb, M. I., Ebtehaj, I., Bonakdari, H., Siddique, R., Heddam, S., & Deo, R. (2018). Rainfall pattern forecasting using novel hybrid intelligent model based ANFIS-FFA. *Water resources management*, 32(1), 105-122.

Appendices A

Source Codes

- ML module

```
#Importing all the required libraries for this project
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
import seaborn as sns
from sklearn.datasets import load_breast_cancer   #Using the dataset via sciki
tlearn.datasets
from sklearn.model_selection import train_test_split

#Loading the datasets
cancer = load_breast_cancer()
df_cancer = pd.DataFrame(np.c_[cancer['data'], cancer['target']], columns= n
p.append(cancer['feature_names'],['target']))

#Training the data and splitting
x = df_cancer.drop(['target'],axis =1)
y= df_cancer['target']
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.2, random_st
ate=5)

#K- nearest neighbours model
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import classification_report , confusion_matrix
knn_model = KNeighborsClassifier(n_neighbors = 5, metric = 'minkowski', p
= 2)
knn_model.fit(x_train, y_train)
y_predict =knn_model.predict(x_test)
cm = confusion_matrix(y_test,y_predict)

from sklearn import metrics
print(classification_report(y_test,y_predict))

Ks = 10
mean_acc = np.zeros((Ks-1))
std_acc = np.zeros((Ks-1))
ConfustionMx = [];
for n in range(1,Ks):

    #Train Model and Predict
```

```
    neigh = KNeighborsClassifier(n_neighbors = n).fit(x_train,y_train)
    yhat=neigh.predict(x_test)
    mean_acc[n-1] = metrics.accuracy_score(y_test, yhat)


    std_acc[n-1]=np.std(yhat==y_test)/np.sqrt(yhat.shape[0])
print( "The best accuracy was with", mean_acc.max(), "with k=", mean_acc.a
rgmax()+1)

#Using SVM Model
from sklearn.svm import SVC
from sklearn.metrics import classification_report , confusion_matrix
from sklearn.svm import SVC
svm_model = SVC(kernel = 'linear', random_state = 0)
svm_model.fit(x_train, y_train)
y_predict =svm_model.predict(x_test)
cm = confusion_matrix(y_test,y_predict)

#model improvisation
min_train =x_train.min()
range_train =(x_train - min_train).max()
x_train_scaled =(x_train-min_train)/range_train

from sklearn.metrics import f1_score
f1_score(y_test, yhat, average='weighted')

from sklearn.metrics import jaccard_similarity_score
jaccard_similarity_score(y_test, yhat)

#using decision tree
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sb
import sklearn
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score, confusion_matrix
from sklearn import tree
import graphviz
from graphviz import Source
dt = DecisionTreeClassifier()
dt.fit(X_train, y_train)
score = accuracy_score(y_test, dt.predict(X_test))
print(score)
```

```
Source(tree.export_graphviz(dt, out_file=None, class_names=['B', 'M'], featur
e_names= X_train.columns))

#using deep learning
import keras
from keras.models import Sequential
from keras.layers import Dense, Dropout, Input, BatchNormalization, Activat
ion
from keras.optimizers import Adam
from keras.models import Model
from keras.optimizers import SGD
from keras.callbacks import ReduceLROnPlateau, EarlyStopping
from IPython.display import Image
from keras.utils.np_utils import to_categorical

seed = 15

# split for train, test
X_train, X_test, y_train, y_test = model_selection.train_test_split(
    X,
    y,
    test_size=0.3,
    random_state = seed,
    shuffle = True)

Y_train = to_categorical(y_train,
                         num_classes = None)

Y_test = to_categorical(y_test,
                        num_classes = None)

# split for train, validation
X_train, X_val, Y_train, Y_val = model_selection.train_test_split(
    X_train,
    Y_train,
    test_size=0.3,
    random_state = seed,
    shuffle = True)

# Upsampling
sm = SMOTE(
    sampling_strategy='not majority',
    k_neighbors = 10,
    n_jobs = 1,
    random_state = 12,
    ratio = 0.7)
```

```
X_res, Y_res = sm.fit_sample(X_train, Y_train)

Y_res = to_categorical(Y_res,
                       num_classes = None)

input_layer = Input(shape=(30,))

x = Dense(
    units = 96,
    kernel_initializer='glorot_uniform',
    use_bias = True,
    bias_initializer='zeros')(input_layer)
x = BatchNormalization()(x)
x = Activation('softmax')(x)

skip1 = x

x = Dense(
    units = 96,
    kernel_initializer='glorot_uniform',
    use_bias = True,
    bias_initializer='zeros')(x)
x = BatchNormalization()(x)
x = keras.layers.Add()([x, skip1])
x = Activation('relu')(x)

skip2 = x

x = Dropout(rate = 0.3)(x)

x = Dense(
    units = 96,
    kernel_initializer='glorot_uniform',
    use_bias = True,
    bias_initializer='zeros')(x)
x = BatchNormalization()(x)
x = keras.layers.Add()([x, skip2])
x = Activation('relu')(x)

x = Dropout(rate = 0.3)(x)

x = Dense(
    units = 96,
    kernel_initializer='glorot_uniform',
    use_bias = True,
    bias_initializer='zeros')(input_layer)
x = BatchNormalization()(x)
```

```
x = keras.layers.Add()([x, skip1])
x = keras.layers.Add()([x, skip2])
x = Activation('relu')(x)

x = Dropout(rate = 0.3)(x)

y = Dense(
    units = 2,
    activation='softmax')(x)

model = Model(
    inputs = input_layer,
    outputs = y)

model.summary()

plt.plot(history.history['acc'])
plt.plot(history.history['val_acc'])
plt.title('model accuracy')
plt.ylabel('accuracy')
plt.xlabel('epoch')
plt.legend(['train', 'validation'], loc='lower right')
plt.show()
plt.plot(history.history['loss'])
plt.plot(history.history['val_loss'])
plt.title('model loss')
plt.ylabel('loss')
plt.xlabel('epoch')
plt.legend(['train', 'validation'], loc='upper left')
plt.show()
predictions = model.predict(X_test)
print('Test accuracy: ' + str(accuracy_score(np.argmax(Y_test, axis=1),np.arg
max(predictions, axis=1))))
print('')
print('')
print(classification_report(np.argmax(Y_test, axis=1),np.argmax(predictions,
axis=1)))
```

Other results

Cascade-forward backprop model for Sulamaniah

Actual rainfall	Predicted rainfall
0.378	0.1952991
0.39225	0.1194367
0.66425	0.1290956
0.44325	0.2242966
0.11025	0.1326103
0.0115	0.1426838
0	0.1694905
0	0.1642859
0	0.1296975
0.1795	0.1326133
0.02975	0.1245664
0.36125	0.0616078
0.26025	0.239272
0.34125	0.4377173
0.462	0.2144288
0.186	0.0827586
0.0335	0.0230168
0.00025	0.0097088
0	0.0146034
0	0.0739542
0.00125	0.0958395
0	0.0199657
0.5105	0.1963305
0.05425	0.1649697

Elman backprop model for Sulamaniah

Actual rainfall	Predicted rainfall
0.378	0.0302991
0.39225	0.0097116
0.66425	0.1547668
0.44325	0.0850731
0.11025	0.1310029
0.0115	0.1023283
0	0.097158
0	0.0308741
0	0.0874139
0.1795	0.1107161
0.02975	0.0158167
0.36125	0.0335907
0.26025	0.2178408
0.34125	0.4625792
0.462	0.202798
0.186	0.0065749
0.0335	0.000198
0.00025	1.55E-06
0	0.0002587
0	0.0545274
0.00125	0.139069
0	2.69E-05
0.5105	0.0254103
0.05425	0.1777569

Feed-forward backprop model for Sulamaniah

Actual rainfall	Predicted rainfall
0.378	0.078995
0.39225	0.05572
0.66425	0.134005

0.44325	0.224372
0.11025	0.125
0.0115	0.130996
0	0.16821
0	0.113063
0	0.101154
0.1795	0.098066
0.02975	0.062619
0.36125	0.04121
0.26025	0.236642
0.34125	0.276581
0.462	0.215748
0.186	0.054178
0.0335	0.01906
0.00025	0.01069
0	0.011321
0	0.053694
0.00125	0.104657
0	0.017871
0.5105	0.195191
0.05425	0.161222

layer recurrent model for Sulamaniah

Actual rainfall	Predicted rainfall
0.378	0.12328998
0.39225	0.02407395
0.66425	0.11013265
0.44325	0.16868761
0.11025	0.1333975
0.0115	0.1401561

0	0.1218291
0	0.07562188
0	0.04608616
0.1795	0.04393569
0.02975	0.02801061
0.36125	0.05570534
0.26025	0.18794357
0.34125	0.28357807
0.462	0.25961561
0.186	0.00878403
0.0335	0.04168239
0.00025	5.27E-05
0	0.00614774
0	0.07405648
0.00125	0.16032495
0	4.27E-05
0.5105	0.13030878
0.05425	0.09512215

Cascade-forward backprop model for Duhok

Actual rainfall	Predicted rainfall
0.359	0.002125
0.2125	0.150753
0.66	2.09E-05
0.43675	0.028592
0.099	0.27011
0.002	0.260476
0	0.08073
0.00325	0.001188
0	0.000114

0.10825	0.000248
0.04825	0.044178
0.3445	0.439002
0.27675	7.34E-06
0.32925	0.206419
0.705	0.303431
0.17125	0.191229
0.0405	0.002858
0.0025	0.053426
0	0.115677
0.002	6.21E-05
0	0.057361
0.004	0.091065
0.09925	0.008848
0.1095	0.079303

Elman backprop model for Duhok

Actual rainfall	Predicted rainfall
0.359	0.05049
0.2125	0.177479
0.66	0.233185
0.43675	0.252189
0.099	0.230316
0.002	0.230453
0	0.067591
0.00325	0.10853
0	0.102021
0.10825	0.089701
0.04825	0.160027
0.3445	0.199337

0.27675	0.121802
0.32925	0.212362
0.705	0.296546
0.17125	0.214504
0.0405	0.13252
0.0025	0.134959
0	0.108671
0.002	0.025685
0	0.047755
0.004	0.074928
0.09925	0.05833
0.1095	0.06548

Feed-forward backprop model for Duhok

Actual rainfall	Predicted rainfall
0.359	0.012915
0.2125	0.185165
0.66	0.20179
0.43675	0.237161
0.099	0.238873
0.002	0.234161
0	0.079539
0.00325	0.081074
0	0.063218
0.10825	0.049963
0.04825	0.095238
0.3445	0.208401
0.27675	0.068353
0.32925	0.207817
0.705	0.263723
0.17125	0.215748

0.0405	0.109016
0.0025	0.139535
0	0.090359
0.002	0.012938
0	0.036221
0.004	0.088938
0.09925	0.016556
0.1095	0.034421

layer recurrentmodel for Duhok

Actual rainfall	Predicted rainfall
0.359	0.004709
0.2125	0.209123
0.66	0.166978
0.43675	0.209524
0.099	0.220286
0.002	0.204433
0	0.085762
0.00325	0.095807
0	0.071667
0.10825	0.051013
0.04825	0.071381
0.3445	0.193102
0.27675	0.06153
0.32925	0.224326
0.705	0.32224
0.17125	0.241798
0.0405	0.140904
0.0025	0.163978
0	0.133567
0.002	0.001302

0	0.01651
0.004	0.095024
0.09925	0.009481
0.1095	0.022036

Cascade-forward backprop model forn Halabja

Actual rainfall	Predicted rainfall
0.348	0.165728
0.2955	0.126671
0.442	0.321438
0.2945	0.161222
0.05	0.285626
0	0.351333
0	0.170317
0	0.140317
0	0.089973
0.115725	0.06116
0.0185	0.089973
0.28325	0.286955
0.2285	0.07692
0.17025	0.17451
0.31075	0.229319
0.209	0.061085
0.025	0.060682
0	0.075182
0	0.123456
0	0.115999
0	0.247724
0	0.397266
0	0.355144
0	0.088255

Elman backprop model forn Halabja

Actual rainfall	Predicted rainfall
0.348	0.218489
0.2955	0.167044
0.442	0.287737
0.2945	0.218428
0.05	0.267867
0	0.333841
0	0.184768
0	0.200354
0	0.181991
0.115725	0.086673
0.0185	0.181991
0.28325	0.273303
0.2285	0.06029
0.17025	0.146551
0.31075	0.213438
0.209	0.161987
0.025	0.130116
0	0.111814
0	0.090995
0	0.068642
0	0.236721
0	0.289012
0	0.276321
0	0.1519

Feed-forward backprop model forn Halabja

Actual rainfall	Predicted rainfall

0.348	0.208224
0.2955	0.13004
0.442	0.235995
0.2945	0.214222
0.05	0.191285
0	0.297689
0	0.17834
0	0.178917
0	0.1365
0.115725	0.119966
0.0185	0.1365
0.28325	0.115228
0.2285	0.052392
0.17025	0.178958
0.31075	0.234463
0.209	0.132649
0.025	0.121582
0	0.084726
0	0.101154
0	0.08215
0	0.24172
0	0.136829
0	0.15508
0	0.10744

layer recurrent model forn Halabja

Actual rainfall	Predicted rainfall
0.348	0.1798959
0.2955	0.1337828
0.442	0.3380428

0.2945	0.1750563
0.05	0.295552
0	0.4503523
0	0.1699327
0	0.1574644
0	0.0958946
0.115725	0.0539533
0.0185	0.0958946
0.28325	0.1201212
0.2285	0.0588584
0.17025	0.1551595
0.31075	0.2134593
0.209	0.1004847
0.025	0.0800526
0	0.0846252
0	0.099654
0	0.0790157
0	0.2406993
0	0.366403
0	0.3268916
0	0.1113967

KNN model

Stations	Actual rainfall	Predicted rainfall
Sulamaniah	0.1340	0.2310
	0.2530	0.2670
	0.3450	0.3260

	0.3120	0.4310
	0.3210	0.4410
	0.3220	0.2560
	0.112	0.2230
	0.4300	0.4210
	0.2680	0.3290
	0.2200	0.2310
	0.2300	0.2410
	0.2670	0.2810
	0.4312	0.4110
	0.4310	0.4670
	0.3270	0.4560
	0.2230	0.2510
	0.2490	0.1570
	0.2230	0.3130
	0.2210	0.2450
	0.4100	0.3510
	0.3240	0.3140
	0.3800	0.2700
	0.1460	0.2450
	0.1210	0.2100

KNN model

Stations	Actual rainfall	Predicted rainfall
Duhok	0.2310	0.3110
	0.2640	0.2570
	0.4210	0.3510

	0.2150	0.3270
	0.4120	0.4330
	0.3420	0.3520
	0.3240	0.3570
	0.4210	0.3530
	0.3250	0.3620
	0.2120	0.3230
	0.3210	0.3310
	0.3340	0.3210
	0.2310	0.2210
	0.3250	0.4670
	0.2120	0.2430
	0.4530	0.4320
	0.3210	0.4370
	0.2450	0.3560
	0.4210	0.4720
	0.4710	0.3120
	0.2560	0.3850
	0.2310	0.2410
	0.3210	0.4230
	0.4330	0.4160

KNN model

Stations	Actual rainfall	Predicted rainfall
Halabja	0.3410	0.3210
	0.3740	0.4670

	0.1830	0.2730
	0.2110	0.2310
	0.2830	0.3840
	0.4320	0.3930
	0.3490	0.4130
	0.3720	0.4210
	0.2740	0.3810
	0.4250	0.4210
	0.4720	0.4130
	0.3270	0.2850
	0.26770	0.2360
	0.2340	0.2150
	0.4870	0.4350
	0.4920	0.4720
	0.3210	0.3420
	0.1530	0.2430
	0.2210	0.1140
	0.2750	0.2670
	0.4120	0.4640
	0.1560	0.1230
	0.1430	0.1570
	0.2250	0.2250

LR model

Stations	Actual rainfall	Predicted rainfall
Sulamaniah	0.1210	0.2140
	0.3310	0.2529

	0.3270	0.2510
	0.2810	0.2820
	0.2250	0.4720
	0.3210	0.2630
	0.3170	0.3210
	0.4170	0.4540
	0.3620	0.3210
	0.3430	0.3720
	0.3720	0.3830
	0.1530	0.2170
	0.1630	0.2420
	0.4410	0.3210
	0.3320	0.2750
	0.3570	0.3950
	0.3140	0.4320
	0.3210	0.3340
	0.2330	0.2140
	0.1630	0.1840
	0.2520	0.2320
	0.4730	0.4520
	0.3820	0.3720
	0.2520	0.2810

LR model

Stations	Actual rainfall	Predicted rainfall
Duhok	0.4530	0.2501
	0.2630	0.2470

	0.1210	0.2130
	0.2730	0.2320
	0.2830	0.3390
	0.3340	0.2890
	0.3320	0.2310
	0.4300	0.420
	0.3210	0.3320
	0.3320	0.3450
	0.3720	0.3790
	0.1230	0.2320
	0.1350	0.2330
	0.4320	0.3570
	0.3220	0.2670
	0.3680	0.3620
	0.3460	0.3720
	0.4510	0.4340
	0.2630	0.2430
	0.1760	0.1960
	0.2430	0.2650
	0.4540	0.4320
	0.2620	0.2420
	0.2310	0.2530

LR model

Stations	Actual rainfall	Predicted rainfall
Halabja	0.4300	0.2410
	0.3210	0.2430

	0.3201	0.3210
	0.4002	0.4420
	0.2220	0.1120
	0.3240	0.3690
	0.3370	0.3210
	0.4510	0.4300
	0.4310	0.4320
	0.3410	0.2550
	0.3430	0.3440
	0.1740	0.2630
	0.1640	0.2210
	0.4580	0.3840
	0.3350	0.2320
	0.3550	0.3730
	0.3320	0.3310
	0.4420	0.4830
	0.2840	0.2610
	0.2360	0.1250
	0.3630	0.3650
	0.3240	0.3620
	0.4320	0.4520
	0.2110	0.4130

SVM model

Stations	Actual rainfall	Predicted rainfall
	0.1100	0.1230

Sulamaniah	0.1840	0.2820
	0.4400	0.2550
	0.3320	0.3320
	0.1730	0.1310
	0.2220	0.2730
	0.2470	0.4310
	0.3110	0.3260
	0.4560	0.4140
	0.3250	0.3780
	0.2870	0.3340
	0.1370	0.1210
	0.2420	0.2630
	0.4740	0.3750
	0.3630	0.3480
	0.4260	0.4430
	0.4560	0.3540
	0.4110	0.3270
	0.2460	0.2670
	0.1320	0.1780
	0.3210	0.2440
	0.3730	0.3770
	0.4540	0.4320
	0.2320	0.31120

SVM model

Stations	Actual rainfall	Predicted rainfall
	0.3720	0.4930

Duhok	0.3730	0.4920
	0.4210	0.4220
	0.2210	0.2230
	0.3210	0.4310
	0.3220	0.3530
	0.2170	0.4310
	0.4310	0.4260
	0.3460	0.4320
	0.3120	0.3580
	0.2540	0.2440
	0.1340	0.1210
	0.3220	0.3230
	0.4210	0.3650
	0.2430	0.2180
	0.3760	0.4210
	0.3260	0.3130
	0.3410	0.3130
	0.2360	0.2420
	0.1740	0.1320
	0.2410	0.2650
	0.2830	0.2670
	0.4340	0.4440
	0.2210	0.3760

SVM model

Stations	Actual rainfall	Predicted rainfall
	0.4210	0.2410

Halabja	0.3310	0.2420
	0.2210	0.4420
	0.4240	0.3510
	0.3240	0.2530
	0.4220	0.3830
	0.2650	0.3410
	0.4540	0.4640
	0.3760	0.4670
	0.3540	0.2780
	0.2540	0.2330
	0.1210	0.2410
	0.2120	0.3430
	0.4210	0.3350
	0.2210	0.2440
	0.3460	0.4310
	0.3360	0.3230
	0.4410	0.4130
	0.2760	0.3120
	0.1620	0.1540
	0.2230	0.1850
	0.1530	0.2370
	0.4040	0.4230
	0.2870	0.3550

DT model

Stations	Actual rainfall	Predicted rainfall
	0.1320	0.1220

Sulamaniah	0.3210	0.2890
	0.2190	0.3210
	0.1210	0.1430
	0.2210	0.1830
	0.3530	0.3760
	0.2320	0.2120
	0.4650	0.4710
	0.2430	0.2310
	0.4220	0.3740
	0.3210	0.2750
	0.1340	0.1230
	0.1720	0.2670
	0.4340	0.3630
	0.2340	0.2210
	0.3730	0.4230
	0.3360	0.2830
	0.3130	0.3430
	0.2660	0.2760
	0.1210	0.2120
	0.3270	0.2570
	0.2560	0.2830
	0.4110	0.3830
	0.2550	0.3110

DT model

Stations	Actual rainfall	Predicted rainfall
	0.3120	0.3220

Duhok	0.3010	0.3300
	0.4220	0.1310
	0.2750	0.3110
	0.1870	0.2210
	0.3750	0.3550
	0.2440	0.2320
	0.4680	0.4540
	0.2550	0.2670
	0.4020	0.3640
	0.3230	0.2320
	0.1440	0.1320
	0.1770	0.2220
	0.4110	0.3550
	0.2350	0.2150
	0.3540	0.4120
	0.3120	0.3230
	0.3450	0.3330
	0.2320	0.2210
	0.1220	0.2450
	0.3230	0.2780
	0.2870	0.3210
	0.4300	0.3230
	0.3150	0.3440

DT model

Stations	Actual rainfall	Predicted rainfall
	0.4210	0.4100

Halabja	0.3210	0.3440
	0.2240	0.2310
	0.3230	0.4110
	0.1210	0.1630
	0.3240	0.2550
	0.3240	0.3420
	0.4550	0.4760
	0.1550	0.2220
	0.4130	0.3780
	0.2430	0.2230
	0.2340	0.2120
	0.1650	0.2210
	0.4430	0.3780
	0.2240	0.2210
	0.3840	0.4260
	0.3230	0.3120
	0.3340	0.3220
	0.2650	0.2760
	0.1680	0.2230
	0.3140	0.2660
	0.2540	0.3150
	0.4230	0.3780
	0.3320	0.3210

Appendices B

Ethics Letter

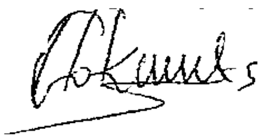
TO THE INSTITUEBOF GRADUATE STUDIES
REFERENCE: TAHIR SHAMSALDDIN ABDALSAMAD (20205133)

I would like to inform you that the above candidate is one of My postgraduate students in Civil Engineering Department. He is taking thesis under my supervision on the thesis entailed: **STATISTICAL AND MACHINE LEARNING TECHNIQUES APPLIED TO THE PREDICTION OF TOTAL RAINFALL IN URBAN CITIES, NORTHERN PART OF IRAQ.** The data used in his study was my own data obtained from experimental work conducted by Ministry of planning and statistics office.

Please do not hesitate to contact me if you have any futher queries or questions.

Thank you very much indeed.

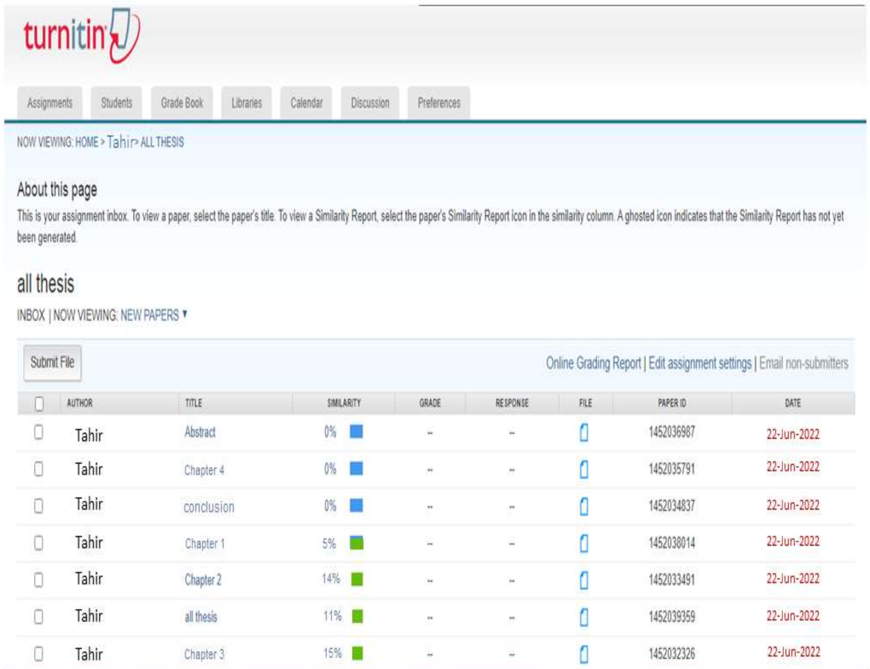
Best regards,



Prof. Dr. Hüseyin Gökçekuş
Dean of Facalty of Civil and Environmental,
Enginnering, Near East Boulevard, ZIP: 99138
Nicosia / TRNC, North Cyprus,
Mersin 10 - Turkey.
Email: huseyin.gokcekus@neu.edu.tr

Appendices C

Similarity Report



AUTHOR	TITLE	SIMILARITY	GRADE	RESPONSE	FILE	PAPER ID	DATE
Tahir	Abstract	0%	--	--		1452036987	22-Jun-2022
Tahir	Chapter 4	0%	--	--		1452035791	22-Jun-2022
Tahir	conclusion	0%	--	--		1452034837	22-Jun-2022
Tahir	Chapter 1	5%	--	--		1452038014	22-Jun-2022
Tahir	Chapter 2	14%	--	--		1452033491	22-Jun-2022
Tahir	all thesis	11%	--	--		1452039359	22-Jun-2022
Tahir	Chapter 3	15%	--	--		1452032326	22-Jun-2022

Prof. Dr. Hüseyin Gökçekuş