M. SABAMEHR

DEVELOPMENT OF AI-DRIVEN DROUGHT MANAGEMENT SYSTEM A CASE STUDY FOR TURKEY

THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
ATILIM UNIVERSITY

MILAD SABAMEHR

A MASTER OF SCIENCE
IN
THE DEPARTMENT OF SOFTWARE ENGINEERING

JUNE 2024

ATILIM UNIVERSITY 2024

DEVELOPMENT OF AI-DRIVEN DROUGHT MANAGEMENT SYSTEM
A CASE STUDY FOR TURKEY

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
ATILIM UNIVERSITY

BY

MILAD SABAMEHR

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER
IN
THE DEPARTMENT OF SOFTWARE ENGINEERING

JUNE 2024

Approval of the Graduate School of Natural and Applied Sciences, Atilim University.

_____

Prof. Dr. Ender KESKİNKILIÇ
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of **Master of Science in Software Engineering, Atilim University**.

_____

Prof. Dr. Ali Yazıcı
Head of Department

This is to certify that we have read the thesis DEVELOPMENT OF AI-DRIVEN DROUGHT MANAGEMENT SYSTEM A CASE STUDY FOR TURKEY submitted by MILAD SABAMEHR and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

_____

Assoc. Prof. Dr. Cansu Çiğdem EKİN
Supervisor

Examining Committee Members:

Asst. Prof. Dr. Güzin TÜRKMEN
Software Eng. Department, Atilim University _____

Assoc. Prof. Dr. Cansu Çiğdem EKİN
Computer Eng. Department, Atilim University _____

Asst. Prof. Dr. Ekrem Çağlar YILMAZ
Computer Eng. Department, Ankara Science University _____

Date:*05/06/2024*

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: Milad Sabamehr

Signature:

ABSTRACT

DEVELOPMENT OF AI-DRIVEN DROUGHT MANAGEMENT SYSTEM A CASE STUDY FOR TURKEY

Sabamehr, Milad

MSc., Department of Software Engineering

Supervisor: Assoc. Prof. Dr. Cansu Çiğdem EKİN

JUNE 2024, 64 pages

Drought presents a significant challenge for developed countries due to industrial growth and pollution. Mitigating this requires innovative methods like Artificial Intelligence (AI). This study introduces an AI-driven drought management system comprising a data management system, prediction system, and PDSI and SPI calculation system. The data management system allows users to analyze agricultural data from various regions of Turkey. The prediction system employs SARIMA, ARIMA, and Prophet algorithms to forecast precipitation and temperature. The best-performing algorithm is selected based on error rates, enabling accurate predictions. These predictions are then saved in the data management system. Leveraging AI-generated data, the PDSI and SPI calculation system forecasts PDSI and SPI for the next two years. Additionally, the system assesses the risk associated with cultivating agricultural products in specific regions by comparing expected weather conditions with PDSI and SPI forecasts. Results indicate that ARIMA is optimal for predicting temperature averages, while SARIMA performs best for forecasting precipitation. For wheat products in Şanlıurfa province in 2024, PDSI indicates an 91% risk level and SPI-3 indicates a 75% risk level.

Keywords: Drought management, Artificial Intelligence (AI), Prediction Systems PDSI (Palmer Drought Severity Index), SPI (Standardized Precipitation Index)

ÖZ

YAPAY ZEKA TABANLI KURAKLIK YÖNETİM SİSTEMİ GELİŞTİRİLMESİ TÜRKİYE İÇİN BİR DURUM ÇALIŞMASI

Sabamehr, Milad

Yüksek Lisans, Yazılım Mühendisliği Bölümü

Danışman: Assoc. Prof. Dr. Cansu Çiğdem EKİN

Haziran 2024, 64 sayfa

Sanayi büyümesi ve kirlilik nedeniyle gelişmiş ülkeler için kuraklık ciddi bir sorun haline gelmektedir. Bu sorunun üstesinden gelmek için yenilikçi yaklaşımlara ihtiyaç vardır, bunlardan biri de yapay zeka (AI) gibi teknolojilerdir. Bu çalışma, bir veri yönetim sistemi, tahmin sistemi ve PDSI ve SPI hesaplama sistemi içeren bir AI tabanlı kuraklık yönetim sistemini tanıtmaktadır. Veri yönetim sistemi, kullanıcıların Türkiye'nin çeşitli bölgelerinden tarımsal verileri analiz etmelerine olanak tanır. Tahmin sistemi, yağış ve sıcaklık tahmin etmek için SARIMA, ARIMA ve Prophet algoritmalarını kullanır. En iyi performans gösteren algoritma, hata oranlarına göre seçilir, böylece doğru tahminler yapılır. Bu tahminler daha sonra veri yönetim sisteminde saklanır. Yapay zeka tarafından oluşturulan verilerden yararlanarak, PDSI ve SPI hesaplama sistemi bir sonraki iki yıl için PDSI ve SPI tahmin eder. Ayrıca, sistem beklenen hava koşullarını PDSI ve SPI tahminleri ile karşılaştırarak belirli bölgelerde tarımsal ürün yetiştirme riskini değerlendirir. Sonuçlar, ARIMA'nın sıcaklık ortalamalarını tahmin etmek için en uygun olduğunu, SARIMA'nın ise yağışı tahmin etmek için en iyi performansı gösterdiğini göstermektedir. 2024'te Şanlıurfa ilinde Buğday ürünleri için PDSI'nin %91 risk seviyesi ve SPI-3'ün %75 risk seviyesi olduğunu göstermektedir.

Anahtar Kelimeler: Kuraklık yönetimi, Yapay Zeka (YZ), Tahmin sistemleri, PDSI (Palmer Kuraklık Şiddet İndeksi), SPI (Standartlaştırılmış Yağış İndeksi)

To …

My Father and Mother,

You are the most valuable things that I have in my life and I can't show that how much I love you. Thank you for being in my life.

You are special to me. Love you.

ACKNOWLEDGMENTS

I am deeply grateful to my advisor, Associate Professor Dr. Cansu Çiğdem EKİN, for her unwavering support and encouragement throughout this study. Her dedicated involvement in every aspect of the process has been instrumental in the completion of this paper. Without her guidance, this achievement would not have been possible. I extend my heartfelt thanks for her continued support and understanding throughout this journey.

Additionally, I would like to express my gratitude to my family members for their constant support and motivation during this study.

TABLE OF CONTENTS

ABSTRACT iii
ÖZ iv
ACKNOWLEDGMENTS vi
TABLE OF CONTENTS vii
LIST OF TABLES x
LIST OF FIGURES xi
LIST OF SYMBOLS/ABBREVIATIONS xiii
CHAPTER 1 1
INTRODUCTION 1
1.1. Statement of the Problem 2
1.2. Purpose of the Study 3
1.3. Significance of The Study 4
1.4. Research Questions 4
1.5. The Structure of This Thesis's Chapters 4
CHAPTER 2 5
LITERATURE REVIEW 5
2.1. Background 5
2.2. Literature Review for Drought Management and Prediction In the Turkiye 5
2.3. Literature Review for Drought Management and Prediction in Foreign Countries 7
2.4. Summary 10
CHAPTER 3 12
BACKGROUND OF THE STUDY 12
3.1. Introduction 12
3.2 Drought Categorization 12
3.2.1 Meteorological Drought 12
3.2.2 Hydrological Drought 13
3.2.3 Agricultural Drought 13
3.2.4 Socio-Economic Drought 14

3.3 Drought Indices 15
3.3.2 Standardized Precipitation Index (SPI) 16
3.3.3 Standardized Precipitation-Evapotranspiration Index (SPEI): 17
3.3.4 Crop Moisture Index (CMI): 17
3.3.5. Surface Water Supply Index (SWSI): 17
3.3.6. Vegetation Health Indices: 17
3.4 Time Series 18
3.4.1 Components of Time series 18
3.5 Time Series Models 19
3.5.5 Prophet 23
3.5.6 Examples of Time Series Usage in Different Sectors 23
3.6. Evaluation Metrics 26
3.6.1 Mean Squared Error (MSE) 26
3.6.2 Root Mean Squared Error (RMSE) 26
3.6.3 Mean Absolute Error (MAE) 26
CHAPTER 4 27
METHODOLOGY 27
4.1. Introduction 27
4.2. Data Collection/Data Extraction 28
4.3. Data Preprocessing 29
4.4. Database Design 30
4.5. Data Management Software 33
4.5.1 Report 34
4.5.2 Data Analysis 34
4.6. Selecting the Best Algorithm 36
4.7. Prediction Software 39
4.8. PDSI and SPI-3 Calculation Software 41
CHAPTER 5 44
RESULTS 44
5.1. Research Question 1 44
5.1.1 Precipitation and Temperature Average Prediction 44
5.2. Research Question 2 48

CHAPTER 6 52
DISCUSSION AND CONCLUSSION 52
6.1. Agricultural Product Analysis 52
6.2. Precipitation 52
6.3. Temperature Average 53
6.4. Predicting PDSI and SPI 54
6.5. Predicting Agricultural Drought by PDSI and SPI 54
6.6. Conclusion 54
6.7.1 Data Quality and Availability 56
REFERENCES 58

LIST OF TABLES

TABLES

Table 3.1 PDSI Index Range 16

Table 3.2 SPI Index range 17

Table 3.3 Parameters of SARIMA and ARIMA models 23

Table 5.1 Comparing the performances of the models for precipitation prediction (2021-2022) 54

Table 5.2 Comparing the performances of the models for temperature average prediction (2021-2022) 54

Table 5.3 Weather Conditions, Prediction, and Observation Consistency Analysis (2022) 57

Table 5.4 wheat product drought index (PDSI) prediction for Sanliurfa (2024) 59

Table 5.5 wheat product drought index (SPI-3) prediction for Sanliurfa (2024).... 60

LIST OF FIGURES

FIGURES

Figure 3.1 (a) Line Graph of Avg LST 2012 [28] 13
Figure 3.2 (b)Line Graph of Avg NDVI 2012[29] 14
Figure 3.3 Line Graph showing NDVI changes w.r.t. LST for the year 2012[29] 14
Figure 3.4 Distribution of (a) water supply and (b) water demand proportions 15
throughout the year [30] 15
Figure 3.5 A time series plot with different components [41] 19
Figure 3.6 Sample STL graph for Decomposition Plot of Daily Carbon Dioxide Data [44] 22
Figure 3.7 Visualization of Indonesia COVID-19 confirmed cases [48] 24
Figure 3.8 The prediction results of different models on SPEI-3[49] 25
Figure 4.1 Steps of research methodology 28
Figure 4.2 Unstructured data before pre-processing 29
Figure 4.3 Structured Data After Pre-Processing 30
Figure 4.4 The main table of the database 32
Figure 4.5 ER Diagram of the Database of the System 32
Figure 4.6 System Functionality Diagram for Agricultural Data Management and Prediction 33
Figure 4.7 Use Case diagram of the data management system. 34
Figure 4.7 Bar chart shows each product percentage to all of the products 35
Figure 4.8. The percentage of product yield rate according to the district. 35
Figure 4.9. auto_arima function in python to find the best model of ARIMA for precipitation data 36
Figure 4.10. auto_arima function in python to find the best model of ARIMA for Temperature average data 37
Figure 4.11. auto_arima function in python to find the best model of SARIMA for precipitation data 38
Figure 4.12. auto_arima function in python to find the best model of SARIMA for Temperature average data 38
Figure 4.13 Parameters adjustment in prophet algorithm for both precipitation and temperature average data 39

Figure 4.14 Form of the prediction system 40
Figure 4.15 Importing required libraries 40
Figure 4.16 creating web form in R 40
Figure 4.17 Checking the existence of files 41
Figure 4.18 Reading uploaded from the form 41
Figure 4.19 ARIMA and SARIMA Functions in R 41
Figure 4.20 Form of the PDSI and SPI-3 software 42
Figure 4.21 PDSI calculation code in R 42
Figure 4.22 SPI-3 calculation in various scales in R 42
Figure 5.2 Error metrics for each Time Series algorithm prediction for temperature average data (2021-2022) 46
Figure 5.3 Actual data and ARIMA prediction for Precipitation(mm) 47
Figure 5.4 Actual data and ARIMA prediction for Temperature Average (Celsius) 47

LIST OF SYMBOLS/ABBREVIATIONS

AI	:	Artificial Intelligence
ML	:	Machine Learning
DL	:	Deep Learning
SPI	:	Standardized Precipitation Index
PDSI	:	Palmer Drought Severity Index
AIC	:	Akaike Information Criterion
SBC	:	Schwarz Bayesian Criterion
RMSE	:	Rout Mean Squared Error

UTeM GCRIS

CHAPTER 1

INTRODUCTION

Nowadays, drought is one of the trending topics in the world that is turned into a challenge for the world. By developing countries and cities worldwide, especially in the economic aspect, governments started to damage the environment such as through the use of fossil fuels, pollution of the seas, unregulated use of fresh water also deforestation for personal purposes. All these factors have contributed to increasing the temperature of the earth, which is known as global warming [1]. As much as it is said, global warming can cause many dangerous and irreparable damages including ice melting and rising sea levels, that could affect the air condition of different regions of the earth and be the cause of other natural disasters. As a result, preventive measures need to be taken to mitigate the occurrence of these natural disasters. Consequently, by determining the damages of global warming and the increasing occurrence of drought, some urgent implementations and innovative solutions are required to mitigate and manage water scarcity. In recent years, Artificial intelligence and AI-based software and technologies have been assisting humans in finding promising solutions to solve these complex environmental worrying trends. Additionally, AI-powered applications can enhance the accuracy of drought prediction based on the previously collected data to aware governments and organizations take the actions that need to be implemented before facing the problems that this worsening trend can create [2].By integrating machine learning algorithms with weather data and ground-based sensors, it becomes easier to detect changes in moisture levels, soil conditions, and the health status of the vegetation. Moreover, AI can contribute to the development of water management systems to warn the organization about precipitation or rainfall of

an area and also water availability, detect leaks, and also in the prediction of precipitation in the coming years or months even days for the selected area or region.[3] This system also can promote and monitor water consumption in agriculture and individual households, keeping records of consumption that can provide useful data for a long time to the government. Furthermore, the system can conduct data analysis to help the government and related organizations implement proactive measures to mitigate impacts more efficiently than manual data analysis. In this thesis, we aimed to explore the problem that the Turkish agriculture department faced when the organization attempted to do drought prediction by using collected data from different parts of the country. This study demonstrates that it is possible to analyze historical data and develop predictive models in a system that is going to show the stakeholders to see the status of the Turkiye, and also it is remarkable to illustrate that integrating cutting-edge AI technologies in the agricultural industry causes a significant boost in the performance of the organization in turning the inconsistency data to valuable knowledge and raising awareness of the condition of agricultural products in Turkey through data analysis provided by the system.

1.1. Statement of the Problem

The occurrence of drought in the countries like Turkiye, caused by global warming and unsustainable development practices, presents critical challenges for water source management and drought prediction. Traditional approaches to drought response are currently unable to monitor the weather condition, water sources, and agricultural product situations. Consequently, there is a pressing need for innovative solutions that harness the potential of artificial intelligence (AI) to enhance drought management, implement proactive measures, and optimize water resource allocation. With Turkiye's rapid urbanization and economic growth, problems arise in allocating water resources to different regions due to limitations in water sources and weaknesses in management. In this study, after h meetings with the Ministry of Agriculture and Forestry and observing the drought prediction systems of the organization, it was noticed that the organization was analyzing and predicting drought and product's

health based on meteorological data, and all reports were manually written and distributed in files. Additionally, there were no system to calculate drought indexes to show the drought condition in different parts of the Turkiye. From the one side , engineers after observing the agricultural products in the farm , they were writing the reports in a xlsx file without any structure and from the other side , there were no prediction system to aware the organization about the precipitation and temperature average in the coming years for the region and calculate the drought indexes to show the drought condition precisely. The ultimate goal of this study is to leverage cutting-edge AI technologies to enhance the performance of the traditional approaches used in environmental or agricultural sciences, promoting innovation and advancement. By integrating advanced and data-driven approaches, this research aims to mitigate the socioeconomic and environmental impacts of drought events by improving the accuracy of drought prediction and monitoring.

1.2. Purpose of the Study

The study aims to dissolve the existing drought management approaches through Innovative and new data-driven technologies. The primary objectives of this research are:

- To calculate the drought indexes such as SPI and PDSI that are imperative in forecasting drought.
- To monitor and analyze the rainfall amount in specific cities of Turkey and compare the precipitation amount in the next two years. Also, in this system, it is possible to predict the next two years' daily rainfall amount.
- To monitor and analyze the average temperature of specific cities of Turkey and compare and analyze the temperature average amount in the two next years. Additionally, this system can predict the next two years' daily temperature average.
- Comparing and predicting the monthly calculated SPI and PDSI in specific year with agricultural products expected environment and result the risky or

not risky environment for products in coming years for specific region in Turkiye.

1.3. Significance of The Study

The significance of this study is focused on the field of drought management, specifically within the Turkey context. By examining the systems that are related to the agriculture field in Turkiye, there is an urgent need to integrate new monitoring and cutting-edge AI-based systems that leverage advanced data-driven approaches. This study seeks to improve the accuracy of drought forecasting, analyzing the coming data from organizations to be more effective in drought mitigation and adaptation strategies.

1.4. Research Questions

The study aims to find answers to the following questions based on the analysis and prediction methods:

1. Research Question 1. Which ML algorithm is more suitable for making predictions for meteorological drought?
2. Research Question 2. Is it possible to predict agricultural drought in different cities of Turkey using meteorological drought index?

1.5. The Structure of This Thesis's Chapters

- Chapter 1: Introduction
- Chapter 2: Related Works
- Chapter 3: Background of the Study
- Chapter 4: Methodology
- Chapter 5: Results
- Chapter 6: Discussion and Conclusions

CHAPTER 2

LITERATURE REVIEW

2.1. Background

In the previous chapter, imperative key points regarding the problem, study objectives, and also the goal of the study were provided. However, there have been many studies carried out on drought prediction, drought management systems, drought monitoring frameworks, and also applied machine learning in drought management systems that are divided into two groups national and international, and they are elaborated in the two sub-sections of 2.2 and 2.3. The provided chapter aims to show how different kinds of studies have been done, exploring and explaining the studies with the same or at least similar topics, illustrating the strengths and weaknesses of the studies in Turkey and also other countries, disclosing their differences and in some cases, their advantages and disadvantages, of the researches, particularly in terms of research methodologies and implementations.

2.2. Literature Review for Drought Management and Prediction In the Turkiye

Cetin and Kayam[4],discusses the impacts of the drought in Turkey, especially in the agricultural sector, and also measures that have been taken by the organizations and government to manage and mitigate the effects of this natural disaster. Also, it discusses the significant impacts of droughts on agricultural products such as wheat, barley, sunflower, beans, and lentils. Following this paper, some papers concentrate on the technological solutions that can be taken to predict drought, for example paper [5]focuses on detecting and establishing correlation between precipitation and climate oscillation indices in the Aras Basin, and also the paper presents a method for detecting drought using various climate indices and the MOWCATL algorithm. It also highlights the important roles of oceanic parameters as antecedents to monitoring drought and shows the correlation between atmospheric and oceanic conditions with

drought occurrences in the Aras Basin. Drought prediction is one of the most useful measures that can be taken by organizations to be aware of the drought happening in the country. In the research [2] focuses on predicting drought in the buyuk Menderes river basin, using the standardized precipitation index (SPI) as a drought indicator. This study employs linear stochastic models namely, Arima and Sarima, in addition, this study shows that the Arima model's performance is very powerful than the other models based on the parameters AIC and SBC. Moreover, this study illustrates the potential of ARIMA models and SPI as valuable tools for drought forecasting and helping the government in water and agricultural product management. It seems most of the researchers in Turkey concentrated mostly on drought analysis based on the various indexes for example paper [6] is analyzing meteorological drought in the kucuk Menderes river basin, located in the western part of Turkey, by using drought indices based on precipitation and temperature data. This study focuses on three drought indices including the dimensionless precipitation anomaly index(DPAI) which measures the departure of precipitation from its normal yearly and the standardized precipitation index(SPI) which is related to the monthly precipitation data to analyze drought at different time scales and also standardized precipitation evapotranspiration index(SPEI) that includes monthly precipitation and temperature data for drought analysis at monthly scales. As much as it is discussed in [6]it can be seen that by using these simple drought index based on the readily available meteorological data, the study successfully identifies the dry periods in the river basin. In addition to this study, also [7] is investigated and analyzed drought conditions in the Seyhan basin , located in the south of Turkey, using remote sensing technology and various vegetation indices. normalized difference vegetation index (NDVI) and the vegetation condition index (VCI) are the two main indices that are used in the study. Nowadays, Artificial intelligence plays an imperative role in different areas and majors such as in agriculture also as it is demonstrated in [2] machine learning time-series algorithms and models was being applied to the drought indices data to predict drought happening in the specific area in different time periods. Following this paper, also the paper [8] aim is to develop a reliable and accurate model for predicting monthly streamflow in the Amasya region. This research combines the wavelet transform and a feedforward backpropagation neural network to achieve higher

accuracy in streamflow predictions. Another drought analysis has been done by the reasearch [9] that focuses on the drought situation in the Igdir plain, located in the eastern part of Turkey, which is known for its suitable soil and weather conditions for agricultural cultivation. This study focused on the effects of low precipitation and high temperatures on both hydrological and meteorological drought in the region.

2.3. Literature Review for Drought Management and Prediction in Foreign Countries

Returning to the international papers and research that have been done till now, we can see that most of the papers are concentrated on the continents such as South Asia, the USA, and mostly Africa. It seems that these parts of the earth are in a war with drought. By looking at and analyzing the papers that have researched the factors that are the cause of the drought, it can be seen that Artificial intelligence and also computer science are playing important roles in the prediction and analysis part of this research. An early warning system was introduced by the paper [10]to propose a composite drought vulnerability indicator for the Pan-Africa region. In addition, this paper highlights and discussed the vulnerability of drought on biophysical and socio-economic factors, and how understanding vulnerability can improve preparedness and mitigate the effects of drought. Study [11] is focused to evaluate the use of European Centre for Medium-Range Weather Forecasts (ECMWF) products in forecasting droughts in East Africa, specifically in the Greater Horn of Africa region. The aim of this study is to illustrate the potential benefits of integrating ECMWF seasonal precipitation forecast and the SPI for more accurate and informative drought prediction in East Africa. Following by the consider paper, [12] is validating and verifying the accuracy and reliability of local knowledge (LTK) indicators used in drought forecasting and early warning in the Zimbabwe. After comparing the result of collected LTK forecast data , and the meteorological data on seasonal rainfall , it is obtained that the majority of LTK indicators used were accurate in forecasting weather and drought conditions compared to the observed data and it might be suitable to use the LTK indicators in formal forecasting systems. South Asia also is one of the most considered parts of earth that research and papers mostly concentrated on the countries there. For example,paper [13]is discussed developing an agricultural drought monitoring index(agCDI) for sri lanka. The purpose of this research is to demonstrate the

performance of integration of multiple input variables including remote sensing data and model-based agroclimatic parameters into a single index. This agCDI is useful to monitor and characterize the spatial and temporal patterns of drought in Sri Lanka. In addition, another paper like [14] is focused on developing Artificial Neural Network (ANN) models for daily rainfall and monthly meteorological drought forecasting in the province of Albay, the Philippines. The strategy of the paper [14] is to use historical rainfall data from nine stations spread across Albay in a period of time and also other parameters such as current-day rainfall, relative humidity, mean temperature, and sunshine duration. The paper highlights the importance of accurate weather prediction in water management, agricultural planning, and disaster mitigation in the region. As it can be considered in the previous papers, cutting-edge technologies like Artificial intelligence are beneficial in drought forecasting around the world and it becomes more popular nowadays. Assessing the performance of the drought forecasting model in the paper [15] is investigating a forecasting model for drought in the northeastern region of Thailand. This study utilizes data on soil moisture and the Normalized Difference Vegetation Index (NDVI) obtained from remote sensing satellite imagery and rainfall data and after gathering data, various types of forecasting models are applied to determine the most suitable model for forecasting soil moisture and NDVI as predictors. Similar to the previous researches [15], [16] also are studies to compare the effectiveness of five data-driven models for forecasting long-term drought conditions in the awash river basin of Ethiopia. This study focuses on forecasting the standardized precipitation index (SPI) which is a meteorological drought index based on precipitation data and after comparing the performance of different models by using various metrics it is found that a new approached which is named as WA-ANN that is a combination of wavelet transforms with ANN outperforms the other models. In addition, data-driven models with uncertainty analysis is recommended for more robust drought forecasting.

In Bangladesh also, the paper [17] aim is to develop an efficient model for forecasting monthly monsoon rainfall in different stations across Bangladesh. This study also used three artificial intelligence (AI) models including artificial neural network(ANN), adaptive neuro-fuzzy inference system (ANFIS) , and genetic algorithm (GA) and the findings suggested that ANFIS and GA models outperform the ANN model and the

statistical MR model in predicting monthly monsoon rainfall for the selected Bangladeshi stations. The study also presents a direction for future research, which includes modeling daily monsoon rainfall data to further benefit various sectors such as agriculture, transportation, sports, and tourism. Indonesia is a country that is located in South Asia and it is famous for its monsoon rainfall and just due to this fact,paper [18] aim is to investigate different data sources and machine learning methods for predicting drought in East Nusa Tenggara province, Indonesia. This study used the standardized precipitation index (SPI) derived from two datasets, namely the Tropical Rainfall Measuring Mission (TRMM) and Modern-Era Retrospective analysis for Research and Applications (MERRA-2), and also applied two machine learning methods including classification and regression tree (CART) and random forest (RF) to predict drought. After comparing the performance of these methods using different datasets and found that Random Forest outperforms CART in predicting drought with higher accuracy. Considering other continents like United States of America, some papers analyzed the drought situation of this part of the earth, as some example we can introduce paper [19] that its aim is to develop a long-term drought forecasting model for the conventional United States (CONUS) using Burg entropy spectral analysis with the maximal overlap discrete wavelet transform and Additionally, this model is focused on forecasting the monthly self-calibrated Palmer Drought Index (scPDSI). The paper also examines the applicability of the suggested forecast model to drought management planning and the ultimate purpose is to provide valuable insights and tools for proactive drought management to prevent socio-economic-environmental losses caused by increasing drought frequency amid climate change. The paper [20] also is another study example that is concentrated on the Nebraska region in the US to monitor drought by using data mining techniques to analyze drought-related data and find associations between drought occurrences and climatic parameters. It also highlighted the potential of data mining as a proactive tool for managing drought and improving the reliability of drought predictions. About the south-eastern Australia continents, paper [21] is designed and developed a monitoring system for agricultural areas in south-eastern Australia by utilizing various remotely-sensed drought factors. This study likewise seeks to compare the performance of three advanced machine learning methods including bias-corrected random forest, multi-layer perceptron

neural network, and support vector machine in reproducing a ground-based drought index called SPEI(Standardized Precipitation Evapotranspiration Index) using remotely-sensed data from satellites such as TRMM and MODIS that are also used in the paper [18] as models. As well as the subjects of the research in the papers [13], [22] are analyzing the relationship between soil moisture and drought in the Xiangjiang River basin to predict agriculture using the Precipitation-Evapotranspiration Index (SPEI).

2.4. Summary

The literature review that is presented in this chapter is showing the various types of research and studies that have been done related to drought prediction, monitoring, and analysis. In the beginning part of the review, it is highlighting the impacts of drought in different sectors such as the economy, socio-economy, and particularly in agriculture in different regions of Turkiye and then it explores the measures and solutions taken or implemented by the organizations and the government to manage and mitigate drought effects in the country. Subsequently, various types of papers and studies have been mentioned and analyzed in this literature review.

In Turkiye, researchers employed different drought indices, such as Standardized Precipitation Index (SPI) and Standardized Precipitation Evapotranspiration Index (SPEI) to predict drought conditions. Additionally, Machine learning algorithms such as ARIMA and SARIMA, deep learning, and neural networks, have been applied on the provided datasets for forecasting and assessing drought in specific regions and time periods. Other types of data like remote sensing data have also been utilized to analyze drought conditions.

Turning to international studies, the review highlights the studies that have been implemented in South Asia, the USA, and Africa. Moreover, Artificial Intelligence and computer science are the decisive tools used in predicting and analyzing drought patterns. The studies in foreign countries are concentrated to create and develop robust models by using a variety of data sources and machine learning algorithms. These studies highlight the importance of early warning systems and the effectiveness of data-driven models in predicting drought and also show the harms of drought.

Overall, the literature review demonstrates the growing interest in using technologies, like Artificial intelligence, remote sensing, and totally cutting-edge technologies for

drought management and prediction. In addition, it is noticed that while there exist several studies related with using AI in drought prediction but there is no study related with AI-driven drought management systems. Mostly studies focused on specific regions and did prediction for the selected area and also there isn't any studies related with an AI-driven system to monitor and analyze the agricultural and meteorological drought in the world so there is a gap that in this study is taken into account.

CHAPTER 3

BACKGROUND OF THE STUDY

3.1. Introduction

Drought is a period or a duration of time that a region or area faced abnormally low precipitation that can lead lack of water sources, and decreased soil moisture, also humans and creatures in the region will not have access to water [23].Also in this paper defined drought as "Drought is one of the most serious and extensive natural hazards in the world". As can be obtained from the definitions of droughts, these can have significant impacts on the ecosystem, agriculture, water supply, hydrology, and socio-economic systems. Additionally, the paper [24] discusses the complex and widespread impacts of drought on various socio-economic sectors due to the role of water in producing goods and services. Also, it highlights the considerable impacts of drought on various types of sectors such as crop productivity, fire hazards, energy production, recreation tourism, livestock and wildlife, land degradation, and more.

3.2 Drought Categorization

There are several different types of droughts, each characterized by the aspect of the environment that is most affected. Here are some of the main types of droughts [1]:

3.2.1 Meteorological Drought

This type of drought is related to the amount of precipitation in a specific region, by comparing the current amount of rainfall with the long-term average in a specific area it could be possible to detect drought in there. Also "Meteorological drought is a natural hazard characterized as below-average precipitation from expected or normal that can affect various economic sectors and lead to period with inadequate streamflow for established uses of the water" [1]. In addition, As an example of research for meteorological drought, paper [25] is focused to apply the Standardized Precipitation

Index (SPI) and Standardized Streamflow Index (SSI) to identify the occurrence and the intensity of droughts in Doce River Basin (DRB), south-eastern Brazil.

3.2.2 Hydrological Drought

Hydrological drought is a kind of drought that refers to a lack of water in rivers, lakes, reservoirs, and groundwater. It is directly related to meteorological drought that the lower amount of precipitation, the lower water levels in reservoirs. Returning to the definition of hydrological drought in paper [26] described this type of drought as "Hydrological drought refers to a lack of water in the hydrological system, manifesting itself in abnormally low streamflow in rivers and abnormally low levels in lakes, reservoirs, and groundwater."

3.2.3 Agricultural Drought

Agricultural drought, as much as it can determine from its name is related to the effects of drought on crops and vegetation. It occurs when there is insufficient soil moisture to support healthy plant growth and development, also "agricultural drought is considered to have set in when the soil moisture availability to plants has dropped to such a level that it adversely affects the crop yield and hence agricultural profitability"[27].Based on the provided information from study [28], more studies have been done by the paper [29] that investigate and analyzed the impact of agricultural drought on the Raichure District in Karnataka, India, using geospatial techniques, specifically remote sensing. This study utilizes geospatial techniques, particularly Remote sensing, to gather data on vegetation stress. Normalized Difference Vegetation Index (NDVI) and Land Surface Temperature (LST) are two key parameters that have been used as indicators of vegetation health and environmental conditions.

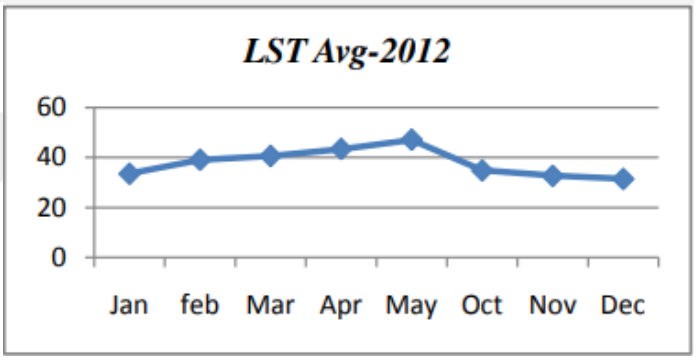


Figure 3.1 (a) Line Graph of Avg LST 2012 [28]

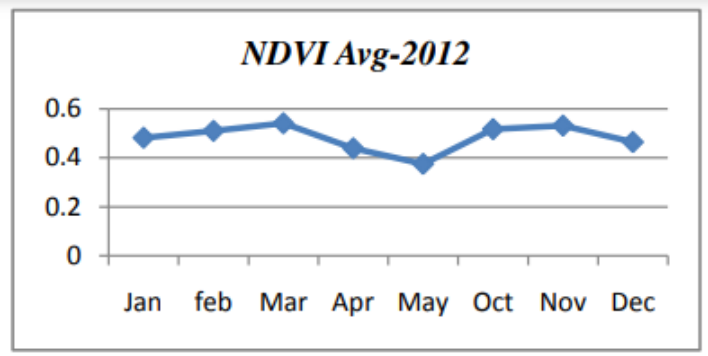


Figure 3.2 (b)Line Graph of Avg NDVI 2012[29]

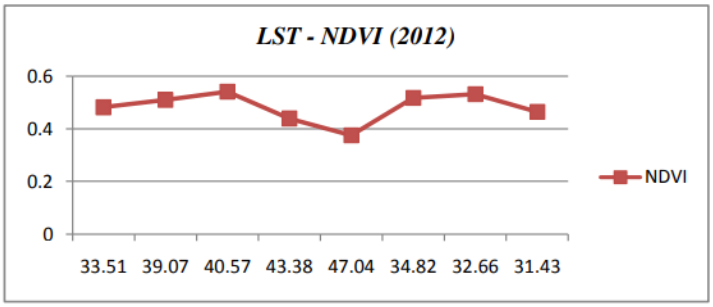


Figure 3.3 Line Graph showing NDVI changes w.r.t. LST for the year 2012[29]

3.2.4 Socio-Economic Drought

Socioeconomic drought is relevant to the impacts of water scarcity on humans and their lives. Also, it focuses on the availability and distribution of water resources to humans including drinking water, industry, and urban areas. Based on the paper [30] ,"Socioeconomic drought occurs when a water shortage is caused by an imbalance between the supply and demand of water resources in natural and human socioeconomic systems". This paper aims to develop a new indicator, the Standardized Supply and Demand Water Index (SSDWI), for assessing socioeconomic drought caused by an imbalance between water supply and demand in natural and human socioeconomic systems. SSDWI is used to analyze the trends of socioeconomic drought in a specific region like China.

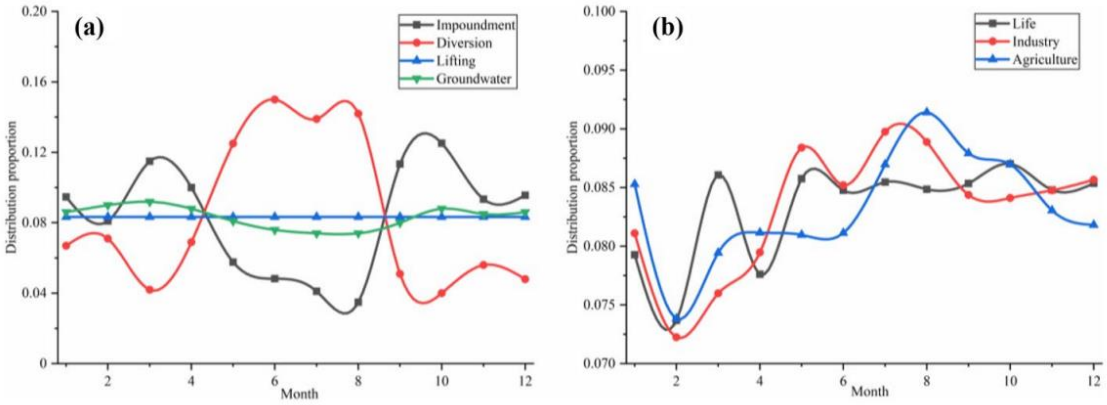


Figure 3.4 Distribution of (a) water supply and (b) water demand proportions throughout the year [30]

3.3 Drought Indices

Drought indices are tools to monitor, quantify and characterize the severity, duration, and spatial extent of drought conditions. With the help of these indices, it is easier to assess and monitor the drought situation of a specific region and also water resource management, agriculture, and disaster preparedness. There are several common drought indices each of which is used, based on the situation with its own methodology and purpose [31]. Based on the study we can introduce the indices of drought:

3.3.1 Palmer Drought Severity Index (PDSI)

The Palmer Drought Severity Index (PDSI) is a well-established and imperative index for drought assessment and monitor drought conditions[32].It is based on soil water balance equation and it is calculated by factors such as precipitation, temperature, and evapotranspiration to show a measure of drought severity[33].This drought index is utilized for drought monitoring and research in various regions. Also, studies have shown the importance of this drought index in providing comprehensive assessment and monitoring of drought conditions especially in agricultural regions and it has been also used in the development of predictive models and forecasting techniques for drought. Studies [33] and [34] PDSI provides numerical ranges that by observing these numbers it is possible to see the drought situation of a region. In the Table 3.1 below it can be seen the PDSI index range:

Table 3.1 PDSI Index Range [32]

Drought Index Range	Moisture Conditions
4.0 <=	Extremely wet
3.00 ~3.99	Very wet
2.00 ~ 2.99	Moderately wet
1.00 ~ 1.99	Slightly wet
0.50 ~ 0.99	Incipient wet spell
0.49 ~ -0.49	Near Normal
-0.50 ~ -0.99	Incipient drought
-1.00 ~ -1.99	Mild drought
-2.00 ~ -2.99	Moderate drought
-3.00 ~ -3.99	Severe drought
<= -4.0	Extreme Drought

3.3.2 Standardized Precipitation Index (SPI)

The Standardized Precipitation Index is a widely used drought index that shows the measure of precipitation variability and drought conditions. This drought index plays a fundamental role in drought climatology analysis and spatial standardization in addition to other drought indices such as the Palmer Drought Severity Index (PDSI)[36]. It is also been used in conjunction with other drought indices such as the Standardized Precipitation-Evapotranspiration Index (SPEI) to assess the drought precisely and accurately. The SPI, which stands for Standardized Precipitation Index focuses on the rainfall amount of a region. It is mainly used to analyze the drought and understand its impact on things like farming. Researchers mostly use it to monitor the effects of the drought on agricultural products and crops like fruits and vegetables. It also helps study how drought patterns change over time. The SPI also is calculated over specific periods, typically the previous. typically, the previous 3, 6, and 12 months, denoted as SPI-3, SPI-6, and SPI-12, respectively. These time scales make it possible to assess the drought conditions in various levels including SPI-3 for short-term related weather drought, SPI-6 for agricultural drought, and SPI-12 for hydrological drought.[37].The SPI-3 and SPI-6 show an effective quantify drought and

prediction crop yields and due to this fact they are useful in agricultural applications.[38] Moreover, SPI-12 give us seasonal estimates of rainfall, helping us understand medium-term trends in precipitation patterns[39].

In the Table 3.2 below it can be seen the SPI range:

Table 3.2 SPI Index range [40]

Drought Index Range	Moisture Conditions
2.00 <=	Extremely wet
1.50 ~1.99	Very wet
1.00 ~ 1.49	Moderately wet
-0.99 ~ 0.99	Near Normal
-1.49 ~ -1.00	Moderately dry
-1.99 ~ -1.50	Severe dry
<= -2.00	Extremely dry

3.3.3 Standardized Precipitation-Evapotranspiration Index (SPEI):

Similar to SPI, SPEI includes both precipitation and potential evapotranspiration (a measure of water loss from soil and plants due to evaporation and transpiration). This index can show a widespread relationship between drought and temperature change.

3.3.4 Crop Moisture Index (CMI):

CMI is a drought index to focus on the soil moisture conditions of crops to monitor the health of vegetation. It considers both precipitation and temperature.

3.3.5. Surface Water Supply Index (SWSI):

SWSI incorporates snowpack, streamflow, and reservoir storage to show the situation of the water supply conditions.

3.3.6. Vegetation Health Indices:

Indices like the Normalized Difference Vegetation Index (NDVI) or the Vegetation Condition Index (VCI) are using remote sensing data and also satellite data to monitor the health condition of vegetation in a region which can indicate the drought situation in there.

3.4 Time Series

"Time series data typically show patterns including trends, seasonal fluctuations, irregular cycles, and sporadic changes in level or variability in the domains of business, economics, environment, medical, and other sciences."[41].Time series data analysis shows patterns like trends, seasonal fluctuations, irregular cycles, and sporadic changes and by analyzing and showing these kinds of data makes it possible for governments, organizations, and researchers to have knowledge about what is happening in different majors [41].

3.4.1 Components of Time series

As it is discussed, a time series is a sequence of data points over successive time intervals. These data points can illustrate various types of patterns and trends or behaviors that can be divided into four main components including [41]:

- **Trend** This component represents the data points movement or direction in a period of time. It indicates whether the data points are generally increasing, decreasing, or staying constant over time. Furthermore, trends can be linear or nonlinear. By recognizing and identifying the trends it could be possible to gain insights into the overall behavior of the time series [41].
- **Seasonal Variations** "Seasonal variations are short-term, typically within less than a year, changes in time series."[41]. Seasonal variation is a repetitive pattern that can happen regularly within a time series. These intervals can be daily, monthly, quarterly, or any other consistent period. This component is mostly used to demonstrate the situation of weather, holidays, cultural events, and other recurring influences.
- **Cyclical Variations** Cyclical variations are long-term patterns that, unlike seasonal variations, it can happen in irregular intervals, and are influenced by economic, political, and societal factors. Business cycles, economic booms, and recessions are examples of cyclical variations. According to the paper [41], this component involves four phases including prosperity, recession, depression, and recovery.
- **Random or Irregular Variations** Random variations are unpredictable fluctuations in the data that do not follow any recognizable patterns. It shows the accidental noise and disturbances in the time series that cannot be attributed

to trends, seasonality, or cyclical influences. Unexpected events, measurement errors or external shocks are the factors that cause irregular variations [41].

"A time series plot is a graph that shows information gathered over time from any process" [41].

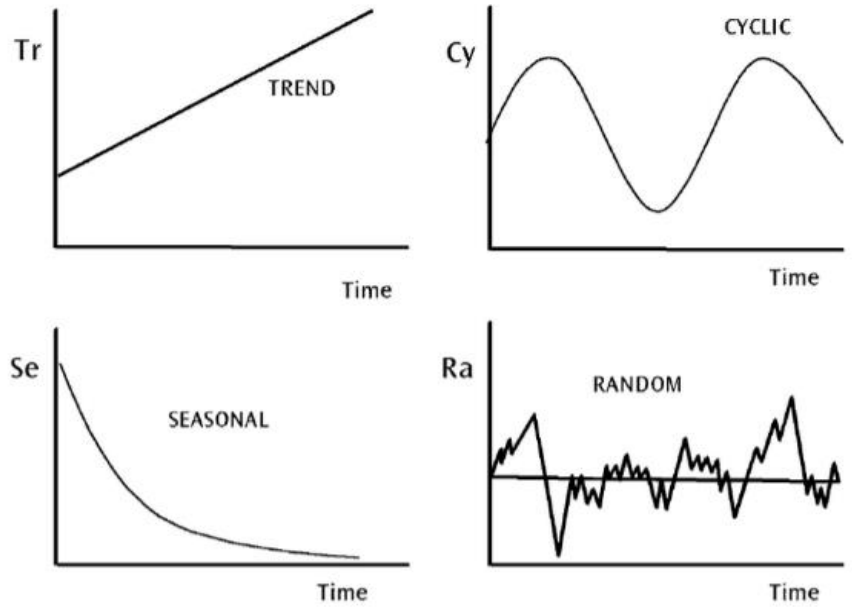


Figure 3.5 A time series plot with different components [41]

3.5 Time Series Models

Time series models recourse to mathematical or statistical frameworks used to illustrate and describe the patterns, trends, and relationships within time series data. "A time series model is said to be linear or non-linear function of past observations" Time series modelling and forecasting has fundamental importance to various practical domains. Thus, a lot of active research works is going on in this subject during several years. Many important models have been proposed in literature for improving the accuracy and effectiveness of time series forecasting. The aim of this dissertation work is to present a concise description of some popular time series forecasting models used in practice, with their salient features. In this thesis, we have described three important classes of time series models, viz. the stochastic, neural networks and SVM based models, together with their inherent forecasting strengths and weaknesses. We have also discussed about the basic issues related to time series modelling, such as stationarity, parsimony, overfitting, etc. Our discussion about different time series

models is supported by giving the experimental forecast results, performed on six real time series datasets. While fitting a model to a dataset, special care is taken to select the most parsimonious one. To evaluate forecast accuracy as well as to compare among different models fitted to a time series, we have used the five performance measures, viz. MSE, MAD, RMSE, MAPE and Theil's U-statistics. For each of the six datasets, we have shown the obtained forecast diagram which graphically depicts the closeness between the original and forecasted observations. To have authenticity as well as clarity in our discussion about time series modelling and forecasting, we have taken the help of various published research works from reputed journals and some standard [42].

3.5.1 Autoregressive Integrated Moving Average (ARIMA) This is a classical model that is the combination of autoregressive (AR) and moving average (MA) components to catch and capture the different aspects of the time series like trends and seasonality. Based on the paper [42], Arima model has three different parameters (p, d, q) that by which it shows different forecasting abilities. By selecting different numbers for the parameters of the ARIMA (p, q, d) model it is possible to gain different kinds of results for data points [42].

$$y_t = \mu + \sum_{i=1}^{p} \quad (\sigma y_{t-1}) + \sum_{i-1}^{q} \quad (\theta \varepsilon_t - 1) + \varepsilon_t \tag{1}$$

μ: *the mean value of the time series*;

p: the number of autoregressive lags;

σ: *autoregressive coefficient* (AR);

q: the number of lags of the moving average process;

θ: *moving average coefficients*(MA);

ε: *the white noise of time series data*;

- The “lag” refers to the number of units in the past that are used to predict the current value of this time series.
- **P:** p represents the number of autoregressive (AR) terms, which is the number of lagged observations included in the model. It is the order of the

autoregressive part of the model, indicating how many past observations are considered in predicting the current value.

- **D:** d represents the degree of differencing required to make the time series stationary. This is the number of differences needed to make the series stationary. Differencing involves computing the differences between consecutive observations to remove trends or seasonal patterns.
- **Q:** q represents the number of moving average (MA) terms, which is the number of lagged forecast errors included in the model. It is the order of the moving average part of the model, indicating how many past forecast errors are considered in predicting the current value.
- **Autoregressive (AR) Component:** The AR component of an ARIMA model captures the relationship between an observation and a certain number of lagged observations (previous values in the time series). The parameter p shows the number of lag observations included in the model.
- **Moving Average (MA) Component:** The MA component of an ARIMA model illustrates the error of the model as a combination of the error terms from previous time points. The parameter q represents the number of lagged forecast errors in the prediction equation.

3.5.2 Exponential Smoothing Models These models are using weighted averages to smooth out noise and emphasize recent data points. To describe the various types of exponential smoothing is discussed like: “Application of exponential smoothing to forecasting time series usually rely on three basic methods: simple exponential smoothing, trend corrected exponential smoothing and a seasonal variation thereof”[43].

3.5.3 Seasonal Decomposition of Time Series (STL) This method decomposes a time series into seasonal, trend and filter the remainder components of which by filtering techniques. Also, another definition described as “STL is a filtering procedure for

decomposing a seasonal time series into three components: trend, seasonal, and remainder. “[44].

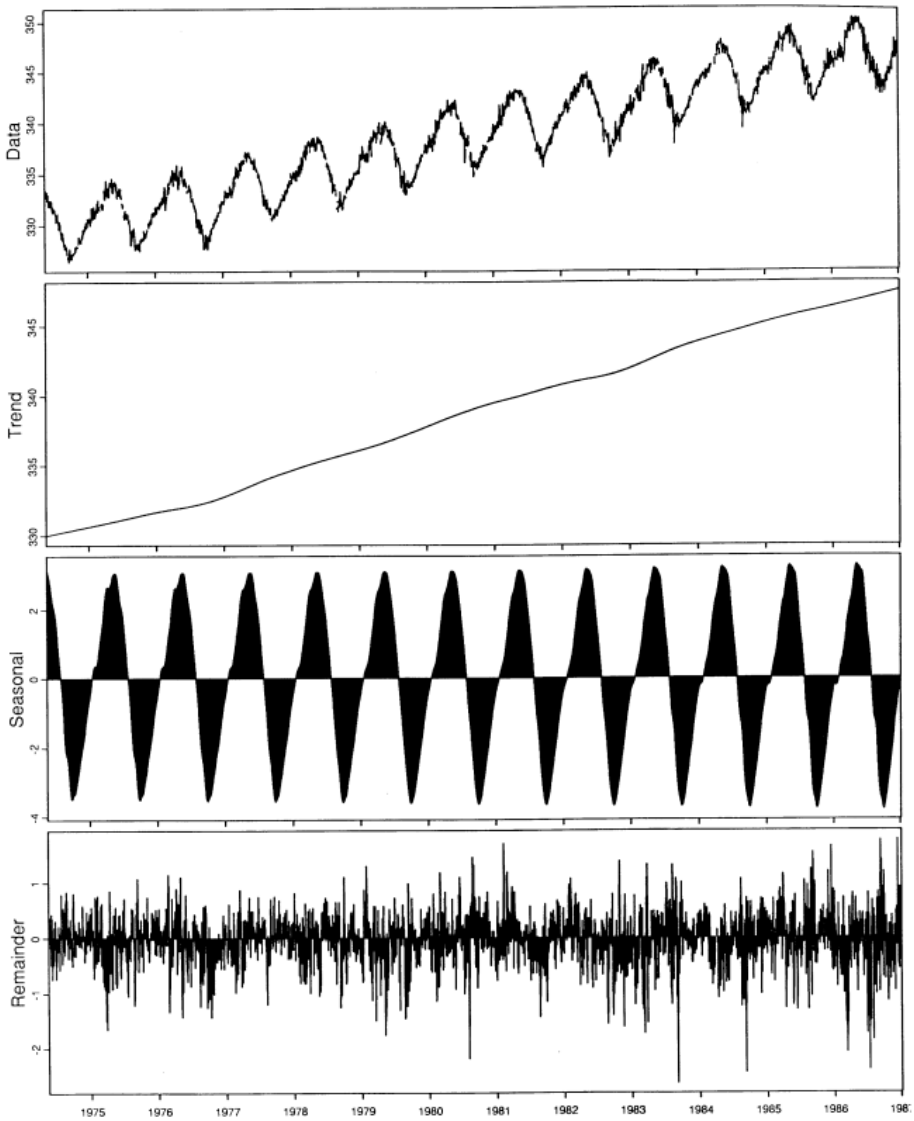


Figure 3.6 Sample STL graph for Decomposition Plot of Daily Carbon Dioxide Data [44]

3.5.4 SARIMA (Seasonal ARIMA) Model The Seasonal Autoregressive Integrated Moving Average (SARIMA) is a widely used time series analysis in various fields like health, climate, and economics. It is a seasonal extension of the ARIMA focusing on seasonal patterns. The SARIMA has been used in various fields like health, climate, and economics. By including seasonal autoregressive and moving average parameters in ARIMA, it turns into the SARIMA model and is effective in capturing and forecasting seasonal variations in data. It is widely used to make predictions in areas like disease outbreaks, climate changes, and economic trends[45]. The Table 3.3 shows the formula of the ARIMA and SARIMA that if non-linear parameters and seasonal order are added to ARIMA parameters it turns to SARIMA model.

Table 3.3 Parameters of SARIMA and ARIMA models

	Linear Parameters	Non-linear Parameters	Seasonal order
ARIMA	(p, q, d)	(P, Q, D)	S
SARIMA	(p, q, d)	(P, Q, D)	S

Moreover, SARIMA has been applied to climate modelling to forecast rainfall patterns, temperature variations, and their impact on drought.[46].

3.5.5 Prophet

This is an open-source tool that was developed by Facebook. It uses a combination of trend, seasonality, and holiday effects to model time series data [47]. In the paper [48], it described prophet as "PROPHET is a procedure for forecasting time series data that was created by Facebook's Core Data science team". Also, this study noticed that the prophet aims to forecast automatically and it gives more ease of use in tuning time series methods. Another advantage of this procedure is that it gives flexibility to our time series data that are changing suddenly or dramatically [48].

3.5.6 Examples of Time Series Usage in Different Sectors

The aim of the paper [48] is to utilize machine learning models for forecasting the trend of COVID-19 in Indonesia. This research focused on comparing the performance

of the two-time series models: Facebook's Prophet forecasting model and the Auto-Regressive Integrated Moving Average (ARIMA) forecasting model. These models are applied to the coronavirus dataset that includes the number of confirmed cases, deaths, and recovered number of COVID-19. The objective of this paper is to compare the performance of these two models and how well these models can predict the future progression of the disease. After applying the models [48], details the prediction of these two models and provides insights into the accuracy and performance of each model as the forecasting horizon progresses.

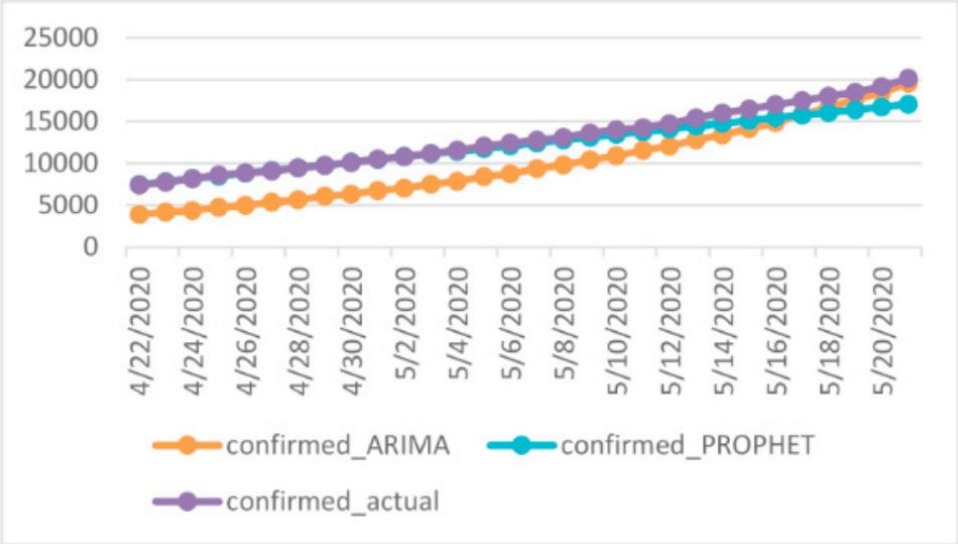


Figure 3.7 Visualization of Indonesia COVID-19 confirmed cases [48]

The paper [49] is focused on drought prediction by different deep-learning models for time series (TSF). It is also evaluating the performance of four deep learning models: RNN, LSTM, GRU, and Transformer, along with a baseline method, on the task of predicting air quality (PM2.5) using hourly measurement from the Beijing Air Quality Dataset. This research provided two main factors including the look-back window and the length of time that are impacting the prediction accuracy. The paper conducts experiments to assess the strengths and weaknesses of these deep learning models for TSF. The main objectives are to determine which model performs best, to identify the optimal look-back window size for accurate predictions, and to understand how the prediction accuracy changes with the length of time into the future being forecasted.

In this study, Arima, Sarima, and Prophet are chosen to analyse and predict the climate conditions and drought in Turkey. The use of ARIMA, SARIMA, and Prophet models

in weather prediction and condition assessment is rooted in their capabilities of effectively analyze and forecast time series data, making them valuable tools for weather forecasting and climate modelling. These models have been applied in different domains such as meteorology, environmental science, and climate research, due to their ability to effectively analyze and forecast time series data, making them valuable tools for weather forecasting and climate modelling. These models are widely used in meteorology, environmental science, and climate research because they can capture and detect complex patterns and trends in weather data.

The ARIMA and SARIMA models are commonly used in weather prediction. They can effectively capture the seasonality variations, trends, and irregular fluctuations in meteorological data. These models have been used to predict climatic parameters such as temperature, rainfall, wind speed, and air quality, drought risks and water management, in the studies accomplished by smith et al. [33], [50], [51], [52].

The Prophet model is popular in weather prediction because it can handle daily, weekly, and yearly patterns, including holidays. This possibility makes it well-suited for capturing and forecasting weather trends. The model has been used to predict climate condition factors including atmospheric temperature, air pollutant levels, health sector and rainfall and this can provide and offer insights for climate researchers, medical and environmental monitoring.[51], [53], [54].

model	index \ site	Beijing	Datong	Fengning	Huailai	Tangshan	Yuxian	Zhangjiakou	Zunhua
RF	MAPE	1.9765	2.3787	2.3224	2.2684	7.2247	12.1292	4.8646	1.6478
	SMAPE	1.4121	1.3475	1.4000	1.4132	1.4070	1.399	1.4273	1.3967
	MAE	**0.7518**	0.6948	0.7020	0.8010	0.6356	**0.6891**	0.7222	0.6861
	MSE	**0.8743**	0.6810	0.7667	0.9444	**0.7387**	0.8073	0.8002	**0.7076**
	R^2	0.4516	0.6210	0.6381	0.6201	**0.3584**	0.4839	0.6036	0.4569
LSTM	MAPE	3.0696	2.1095	**0.3630**	3.1725	12.1649	11.3911	4.8460	2.3103
	SMAPE	1.2544	1.1406	1.2767	**0.6491**	1.1997	1.1926	**1.0394**	1.1758
	MAE	1.1207	0.7811	0.7647	**0.2874**	0.7235	0.8318	**0.7205**	0.8944
	MSE	1.9037	0.9042	0.7776	1.2331	0.9602	1.0605	0.8724	1.4151
	R^2	0.2643	0.5016	0.6258	0.5600	0.2869	0.4350	0.5721	0.2059
WNN	MAPE	2.0856	**1.7870**	1.9017	3.1999	**4.1885**	6.8363	**3.4948**	**1.1536**
	SMAPE	**1.1019**	**1.0197**	**1.0303**	1.0986	**1.1023**	**1.1196**	1.1414	**1.1285**
	MAE	0.7949	**0.6605**	**0.6296**	0.8013	**0.6042**	0.7206	0.7341	0.6877
	MSE	1.0019	**0.6591**	**0.6261**	0.9262	0.7659	0.8968	**0.7965**	0.7677
	R^2	0.4420	**0.6459**	0.7085	0.6290	0.3577	0.4720	**0.6081**	0.4410
SVR	MAPE	**1.9655**	2.2615	2.4533	2.1187	7.2458	11.6990	4.3015	1.7521
	SMAPE	1.3897	1.3939	1.4130	**0.6491**	1.4297	1.4048	1.4547	1.4047
	MAE	0.7715	0.6865	0.6940	0.7739	0.7067	**0.7103**	0.7612	**0.6800**
	MSE	0.8874	0.7145	0.7144	**0.8584**	0.8727	0.8195	0.8747	0.7468
	R^2	**0.4534**	0.5973	**0.6595**	**0.6915**	0.2433	0.4676	0.5547	**0.4873**

Figure 3.8 The prediction results of different models on SPEI-3[49]

3.6. Evaluation Metrics

In this section performances metrics of study that are used in the CHAPETR 5 discussed.

3.6.1 Mean Squared Error (MSE)

Mean Squared Error (MSE) is a widely used metric in statistic and machine learning to measure the average of the squares of the errors or deviations. İt is used in various studies and fields such as economies, engineering and machine learning to calculate the quality of predictor [55]. It takes values between 0 and 1 that when the values is close to 0 it shows a better accuracy of the model.

$$MSE = \frac{1}{n}\sum_{i=0}^{n} \quad (\hat{g}_i - g_i)^2 \tag{2}$$

3.6.2 Root Mean Squared Error (RMSE)

Root Mean Squared Error (RMSE) is a measure of the differences between values predicted by a model or estimator and the actual values. It is the squared root of the average of the squared differences between predicted and actual values[55]. It takes values between 0 and 1 that when

the values is close to 0 it shows a better accuracy of the model.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} \quad (\hat{g}_i - g_i)^2}{N}} \tag{3}$$

3.6.3 Mean Absolute Error (MAE)

The Mean Absolute Error (MAE) is another evaluation metric for regression models. It measures the average absolute difference between the predicted values and actual values[55]. It takes values between 0 and 1 that when the values is close to 0 it shows a better accuracy of the model.

$$MAE = \frac{\sum_{i=1}^{N} \quad |\hat{g}_i - g_i|}{N} \tag{4}$$

CHAPTER 4

METHODOLOGY

4.1. Introduction

The objective of this study is to design and develop a new system for the Ministry and Forestry organization of Turkiye to monitor the health of the vegetation and agricultural products in different regions of Turkiye and also predict drought in specific regions of Turkey. After collaborating and meeting with the organization, it is found that some data are handled manually in this organization that should be upgraded by cutting-edge such as Artificial Intelligence to make the monitoring process faster and do prediction for precipitation and other factors that are responsible and play vital roles in drought happening. The designed system is a data-management system that is integrated with an AI-based system that makes the prediction about drought happening for coming next two years by predicting meteorological data and a PDSI and SPI calculation system that is calculating drought indexes like the Standardized Precipitation Index (SPI) and Palmer Drought Severity Index (PDSI) and with temperature average and precipitation and then returns the predicted visualized data for the data system that it makes it possible for the administrator and user of the system to see the result of the prediction by the second third systems. In the data-management system by inputting the specifications of the agricultural products including Province, District, Cultivation area, Development and Maintenance Status, and Observation Status, the system makes it easy to manage these data to the organization such as giving report based on each specification, manage the given data and analyze the data that are coming from the database by showing the analyzed data in various types of bar charts based on the different specifications that we are choosing by the form before the system start to analyze the data in database. All of the research steps of this study are presented in Figure 4.1 in the below.

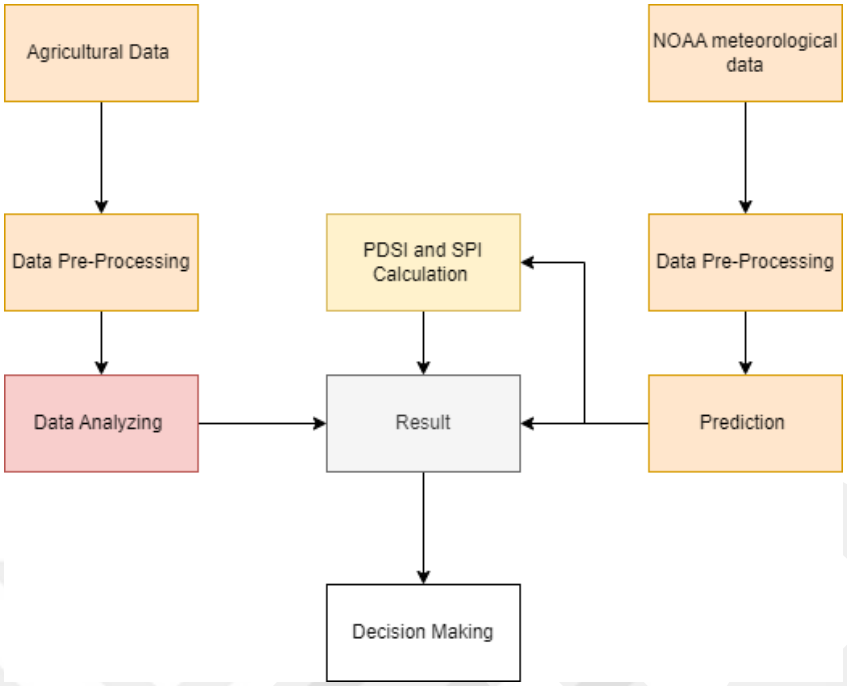


Figure 4.1 Steps of research methodology

4.2. Data Collection/Data Extraction

There two kinds of data that have been used in this study. The first one is the data provided by the ministry and forestry organization of the Turkiye to analyze by the designed system and the second one is meteorological data obtained from NOAA to do prediction for the next coming two years.

The data related to the status of the vegetation was provided and presented by the Forestry and Ministry organization of Turkey considering our request. The coming data formats were in the xlsx format and they were monthly reports that the agriculture engineers after a careful observation wrote the report for the organization in each month of the year. The columns of these files included the Cultivation area, Crop Yield Status of each agricultural products, Development and Maintenance status of each agricultural product and Observation status provided by the professions and in the rows

of the files, there are the names of the districts and the products that are yielding in each of these districts.

Meteorological data of this study was extracted from The Global Historical Climatology Network (GHCN) is a database maintained by the National Centers for Environmental Information (NCEI), which is a part of the National Oceanic and Atmospheric Administration (NOAA) in the United States. GHCN collects, quality-controls, and provides access to historical climate data from thousands of weather stations worldwide.

GHCN-Daily is a database that addresses the critical need for historical daily temperature, precipitation, and snow records over global land areas. This dataset includes Station, Name, Latitude, Longitude, Elevation, Date, Precipitation (PRCP), Snow depth(SNWD), Temperature Average (TAVG), Maximum Temperature (TMAX) and Minimum Temperature (TMIN) also it includes 188156 rows of data and 11 columns related with the daily meteorological data of all the stations in the Turkey.

4.3. Data Preprocessing

	Ekim Alanı (Da.)	Çıkış Durumu	Gelişme ve Bakım Durumu	Gözlem Durumu
AKÇAKALE İLÇESİ				
BUĞDAY (SULU)	146,286		4-5 yaprak oluşmuştur.	Aylık yağış yeterli olup yıllık yağış ortalamanın altındadır.
BUĞDAY (KURU)	185,000	(30.10.2021) Çıkış var	3-4 yaprak oluşmuştur.	Aylık yağış yeterli olup yıllık yağış ortalamanın altındadır.
ARPA (SULU)	2,905		4-5 yaprak oluşmuştur.	Aylık yağış yeterli olup yıllık yağış ortalamanın altındadır.
ARPA (KURU)	14,000	(30.10.2021) Çıkış var	3-4 yaprak oluşmuştur.	Aylık yağış yeterli olup yıllık yağış ortalamanın altındadır.
MISIR (II.ÜRÜN)		(30.06.2020)	HASAT EDİLDİ.	Aylık yağış yeterli olup yıllık yağış ortalamanın altındadır.
PAMUK				Aylık yağış yeterli olup yıllık yağış ortalamanın altındadır.
MERCİMEK	5,747	(30.10.2021) Çıkış var	3-4 yaprak oluşmuştur.	Aylık yağış yeterli olup yıllık yağış ortalamanın altındadır.

Figure 4.2 Unstructured data before pre-processing

The provided data by the organization was in an xlsx file and it was not able to handle the data as shown in Figure 4.2 consequently by using the Python programming language, the data changed to a CSV file and we cleaned the blank and blank columns, and rows to turn this unstructured and unusable data into usable and valuable data.

After pre-processing the data, it changed to Figure 4.3 below.

id	Urun	EkimAlani	CikisDurumu	GelismeBakim	GozlemDurumu	il	ilce	Ay	yil	
2	BUĞDAY (SULU)	146286		4-5 yaprak oluşmuştur	Aylık yağış yeterli olup yıllık yağış ortalamanın altındadır.	Şanlıurfa	AKÇAKALE	Ocak	2022	
3	BUĞDAY (KURU)	185000	Çıkış var	3-4 yaprak oluşmuştur	Aylık yağış yeterli olup yıllık yağış ortalamanın altındadır.	Şanlıurfa	AKÇAKALE	Ocak	2022	
4	ARPA (SULU)	2905		4-5 yaprak oluşmuştur	Aylık yağış yeterli olup yıllık yağış ortalamanın altındadır.	Şanlıurfa	AKÇAKALE	Ocak	2022	
5	ARPA (KURU)	14000	Çıkış var	3-4 yaprak oluşmuştur	Aylık yağış yeterli olup yıllık yağış ortalamanın altındadır.	Şanlıurfa	AKÇAKALE	Ocak	2022	
6	MISIR (II.ÜRÜN)			Hasat edildi	Aylık yağış yeterli olup yıllık yağış ortalamanın altındadır.	Şanlıurfa	AKÇAKALE	Ocak	2022	
7	PAMUK				Aylık yağış yeterli olup yıllık yağış ortalamanın altındadır.	Şanlıurfa	AKÇAKALE	Ocak	2022	
8	MERCİMEK	5747	Çıkış var	3-4 yaprak oluşmuştur	Aylık yağış yeterli olup yıllık yağış ortalamanın altındadır.	Şanlıurfa	AKÇAKALE	Ocak	2022	
9	BUĞDAY (SULU)	1192		Ekim bitti	Aylık yağış yeterli olup yıllık yağış ortalamanın altındadır.	Şanlıurfa	Birecik	Ocak	2022	

Figure 4.3 Structured Data After Pre-Processing

Returning to the data related to precipitation data gained from NOAA, the format of the date of file was changed to a readable format to be able to give this data for a time series model and algorithms. As the files included all of the stations in Turkey, by Python the file sliced to the city that our agricultural products report was about the city and the predictions are based on the city's meteorological data for the next two years.

4.4. Database Design

After pre-processing and making the data structured, the database was designed and prepared for the newly designed system. After the Preprocessing stage, it is necessary to make these data structured and import to the MySQL database. But there are some stages that also need to be implemented to handle this Big data more efficiently and easily. By creating a relational database, it is able to handle these data more easily and quickly. In the system database, we created several tables for each of the features such as Ay(Month) which shows the month of the obtained data from, cikisdurumu(crop yield status) which shows the cultivation and growth situation of the agricultural products, Gelismevebakim(Development and Maintenance) also shows the progress of growth of an agriculture product. In addition to these data, There are iller table and ilce table that show the provinces and districts of each province respectively and

moreover, there exists a urun table which shows the agricultural products of the turkey. So by reading the Excel files, the information is derived and listed.

In this section, the various tables of this system will be discussed.

- **Month(Ay) Table:** In this table, there are two columns ID and the name of the month in Turkish that is named as Ay_adi column which includes (1. Ocak, 2. Subat, 3. Mart, 4. Nisan, 5. Mayis, 6. Haziran, 7. Temmuz, 8. Augustos, 9. Eylul, 10. Ekim, 11. Kasim and 12. Aralik)
- **Crop Yield Status (Cikisdurumu) Table**: In this table, there are two columns ID and names of the status including (1. Yeterli Nem Yok, 2. Zayif, 3. iyi, 4. Cok iyi, 5. Cikis var)

- **Development and Maintenance (Gelismevebakim):** In this table, there are two columns ID and name of the categories of the development and maintenance including (1. Ekim Bitti 2. Cikis 3. 2-3 Yaprak olusmustur 4. Kardeslenme 5. Sapa Kalkma 6. Basaklanma 7. Ciceklenme 8. Olgunlasma 9. Hasat edildi 10. 4-5 yaprak olusmustur 11. 3-4 Yaprak Olusmustur)

- **Provinces(iller):** This table includes the ID column and name of the provinces of Turkey such as (1. Adana, 2. Adiyaman, 3. Afyonkarahisar, 4. Agri , 5. Aksaray , 6. Amasya, 7. Ankara, 8. Antalya, 9. Ardahan , 10. Artvin, etc.)

- **Districts(ilceler):** This table includes the ID column and, Province_id(il_id) which shows the id of the province and also the name of the district of that province)

- **Agricultural Products (Urun):** This table includes the ID and name of the agricultural products for example (BUGDAY (SULU), BUGDAY(KURU), ARPA(SULU), ARPA(KURU), MISIR, PAMUK, MERCIMEK, ANTEPFISTIGI, MERCIMEK and etc)

Main_Table: This table shown in the Figure includes the ID of each feature and also the year(Yil) and also Ekim_Alani column which shows the cultivation area.

id	Urun_id	Ekim_Alani	cikisdurumu_id	gelismebakim_id	Gozlem_Durumu	ilceler_id	il_id	Ay_id	yil
2	15	146286		10	Aylık yağış yeterli olup yıllık yağış ortalamanın ...	858	71	1	2022
3	16	185000	5	11	Aylık yağış yeterli olup yıllık yağış ortalamanın ...	858	71	1	2022
4	17	2905		10	Aylık yağış yeterli olup yıllık yağış ortalamanın ...	858	71	1	2022
5	18	14000	5	11	Aylık yağış yeterli olup yıllık yağış ortalamanın ...	858	71	1	2022
6	19			9	Aylık yağış yeterli olup yıllık yağış ortalamanın ...	858	71	1	2022
7	20				Aylık yağış yeterli olup yıllık yağış ortalamanın ...	858	71	1	2022
8	21	5747	5	11	Aylık yağış yeterli olup yıllık yağış ortalamanın ...	858	71	1	2022
9	15	1192		1	Aylık yağış yeterli olup yıllık yağış ortalamanın ...	859	71	1	2022
10	16	3400	5	1	Aylık yağış yeterli olup yıllık yağış ortalamanın ...	859	71	1	2022

Figure 4.4 The main table of the database

Consequently, a relational database was created in the database to make the data management process easier and quickly.

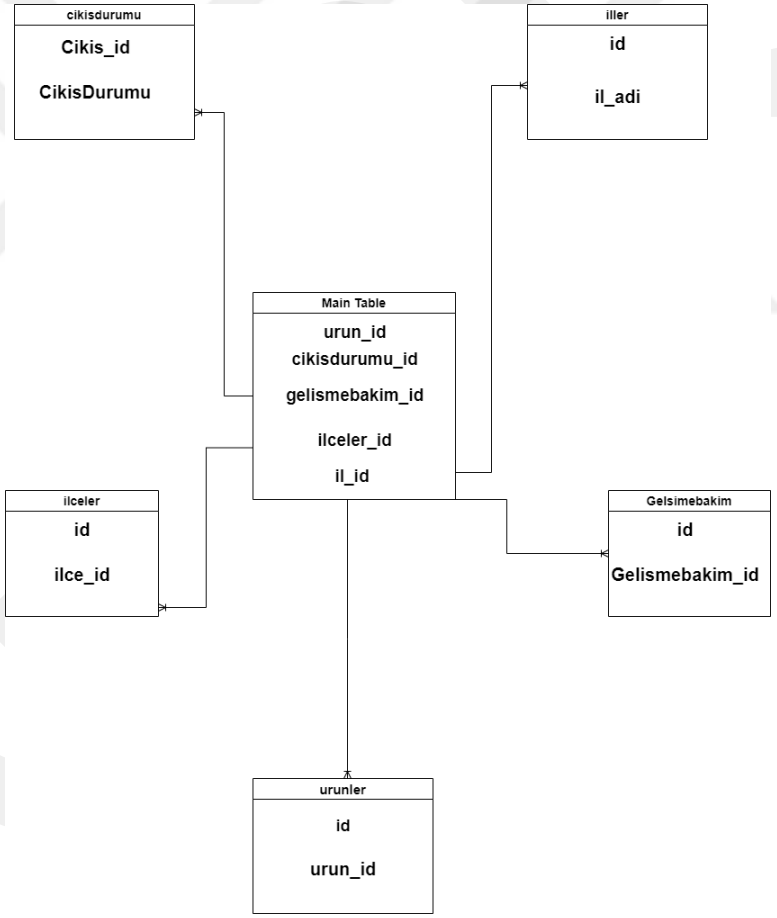


Figure 4.5 ER Diagram of the Database of the System

Also, in the Figure 4.6, it is possible to see the functionality of the system.

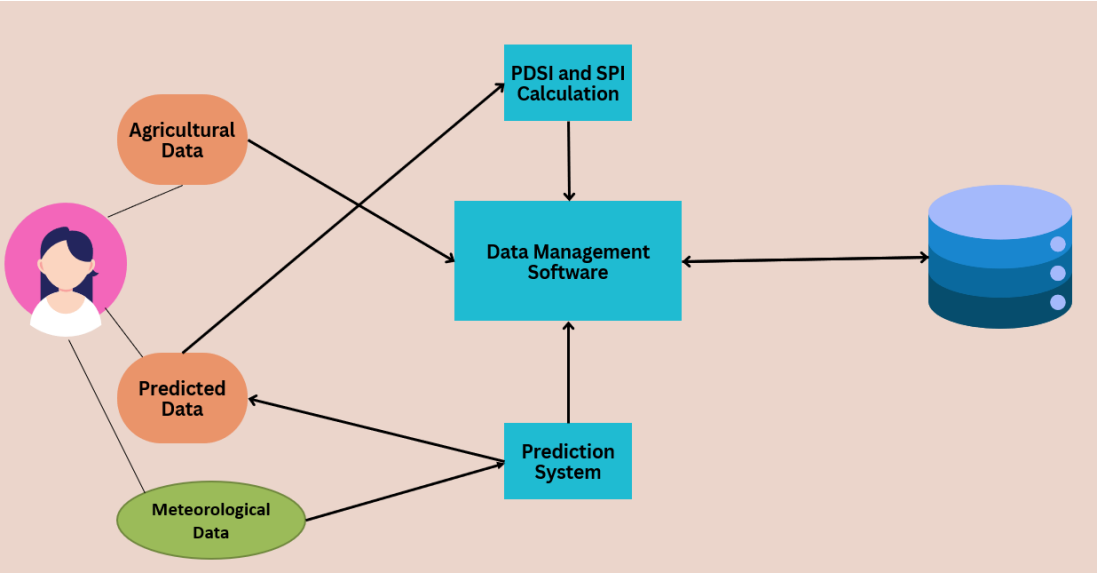


Figure 4.6 System Functionality Diagram for Agricultural Data Management and Prediction

4.5. Data Management Software

In the data management part of the system, it is possible to retrieve the data from the database and all of the data are shown in the table. Also, users can insert, update, and delete a record in the database. Figure 4.7 shows the Use Case diagram of the Data management software.

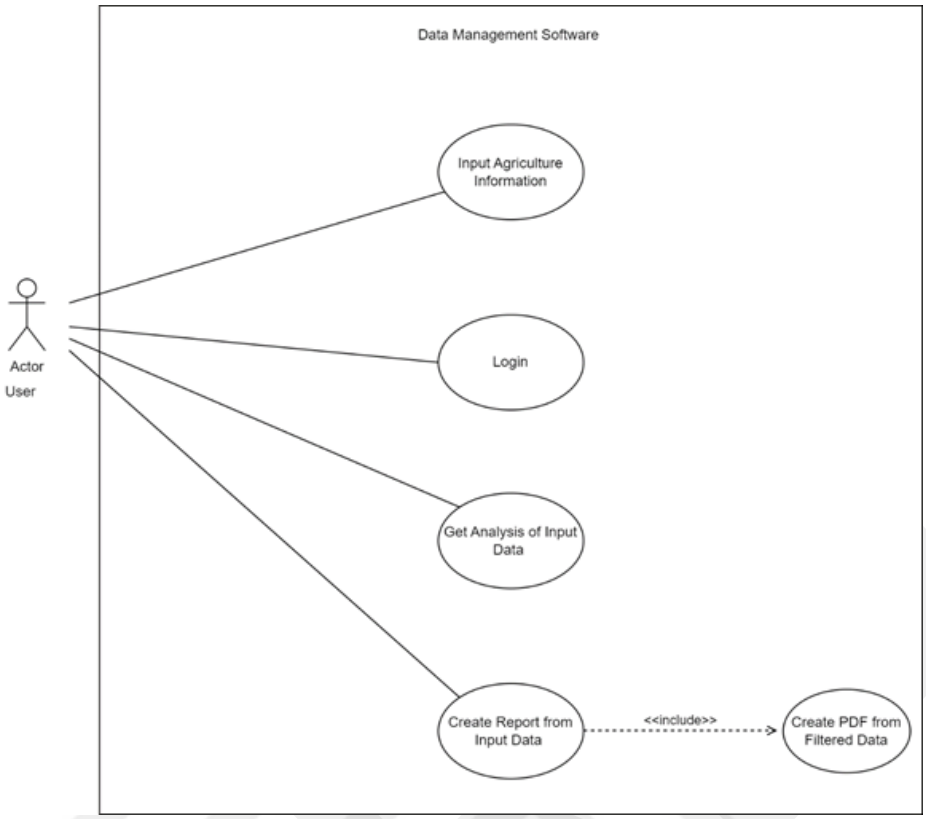


Figure 4.7 Use Case diagram of the data management system.

4.5.1 Report

In the system a feature or ability has been designed to manage or retrieve the data by giving groups of attributes and specifications and the system will retrieve the data related to them. As an example, it can be said that by inputting the product name and the year this option will retrieve a generated pdf file to the user of the system by specifying the data based on the provided year and product.

4.5.2 Data Analysis

This system also includes a data analyzing system that retrieves data from the database by function in the controller and then shows the retrieved data in different subjected charts based on the provided year and province of Turkey. In this section user by entering the year and name of the city system takes the user to the data-analyzing page

5 bar charts show the statistical analysis of the status of the agricultural products in the province. Bar charts give information about all of the agricultural products that are yielding, the crop-yielding status of the products, crops that are yielding in the districts, and the development and maintenance status of the agricultural products in the region that are going to help the organization to know the status of the area from an agricultural perspective.

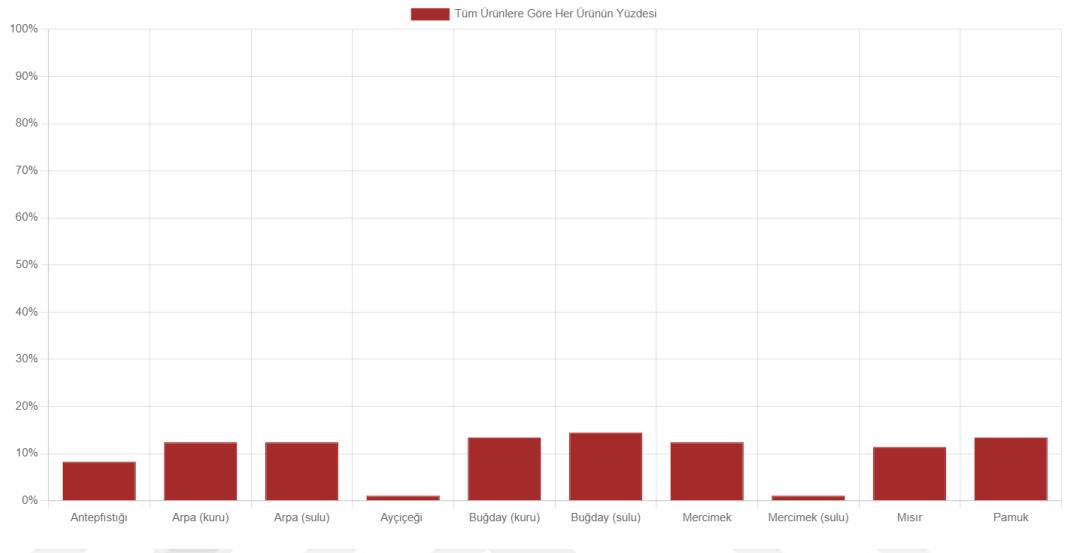


Figure 4.7 Bar chart shows each product percentage to all of the products

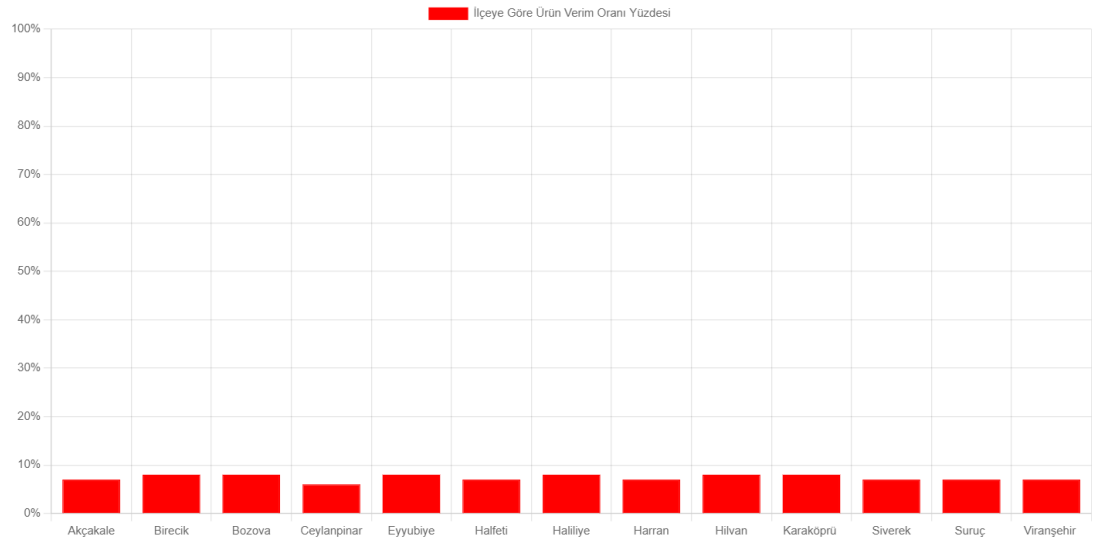


Figure 4.8. The percentage of product yield rate according to the district.

4.6. Selecting the Best Algorithm

Before using machine learning algorithm in the prediction system, both of which have been fitted to the 7 years meteorological data between 2013 and 2020 and after training the data, 2021 and 2022 predicted with the model. With the calculation of the MSE, RMSE and MAE metrics according to the test data (real data) for 2021-2022 and predicted data, best performed algorithm selected. (all of these processes have been implemented in python programming language)

In the fine-tuning section of the selecting the best ARIMA models according to the parameters (p, q, d), a function named as auto_arima () used to choose the best model. This function is from the pmdarima library of the python, which automates the process of building ARIMA models. This function systematically searches through combinations of ARIMA parameters to find the best model according to a specified criterion, typically the Akaike Information Criterion (AIC). It iterates through various combinations of ARIMA parameters (p, q, d) and optionally seasonal parameters (P,d,q,m) if seasonality is considered. In the parameters of the auto_arima function if the seasonality parameter is set as false the seasonality parameter is not going to be considered. Figures 4.9 and 4.10 are showing the functionality and usage of auto_arima function in python to choose the best model of ARIMA for predicting Temperature average and precipitation data.

```
from pmdarima import auto_arima
stepwise_fit1 = auto_arima(train['PRCP'], trace=True,seasonal=False, suppress_warnings=True)
stepwise_fit1.summary()
```

```
Performing stepwise search to minimize aic
 ARIMA(2,0,2)(0,0,0)[0]             : AIC=inf, Time=1.88 sec
 ARIMA(0,0,0)(0,0,0)[0]             : AIC=448.816, Time=0.05 sec
 ARIMA(1,0,0)(0,0,0)[0]             : AIC=414.216, Time=0.05 sec
 ARIMA(0,0,1)(0,0,0)[0]             : AIC=420.780, Time=0.06 sec
 ARIMA(2,0,0)(0,0,0)[0]             : AIC=415.097, Time=0.08 sec
 ARIMA(1,0,1)(0,0,0)[0]             : AIC=413.068, Time=0.12 sec
 ARIMA(2,0,1)(0,0,0)[0]             : AIC=417.317, Time=0.25 sec
 ARIMA(1,0,2)(0,0,0)[0]             : AIC=inf, Time=0.40 sec
 ARIMA(0,0,2)(0,0,0)[0]             : AIC=421.477, Time=0.08 sec
 ARIMA(1,0,1)(0,0,0)[0] intercept   : AIC=400.593, Time=0.14 sec
 ARIMA(0,0,1)(0,0,0)[0] intercept   : AIC=398.930, Time=0.10 sec
 ARIMA(0,0,0)(0,0,0)[0] intercept   : AIC=409.524, Time=0.03 sec
 ARIMA(0,0,2)(0,0,0)[0] intercept   : AIC=400.758, Time=0.12 sec
 ARIMA(1,0,0)(0,0,0)[0] intercept   : AIC=400.210, Time=0.08 sec
 ARIMA(1,0,2)(0,0,0)[0] intercept   : AIC=402.381, Time=0.35 sec

Best model:  ARIMA(0,0,1)(0,0,0)[0] intercept
Total fit time: 4.515 seconds
```

Figure 4.9. auto_arima function in python to find the best model of ARIMA for precipitation data

```
from pmdarima import auto_arima
stepwise_fit3 = auto_arima(train['TAVG'], trace=True,seasonal=False, suppress_warnings=True)
stepwise_fit3.summary()
```

```
Performing stepwise search to minimize aic
 ARIMA(2,0,2)(0,0,0)[0]             : AIC=518.334, Time=0.24 sec
 ARIMA(0,0,0)(0,0,0)[0]             : AIC=883.156, Time=0.06 sec
 ARIMA(1,0,0)(0,0,0)[0]             : AIC=602.431, Time=0.12 sec
 ARIMA(0,0,1)(0,0,0)[0]             : AIC=764.693, Time=0.24 sec
 ARIMA(1,0,2)(0,0,0)[0]             : AIC=536.696, Time=0.26 sec
 ARIMA(2,0,1)(0,0,0)[0]             : AIC=533.897, Time=0.15 sec
 ARIMA(3,0,2)(0,0,0)[0]             : AIC=inf, Time=1.09 sec
 ARIMA(2,0,3)(0,0,0)[0]             : AIC=517.058, Time=0.52 sec
 ARIMA(1,0,3)(0,0,0)[0]             : AIC=522.912, Time=0.23 sec
 ARIMA(3,0,3)(0,0,0)[0]             : AIC=518.978, Time=0.40 sec
 ARIMA(2,0,4)(0,0,0)[0]             : AIC=517.898, Time=0.57 sec
 ARIMA(1,0,4)(0,0,0)[0]             : AIC=517.959, Time=0.43 sec
 ARIMA(3,0,4)(0,0,0)[0]             : AIC=419.068, Time=0.84 sec
 ARIMA(4,0,4)(0,0,0)[0]             : AIC=inf, Time=1.23 sec
 ARIMA(3,0,5)(0,0,0)[0]             : AIC=inf, Time=1.00 sec
 ARIMA(2,0,5)(0,0,0)[0]             : AIC=inf, Time=0.73 sec
 ARIMA(4,0,3)(0,0,0)[0]             : AIC=inf, Time=0.87 sec
 ARIMA(4,0,5)(0,0,0)[0]             : AIC=inf, Time=0.97 sec
 ARIMA(3,0,4)(0,0,0)[0] intercept   : AIC=400.953, Time=0.96 sec
 ARIMA(2,0,4)(0,0,0)[0] intercept   : AIC=inf, Time=0.85 sec
 ARIMA(3,0,3)(0,0,0)[0] intercept   : AIC=inf, Time=0.89 sec
 ARIMA(4,0,4)(0,0,0)[0] intercept   : AIC=412.042, Time=0.97 sec
 ARIMA(3,0,5)(0,0,0)[0] intercept   : AIC=inf, Time=0.94 sec
 ARIMA(2,0,3)(0,0,0)[0] intercept   : AIC=inf, Time=0.88 sec
 ARIMA(2,0,5)(0,0,0)[0] intercept   : AIC=inf, Time=0.90 sec
 ARIMA(4,0,3)(0,0,0)[0] intercept   : AIC=423.474, Time=0.90 sec
 ARIMA(4,0,5)(0,0,0)[0] intercept   : AIC=447.359, Time=1.31 sec

Best model:  ARIMA(3,0,4)(0,0,0)[0] intercept
Total fit time: 18.572 seconds
```

Figure 4.10. auto_arima function in python to find the best model of ARIMA for Temperature average data

If the seasonality parameter in the function set as true it means that the model is SARIMA. Figures 4.11 and 4.12 are showing the functionality and usage of auto_arima function in python to choose the best model of SARIMA algorithm for predicting Temperature average and precipitation data.

```
from pmdarima import auto_arima
stepwise_fit2 = auto_arima(train['PRCP'], trace=True,seasonal=True, suppress_warnings=True, m=12)
stepwise_fit2.summary()
```

```
Performing stepwise search to minimize aic
 ARIMA(2,0,2)(1,0,1)[12] intercept   : AIC=396.849, Time=1.71 sec
 ARIMA(0,0,0)(0,0,0)[12] intercept   : AIC=409.524, Time=0.06 sec
 ARIMA(1,0,0)(1,0,0)[12] intercept   : AIC=398.457, Time=0.21 sec
 ARIMA(0,0,1)(0,0,1)[12] intercept   : AIC=397.644, Time=0.15 sec
 ARIMA(0,0,0)(0,0,0)[12]             : AIC=448.816, Time=0.03 sec
 ARIMA(2,0,2)(0,0,1)[12] intercept   : AIC=401.730, Time=0.51 sec
 ARIMA(2,0,2)(1,0,0)[12] intercept   : AIC=401.190, Time=0.52 sec
 ARIMA(2,0,2)(2,0,1)[12] intercept   : AIC=inf, Time=2.70 sec
 ARIMA(2,0,2)(1,0,2)[12] intercept   : AIC=398.573, Time=2.36 sec
 ARIMA(2,0,2)(0,0,0)[12] intercept   : AIC=403.450, Time=0.30 sec
 ARIMA(2,0,2)(0,0,2)[12] intercept   : AIC=403.561, Time=1.13 sec
 ARIMA(2,0,2)(2,0,0)[12] intercept   : AIC=402.089, Time=1.28 sec
 ARIMA(2,0,2)(2,0,2)[12] intercept   : AIC=inf, Time=2.94 sec
 ARIMA(1,0,2)(1,0,1)[12] intercept   : AIC=inf, Time=1.13 sec
 ARIMA(2,0,1)(1,0,1)[12] intercept   : AIC=inf, Time=1.20 sec
 ARIMA(3,0,2)(1,0,1)[12] intercept   : AIC=393.781, Time=1.43 sec
 ARIMA(3,0,2)(0,0,1)[12] intercept   : AIC=400.233, Time=0.75 sec
 ARIMA(3,0,2)(1,0,0)[12] intercept   : AIC=399.503, Time=0.82 sec
 ARIMA(3,0,2)(2,0,1)[12] intercept   : AIC=inf, Time=3.06 sec
 ARIMA(3,0,2)(1,0,2)[12] intercept   : AIC=394.576, Time=2.89 sec
 ARIMA(3,0,2)(0,0,0)[12] intercept   : AIC=402.403, Time=0.49 sec
 ARIMA(3,0,2)(0,0,2)[12] intercept   : AIC=401.935, Time=1.77 sec
 ARIMA(3,0,2)(2,0,0)[12] intercept   : AIC=400.186, Time=2.01 sec
 ARIMA(3,0,2)(2,0,2)[12] intercept   : AIC=inf, Time=3.53 sec
 ARIMA(3,0,1)(1,0,1)[12] intercept   : AIC=inf, Time=1.28 sec
 ARIMA(4,0,2)(1,0,1)[12] intercept   : AIC=inf, Time=1.56 sec
 ARIMA(3,0,3)(1,0,1)[12] intercept   : AIC=inf, Time=1.45 sec
 ARIMA(2,0,3)(1,0,1)[12] intercept   : AIC=inf, Time=1.42 sec
 ARIMA(4,0,1)(1,0,1)[12] intercept   : AIC=inf, Time=1.71 sec
 ARIMA(4,0,3)(1,0,1)[12] intercept   : AIC=inf, Time=1.33 sec
 ARIMA(3,0,2)(1,0,1)[12]             : AIC=inf, Time=1.85 sec

Best model:  ARIMA(3,0,2)(1,0,1)[12] intercept
Total fit time: 43.742 seconds
```

Figure 4.11. auto_arima function in python to find the best model of SARIMA for precipitation data

```
from pmdarima import auto_arima
stepwise_fit4 = auto_arima(train['TAVG'], trace=True,seasonal=True, suppress_warnings=True, m=12)
stepwise_fit4.summary()
```

```
Performing stepwise search to minimize aic
 ARIMA(2,0,2)(1,1,1)[12] intercept   : AIC=334.726, Time=1.26 sec
 ARIMA(0,0,0)(0,1,0)[12] intercept   : AIC=385.317, Time=0.13 sec
 ARIMA(1,0,0)(1,1,0)[12] intercept   : AIC=343.506, Time=0.36 sec
 ARIMA(0,0,1)(0,1,1)[12] intercept   : AIC=inf, Time=1.05 sec
 ARIMA(0,0,0)(0,1,0)[12]             : AIC=384.710, Time=0.07 sec
 ARIMA(2,0,2)(0,1,1)[12] intercept   : AIC=inf, Time=1.83 sec
 ARIMA(2,0,2)(1,1,0)[12] intercept   : AIC=344.806, Time=1.25 sec
 ARIMA(2,0,2)(2,1,1)[12] intercept   : AIC=335.586, Time=3.31 sec
 ARIMA(2,0,2)(1,1,2)[12] intercept   : AIC=inf, Time=4.88 sec
 ARIMA(2,0,2)(0,1,0)[12] intercept   : AIC=387.154, Time=0.37 sec
 ARIMA(2,0,2)(0,1,2)[12] intercept   : AIC=332.738, Time=2.20 sec
 ARIMA(1,0,2)(0,1,2)[12] intercept   : AIC=330.757, Time=1.52 sec
 ARIMA(1,0,2)(0,1,1)[12] intercept   : AIC=inf, Time=1.40 sec
 ARIMA(1,0,2)(1,1,2)[12] intercept   : AIC=inf, Time=4.04 sec
 ARIMA(1,0,2)(1,1,1)[12] intercept   : AIC=332.729, Time=0.79 sec
 ARIMA(0,0,2)(0,1,2)[12] intercept   : AIC=328.804, Time=1.09 sec
 ARIMA(0,0,2)(0,1,1)[12] intercept   : AIC=inf, Time=0.90 sec
 ARIMA(0,0,2)(1,1,2)[12] intercept   : AIC=inf, Time=4.01 sec
 ARIMA(0,0,2)(1,1,1)[12] intercept   : AIC=330.927, Time=0.56 sec
 ARIMA(0,0,1)(0,1,2)[12] intercept   : AIC=330.653, Time=1.15 sec
 ARIMA(0,0,3)(0,1,2)[12] intercept   : AIC=330.762, Time=1.20 sec
 ARIMA(1,0,1)(0,1,2)[12] intercept   : AIC=330.067, Time=1.34 sec
 ARIMA(1,0,3)(0,1,2)[12] intercept   : AIC=332.739, Time=3.04 sec
 ARIMA(0,0,2)(0,1,2)[12]             : AIC=334.878, Time=0.82 sec

Best model:  ARIMA(0,0,2)(0,1,2)[12] intercept
Total fit time: 38.655 seconds
```

Figure 4.12. auto_arima function in python to find the best model of SARIMA for Temperature average data

For prophet algorithm tuning, it is possible to tune the parameters such as changepoint, seasonality prior and seasonality mode. Figure 4.13 is showing the parameters tuning in prophet algorithm both for precipitation and temperature average data.

```
# Define the parameter grid for grid search
param_grid = {
    'changepoint_prior_scale': [0.001, 0.01, 0.1, 0.5],
    'seasonality_prior_scale': [0.1, 1.0, 10.0],
    'seasonality_mode': ['additive', 'multiplicative']
}

# Generate all combinations of parameters
all_params = [dict(zip(param_grid.keys(), v)) for v in itertools.product(*param_grid.values())]

# Initialize a DataFrame to store the results
results = pd.DataFrame()

# Loop through each combination of parameters
for params in all_params:
    # Initialize and fit the Prophet model with the current parameters
    model = Prophet(**params).fit(df)
```

Figure 4.13 Parameters adjustment in prophet algorithm for both precipitation and temperature average data

4.7. Prediction Software

The prediction system of the data management system is integrated with the system and it is like a sub-system to the system to provide the system with the prediction charts and statistical representations of precipitation and temperature averages of the specific area in Turkey. By the way, user of the system inputs the dataset of the area related to the precipitation and temperature average and the system returns the results to the Data management system. In the data management system, it is possible to find that charts by choosing the name of the city from the form page. This is the prediction system that is implemented in R programming language by using R shiny framework. Figure 4.14 is showing the form of the prediction system.

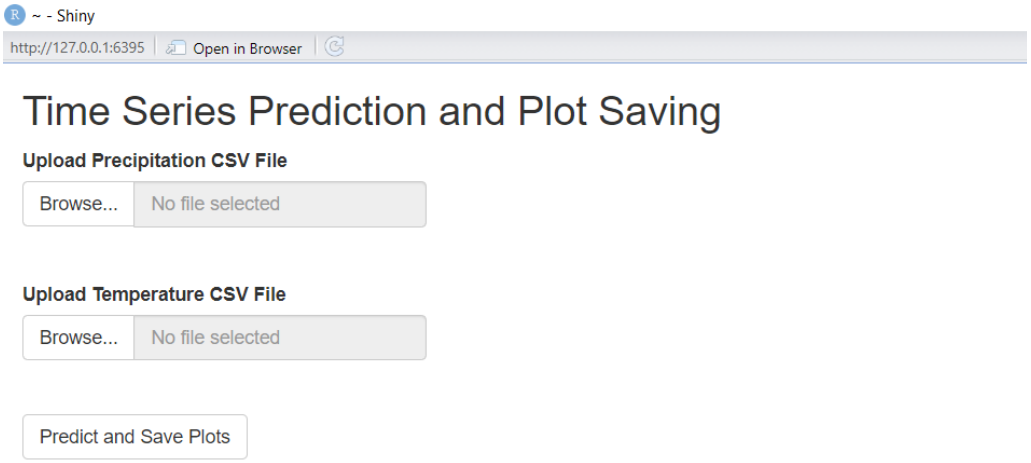


Figure 4.14 Form of the prediction system

4.7.1. In the first stage as it is shown in Figure 4.15 before implementing the algorithm, important libraries imported:

```
library(shiny)
library(forecast)
library(plotly)
library(htmlwidgets)
```

Figure 4.15 Importing required libraries

4.7.2. After Importing the required libraries, it is necessary to design our input form of the R shiny to make it possible for the user to input the required CSV files as shown in the Figure 4.16:

```
ui <- fluidPage(
  titlePanel("Time Series Prediction and Plot Saving"),
  fileInput("precipitation_csv_file", "Upload Precipitation CSV File"),
  fileInput("temperature_csv_file", "Upload Temperature CSV File"),
  actionButton("predict_button", "Predict and Save Plots"),
  # Plotly plots
  plotlyOutput("precipitation_plot"),
  plotlyOutput("temperature_plot")
)
```

Figure 4.16 creating web form in R

4.7.3. In this step it is necessary to check the exitance of the files in that the system returns a warning message to the user of the system as shown in the Figure 4.17:

```
server <- function(input, output, session) {
  precipitation_data <- NULL
  temperature_data <- NULL

  observeEvent(input$predict_button, {
    # Check if a CSV file for precipitation is uploaded
    if (is.null(input$precipitation_csv_file)) {
      return(NULL)
    }

    # Check if a CSV file for temperature is uploaded
    if (is.null(input$temperature_csv_file)) {
      return(NULL)
    }
```

Figure 4.17 Checking the existence of files

4.7.4. After reading the uploaded files shown in the Figure 4.18, system provides the

```
# Read the uploaded CSV files
precipitation_data <- read.csv(input$precipitation_csv_file$datapath)
temperature_data <- read.csv(input$temperature_csv_file$datapath)
```

Figure 4.18 Reading uploaded from the form

uploaded files to arima () function to do prediction. In the arima function when the seasonal attribute sets to true as shown in the Figure 4.19, this function will work as SARIMA algorithm and Fitting the models on the data, and creating future months for prediction if it is not set, it works ARIMA function.

```
# Fit ARIMA model for temperature
arima_temp <- auto.arima(temperature_data$TAVG)
# Forecast temperature for the next 24 months
forecast_temp_data <- forecast(arima_temp, h = 24)

# Fit SARIMA model for precipitation
sarima_precip <- auto.arima(precipitation_data$PRCP,seasonal = TRUE,period = 12)
# Forecast precipitation for the next 24 months
forecast_precip_data <- forecast(sarima_precip, h = 24)
```

Figure 4.19 ARIMA and SARIMA Functions in R

4.8. PDSI and SPI-3 Calculation Software

After predicting the next two years' data these are going to be saved in CSV data frame. For calculating PDSI and SPI-3, there is also a system designed. In this system, by uploading the generated predicted data and inputting the latitude of the region and year to start calculation with the form shown in the Figure 4.20, it calculates the PDSI and SPI-3 and present it to the user.

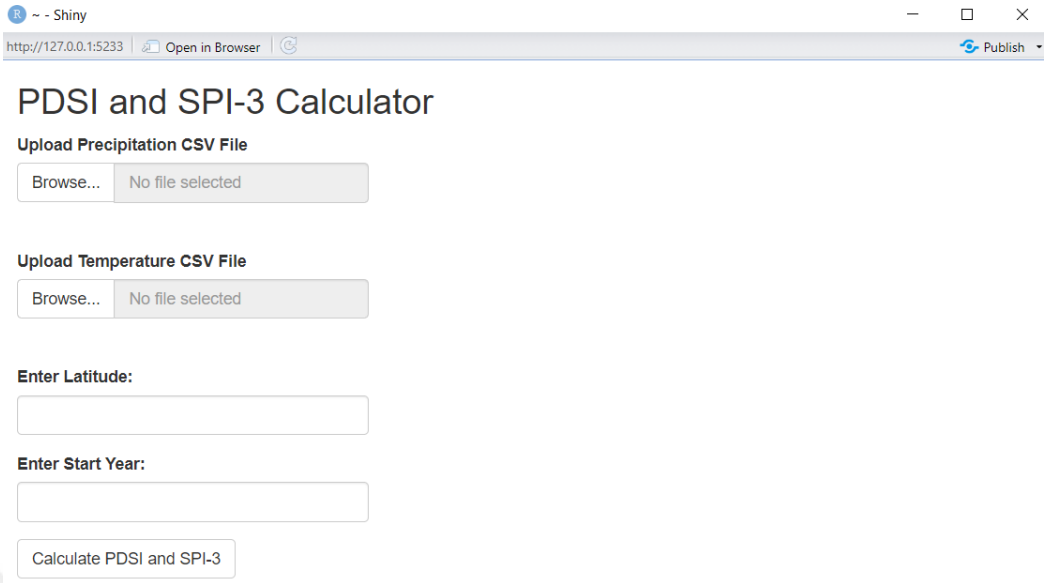


Figure 4.20 Form of the PDSI and SPI-3 software

For calculating PDSI firstly it is necessary to calculate the PE (potential evapotranspiration) attribute and then by giving the PE and precipitation data to pdsi () function as shown in the Figure 4.21 it calculates the pdsi values.

```
# Calculate PDSI
PE <- thornthwaite(temperature_data$x, lat = lat)
pdsi <- pdsi(precipitation_data$x, PE, start = start_year, sc = FALSE)$X
```

Figure 4.21 PDSI calculation code in R

To calculate SPI-3, the function presented in the Figure 4.22, uses P (monthly precipitation of the area) and scale metrics to calculate SPI for 3-month, for the next coming two years.

```
# Calculate SPI-3
precip_data <- precipitation_data$x
long_term_avg <- mean(precip_data)
precip_anomalies <- precip_data - long_term_avg
cumulative_anomalies <- numeric(length(precip_anomalies))
for (i in 3:length(precip_anomalies)) {
  cumulative_anomalies[i] <- sum(precip_anomalies[max(1, i - 2):i])
}
spi_3 <- cumulative_anomalies / sd(precip_anomalies)
```

Figure 4.22 SPI-3 calculation in various scales in R

In this study, PDSI and SPI were used to assess the drought conditions of the specific cities in the Turkiye for two reasons:

- ✓ The first is that by observing the provided documents by the forestry and ministry organization of the Turkiye it can be seen that the organization used PDSI and SPI to assess the drought situation in the Turkiye.
- ✓ The second one is related to the majority usage of these two drought indexes in several regions and research. It is clear that the PDSI, a commonly used drought index, has been employed to assess meteorological drought in various regions, including Turkey [56]. Both PDSI and SPI-3 have been used as indicators of developing drought conditions [57]. And are being used in monitoring drought and both of which play important roles in assessing drought in Turkey [58], [59].

CHAPTER 5

RESULTS

In this study, the results are based on two data analyses and prediction (Time Series data) by using different ML time series models such as ARIMA, SARIMA, and Prophet.

5.1. Research Question 1: Which ML algorithm is more suitable for making predictions for meteorological drought?

This study utilizes three different algorithms/models for prediction: 1. ARIMA, 2. SARIMA, and 3. Prophet. The research question 1 seeks to determine the best performance algorithm. In this study machine learning evaluation metrics such as MSE, RMSE, and MAE have been chosen to compare the performance of the models.

5.1.1 Precipitation and Temperature Average Prediction

After doing prediction by these three time series algorithms for both of the datasets the best algorithms were chosen to use in the system. In this study, error metrics have been considered to compare the algorithm's performances. Error metrics refer to how close the predicted and real data are. They are typically measured by comparing the predictions made by the model to the ground truth labels. These metrics are expressed as a percentage, where higher percentages indicate better performance. MSE(Mean Squared Error), RMSE(Root Mean Squared Error), and MAE(Mean Absolute Error) are the metrics that have been chosen to compare the accuracy of the algorithms used in the study. Tables 5.1 and 5.2 below show the MSE, RMSE and MAE metrics of the models. The data spanning from 2013 to 2020 was utilized for prediction purposes. Error metrics were computed using two years of test data (2021-2022). As it is shown in the Table 5.1 SARIMA algorithm outperformed the other algorithms in the case of

predicting precipitation. Conversely, Table 5.2 demonstrates that the ARIMA algorithm performed best among the algorithms in predicting average temperature.

Table 5.1 Comparing the performances of the models for precipitation prediction (2021-2022)

Model	**MSE**	**RMSE**	**MAE**
Arima	1.79	1.33	1.27
Sarima	1.52	1.23	1.027
Prophet	2.76	1.66	1.42

Table 5.2 Comparing the performances of the models for temperature average prediction (2021-2022)

Model	**MSE**	**RMSE**	**MAE**
Arima	4.24	2.05	1.72
Sarima	6.09	2.46	1.98
Prophet	8.67	2.94	2.26

Additionally, Figures 5.1 and 5.2 are showing each algorithm and its error ratings which is another illustration to show the results of the comparison between the algorithms.

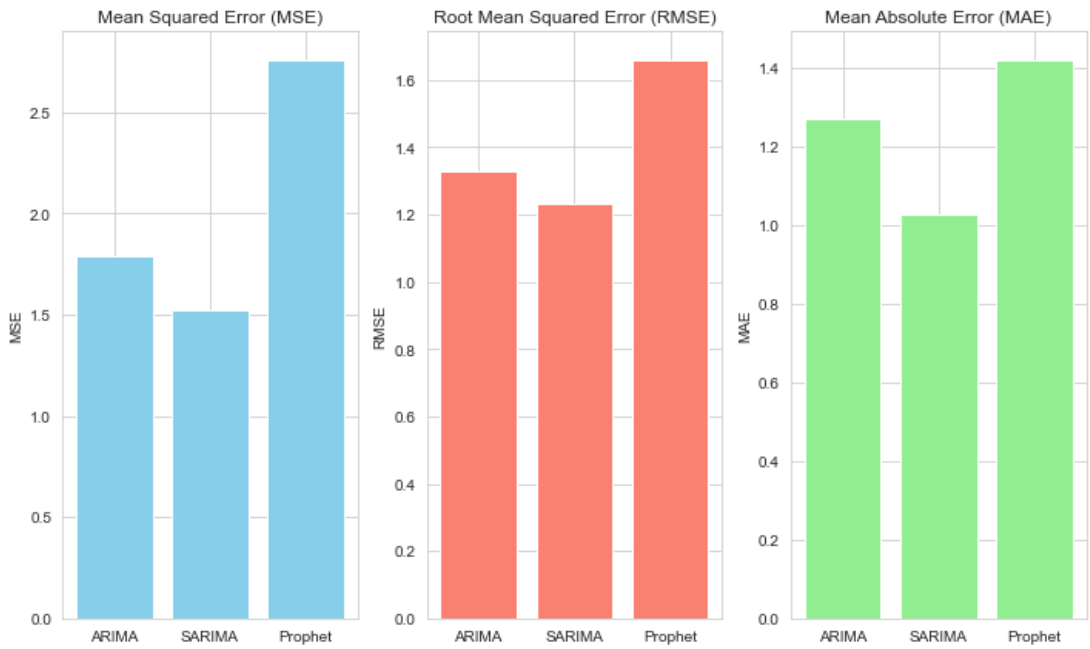


Figure 5.1 Error metrics for each Time Series algorithm prediction for precipitation data (2021-2022)

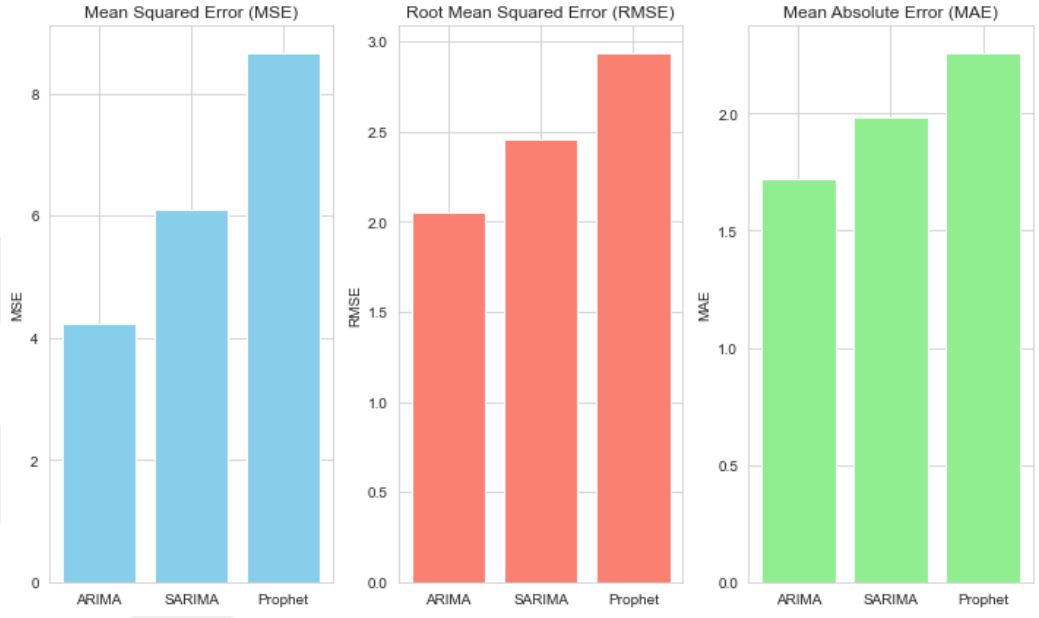


Figure 5.2 Error metrics for each Time Series algorithm prediction for temperature average data (2021-2022)

Figures 5.3 and 5.4 illustrate the comparison between actual and predicted values for precipitation (PRCP) and average temperature over the period from January 2021 to

January 2023. Figure 5.3 displays the actual and predicted precipitation data using the SARIMA algorithm, while Figure 5.4 shows the actual and predicted average temperature data using the ARIMA algorithm.

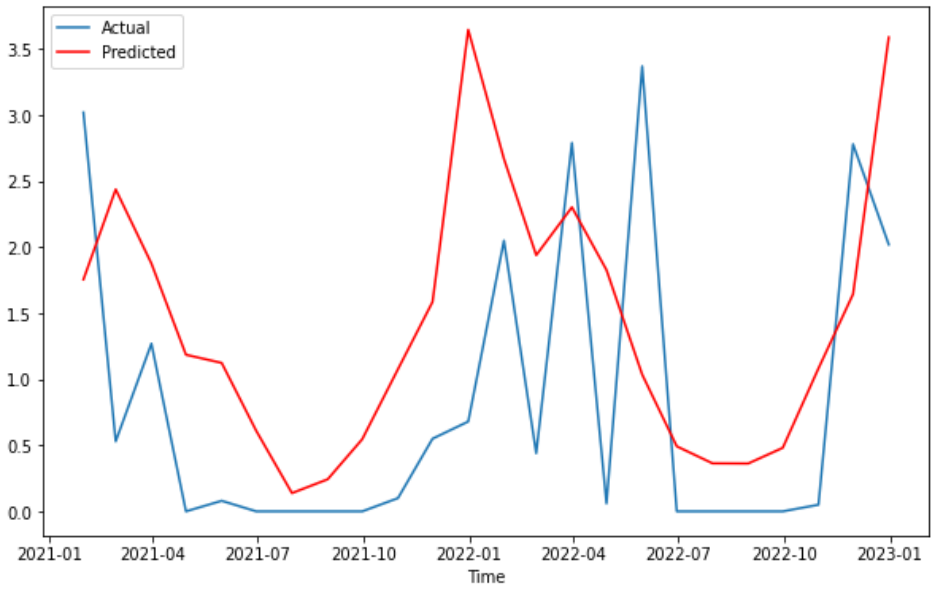


Figure 5.3 Actual data and ARIMA prediction for Precipitation(mm)

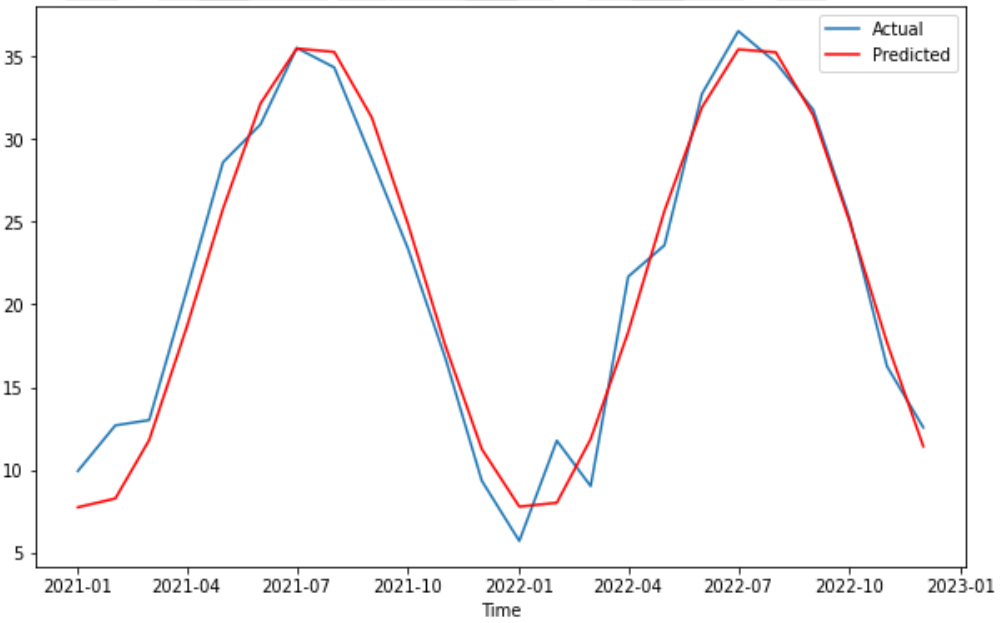


Figure 5.4 Actual data and ARIMA prediction for Temperature Average (Celsius)

5.2. Research Question 2: Is it possible to predict agricultural drought in different cities of Turkey using meteorological drought index?

To answer this question, the province of Sanliurfa was selected for analysis in the year 2022. Given the diverse range of agricultural products cultivated in each province, wheat was chosen as a representative example for examination, aligning with the study's focus on understanding the correlation and impacts of meteorological drought on agricultural drought. After forecasting the data by AI for two years, the predicted data was used in the PDSI and SPI calculation system to calculate the selected year's PDSI and SPI.

In the table below, we compared the expected weather conditions in different months of 2022 and the system's calculation for PDSI and SPI-3 of the area in each month with the reports of the engineer observations provided by the Ministry and Forestry Organization of Turkey. In the result section, we determine the accuracy of our system's predictions and calculations and its eligibility to provide this report to the organization. As seen from Table 5.3, 8 out of 12 months predicted accurately for SPI-3 index in 12 months in the 2022, while 7 months were accurately predicted for PDSI index. So, it can be said that our system's prediction is 66% correct for SPI-3 and 58.3% for PDSI.

Table 5.3 Weather Conditions, Prediction, and Observation Consistency Analysis (2022)

Month	**Calculated (PDSI)**	**Calculated (SPI-3)**	**Engineering Observation**	**Result consistency for SPI-3**	**Result consistency for PDSI**
January	Severe drought	Extremely Wet	Monthly rainfall is sufficient	SPI-3 and observation are consistent	Prediction and observation are not consistent
February	Severe drought	Severely dry	Monthly rainfall is sufficient	SPI-3 and observation are not consistent	PDSI and observation are not consistent

Table 5.3 (continued)

March	Severe drought	Severely dry	Monthly rainfall is sufficient	SPI-3 and observation are not consistent	PDSI and observation are not consistent
April	Extreme drought	Extreme dry	Monthly rainfall is below normal	SPI-3 and observation are consistent	PDSI and observation are consistent
May	Extreme drought	Extreme dry	Monthly rainfall is below normal	SPI-3 and observation are consistent	PDSI and observation are consistent
June	Extreme drought	Extreme dry	Insufficient rainfall/drought conditions	SPI-3 and observation are consistent	PDSI and observation are consistent
July	Severe drought	Extreme dry	Insufficient rainfall/ drought conditions	SPI-3 and observation are consistent	PDSI and observation are consistent
August	Moderate drought	Extreme dry	Insufficient rainfall/ drought conditions	SPI-3 and observation are consistent	PDSI and observation are consistent
September	Moderate Drought	Extreme dry	Monthly rainfall is below normal/ drought conditions	SPI-3 and observation are consistent	PDSI and observation are consistent
October	Extreme Drought	Extreme dry	Monthly rainfall is below normal	SPI-3 and observation are consistent	PDSI and observation are consistent
November	Extreme Drought	Extreme dry	Monthly rainfall is normal	SPI-3 and observation are not consistent	PDSI and observation are not consistent
December	Extreme Drought	Extreme dry	Monthly rainfall is normal	SPI-3 and observation are not consistent	PDSI and observation are not consistent

Comparing the Expected Weather Conditions: By comparing the expected weather conditions and PDSI and SPI-3 predictions, it can be determined if the agricultural

environment is risky for agricultural product or not in the coming 2 years. For example, in the table 5.3, from January to May in 2022 while wheat was expecting humid weather condition PDSI and SPI-3 indexes were showing dry weather condition for the region and this can be showing a risky condition for the area to cultivate wheat product.

For example, in the Tables 5.4 and 5.5, for each drought indexes like SPI-3 and PDSI has respective tables to compare and predict agricultural drought for 2024.

Table 5.4 wheat product drought index (PDSI) prediction for Sanliurfa (2024)

Month	Expected PDSI	Calculated (PDSI)	Result
January	Incipient Wet Spell	Near Normal	No Risk
February	Incipient Wet Spell	Incipient drought	Risky
March	Incipient Wet Spell	Mild drought	Risky
April	Slightly Wet	Severe drought	Risky
May	Slightly Wet	Extreme drought	Risky
June	Slightly Wet	Extreme drought	Risky
July	Incipient Wet Spell	Extreme drought	Risky
August	Near Normal	Severe drought	Risky
September	Near Normal	Severe drought	Risky
October	Near Normal	Severe drought	Risky
November	Near Normal	Moderate drought	Risky
December	Incipient Wet Spell	Moderate drought	Risky

Table 5.5 wheat product drought index (SPI-3) prediction for Sanliurfa (2024)

Month	Expected SPI-3	SPI-3	Result
January	Near Normal	Extremely Wet	Risky
February	Near Normal	Extremely Wet	Risky
March	Near Normal	Extremely Wet	Risky
April	Near Normal	Moderately Wet	No Risk
May	Near Normal	Near Normal	No Risk
June	Near Normal	Moderately dry	No Risk
July	Near Normal	Extremely dry	Risky
August	Near Normal	Extremely dry	Risky
September	Near Normal	Extremely dry	Risky
October	Near Normal	Extremely dry	Risky
November	Near Normal	Extremely dry	Risky
December	Near Normal	Severe dry	Risky

In table 5.4, the comparison is between PDSI drought index and expected weather conditions shows that over 91% the system indicates risky weather condition for wheat product in 2024 and 9% indicates no risk for agriculture. Additionally, for the SPI-3 drought index shown in the Table 5.5, which is comparing the expected weather conditions and SPI-3 range, SPI-3 indicates a 75% risky weather condition environment for cultivating agriculture product and 25% normal or not risky conditions.

CHAPTER 6

DISCUSSION AND CONCLUSSION

In this chapter, we explain the findings of this research and introduce the designed system, how it predicts and analyzes data obtained from various organizations and sources.

6.1. Agricultural Product Analysis

In this study, we used the data provided by the Ministry and Forestry Organization of the Turkey. These data were created and collected by the agricultural engineers observing the products cultivated on the farms. The data includes the agricultural product, month, year, crop yield status and development status and also the province, districts. By structuring and making these data usable and by inputting the data into the system's database, it can show the all of the product's status, amounts in bar charts to the user of the system, providing insights about the situation of the products in the selected month and region of the Turkey.

6.2. Precipitation

For predicting precipitation, we needed data and by searching for data in various types of sources and organization we found the data from GHCN, a database managed by the National Oceanic and Atmospheric Administration (NOAA). After receiving data, we used ML time series algorithms to make predictions since as our data was time series type. We used ARIMA (Auto Regressive Integrated Moving Average), SARIMA (Seasonal Autoregressive Integrated Moving Average) and prophet for examining our model to find the best fitted algorithm. There are different metric results for each algorithm since we calculate the error rate by RMSE, MSE and MAE. Therefore, after calculating the error rate, it showed that the SARIMA algorithm

resulted with the lowest error rate RMSE:1.52, MSE: 1.23 and MAE:1.027 is the best algorithm choose to do prediction precipitation in the system.

In similar studies for predicting precipitation with time series algorithms such as ARIMA, SARIMA, and Prophet, several references provide valuable insights. In the similar studies [46], [60], [61], [62] ARIMA and SARIMA models have been extensively used in forecasting precipitation patterns. Additionally, for instance, The SARIMA model has been used to predict rainfall in specific such as Embu County in Umiam, Meghalaya in India [46]. Also, in the study [63], by comparing the ARIMA and SARIMA in forecasting rainfall, SARIMA showed superior in modeling and forecasting rainfall.

6.3. Temperature Average

In predicting the temperature average for the regions, we used data provided by the NOAA organization. The data includes precipitation and daily temperature average information of the region. After receiving the data, we again used ML time series algorithms ARIMA, SARIMA, and Prophet to choose the best model for making predictions. In this case, we also considered MSE, RMSE, and MAE as metrics to compare the performances of algorithms. The calculation showed that ARIMA, with a low error rate (RMSE: 2.05, MSE: 4.24, and MAE: 1.72), is the best algorithm for predicting the temperature average.

In other studies, about forecasting average temperatures, like in study [64], ARIMA was employed in forecasting monthly temperature and humidity in the Taihu Lake region of China, emphasizing its utility in hydrological studies. Furthermore, paper [65] suggested that ARIMA, coupled with the Standardized Temperature Index (STI), presented considerable results in forecasting temperature, showing the potential of ARIMA in climate change modeling. Additionally, studies like [66], [67] applied the ARIMA model in forecasting pipe surface temperature and soil temperature, respectively, highlighting its application in predicting temperature-related variables.

6.4. Predicting PDSI and SPI

The aim of this study is to predict drought indexes for the coming years to show the organization that in which date which regions of the Turkey are at risk of drought. By predicting the average temperature and precipitation for the coming years, the system calculates the PDSI and SPI for the future years and this can show the drought status of the Turkey.

6.5. Predicting Agricultural Drought by PDSI and SPI

We compared the data provided by the Ministry and Forestry Organization of the Turkey, collected and reported by the agricultural engineers, with the calculated PDSI and SPI for a specific region. By assessing the drought conditions for the region and the weather conditions necessary for the agricultural products and cultivation, we determine how appropriate the region is for the products. This innovative idea can help the organization to mitigate the harmful damages that drought can create in the area.

6.6. Conclusion

The main aim of this study is to design an AI-based drought assessment and prediction system for the Ministry of Forestry in Turkey. The data management software initially manages the data provided by farmers and allows users to filter it based on specified criteria. Additionally, it analyzes user-provided reports and generates charts depicting various agricultural metrics, such as crop yield status by province, percentage of specific products in agricultural output, and identification of districts with the highest product yields.

Secondly, the prediction software, incorporates a forecasting component utilizing the SARIMA and ARIMA algorithms respectively, which have demonstrated superior performance compared to Prophet algorithm. Also, it forecasts Monthly precipitation and average temperatures based on user-input time series data. The generated prediction charts are visible to all users within the management system. Furthermore, PDSI-SPI calculation software calculates the Palmer Drought Severity Index (PDSI) and Standardized Precipitation Index (SPI) from the predicted data for the next two years, aiding in drought assessment and prediction efforts. Consequently, by these obtained data, it is possible to the user to have knowledge about the Meteorological

drought situation in the coming years for the selected region. By comparing a specific product's expected weather condition and the weather condition that will happen, it is also possible to predict agricultural drought in the specific region for coming years. To predict drought indices such as Palmer Drought Index (PDSI) and the standardized Precipitation Index (SPI), various models including ARIMA, SARIMA and Prophet have been employed in different studies. In the study [68], Applied ARIMA and SARIMA models to predict drought by using SPI drought index in Kansabati River of the India and also the research [69] used ARIMA and SARIMA models to forecast drought events in Afghanistan based on the SPI. Also in the paper [70], İt is considered that Prophet model is a successful model for researchers to use in predicting drought indices. From the perspective of developing AI-driven systems a comprehensive approach integrating advanced technologies and scientific methodologies is essential. Studies such as [3], [71], emphasize the importance of accurate drought forecasting for proactive drought management plans. Integrating machine learning models like those used for predicting soil moisture droughts such as in the study [72], can provide noticeable insights into drought prediction mechanisms. Therefore, our study is illustrating a new developed AI-driven softwares that is more innovative than previous studies mentioned. In this study we are predicting PDSI and SPI-3 that are critical drought indexes in showing drought conditions for a region , also by integrating the reports provided by the engineers in the farms and meteorological predictions, there is a possibility to show the percentage of drought happening in a province and how much it is risky.

6.7. Limitations

This study faced some limitations during the implementations or maybe can be faced in the future studies.

Some potential limitations of this study include:

6.7.1 Data Quality and Availability

The effectiveness of AI-driven systems is based on the quality of the historical data. In this study, the data received from the Ministry and Forestry Organization of Turkey was unstructured and difficult to manage. Due to this fact, preprocessing methods were applied to the data to make it structured. On the other hand, due to the lack of meteorological data in the organization, the data was obtained from NOAA (National Oceanic and Atmospheric Administration), though the received data included several details that also required preprocessing methods to be applied to the data again.

6.8. Future Works

For the future studies related with development of AI-powered systems to predict and assess the drought, in addition to the methods and indices have been used in this study, some other methods, models and indices should be used to predict drought, For instance, in addition to the drought indices such as SPI and PDSI, used in this study,[73] SPEI used in to predict drought in the Ankara, Turkey region. Also [73] predicted the drought for the next 24 years. Therefore, the next study will consider SPEI drought index additionally and extend the prediction duration up to 24 years.

In research [56], advanced fuzzy logic models called co-active neuro fuzzy inference system (CANSIF) is introduced to predict SPI drought indexes and it demonstrates the efficacy of AI in predicting meteorological drought compared to other models. This can be another study case for the next studies to use fuzzy logic models in AI-powered drought prediction systems to enhance the drought forecasting. Moreover, the work of [74] On drought prediction using ANN (Artificial Neural Networks) and support vector regression are the examples of the other AI methodologies than can use in predicting drought conditions. Holt-winter algorithm can be another case to use in this study. The Holt-Winters method, known for its effectiveness in forecasting time series data with trend and seasonality, can be highly beneficial for drought prediction [43].

Therefore, the next phase research can focus on using other algorithms in the system to improve the accuracy of the system in predicting precipitation and temperature average.

REFERENCES

[1] "Definition of Drought | Did You Know? | National Centers for Environmental Information (NCEI)." Accessed: May 10, 2024. [Online]. Available: https://www.ncei.noaa.gov/access/monitoring/dyk/drought-definition

[2] Ö. F. Durdu, "Application of linear stochastic models for drought forecasting in the Büyük Menderes river basin, western Turkey," *Stoch Environ Res Risk Assess*, vol. 24, no. 8, pp. 1145–1162, Dec. 2010.

[3] L. Zhang, J. Xiao, Y. Zhou, Y. Zheng, J. Li, and H. Xiao, "Drought events and their effects on vegetation productivity in China," *Ecosphere*, vol. 7, no. 12, p. e01591, Dec. 2016.

[4] O. Cetin and Y. Kayam, "The Impacts of Drought and Mitigation Strategies in Turkey", Accessed: May 10, 2024. [Online]. Available: https://www.academia.edu/23517953/The_Impacts_of_Drought_and_Mitigation_Strategies_in_Turkey

[5] E. Topçu, "Drought Monitoring Using MOWCATL Data Mining Algorithm in Aras Basin, Turkey," *Earth sci. res. j.*, vol. 26, no. 2, pp. 183–196, Sep. 2022.

[6] E. Eris, Y. Cavus, H. Aksoy, H. I. Burgan, H. Aksu, and H. Boyacioglu, "Spatiotemporal analysis of meteorological drought over Kucuk Menderes River Basin in the Aegean Region of Turkey," *Theor Appl Climatol*, vol. 142, no. 3, pp. 1515–1530, Nov. 2020.

[7] M. Dikici, "Drought Analysis for the Seyhan Basin with Vegetation Indices and Comparison with Meteorological Different Indices," *Sustainability*, vol. 14, no. 8, p. 4464, Apr. 2022.

[8] O. M. KatiPoğlu, "Monthly streamflow prediction in Amasya, Türkiye, using an integrated approach of a feedforward backpropagation neural network and discrete wavelet transform," *Model. Earth Syst. Environ.*, vol. 9, no. 2, pp. 2463–2475, Jun. 2023.

[9] S. Yaltı and H. Aksu, "Drought Analysis of Iğdır Turkey," *Turkish Journal of Agriculture - Food Science and Technology*, vol. 7, no. 12, Art. no. 12, Dec. 2019.

[10] G. Naumann, P. Barbosa, L. Garrote, A. Iglesias, and J. Vogt, "Exploring drought vulnerability in Africa: an indicator based analysis to be used in early warning systems," *Hydrol. Earth Syst. Sci.*, vol. 18, no. 5, pp. 1591–1604, May 2014.

[11] E. Mwangi, F. Wetterhall, E. Dutra, F. Di Giuseppe, and F. Pappenberger, "Forecasting droughts in East Africa," *Hydrol. Earth Syst. Sci.*, vol. 18, no. 2, pp. 611–620, Feb. 2014.

[12] B. Chisadza, M. J. Tumbare, W. R. Nyabeze, and I. Nhapi, "Validation of local knowledge drought forecasting systems in the Limpopo River Basin in Southern Africa," *Disaster Prevention and Management*, vol. 23, no. 5, pp. 551–566, Oct. 2014.

[13] Y. Bayissa *et al.*, "Developing a Combined Drought Index to Monitor Agricultural Drought in Sri Lanka," *Water*, vol. 14, no. 20, p. 3317, Oct. 2022.

[14] S. C. Caress, A. A. Belen, I. J. Esguerra, H. D. Wacan, F. D. Poso, and M. B. Solomon, "Rainfall And Meteorological Drought Forecasting in Albay, Philippines Using Artificial Neural Network," in *2021 IEEE 13th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM)*, Manila, Philippines: IEEE, Nov. 2021, pp. 1–6.

[15] N. Rachpibool and S. Kajornkasirat, "Assessment of Drought Forecasting Model for Northeastern, Thailand," in *2022 37th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC)*, Phuket, Thailand: IEEE, Jul. 2022, pp. 116–119.

[16] A. Belayneh, J. Adamowski, B. Khalil, and B. Ozga-Zielinski, "Long-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet neural network and wavelet support vector regression models," *Journal of Hydrology*, vol. 508, pp. 418–429, Jan. 2014.

[17] S. Banik, M. Anwer, A. F. M. K. Khan, R. A. Rouf, and F. H. Chanchary, "Forecasting Bangladeshi monsoon rainfall using neural network and genetic algorithm approaches," *The International Technology Management Review*, vol. 2, no. 1, pp. 1–18, Oct. 2009.

[18] H. Kuswanto and A. Naufal, "Evaluation of performance of drought prediction in Indonesia based on TRMM and MERRA-2 using machine learning methods," *MethodsX*, vol. 6, pp. 1238–1251, 2019.

[19] J. Han and V. P. Singh, "Long-Lead Drought Forecasting Across the Continental United States Using Burg Entropy Spectral Analysis With a Multiresolution Approach," *Water Resources Research*, vol. 59, no. 4, p. e2022WR034188, Apr. 2023.

[20] T. Tadesse, D. A. Wilhite, S. K. Harms, M. J. Hayes, and S. Goddard, "Drought Monitoring Using Data Mining Techniques: A Case Study for Nebraska, USA," *Natural Hazards*, vol. 33, no. 1, pp. 137–159, Sep. 2004.

[21] P. Feng, B. Wang, D. L. Liu, and Q. Yu, “Machine learning-based integration of remotely-sensed drought factors can improve the estimation of agricultural drought in South-Eastern Australia,” *Agricultural Systems*, vol. 173, pp. 303–316, Jul. 2019.

[22] Y. Tian, Y.-P. Xu, and G. Wang, “Agricultural drought prediction using climate indices based on Support Vector Regression in Xiangjiang River basin,” *Science of The Total Environment*, vol. 622–623, pp. 710–720, May 2018.

[23] Q. Zhang *et al.*, “Causes and Changes of Drought in China: Research Progress and Prospects,” *J Meteorol Res*, vol. 34, no. 3, pp. 460–481, Jun. 2020.

[24] M. V. K. Sivakumar, R. Stefanski, M. Bazza, and S. A. Zelaya, “Preface,” *Weather and Climate Extremes*, vol. 3, pp. 2–3, Jun. 2014.

[25] E. T. D. Jesus, J. D. S. Amorim, R. Junqueira, M. R. Viola, and C. R. D. Mello, “Meteorological and hydrological drought from 1987 to 2017 in Doce River Basin, Southeastern Brazil,” *RBRH*, vol. 25, p. e29, 2020.

[26] A. F. Van Loon, “Hydrological drought explained,” *WIREs Water*, vol. 2, no. 4, pp. 359–392, Jul. 2015.

[27] A. Filipović, “Water Plant and Soil Relation under Stress Situations,” in *Soil Moisture Importance*, IntechOpen, 2020.

[28] F. Mannocchi, F. Todisco, and L. Vergni, “Agricultural drought: indices, definition and analysis,” 1978.

[29] S. Sruthi and M. A. M. Aslam, “Agricultural Drought Analysis Using the NDVI and Land Surface Temperature Data; a Case Study of Raichur District,” *Aquatic Procedia*, vol. 4, pp. 1258–1264, 2015.

[30] J. Zhou, X. Chen, C. Xu, and P. Wu, “Assessing Socioeconomic Drought Based on a Standardized Supply and Demand Water Index,” *Water Resour Manage*, vol. 36, no. 6, pp. 1937–1953, Apr. 2022.

[31] A. Zargar, R. Sadiq, B. Naser, and F. I. Khan, “A review of drought indices,” *Environ. Rev.*, vol. 19, no. NA, pp. 333–349, Dec. 2011.

[32] S. M. Vicente-Serrano, S. Beguería, and J. I. López-Moreno, “A Multiscalar Drought Index Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration Index,” *Journal of Climate*, vol. 23, no. 7, pp. 1696–1718, Apr. 2010.

[33] A. Dai, K. E. Trenberth, and T. Qian, “A Global Dataset of Palmer Drought Severity Index for 1870–2002: Relationship with Soil Moisture and Effects of Surface Warming,” *Journal of Hydrometeorology*, vol. 5, no. 6, pp. 1117–1130, Dec. 2004.

[34] A. Alkan, “Drought Forecasting using Palmer Drought Severity Index with Wavelet Transform Technique and Machine Learning Methods,” *Int.J.Res.Publ.Rev.*, vol. 04, no. 01, pp. 2177–2185, 2023.

[35] D. Rastogi, K. Kurte, V. Reshniak, and O. Kotevska, “AI-Based Approach for Advancing the Understanding of Spatiotemporal Drought Characteristics,” AI4ESP--1111, 1769665, Apr. 2021.

[36] H. Wu, M. J. Hayes, D. A. Wilhite, and M. D. Svoboda, “The effect of the length of record on the standardized precipitation index calculation,” *Int. J. Climatol.*, vol. 25, no. 4, pp. 505–520, Mar. 2005.

[37] Y.-W. Choi *et al.*, “Future changes in drought characteristics over South Korea using multi regional climate models with the standardized precipitation index,” *Asia-Pacific J Atmos Sci*, vol. 52, no. 2, pp. 209–222, May 2016.

[38] P. Pandya, R. Kumarkhaniya, R. Parmar, and P. Ajani, “Meteorological Drought Analysis Using Standardized Precipitation Index,” *Curr. World Environ*, vol. 15, no. 3, pp. 477–486, Dec. 2020.

[39] N. R. Patel, P. Chopra, and V. K. Dadhwal, “Analyzing spatial patterns of meteorological drought using standardized precipitation index,” *Meteorological Applications*, vol. 14, no. 4, pp. 329–336, Dec. 2007.

[40] N. B. Guttman, “ACCEPTING THE STANDARDIZED PRECIPITATION INDEX: A CALCULATION ALGORITHM [1],” *J American Water Resour Assoc*, vol. 35, no. 2, pp. 311–322, Apr. 1999.

[41] R. Shumway and D. Stoffer, "Time series analysis and its applications", *Springer Texts in Statistics*, 2000.

[42] R. Adhikari and R. K. Agrawal, “An Introductory Study on Time Series Modeling and Forecasting.” arXiv, Feb. 26, 2013. Accessed: May 10, 2024. [Online]. Available: http://arxiv.org/abs/1302.6613

[43] B. Billah, M. L. King, R. D. Snyder, and A. B. Koehler, “Exponential smoothing model selection for forecasting,” *International Journal of Forecasting*, vol. 22, no. 2, pp. 239–247, Apr. 2006.

[44] R. B. Cleveland, W. S. Cleveland, and I. Terpenning, “STL: A Seasonal-Trend Decomposition Procedure Based on Loess,” *Journal of Official Statistics*, p. 3, Mar. 1990.

[45] D. Alba and A. E. Muñoz-Zavala, “A comparison between SARIMA models and Feed Forward Neural Network Ensemble Models for Time Series Data,” *RCS*, vol. 92, no. 1, pp. 9–22, Dec. 2015.

[46] P. P. Dabral and I. Tabing, “Modelling and Forecasting of Monthly Rainfall and Temperature Time Series Using SARIMA for Trend Detection- A Case Study of Umiam, Meghalaya (India),” *IJECC*, pp. 155–172, Nov. 2020.

[47] S. J. Taylor and B. Letham, “*Forecasting at scale*,” PeerJ Inc., e3190v2, Sep. 2017.

[48] C. B. Aditya Satrio, W. Darmawan, B. U. Nadia, and N. Hanafiah, “Time series analysis and forecasting of coronavirus disease in Indonesia using ARIMA model and PROPHET,” *Procedia Computer Science*, vol. 179, pp. 524–532, Jan. 2021.

[49] W. Tian, J. Wu, H. Cui, and T. Hu, “Drought Prediction Based on Feature-Based Transfer Learning and Time Series Imaging,” *IEEE Access*, vol. 9, pp. 101454–101468, 2021.

[50] R. D. Patil and O. S. Jadhav, “Predictive Modelling of Rainfall Data for Aurangabad Region by using ARIMA Method,” *IJET*, vol. 7, no. 4.10, p. 1085, Oct. 2018.

[51] D. G, “Statistical Modelling of Maximum and Minimum Temperature in Thiruvananthapuram District,” *IJRASET*, vol. 7, no. 7, pp. 1270–1274, Jul. 2019.

[52] E. Zhao, J. Zhao, L. Liu, Z. Su, and N. An, “Hybrid Wind Speed Prediction Based on a Self-Adaptive ARIMAX Model with an Exogenous WRF Simulation,” *Energies*, vol. 9, no. 1, p. 7, Dec. 2015.

[53] J. Shen, D. Valagolam, and S. McCalla, “Prophet forecasting model: a machine learning approach to predict the concentration of air pollutants ($PM_{2.5}$, PM_{10}, O_3, NO_2, SO_2, CO) in Seoul, South Korea,” *PeerJ*, vol. 8, p. e9961, Sep. 2020.

[54] V. Tulshyan, D. Sharma, and M. Mittal, “An Eye on the Future of COVID-19: Prediction of Likely Positive Cases and Fatality in India over a 30-Day Horizon Using the Prophet Model,” *Disaster med. public health prep.*, vol. 16, no. 3, pp. 980–986, Jun. 2022.

[55] G. James, D. Witten, T. Hastie, and R. Tibshirani, (2021, July 28), *An Introduction to Statistical Learning*, (2nd ed.), [Online]. Available: https://link.springer.com/book/10.1007/978-1-0716-1418-1. Accessed: Apr. 20, 2024.

[56] A. Malik, A. Kumar, P. Rai, and A. Kuriqi, "Prediction of Multi-Scalar Standardized Precipitation Index by Using Artificial Intelligence and Regression Models," *Climate*, vol. 9, no. 2, p. 28, Feb. 2021.

[57] E. J. Burke, S. J. Brown, and N. Christidis, "Modeling the Recent Evolution of Global Drought and Projections for the Twenty-First Century with the Hadley Centre Climate Model," *Journal of Hydrometeorology*, vol. 7, no. 5, pp. 1113–1125, Oct. 2006.

[58] J. Jacobi, D. Perrone, L. L. Duncan, and G. Hornberger, "A tool for calculating the Palmer drought indices: Technical Note," *Water Resour. Res.*, vol. 49, no. 9, pp. 6086–6089, Sep. 2013.

[59] M. W. Raphael, M. M. Benedict, and M. R. James, "Analysis of spatial and temporal drought variability in a tropical river basin using Palmer Drought Severity Index (PDSI)," *Int. J. Water Res. Environ. Eng.*, vol. 9, no. 8, pp. 178–190, Aug. 2017.

[60] Chang, "SEASONAL AUTOREGRESSIVE INTEGRATED MOVING AVERAGE MODEL FOR PRECIPITATION TIME SERIES," *Journal of Mathematics and Statistics*, vol. 8, no. 4, pp. 500–505, Apr. 2012.

[61] A. K. Mishra and V. R. Desai, "Drought forecasting using stochastic models," *Stoch Environ Res Ris Assess*, vol. 19, no. 5, pp. 326–339, Nov. 2005.

[62] T. Papalaskaris, "Forecasting Monthly Rainfall of Karyes, Chios Island, Greece, Central-Eastern Mediterranean Basin Using the ARIMA Method," in *The 4th EWaS International Conference: Valuing the Water, Carbon, Ecological Footprints of Human Activities*, MDPI, Sep. 2020, p. 68.

[63] C. C. Nwokike and E. W. Okereke, "Comparison of the Performance of the SANN, SARIMA and ARIMA Models for Forecasting Quarterly GDP of Nigeria," *ARJOM*, pp. 1–20, May 2021.

[64] J. Zhou, D. Wang, S. S. Band, E. Mirzania, and T. Roshni, "Atmosphere air temperature forecasting using the honey badger optimization algorithm: on the warmest and coldest areas of the world," *Engineering Applications of Computational Fluid Mechanics*, vol. 17, no. 1, p. 2174189, Dec. 2023.

[65] L. S. Al Blooshi, S. Alyan, N. J. Joshua, and T. S. Ksiksi, "Modeling Current and Future Climate Change in the UAE using Various GCMs in MarksimGCMR," *TOASCJ*, vol. 13, no. 1, pp. 56–64, Dec. 2019.

[66] E. E. Sampson, "Development of a Technique for Improving Access to Farm Practice Information Based on Monitored Temperature Conditions," *IJSTS*, vol. 2, no. 6, p. 165, 2014.

[67] K. Thiyagarajan, S. Kodagoda, and N. Ulapane, *"Short-term Time Series Forecasting of Concrete Sewer Pipe Surface Temperature,"* presented at the 15th IEEE Conference on Industrial Electronics and Applications (ICIEA), Sep. 17, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:219095249

[68] M. H. Shaukat *et al.*, "Monthly drought prediction based on ensemble models," *PeerJ*, vol. 8, p. e9853, Sep. 2020.

[69] R. Rezaiy and A. Shabri, "Using the ARIMA/SARIMA Model for Afghanistan's Drought Forecasting Based on Standardized Precipitation Index," *MATEMATIKA*, Dec. 2023.

[70] T. Toharudin, R. S. Pontoh, R. E. Caraka, S. Zahroh, Y. Lee, and R. C. Chen, "Employing long short-term memory and Facebook prophet model in air temperature forecasting," *Communications in Statistics - Simulation and Computation*, vol. 52, no. 2, pp. 279–290, Feb. 2023.

[71] J. Han and V. P. Singh, "A review of widely used drought indices and the challenges of drought assessment under climate change," *Environ Monit Assess*, vol. 195, no. 12, p. 1438, Dec. 2023.

[72] F. Huang *et al.*, "Towards interpreting machine learning models for predicting soil moisture droughts," *Environ. Res. Lett.*, vol. 18, no. 7, p. 074002, Jul. 2023.

[73] A. Danandeh Mehr, A. U. Sorman, E. Kahya, and M. Hesami Afshar, "Climate change impacts on meteorological drought using SPI and SPEI: case study of Ankara, Turkey," *Hydrological Sciences Journal*, vol. 65, no. 2, pp. 254–268, Jan. 2020.

[74] R. B. W. Vodounon, H. Soude, and O. Mamadou, "Drought Forecasting in Alibori Department in Benin using the Standardized Precipitation Index and Machine Learning Approaches," *IJACSA*, vol. 13, no. 12, 2022.