

FORECASTING SOIL MOISTURE BASED ON EVALUATION OF TIME SERIES ANALYSIS

*Thesis submitted in partial fulfillment of the requirements for the award of
degree of*

Master of Engineering
in
Computer Science and Engineering

Submitted By
Sukhwinder Singh
(Roll No. 801732052)

Under the supervision of:
Dr. Parteek Kumar
Associate Professor

Dr. Sanmeet Kaur
Assistant Professor



THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

COMPUTER SCIENCE AND ENGINEERING DEPARTMENT
THAPAR INSTITUTE OF ENGINEERING AND TECHNOLOGY
PATIALA – 147004

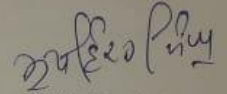
June 2019

THAPAR
Electronics
(To be filled
ing & Technology
following thesis
Repository
SANMEET Kaur

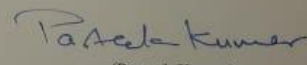
CERTIFICATE

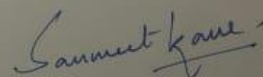
I hereby certify that the work which is being presented in the thesis entitled, "Forecasting Soil Moisture based on Evaluation of Time Series Analysis", in partial fulfillment of the requirements for the award of degree of Master of Engineering in *Computer Science and Engineering* submitted in Computer Science and Engineering Department of Thapar Institute of Engineering and Technology, Patiala, is an authentic record of my own work carried out under the supervision of *Parteek Kumar* and *Sanmeet Kaur* and refers other researcher's work which are duly listed in the reference section.

The matter presented in the thesis has not been submitted for the award of any other degree of this or any other University.


(Sukhwinder Singh)

This is to certify that the above statement made by the candidate is correct and true to the best of my knowledge.


(Parteek Kumar)
Associate Professor,
Computer Science and Engineering Department


(Sanmeet Kaur)
Assistant Professor,
Computer Science and Engineering Department

ACKNOWLEDGMENT

First of all, I would like to thank the Almighty, who has always guided me to work on the right path of the life. This work would not have been possible without the encouragement and able guidance of my supervisor **Dr. Parteek Kumar** and **Dr. Sanmeet Kaur**. I thank my supervisors for their time, patience, discussions and valuable comments. Their enthusiasm and optimism made this experience both rewarding and enjoyable. I am also thankful to **Ms. Sawinder Kaur**, Ph.D. scholar in the Computer Science and Engineering Department for her guidance and valuable time.

I would like to express my gratitude to **Dr. Maninder Singh**, Head of Computer Science and Engineering Department and **Dr. Ashutosh Mishra**, P.G. coordinator for their constant motivation and encouragement. I am also thankful to the entire faculty and staff members of the Computer Science Department for their direct-indirect help, cooperation, love, and affection. Last but not least, I would like to thank my parents for their wonderful love and encouragement, without their blessings none of this would have been possible. I would also like to thank my close friends for their constant support.

Sukhwinder Singh

ABSTRACT

Precision agriculture is a technique that is incorporated to produce high crop yield with the best utilization of available resources. Traditional farming is adversely affected due to improper resource management. In order to overcome the efforts of a farmer, a model for the soil moisture forecasting has been proposed to deliver better results of farming. The proposed model uses the Auto-Regressive Integrated Moving Average (ARIMA) and Long short-term memory (LSTM) to predict soil moisture. The models are trained on a dataset acquired from IIT Kanpur agricultural site. For analyzing the performance of the model Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) have been used as performance metrics. This study reveals that the LSTM model performs better in contrast with the ARIMA model. It paving way for the early prediction of the soil moisture that can be used with other advanced innovative irrigation techniques.

TABLE OF CONTENTS

CERTIFICATE	I
ACKNOWLEDGMENT	II
ABSTRACT	III
TABLE OF CONTENTS	V
LIST OF TABLES	IX
LIST OF FIGURES	XII
Chapter-1: Introduction	
1.1 Agriculture	1
1.1.1 Background	2
1.1.2 Agriculture in India	4
1.2 Type of Agriculture	4
1.2.1 Traditional Agriculture	4
1.2.2 Regenerative Agriculture	5
1.2.3 Precision Agriculture	6
1.3 Evolution of Precision agriculture	7
1.4 The motive of precision agriculture	8
1.5 Impact of Precision Agriculture on Industry	9
1.6 Challenges in Precision Agriculture	12
1.7 Study of Soil moisture for Agriculture	13
1.7.1 Role of soil moisture in Precision Agriculture	14
1.7.2 Methods of measuring the Soil Moisture	15
1.7.2.1 Infrared Moisture Balance Technique	15
1.7.2.2 Neutron Scattering Method	15
1.7.2.3 Optical Technique	15
1.8 Time Series	16
1.8.1 Components of time Series	17

1.8.2	Type of Time Series	17
1.8.2.1	Non-Stationary Time Series	17
1.8.2.2	Stationary Time Series	17
1.8.3	Applications of Time Series	19
1.9	Thesis Outline	20

Chapter-2: Literature Survey

2.1	Precision Agriculture	21
2.2	Soil Moisture Prediction for Precision Agriculture	22
2.3	Related Work on the Precision Agriculture	24
2.4	Related Work on Time Series	26
2.5	Methods and Tools used	28

Chapter-3: Problem Statement

3.1	Research Gap	31
3.2	Objectives of the proposed work	32

Chapter-4: dataset

4.1	MET Dataset	33
4.2	Snow and Rain Transition dataset	34
4.3	Michigan University dataset	36
4.4	IMD Pune Soil-based Dataset	38
4.5	IMD Pune Weather-based Dataset	40
4.6	IIT-K Dataset	43

Chapter-5: Models for Building Soil Moisture Prediction System

5.1	Time-Series	47
5.1.1	Working of the Time series models	47
5.1.2	Autocorrelation and Partial Autocorrelation Functions	49
5.2	Models based on Time Series Analysis	50

5.2.1 Automatic Regressive	50
5.2.2 Moving Average	51
5.2.2.1 Centered Moving Average	51
5.2.2.2 Trailing Moving Average	52
5.2.3 ARMA	52
5.2.4 ARIMA	53
5.2.4.1 Working of the ARIMA Model	53
5.2.5 LSTM	55
5.2.5.1 Working of the LSTM Model	53
5.3 Validation Phase	58
5.4 Chapter Summary	59

Chapter 6: Implementation of Models for Moisture Prediction

System

6.1 System Components	60
6.2 Implementation of the ARIMA model	61
6.3 Parametrize the ARIMA model	63
6.4 Implementation of the LSTM model	68
6.5 Parametrize the LSTM model	70
6.6 Chapter summary	71

Chapter 7: Results and Discussion

7.1 Evaluation of results for ARIMA model	72
7.2 Evaluation of results for LSTM model	75
7.3 Comparison analysis of the ARIMA and LSTM	79
7.4 Chapter Summary	80

Chapter 8: Conclusion and Future Work

8.1 Work Accomplished	81
8.2 Conclusion	81

8.3 Social Benefits	82
8.4 Future Work	82
References	84
List of Publications	87
Plagiarism Report	88

LIST OF TABLES

Table No.	Caption	Page Number
Table 2.1	Summary of the related work done	28
Table 4. 1	MET Dataset of the Patiala city	33
Table 4.2	Snow and rain transition dataset	35
Table 4.3	Michigan University dataset	37
Table 4.4	IMD Pune soil-based dataset	39
Table 4.5	IMD Pune weather-based dataset	41
Table 4.6	IIT-K Dataset	43
Table 4.7	IIT-K processed dataset	44
Table 4.8	Summary of all dataset	45
Table 5.1	Differencing order of the sample observations	48
Table 6.1	Parametric values of ACF and PACF for this experiment	67
Table 6.2	Values of p, q, and d used in ARIMA	68
Table 6.3	Parameter for the LSTM model	70
Table 7.1	Performance metrics evaluation of ARIMA	75
Table 7.2	Performance metrics evaluation of LSTM	78
Table 7.3	Comparison of the error rates of ARIMA and LSTM	79

LIST OF FIGURES

Figure No.	Caption	Page Number
Figure 1.1	Wheat crop	1
Figure 1.2	History of the Agriculture	3
Figure 1.3	Traditional Agriculture	5
Figure 1.4	Regenerative Agriculture	6
Figure 1.5	Precision Agriculture	7
Figure 1.6	A tractor with the GPS enabled device	10
Figure 1.7	Automatic robot for agriculture	11
Figure 1.8	A drone device	11
Figure 1.9	Time series forecasting	16
Figure 1.10	Time series components	17
Figure 1.11	Non-stationary Time series	18
Figure 1.12	Stationary Time series	19
Figure 2.1	Precision Agriculture and its technologies	21
Figure 2.2	Fetching soil moisture information	23
Figure 2.3	Workflow of the soil moisture predictions using different models	23
Figure 4.1	MET Dataset of the Patiala city	34

Figure 4.2	Snow and rain transition dataset	36
Figure 4.3	Michigan University dataset	38
Figure 4.4	IMD Pune soil-based dataset	40
Figure 4.5	IMD Pune weather-based dataset	42
Figure 4.6	IIT-K processed Dataset	44
Figure 5.1	Flowchart of the time series	47
Figure 5.2	Example of Autocorrelation	49
Figure 5.3	Example of partial autocorrelation	50
Figure 5.4	Flowchart of ARIMA	54
Figure 5.5	Architecture of an LSTM	56
Figure 5.6	Flowchart of the LSTM	57
Figure 5.7	Process to perform Validations	58
Figure 6.1	Workflow of the ARIMA model	62
Figure 6.2	Autocorrelation plot of the soil moisture at 10m	63
Figure 6.3	Partial-Autocorrelation plot of the soil moisture at 10m	64
Figure 6.4	Autocorrelation plot of the soil moisture at 25 m	64
Figure 6.5	Partial-Autocorrelation plot of the soil moisture at 25m	65
Figure 6.6	Autocorrelation plot of the soil moisture at 50 m	65
Figure 6.7	Partial-Autocorrelation plot of the soil moisture at 50 m	66
Figure 6.8	Autocorrelation plot of the soil moisture at 80 m	66
Figure 6.9	Partial-Autocorrelation plot of the soil moisture at 80 m	67

Figure 6.10	Workflow of the LSTM model	69
Figure 7.1	Actual and Predicted soil moisture at 10 m	73
Figure 7.2	Actual and Predicted soil moisture at 25 m	73
Figure 7.3	Actual and Predicted soil moisture at 50 m	74
Figure 7.4	Actual and Predicted soil moisture at 80 m	74
Figure 7.5	Actual and Predicted soil moisture at 10 m.	76
Figure 7.6	Actual and Predicted soil moisture at 25 m	77
Figure 7.7	Actual and Predicted soil moisture at 50 m	77
Figure 7.8	Actual and Predicted soil moisture at 80 m	78
Figure 7.9	Graphical comparison of ARIMA and the LSTM	80

LIST OF ABBREVIATIONS

WSN	Wireless Sensor Networks
NGOs	Non-Governmental Organization
SMOS	Soil Moisture Ocean Salinity
GPS	Global Positioning System
AI	Artificial Intelligence
UID	Unique Identity
GSM	Global System for Mobile Communications
IR	Infra-Red
WI	Willmott's Index
BPNN	Back Propagation Neural Network
M-ARIMA	MapReduce based Autoregressive Integrated Moving Average
M-KNN	MapReduce-K-Nearest Neighbors
M-HM	MapReduce-based Hybrid Model
FFNN	Feed-Forward Neural Network
ANN	Artificial Neural Network
MLR	Multiple Linear regression
MODIS	Moderate Resolution Imaging Spectroradiometer
AOD	Aerosol Optical Depth

CNN	Convolutional Neural Network
SCG	Scaled Conjugate Gradient
BFGS	Broyden Fletcher Goldfarb Shanno
MRF	Markov Random Field
SPSS	Statistical Package for the Social Sciences
IoT	Internet of Things
CMD	Command Prompt

1.1 Agriculture

Agriculture is an art and science of producing crops and livestock for economic purpose. It is the raising of plant life from the soil for the use of mankind and their livestock. The word 'Agriculture' is derived from the Latin language, the 'Ager' means Land or field and 'Culture' means cultivation. Agriculture is the milestone in the history of human civilization, this agribusiness helps mankind to shift from one location to another and starts its new farming business in the new place. It is one of the oldest and prime activities of the human being. Agriculture has remained an important source of land. Despite growing urbanization and industrialization in the world, more than fifty percent of the population is still working in the agriculture sector. In developing countries like India, China the agriculture sector plays a major source of employment and it has contributed to the national economy. The basic aim of agriculture is to raise stronger and more fruitful crops and plants and helps these crops and plants for their growth by improving the soil and supplying the water. It is the process of production of the fruits, vegetables, fiber, and many other desired products by the cultivation of certain plants and herbs.



Figure 1.1: Wheat crop

[<https://lamalfa.house.gov/issues/agriculture>]

It includes the preparation of plants and animal products for people to use and distribute them to the markets. An example of the agriculture depicted in Figure 1. This illustrates a plant of the wheat crop in the field and it is ready to harvest.

1.1.1 Background

Evidence suggests that in the sixth millennium BC the irrigation first appeared in Egypt and Mesopotamia. Various groups of people start digging the canal and it helps the farmers and native people by regulating the flow of the rivers, such as the Nile (a river in Egypt.). Every year the Nile river had had disastrous effects on farmers, it washes away the crops, swaps the fields and alternatively when the water is in insufficient quantity, then the lands are dried up and the crops are killed. This canal networks to protect the crops are developed as sophisticated irrigation systems. The Qanats is the oldest method of irrigation, it is the man-made underground streams used for the irrigation. This methodology is still in use in the parts of the Middle-East. In the Bronze Age, civilization all over the world invents advanced metalworking techniques, it creates ever-stronger agricultural equipment such as Axes, Plow. People continue the plants to use as a food source and domesticating the animals to make a useful dairy product. Afterward, in the middle age, the farmers of the Europe region start farming with the complex and strong irrigation systems such as water-raising machines, dams, and the reservoirs. These inventions provide too much help in the agriculture sector and boost the efficiency of agriculture [29].

In the 17th century, Britain has a “revolution”, that improved the agricultural method and techniques, it has a variety of new developments such as Horse Drawn seed planter. These developments increase agricultural production in the Europe region. Afterward, farmers develop new methods of irrigation such as crop rotation, plating new crops like turnip, corn, potatoes, etc [30].

The developments in the agricultural sectors are improving over the years. Irrigation equipment's such as Plow and other device are improved, there was the new machine in the agriculture industry called as mechanical combine harvester (It is the machine that harvests the crop from the field) was invented in the 18th century. The steam-powered engine is the first tractor, and it is designed to pull the agriculture

equipment's but it was not a worthy development due to the cost of the tractor. Afterward, the gasoline-powered tractor was invented. Hence, in the last century, these advanced technologies increase the level of farming and improve the agriculture sector to produce more crops and dairy products. Apart from this, a technology of the hybrid seed is developed, in this era, the particular hybrid corn revolutionized agriculture is determined. These hybrids improve the production of the crops, it increases the productivity of the crop. The plants and fruits which are harder and uniform to grown now become easier and simpler. In the second half of the 20th century, the hybrid seeds contributed to the increasing agriculture output [31]. An example of an agriculture life cycle is depicted in Figure 1.2.

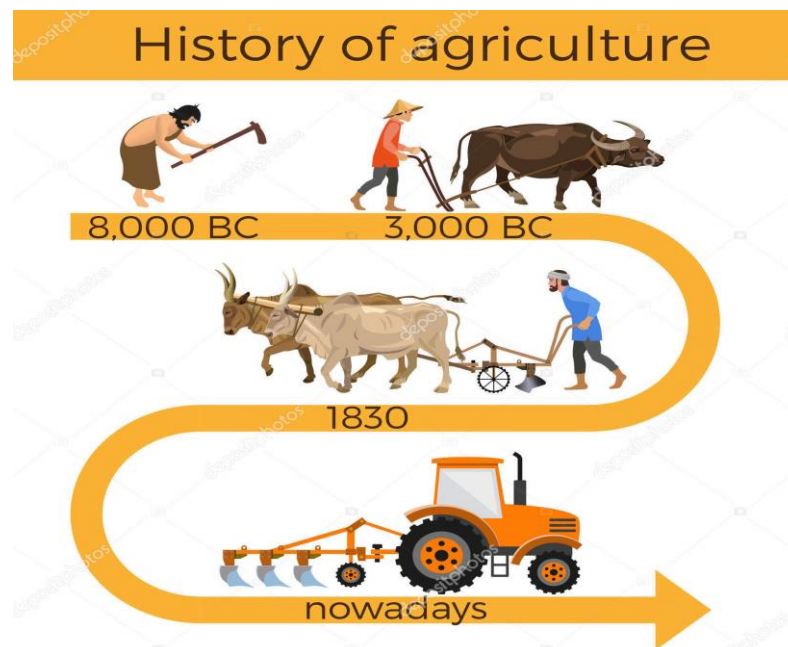


Figure 1.2: History of the agriculture

[<https://depositphotos.com/186974014/stock-illustration-history-of-agriculture.html>]

In the past 10 years, a huge controversy has raised over the use of fertilizers, chemical pesticides on farms. The increase in awareness about the impact of the use of these chemical products on human health forced the government authorities to change the rules and regulation of the agricultural industry.

These tremendous change in the invention gave birth to precision agriculture. Precision agriculture makes agriculture in an automatic way. The farmer can monitor

his own crops from his smartphone, remote device and television. The more about precision agriculture is discussed in the latter section of this chapter [32].

1.1.2 Agriculture in India

India is an agricultural country. In India, more than seventy percent of the population depends on the agriculture business, the nation has its one-thirds income from the agriculture sector, mostly the economy of India is depending on the agriculture. The development in the agriculture sector leads to the economic welfare of the country.

Agriculture in India remained underdeveloped for a long time. It is hard to produce enough food for the entire population of India. India had to import food-grains from other nations such as Pakistan, Australia in the past times. During late 1900's India has a Green Revolution to improve the agriculture sector and India produces more food-grains than it's a requirement, even it also do an export of the agriculture products such as Tea, Mangoes to the European Nations. These improvements are held by the amendments in the policies during the five-year plan of the governments. Now, India is a self-sufficient nation in the food-grain market. Today, India has ranked in top five in the production of groundnuts and the tea in the world, second in the production of the jute, sugarcane, rice, and the oilseeds. In the era of pre-Independence of India, the Indian agricultural sector depends upon the rains, sometimes India has insufficient production of the food grain due to the low monsoon and when there was a good monsoon then India had a good yield of the crops. However, after independence, India changes its policies, develops its new irrigation equipment's and has a huge market in the worldwide agriculture sector [33].

1.2 Types of agriculture

Agriculture has a vital role in the economic development of the nation. The principal means of livelihood is agriculture for more than 50 percent of rural households. In the below passages, we will go through the different types of agricultural methods.

1.2.1 Traditional Agriculture

Traditional agriculture is the original type of agriculture system and the most ancient food production system that has been practiced for thousands of years. It is the main source of improving farming technology like organic and modern farming. The

traditional agriculture uses a farming method such as compost/dung fertilizer, natural weed, crop rotations, and the biological pest control to protect the ecosystem. Today, there are only a few numbers of the farmers are doing traditional farming and in the under developing provinces. Besides this most of the nations prefers the advanced technologies, pesticides, and chemical fertilizer. Figure 1.3 illustrates a traditional agriculture technique; it depicts that a farmer is plaining the soil and make it ready for harvesting the crop with the use of two oxen.



Figure 1.3: Traditional Agriculture

[https://commons.wikimedia.org/wiki/File:Traditional_Farming_Methods_and_Equipments.jpg]

1.2.2 Regenerative Agriculture

The Regenerative Agriculture is the practices of the grazing the animals and does farming among other benefits such as reversing the climate change to rebuild the organic matter of the soil and restoring the soil properties that are degraded while doing long-time farming. It is beneficial for the soil on removing the excess carbon from the soil and improving the water cycle. The regenerative agriculture is the management of the agriculture resources to maintain the economic, reap sustainable environmental and social benefits. The soil holds the power of photosynthesis for the plants, improves its fertility, close the carbon cycle and improves the nutrient density during this process.



Figure 1.4: Regenerative Agriculture

[<http://www.aaronniederhelman.com/index.php/2016/02/16/regenerative-agriculture-resilient-in-nature/>]

In this agriculture, the land is freed for the time and there is no cultivation of the crop is going for a year or more than a year. It not only increases the health and biodiversity of the crop but also improves the biodiversity at both above and below the level in the soil surface. It also improves the atmospheric CO₂, and reverse the civilized threatened to soil properties by a human. It decreases the damaging effects on soil from tillage. In Figure 1.4, a Farmer left his fields for more than a year, to increase the fertility of the soil and gains the carbon to the soil. The reason for not to do farming for a year or more than that is to gain the productivity of the soil. The fertility level of the soil is improved and the crop which grown here after a gap has high hygiene and a good number of the production. Organic farming is an example of this, a farmer left his crop to decay the harmful bacteria from the soil [34].

1.2.3 Precision Agriculture

Precision Agriculture is a developing approach to farm management to ensure the productivity of the crop and the soil. This agriculture uses IT (Information technology) and communication device to optimize the health and the productivity of the crop. The precision agriculture devices mainly rely on specialized equipment based on farming, software, and IT services. It includes the real-time data of the environment such as air, humidity, and temperature along with other relevant information like weather conditions, history, availability of the equipment and the cost of the labor.

The Predictive analytics software such as IBM's SPSS (Statistical Package for the Social Sciences) is used on the provided data by the farmer with crop-related

information about the optimal planting times, harvesting times, crop rotation and soil management. The soil moisture content, soil temperature, and the surrounding air are measured using the sensors. Robots and the satellite drones provide details with the real-time images of individual plants. The information from these devices can be processed and integrate with sensor and other data to guide the equipment to take an action according to such as precisely what fields to water and when or where to plant a crop. In the last years, precision agriculture is so evolved and could support the IT and other technology devices required to fully implement and benefit precision agriculture. Today the smart service like smart sensors, cloud-based IoT (Internet of Things), mobile apps makes the precision agriculture possible for the farming societies and for the low-income farmers. An example of the precision agriculture is depicted in Figure 1.5, That a farmer uses a tractor and a slider, the slider has sensors which are installed within the slider. It helps to maintain the soil at a balanced level and it also helps the farmer by reducing efforts and capital. The balanced level of the soil also good for the field while watering the crop [35, 36].



Figure 1.5: Precision Agriculture

[<https://www.fginsight.com/news/news/we-need-technology-and-precision-farming-if-we-want-to-improve-productivity-79627>]

1.3 Evolution of Precision Agriculture

Precision agriculture is the management of agricultural resources using modern equipment. The farmer can manage crops as per their requirements, the water level to the crops, fertilizer and nutrients levels of the soil should be measured to meet the

requirement. Such type of technology developed a channel to optimize the nutrients and water for the crop and protects the crops from the insects. It required a dataset to detects the behavior of the crop and moisture level of the crops, it includes values that can be fetched using different WSN (Wireless Sensor Networks) such as soil properties using soil based sensors, crop data from the revenue department of the native region and other environmental factors like solar radiation using solar meter, air velocity, and humidity. These several values are recorded using different types of sensors. These several sensors and its attributes are depicted as follows:

- Soil-based attributes: Soil texture, structure, soil moisture, and soil nutrients are required attributes to improve precision agriculture. This information is fetched using different soil-based sensors.
- Crop-based attributes: The attributes required for the crop analysis are the count of plants, crop tissue nutrient status, crop stress, *etc.* It has the number of the seeds implanted in the field, several plants grow and the number of productive plants. It records all the information about the crop during its whole life cycle.
- Crop Yield: It represents the number of crop production in the previous years. The total crop harvested and the total number of the yield produced every year is counted. The NGOs (Non-Governmental Organisation) and the Government agencies are taking care of the information of the crop and its numbers.
- Climatic factors: The role of the climate is very vital in precision agriculture. The climate factor includes the temperature, humidity, rainfall, solar radiation, wind direction, and velocity are also are records to note the changed in climate behavior. These values can be obtained using online tools such as weather-based forecasting tools, SMOS (Soil Moisture Ocean Salinity).

These above-discussed factors are enough to generate the data for the crop and its related attributes. These factors are also helpful to determine the health of the crop and the fertility level of the soil.

1.4 The motive of Precision Agriculture

As the population of the world grows, the farmers must grow more food and products to fulfill the requirement of the world, but it is required to have the arable land for farming. The growth of the population leads to an increase in the residential sector, that directly impacts on the agriculture sector. The Precision agriculture uses the

sensors, GPS (Global Positioning System), Mapping Tools, robots and the data-analytics tools to customize the care of the crops without increasing the labor and its cost. The camera-equipped drones and the fixed sensors or the robot based sensors send the images and data wirelessly to the server or local computer on individual plants, having information about the properties of the crop such as leaf shape, stem size and the soil moisture around the plant, which determines the health and stress level of the crop. Therefore, precision agriculture can improve time management, decreases the usage of water and chemical, and produce healthier crops with higher yields. Its benefits farmers' capital and conserves resources while reducing chemical runoff. There are several other impacts of precision agriculture on the environment and these are described as follow:

- The productivity of agriculture is enhanced.
- Efficient use of water and other resources like organic products.
- The degradation of the soil in cultivable land shall be prevented.
- Reduction in the volume of chemical material usage in crop production.
- Sharing the results of modern farm practices to improve the quality, quantity and reduced cost of production in agricultural crops to researchers and developers.

1.5 Impact of precision Agriculture on Industry

Precision agriculture is the combination of the software and hardware-based products. The business objectives of on-time delivery of the product and extended supply-chain systems are threatening the traditional engineering and the manufacturing process. Groups of the products are under pressure to increase product delivery and manage the growing complexity of distributed organizations and supply chains. The future of agricultural products to meet global food production entirely depend upon the development of precision agriculture technology.

1.5.1 Precision Agriculture techniques

The Precision Agriculture is made-up of different technologies to work in the farms, these are mainly embedded with the automatic or manual setting and these are described below:

- **GPS** to monitor the health and productivity of the crop. The GPS system provides information in real time like longitude and latitude of the location and shares

most of the precise information of crop and soil contents. Furthermore, the GPS device is used for navigation and positioning assistance through large areas of the fields where farmers need high levels of accuracy. This innovative technology also assists the tractor driver to control steers of the tractor and follows the recommended pattern of the fields in an automatic mode. An example of GPS based device is illustrated in Figure 1.6. A farmer doing farming with the help of a GPS system to monitor the requirement of the soil pavement for the corp.



Figure 1.6: A Tractor with the GPS enabled device.

[\[https://www.green-diamond.ca/new-equipment/agricultural-equipment/precision-ag/4240-universal-display/\]](https://www.green-diamond.ca/new-equipment/agricultural-equipment/precision-ag/4240-universal-display/)

- **Autonomous vehicles** are lighter and smaller vehicles that are equipped with sensors and driverless technology to replace heavier traditional agriculture vehicles such as water cannon. The compaction brought about lower water and supplement maintenance that diminished crop production. Using autonomous vehicles, the productivity of the crop will be increased as a result of higher quality soil. Also, the costs will decrease when there is no driver involved. In Figure 1.7, it depicts an automatic device used in farming to level the soil and field. It works with mobile and the cloud service. This device is completely a field robot which provides more efficient results.



Figure 1.7 Automatic robot for agriculture

[\[https://www.oemoffhighway.com/trends/equipment-launches/agriculture/press-release/20983953/fendt-fendt-develops-robot-system-for-planting-and-accurate-documentation\]](https://www.oemoffhighway.com/trends/equipment-launches/agriculture/press-release/20983953/fendt-fendt-develops-robot-system-for-planting-and-accurate-documentation)



Figure 1.8: A drone device

[\[https://www.bainbridgeconsulting.com/blog/ai-in-agriculture-a-smart-farm-introduction/\]](https://www.bainbridgeconsulting.com/blog/ai-in-agriculture-a-smart-farm-introduction/)

- **Drones** are fly over large fields of the crops; it assists farmers by spotting problem areas in huge fields. It is an easier way to keep an eye on the fields in a single hand, on the other hand, the people must go through the whole fields and note down the observation of the crop manually. It also provides help in the irrigation-based equipment to control them. A drone device is illustrated in

Figure 1.8. It is estimated that more than 75 million agricultural IoT devices will be used by 2020, that creates huge opportunities in agricultural businesses.

- **Installing different sensors to tractors** to optimize the amount of fertilizer for the best results. All fields of the same crop respond differently to fertilizers; therefore, it is not cost-effective or efficient to apply a standard amount of fertilizer to the entire area. Now, with site-specific, the advanced precision fertilization techniques such as satellite imagery, ground-based sensors, and drone monitoring to collect information, precision fertilizer application is a smarter option for farmers.
- **Analytics software** accesses crop data and soil related information acquired using several technologies. It provides decisions to farm management about optimal planting of the crop, crop rotation patterns, and harvesting times as well as management of the soil.

The precision agriculture is developing day by day, there are many start-up companies that are working to increase the level of agriculture. Several new technologies are developed by them such as The AgNext, a start-up in Mohali (India) develop a protocol for agriculture using AI (Artificial Intelligence) to improve the food quality.

Furthermore, the Fasal, an institute in Bangalore (India) works on the prediction of the crop growth, revenue and natural resources management. Hence, precision agriculture boosting the economic standard of the industry. The Strider, a Brazil-based start-up working on the analysis of the diseases to the crop. It predicts crop health using the different attributes of the plant.

1.6 Challenges in Precision Agriculture

The cycle of precision agriculture goes through different phases. There are different challenges in precision agriculture. These challenges are depicted in the following points.

Connectivity in rural areas: A strong internet connection is required to send/receive the signal from the server. While in rural areas strong, reliable internet connectivity is not available. The disturbing internet connection may create hindrance for an end to end.

Variability in the soil moisture: Different crops have different behaviors. It depends upon the availability of the nutrients to the crops. Some crops consume too much water for the growth and some use a low level of the water to grow. It is very hard to maintain the balance between the different crops in the same field. So, it becomes harder to record all different values of the crop.

Configuration problem: Several sensors are deployed at various locations of the fields. In countries like India, Canada, the lands, atmosphere and the environment are different in every provenance of the country. Hence, the same type of sensors can't be deployed in all regions, this variability leads to the problem of the configurations of the software and hardware for precision agriculture.

1.7 Study of Soil moisture for Agriculture

The moisture content or water content is the amount of water dissolved in a material, such as soil (called soil moisture), ceramics rock, wood, or fruit. The moisture content is widely used in a range of technical and scientific areas and is expressed in a ratio, where it can be in the range from zero to the value of the porosity of the material at immersion. The zero value in the soil moisture depicts that the soil is completely dry like as drought. It is measured in the volumetric units or in mass units. Soil moisture is an important variable that controls the water exchange and heat energy between the surface of the land and the atmospheric evaporation. It is the percentage of the water filled in the soil and it is affected by temperature, precipitation, humidity, soil characteristic and many more. It can be given on a volumetric or mass basis. Soil moisture is a key variable in controlling the exchange of water and heat energy between the land surface and the atmosphere through evaporation and plant transpiration. It is the water stored in the soil and is affected by precipitation, temperature, soil characteristics, and more. The health of our crops directly depends upon enough supply of moisture and soil nutrients, among other required things. As the level of the moisture declines, the growth of the plants and its normal function are disrupted, and the number of crop yields is reduced. And, as there are climatic changes the moisture availability is becoming more variable.

The amount of soil water or moisture for the crop is very important to know because of certain reasons and these are as follows:

- Soil water effects the temperature of the soil, the biological processes of crop and soil directly depend upon water.
- Water is essential for photosynthesis of the plant.
- Microorganisms in the soil require water for their metabolic activities.
- The soil moisture serves as a solvent and carrier of food nutrients for plant growth.
- Soil moisture is a principal constituent of the growing plant.
- Soil moisture acts as a nutrient itself.
- Soil water helps to increases the biological and chemical activities of soil.
- The yield of a crop is more often determined by the amount of water available in the soil rather than the deficiency of other food nutrients.

1.7.1 Role of Soil Moisture in Precision Agriculture

The soil moisture plays an important role in precision agriculture. There are several impacts of the soil moisture on agriculture. This soil moisture improves the fertility of the soil and helps to save the water. The more about the role of soil moisture in precision agriculture is discussed below:

- It increases the fertility of the soil: The moisture level of the soil depicts the fertility of the soil. The precision agriculture provide helps to maintain it. The soil moisture sensor depicts the moisture value and using the precision technology the fertility of the soil should be increased or decreased, that resulted with the increase in the productivity of the crop and high hygienic value of the crop. On the other hand, the high fertility of the soil leads to the birth of bushes in the crop's fields that attack the crop and buried the crop.
- It provides proper irrigation timing: The soil moisture depicts the irrigation timing of the water and fertilizing for the crop. Different seasons have different crops with their own soil moisture readings when the soil moisture values decrease then it is required to irrigate the field. Alternatively, if it increases then stop irrigating the field and wait to reach soil moisture at the recommended level.
- It depicts the requirement of the water: In precision agriculture, the soil moisture maintains the required level of the water to the crop. The soil moisture readings depict the quantity of the water required for the crop. It saves the water from the wastage using the estimation of the required water. It also suggests the farmer to

arrange this quantity of the water to this crop and reduces the expenditure of the farmer by depicting the required water quantity.

1.7.2 Methods of measuring the Soil Moisture

The soil moisture estimation played an important role in successful farm management. Modern techniques are used to measure the soil moisture employ electrical properties of the soil (viz., impedance, soil resistivity, dielectric constant, and, capacitance), soil moisture potential, radioactive, and infrared rays' techniques like neutron scattering, gamma attenuation, and optical techniques. These techniques determining the content of the soil moisture that can be used for both laboratory and as its applications.

1.7.2.1 Infrared moisture balance technique

This technique is mainly based on the principle of electro balancing combined with the infrared heating technique to ascertain the soil moisture content. It is controlled by the incorporated advanced microprocessor, for achieving high accuracy, reliability and rapid measurement (approximately 15–20 min) of the content of the soil moisture. The small sample weight (2–5 g), high instrumental costs and the application only for laboratory measurements and experiments are the major limitations of this instrument.

1.7.2.2 Neutron Scattering Method

It is a soil moisture meter, utilizing the concept of neutron scattering. Fast neutrons with very high energy (million electron volts) are injected into the soil from a source of radioactive material at the measurement location. The neutrons are slowed by impact with nuclei of hydrogen atoms present in the soil. These hydrogen atoms present in the soil in terms of pore water and vapors. The average energy loss due to the collision of neutrons with hydrogen atoms is considerable for the measurement. This loss is directly a measure of soil moisture content. It is measured in terms of the total number of slow neutrons counted per unit of time averaged over a volume of soil. The measured counts give the volumetric soil moisture content using the calibrated equation.

1.7.2.3 Optical techniques

Such technologies are based on the change in the reflected light and characteristics of the incident and when it passes through the soil mass. These methods include the use of fiber optic sensors, polarized light, and near-infrared sensors. However, the

practical applications of these sensors for soil moisture measurement are yet to be established.

The thesis work is on the prediction of the soil moisture using the time series analysis. The above part discusses precision agriculture and the importance of soil moisture in precision agriculture. In the next section, the time series analysis is introduced.

1.8 Time series

Time series is the collection of real-time data with respect to the time. It is the series of observations that are recorded in a particular time interval. Most of the business institutes focus on the time series data to analyze the number of the sales production of a year, website traffic, stock monitoring and many more. It is the sequence of data points measured at a regular time-intervals over a period. The data must be in the regular form, there should not be any gap of the time in between. In time series statistical methods are used to analyze the data and extract meaningful insights about the data. A simple time series forecasting can be observed in Figure 1.9 [37].

The data points are collected over a period. These data points (past values) are analyzed to forecast a future. Obviously, it is time-dependent. Time Series Analysis helps us to recognize the major components in a time series data. Mathematically, a time series is given by equation (1.1).

$$y = f(t) \tag{1.1}$$

where y is the phenomena at any given time t . Thus, y can be regarded as a function of

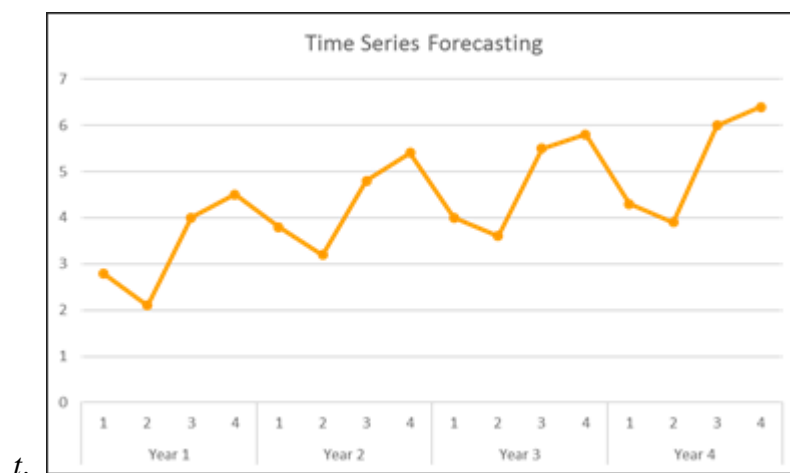


Figure 1.9: Time series forecasting

1.8.1 Components of time series

In the time series, there are four components that are required to analyze the data and forecasts the next observation. These components are necessary component for the working of the Time Series. A graphical representation of time series components is depicted in Figure 1.10 and these components are described below.

Trend: Trend is the increase or decrease in the series over a period, it persists over a long period of time.

Seasonality: Regular pattern of up and down fluctuations. It is a short-term variation occurring due to seasonal factors.

Cyclicality: It is a medium-term variation caused by circumstances, which repeat in irregular intervals.

Irregularity: It refers to variations which occur due to unpredictable factors and do not repeat patterns.



Figure 1.10: Time series components

1.8.2 Type of Time Series

The Time series is mainly of two types such as non-stationary time series and the stationary time series. The more about both of the time series are discussed below:

1.8.2.1 Non-Stationary Time Series

The data set is made of observations and a non-stationary time series data show seasonal fluctuations, trends, and other structures that depend on the index value of the time. Forecasting methods in the time series analysis are concerned with making

non-stationary time series based data to the stationary dataset by observing the mean, variance values of the dataset. The classical time series analysis and forecasting methods are concerned with making non-stationary time series data stationary by identifying and removing trends and removing stationary effects. The Non-stationary dataset is basic dataset data is acquired for the experiment or the collection of observation. Such as closed prices of the stock, traffic observations of the road are examples of the non-stationary dataset. An example of the non-stationary dataset using a dataset of shampoo sale is depicted in Figure 1.11.

1.8.2.2 Stationarity Data in Time Series

A stationary dataset has the property that the mean, variance and autocorrelation of the dataset do not change over time. It consists of the flat looking series, constant variance over time, without trend, a constant autocorrelation structure over time and no periodic fluctuations. In the stationary time-series dataset, it can be easier to build the model. Statistical modeling methods assume or require the time series data to be stationary. The data can be converted to stationary using two different approaches such as Differencing and Transformation. These methods are discussed in Chapter 5. A stationary timeline is plotted in Figure 1.12, it depicts that whole data is linear and with a unique variance.

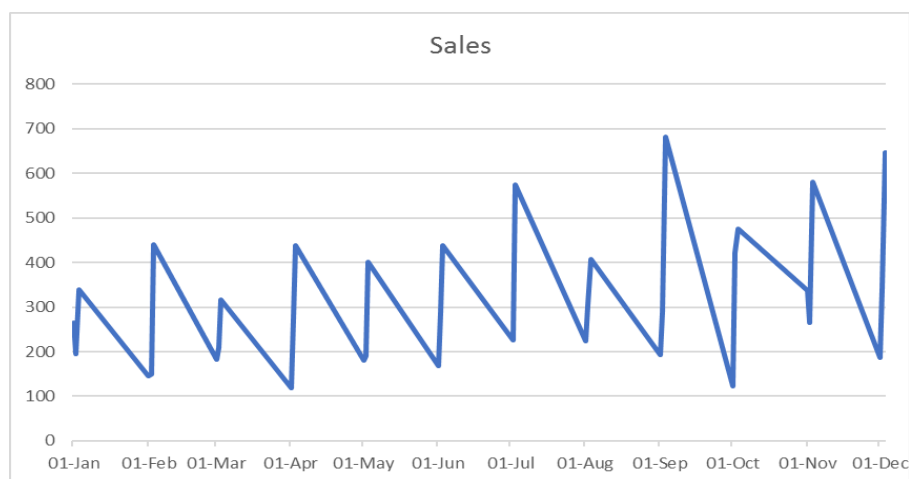


Figure 1.11: Nonstationary Time series

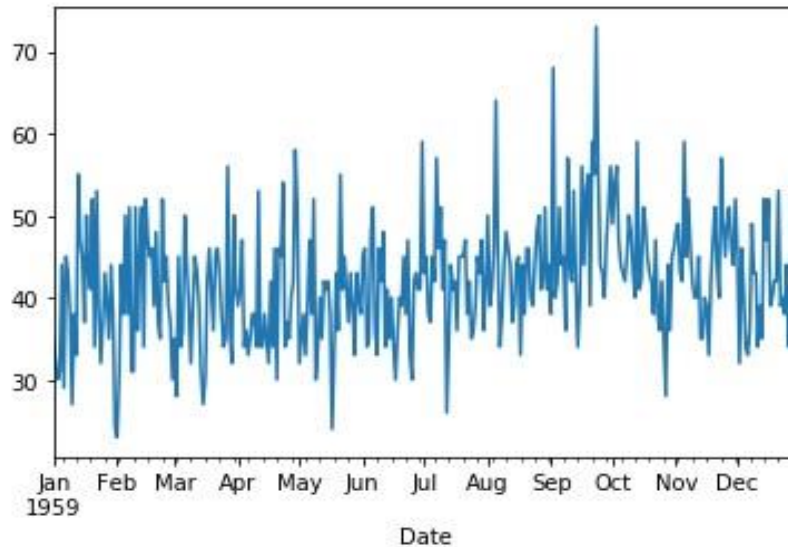


Figure 1.12: Stationary Time series

In the time series, the stationary dataset is required to build the model. It has a unique variance, single mean value over the whole dataset. The trend in the data value is removed using the lag value. The lag value is the degree of the differencing through which the observations are shifted by a lag or more than a lag value.

1.8.3 Applications of Time Series

The applications of time series models are manifold, it includes sales forecasting, weather forecasting, inventory studies and more. Some of the basic types of application of time series analysis are depicted below:

- **Descriptive Analysis:** It determines trends and patterns of the future using graphs and other tools.
- **Forecasting:** It is mostly used in financial and business forecasting based on historical patterns and trends.
- **Explanative Analysis:** It is used to measure the cross-correlation or relationship between the two time-series based dataset and their dependency on one another.

The biggest advantage of the time series analysis is that it helps to predict the future based on the historical data trends.

1.9 Thesis Outline

The goal is to measure the error rate of the two-time series models. The ARIMA and the LSTM models are used to compute and predict the soil moisture. In the first chapter of the thesis the introduction of precision agriculture, soil moisture, and the time series are discussed. Chapter 2 highlights the related works done in the field of precision agriculture, soil moisture, and the time series. Chapter 3 discusses the different type of datasets used in existing technologies. In chapter 4, the problem statement and the objectives of the thesis are described. In Chapter 5, the working of the proposed system has been explained. Chapter 6 illustrates the design and implementation of the proposed system. The obtained results have been highlighted in Chapter 7 followed by a conclusion and future scope in Chapter 8.

Chapter-2

Literature Survey

In this section, several existing approaches such as forecasting of soil moisture with respect to the usage of water, precipitation and other parameters of crops have been discussed. Apart from this, several different studies on time series with different types of datasets are also described in the latter part of this Chapter.

2.1 Precision Agriculture

Precision agriculture is the most innovative technology in today's era. In the research of precision agriculture, several technologies such as GPS, Drones, and WSN are collaborated to increase the production of agriculture. Many researchers' efforts to improve the level of agriculture and most of them work to improve the yield production of the crops. Figure 2.1 illustrates the different precision agriculture-based techniques that are used in the current agriculture sector.

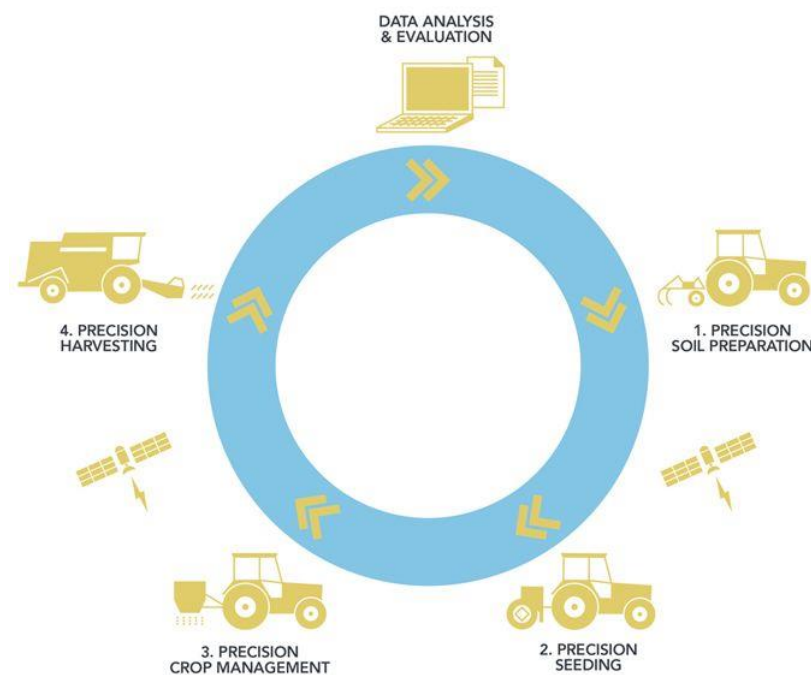


Figure 2.1: Precision Agriculture and its technologies
[<https://in.pinterest.com/pin/164240717635258207/?lp=true>]

- **Data Analysis and Evaluation:** The data based on the agriculture sector is fetched. It has several parameters used in the agriculture such as environmental

attributes like weather, moisture, sun-timing and many more and other parameters of the crops such as soil moisture, soil temperature, soil type, *etc.*

- **Precision Soil Preparation:** The soil is maintained for the special crop by utilizing the equipment such as cultivator or plows to level and arable the farms as per the crop requirements. The analyst recommends the farmers to use the resources for the crop in an efficient way that maintains a ratio between the crop and the resources required for the crop.
- **Precision seeding:** Seeds of the crops are managed in a way that helps to increase the yield production of the crop.
- **Precision Crop Management:** The available resources are used in a proper way that not only increases the production of the crop but also decreases the utilization of the resources such as water, fertilizer and the pesticides.
- **Precision Harvesting:** The ready crop is cut by the advanced harvesting machines; these machines are highly trained with their models and it cut the crop with best efforts. These crops are directly forwarded to the market without doing any other process because most of the processing is done by a single harvesting machine.

2.2 Soil Moisture Prediction for Precision Agriculture

The soil moisture prediction is ongoing research in the fields of precision agriculture. It uses machine learning and the deep learning models to forecasts the moisture of the soil. In precision agriculture, several technologies are used to acquire the soil moisture reading from the earth surface and for forecasting of the data, many researchers use their different strategy to forecast it. The soil moisture and reading of other attributes are fetched using different techniques and one of the most used technique is depicted in Figure 2.2. It illustrates the working of the soil moisture sensor. The sensor of the soil moisture is attached to Arduino. The sensor starts blinking when it is not processing the information of the soil moisture and the sensors stop blinking and run ideally when deployed in the bucket, where the buckets are full of the two types of different soil. A simple workflow of how the soil moisture works in Information-Technology is depicted in Figure 2.3. It has an Arduino device, sensor, connector and two buckets with different soil.



Figure 2.2: Fetching soil moisture information

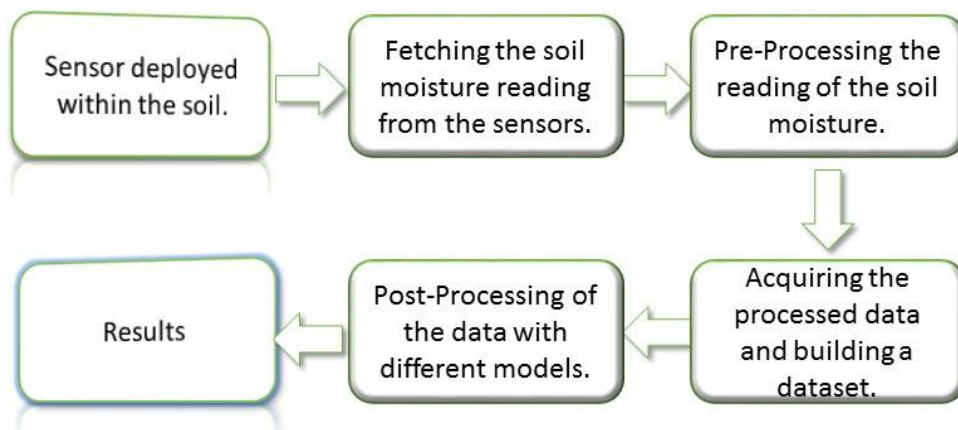


Figure 2.3: Workflow of the soil moisture prediction using different models

This workflow of the soil moisture prediction is used by most of the researchers.

- Soil-based sensors are installed within the soil.
- WSN and cloud service are used to communicate the information from the fields to the local data centers. The IoT based devices are used to make an end to end communication.
- Required data is fetched from the local servers and process the data using feature extraction and scaling of it.
- The processed data is used to build the model, where the model can be any such as Decision Tree, LSTM and SVM (Support Vector Machine).
- Prediction of the data is obtained using the prediction method of the model and after validating the results with the validated data the results are fetched.

2.3 Related Work on the Precision Agriculture

Zhu *et al.* (2011) discuss the sensor technology used for smart farming. Researchers designed an application in C programming language for the environment monitoring system based on WSN. This proposed system has consisted of the software layer and the hardware layer. Within the hardware layer, it had a data acquisition module to fetch data and a processing module to process the information. A wireless communication module used to transmit the data from the node to the server. After acquiring the data from the real environment using sensors there is only $\pm 1^{\circ}\text{C}$ change in temperature in Actual and predicted data [1].

Mat *et al.*, (2015) develops a WSN technology to achieves the proper irrigations results. Crops are monitored using soil moisture sensors and controlled using a closed-loop system. The results are computed using two different strategies such as scheduled system and the automated system. This study reveals that the automated strategy saves 1500 ml of water per day [2].

Li Tan (2016) develops a cloud-based framework of decision support and automation systems in Ruby. This framework is composed of three layers such as data importing from the agriculture-based sensors, decision module, and the software-based control unit to manage the equipment. The sensors export data of time stamp, UID (Unique Identity) key, the position of the sensor, movement of the sensor and recorded information from the environment. This experiment is tested on AWS (Amazon Web Services) [3].

Zhang *et al.* (2017) proposed an algorithm for an automatic valve. The valve is used to open or close automatically with respect to the moisture level within the soil. Soil sensors and other real-time devices monitor and collect the data about the growth of the crops and upload it to the data center. After mining and extraction of the data, it forecasts the water required for crops in different periods and makes an automatic irrigation system [4].

Sreeram *et al.* (2017) develop a prototype to prevents the crops from the flooding, security from the animal grazing and overwatering to the crop. It monitors and controlled various activities using GSM (Global System for Mobile Communications) and DTMF (Dual-tone multi-frequency signaling) technology. The soil moisture sensor, IR (Infra-Red) sensors and other sensors tell the behavior of the field and the Arduino is used to take action accordingly. The results of the activities are displayed on the smartphone [5].

H L *et al.* (2018) uses machine learning and big data tools to predict ryegrass in the wheat crop. Ryegrass is a disease that slowly decays the life of plants' crops. This study is experimented in the Australia region by the University of Australia with the help of the 173 villages. The researchers use the MRF (Markov Random Field) model to measure the density of the diseases in the region [6].

Lenka *et al.* (2015) developed an optimization technique for the prediction of soil moisture. Gradient descent with momentum based Neural Networks is used to predict the contents of the soil moisture before an hour in advance. The results of this study are evaluated using RMSE, MSE as performance metrics and these have 0.039622 and 0.19905 resulted values, which are high on the accuracy of the model [7].

Mohaptra *et al.* (2016) depict the models to predict the SMC (Soil Moisture Content) on an hourly basis. Two optimization techniques such as BFGS (Broyden Fletcher Goldfarb Shanno) and the SCG (Scaled Conjugate Gradient) are proposed using the neural network. The Soil moisture content is predicted with respect to another parameter like soil temperature, soil type, timestamp, environment temperature, environment humidity, Ultra Violet (UV) index, Air flow rate, and Air CO₂. The predicted value of the soil moisture is compared with the validated dataset and the MSE, RMSE and the R-squared are used to compute the performance of the model which resulted in a low error rate [8].

Guevara *et al.* (2017) proposed an application on vegetable crops using Data Mining techniques in Java-based platform. Researchers of Ecuador divides the application into two phases. In the first phase model is built using classification-based Decision Tree and the data is predicted in the next phase. The soil moisture is predicted with respect to the relative humidity, temperature, luminosity, and carbon-dioxide parameters. The predicted results are 96 % accurate [9].

Han *et al.* (2018) depict the estimation of the soil moisture level using the CART algorithm. A CART is a Classification and Regression Tree algorithm. The soil moisture is estimated at a high spatial resolution of 1 km distance from earth surface using meteorological satellites. It uses precipitation, temperature, soil properties to predicts the soil moisture. The results of this study have high accuracy and high spatial resolutions [10].

Priya *et al.* (2018) use Naive Bayes classifier integration of the MapReduce paradigm. Researchers recommended a crop such as Chili, Rice, Cotton, Maize to the farmer in the Telangana region of India. They use several parameters like rainfall, soil moisture,

temperature, and atmospheric pressure. MapReduce integrated with Naive Bayes paradigm predict the behavior of the different crop with respect to the given attributes [11].

Sobayo *et al.* (2018) estimate the soil moisture using the CNN (Convolutional Neural Network), it integrates with the thermal images using the drone devices with higher spatial resolution [12].

Prasad *et al.* (2019) develop a multivariate predictive model to forecasts the weekly soil moisture by hybridizing the Feed Forwarded Neural Network. The data for the model is collected from the four different sites of New Queensland of Australia and it consists of the 14 different attributes of the agriculture from the year 2007 to 2016. The accuracy of the model is computed using the Pearson's correlation coefficient (r), MAE, RMSE, MAPE, WI (Willmott's Index), the Legates-McCabe Index (L), and the Nash-Sutcliffe coefficient [13].

The above studies discussed the existing technologies used in precision agriculture. These studies are WSN based that depicts the fetching of the dataset for the soil and prediction of the soil moisture using the different models of machine learning and deep learning. Furthermore, this thesis uses the time series analysis to evaluate the soil moisture. There are several studies that use the time-series-based models are explored, these studies are discussed in the next section.

2.4 Related Work on Time Series

Soni *et al.* (2014) apply ARIMA model to simulate monthly mean Terras MODIS (Moderate Resolution Imaging Spectroradiometer), AOD550 nm (Aerosol Optical Depth) on a dataset of the eight sites of Indian-Gangetic plain from Mar 2000 to Mar 2012. This study analyzes the data in annual and seasonal variation and several AODs of all the sites are compared with their respective terms [14].

Bari *et al.* (2015) proposed an ARIMA model to forecast the monthly rainfall using data from 1980 to 2010 of Sylhet, Bangladesh. The data from 1986 to 2006 is used for the training data set and the remaining is used as a validated dataset. The validity of this dataset is computed using the Box and Jenkins method and the ARIMA (1,1,1) perform better from all respected parameter of p,d,q . [15]

Taki *et al.* (2016) compare the MLR (Multiple Linear regression) and the ANN (Artificial Neural Network) to predict the room air, roof temperature and the loss of

energy in a semi-solar based greenhouse designed in the Azerbaijan, Iran. The dataset used by the researcher has several environment variables such as room temperature, inside humidity, inside soil temperature, outside air temperature and the wind speed. The statistical results show that MLR models depict better results over the ANN model [16].

C. Narendra Babu and B. Eswara Reddy (2016) proposed a hybrid ARIMA-ANN model. The sunspot data, stock market data, and the electricity price data. The performance metrics such as MAE, MSE are used to compute the error rate of the model, which depicts that both models perform better in one-step-lag order instead of three or five-step-lag order of the data [17].

Zhang *et al.* (2017) proposed a model for the water in the Hetao, China. The researcher uses irrigation based monthly dataset having attributes of timestamp, evaporation, temperature, water diversion, precipitation, and water table depth of 14 years where data of 2000 to 2011 is used as training dataset and the values of 2012-13 are used as testing dataset of the model. This study reveals that dropout prevents the system from overfitting and the performance metrics R2 and RMSE depicts that LSTM model performs better than FFNN (Feed-Forward Neural Network) and the dual LSTM [18].

Zhang *et al.* (2018) use ARIMA model for trend analysis and forecasting of data of PM2.5 in the air in Fuzhou, China to find the change value in the air pollution. It includes data of two years (2014 to 2016) for modeling and analysis and uses the next year 2017 for the validation of the model. This study has no actual increase in the trend for the year 2017 as expected. In fact, there is a decrease in trend due to new policies in the nation [19].

Qing *et al.* (2018) predicted the hourly based weather data using the LSTM model. The study uses a dataset of Cape Verde in the Africa Continent from Mar 2011 to Dec 2013 and has a gap of 6 months from September 2012 to December 2012 with features of temperature, humidity, visibility, wind speed, and weather type. It predicted the last six months' data with 18.34% more accuracy as compared with BPNN (Back Propagation Neural Network) [20].

Bendre *et al.*, (2019) discussed the predictive analytics approaches to predicts the weather station data of 100 Years of 39 cities of Maharashtra, India such as temperature, humidity and more variable using time series. The neural network based on the MapReduce paradigm is used with ARIMA in this study. M-ARIMA

(MapReduce based Autoregressive Integrated Moving Average), M-KNN (MapReduce K-Nearest Neighbors) and the M-HM (MapReduce-based Hybrid Model) are proposed to increase the accuracy. The performance measures such as MAD (Mean Absolute Dimension), MSE, RMSE, and the MAPE are used to compute the accuracy [21].

2.5 Methods and Tools used

Several researchers perform different operations and experiment using various tools and the methodologies in the fields of Precision agriculture and the time series. The WSN framework is used to fetch the reading of the soil moisture [1]. Furtherly, a model is proposed to control the irrigation equipment's and save the water usage by forecasting the requirements of the water [2, 4, 5]. Machine learning based models such as MRF, Decision Tree, and Naïve Bays are used to predict the diseases in the crop, level of the soil moisture, and recommends a crop to the farmer respectively [6, 9, 10, 11]. Apart from this neural-network based approaches such as LSTM, BiLSTM, FFNN are used to forecasts the soil moisture [7, 8, 12]. In time series analysis, an ARIMA model is used to forecasts the data of different studies [14, 15, 19, 21]. Moreover, the LSTM model is used to build the model and predict the information of the precision agriculture-based data like soil moisture [16, 17, 18, 20]. Table 2.1 summarizes various methodologies used by researchers in the field of Precision Agriculture and Time series.

Table 2.1: Summary of the related work done

Author	Dataset	Algorithm	Proposed Scheme
Zhang <i>et al.</i> [13]	Soil moisture, Water Level, Video-Image Data	Decision Model	Real-time automatic valve based on water requirement is built.
Lenka <i>et al.</i> [16]	Temperature, Humidity, Soil Type, Air flow rate, UV index, Soil moisture, Soil temperature, Carbon-dioxide	Neural Networks	Prediction of the soil moisture contents before an hour.
Mohapatra and Lenka [17]	Soil moisture content, soil temperature, soil type, timestamp, environment temperature, environment humidity, Ultra Violet (UV), Sunlight intensity,	Scaled Conjugate Gradient (SCG) and Broyden Fletcher Goldfarb Shanno (BFGS)	Predict the Soil Moisture Content (SMC) on an hourly basis.

	Sunrise time, UV index, Air flow rate and Air CO ₂		
Guevara <i>et al.</i> [18]	Soil moisture, Relative Humidity, Temperature, Luminosity, Carbon-Dioxide	Decision Tree	Analysis of different crop data and its growth in a special Green environment.
Han <i>et al.</i> [19]	Precipitation, Land Surface Temperature, Evapotranspiration, biophysical, elevation, Topography, vegetation condition, Soil texture type, Soil water content.	CART Decision Tree	Analysis and Estimation of the soil moisture
Priya <i>et al.</i> [20]	Rainfall, Soil moisture, Temperature, Atmospheric pressure, Crop type	Naïve Bays classifier	Analysis and prediction for the best crop of the season.
Soni <i>et al.</i> [23]	Quarterly Aerial data of 8 Indian Cities	ARIMA	Statistical analysis of Ganga plain site of 12 years
Zhang <i>et al.</i> [28]	Month, Pm _{2,5} , SO ₂ , NO ₂ , PM _{1.0} . MCR, MMT, MMH	ARIMA	Forecasting of PM _{2,5} for measuring the effects of pollution.
Qing <i>et al.</i> [29]	Temperature, dew point, Humidity, Visibility, Wind speed, Weather	LSTM	Prediction of hourly based data of solar irradiance

In this chapter, several works are done in the field of precision agriculture and the time series are discussed. It is cleared that the soil moisture is an important variable of the precision agriculture to forecast the level of the water, soil moisture content, diseases to the crop and many more. It is forecasted by the many researchers with their own approaches, but in the end, their mission is to save the water and improve the scope of agriculture. In this chapter, different type of the methods are discussed such as Decision Tree, ARIMA and the LSTM with their dataset and the results and in this research, a comparison of the two models is proposed on the same dataset and the results are computed using several performance metrics such as MSE, RMSE, MAE and the MAPE.

In the next chapter, the problem statement is defined. The problems with the existing system will be discussed. The objectives of the proposed work will also be explained in the next chapter.

The precision agriculture is a vast and new technology towards the era of the modern generation. Most of the scientist and researchers are focused on this concept to improve the productivity and standard of the agriculture sector. New applications are developed to increase the health and the production of the crops. Such applications provide support for the farmer, government and other bodies that perform operations in the public sector. The Precision Agriculture is information and technology-based farm management framework to recognize, examine and manage the fields by using all practices of yield generation in right place at right time and in the right way for optimum benefits, sustainability and protection of the land resource.

In traditional agriculture, the farmers are unaware of the weather conditions. They use various available resources for the watering and fertilizing without taking into consideration the requirements of the crops. Moreover, In India population is drastically increasing every year, according to the World Bank report, there are 1.1% inclines in India's population and the census report from the previous year's depicts that there may be a vast change in the number of populations in the upcoming years. This leads to deforestation because the land of the agriculture sector is acquired for commercial and residential purpose. The shortage of land has been fixed using the precision agriculture that helps the farmers and governments by providing more healthier crops in the limited areas of the fields. Precision agriculture maintains a ratio between the food and the individual so that everyone can have dinner before his bed. It helps the farmers, local retailer, the environmentalists such as NGOs in a way to improve the water level, fertilization of the crops for farmers by measuring the soil attributes.

Before precision agriculture, the water and other resources are utilized in an unappropriated way for the cropping. New farmer has not enough knowledge and experience about farming, so he uses the assumptions to grow the crop. This unexperienced farming does waste of water and other resources and this affects the environment and the soil by the over usage or less usage. This unexpected water wastage can be utilized with the use of precision agriculture and scheduling mechanism. It has IT-enabled devices that fetches the information from the

environment using WSN and share it to the cloud servers for the processing and the future use. The acquired data is processed with the advanced analytics approaches such as machine learning and the deep learning to control the irrigation equipment and forecasts the soil moisture or other required attributes.

This precision agriculture improves the role of agriculture in mankind using IT devices. An advanced module is developed for the crops that predict the soil moisture. This forecasting of the soil moisture helps the farmer by early prediction of the resource's requirements.

3.1 Research Gap

The approaches used in [1, 2, 5] suggest the sensor and spatial based technology for fetching the reading of the soil moisture from the earth surface. The proposed approaches in [2, 4] develop the prototypes for the agriculture equipment to take a recommended decision according to the rules and regulations set for the prototype. The work proposed in [3] uses the cloud service for the irrigation and provide decision support to the equipment of agriculture.[6, 9, 10, 12] uses the machine learning model such as Decision Tree and the Naïve Bays to predict the soil moisture while the study [7, 13] uses the deep learning models to forecasts the soil moisture. From these above studies, the phenomenon of the agriculture-based dataset is acquired. Hence, from the view of the above studies, a soil moisture-based dataset is imported for this experiment.

The work done in [12, 13, 19, 21] suggests a statistical model like ARIMA for the prediction of the rainfall and the soil moisture, while the [16, 18, 20] uses the deep learning model for the weather forecasting and the soil moisture. [17] use both models such as ARIMA and Neural Network to compare the prediction of the several datasets. These studies use different performance metrics such as RMSE, MSE, and MAE for the evaluation of the model. This thesis mainly compares the statistical model “ARIMA” and the deep learning-based model “LSTM” for the forecasting of the soil moisture and recommend the best model for the experiment based on the soil moisture.

3.2 Objectives of the proposed work

This section describes the definition of the proposed study. To develop soil moisture prediction system following objective have been framed.

- To collect and analyze the existing dataset available for the development of smart irrigation system.
- To perform pre-processing on the selected dataset for time series analyze for prediction of the soil moisture.
- To build the ARIMA and the LSTM based soil moisture predictions model.
- To test and validate the proposed models.
- To compare both the models and recommends the most appropriate model for the soil moisture prediction system.

Available Dataset for Soil Moisture Prediction System

The experimentation for a study requires a dataset to analyze the problem. Several agencies in the world use information from farmers, consumers, developers, and technicians to create a formatted dataset. For experimentation, several datasets from different agencies are acquired and studied. These datasets are based on the recorded values from the IoT based devices such as soil moisture sensors, humidity sensors, and many more. There are several types of datasets that are discussed in the below sections.

4.1 MET Dataset

The MET dataset is owned by the Government of India. The meteorological department of India is acquiring information about several locations. It includes meteorological information of all states of India. MET dataset is acquired on a monthly and quarterly basis from 1901 to 2002. The aim of this dataset is to record the information of all the atmospheric terms such as temperature, Precipitation, Evapotranspiration, *etc.* In Table 4.1, the environment temperature is recorded in °C, it has a minimum temperature, maximum temperature and the average temperature (that is the mean of the minimum temperature and maximum temperature). Precipitation is the ratio of water particle in the atmospheric air. Vapour pressure and the cloud cover is related to the cloudy weather, Cloud cover records the percentage of the area in the sky covers that have clouds. The evapotranspiration is the term that records the value of the water in the soil. The screenshot of this dataset with their values is depicted in Figure 4.1.

Table 4.1: MET dataset

Attributes	unit	Duration
Min Temperature (Environment)	°C	1901-2002 (Monthly readings)
Max Temperature (Environment)	°C	1901-2002 (Monthly reading)
Avg Temperature (Environment)	°C	1901-2002 (Monthly Reading)
Precipitation	NA	1901-2002 (Monthly Reading)

Vapour Pressure	NA	1901-2002 (Monthly Reading)
Wet Day Frequency	NA	1901-2002 (Monthly Reading)
Evapotranspiration	NA	1901-2002 (Monthly Reading)
Cloud Cover	NA	1901-2002 (Monthly Reading)

1	Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2	1901	19.929	21.649	28.812	34.584	38.844	41.711	37.061	35.516	35.795	35.073	29.035	23.15
3	1902	22.349	24.569	30.677	35.782	39.595	39.688	36.465	36.155	35.016	33.428	28.356	22.068
4	1903	20.837	22.635	27.136	33.989	38.951	42.098	38.102	35.187	36.004	34.101	28.058	22.386
5	1904	20.202	23.684	27.858	36.052	39.336	40.96	36.361	34.976	35.057	33.679	27.652	22.992
6	1905	19.067	18.747	24.926	32.463	40.702	41.456	37.364	36.597	36.395	34.414	29.145	22.871
7	1906	19.818	21.119	26.654	34.848	41.744	39.982	37.448	34.018	34.668	34.351	29.523	23.963
8	1907	22.682	20.729	25.12	32.103	37.764	39.81	37.98	34.568	36.548	34.469	28.852	22.307
9	1908	21.375	23.63	28.825	36.022	39.766	41.873	35.925	32.98	35.136	33.679	27.456	22.455
10	1909	19.963	21.989	29.135	33.815	39.839	38.76	34.776	33.884	34.581	34.193	29.263	22.21
11	1910	20.804	22.964	28.715	34.219	40.248	39.754	35.137	33.825	35.148	32.972	27.371	22.242
12	1911	21.735	23.569	26.298	34.868	41.044	39.829	38.376	36.509	35.399	34.096	26.107	22.162
13	1912	21.29	24.387	28.438	34.726	40.014	41.622	36.365	33.993	34.169	33.925	27.202	22.232
14	1913	21.858	22.411	25.961	36.298	37.913	37.775	35.148	34.855	35.666	34.165	27.613	21.785
15	1914	22.674	22.117	27.706	34.343	39.373	39.512	35.18	34.805	34.731	31.675	28.143	21.999
16	1915	21.082	21.626	28.823	35.562	42.349	40.504	38.177	35.96	35.974	34.835	28.824	23.006
17	1916	21.921	22.609	31.72	36.953	39.79	39.62	35.433	33.893	34.722	33.768	27.186	22.016
18	1917	20.997	23.907	28.487	31.971	35.532	38.976	34.953	33.635	33.101	32.21	26.618	22.364
19	1918	20.597	23.766	28.34	32.699	41.714	39.594	38.445	35.904	35.79	34.157	28.001	21.896
20	1919	21.075	22.383	29.077	34.216	38.863	42.478	35.72	34.188	35.095	33.681	27.793	21.822
21	1920	21.478	21.844	28.087	34.721	36.36	39.451	36.025	35.263	36.089	34.818	29.402	23.847
22	1921	21.704	24.358	31.868	38.651	43.137	41.408	37.744	34.356	35.002	32.987	28.12	23.861
23	1922	20.234	24.092	30.52	35.9	40.569	39.89	36.304	34.658	34.212	32.896	27.841	22.115
24	1923	20.896	21.774	29.433	36.542	39.262	42.155	35.877	33.364	35.061	32.805	27.819	22.766
25	1924	20.466	22.883	29.917	36.21	36.672	42.48	37.08	34.474	33.503	33.355	27.787	23.131
26	1925	19.242	21.5	29.814	37.618	39.971	38.099	34.477	34.231	35.589	33.416	26.501	22.362
27	1926	21.451	24.272	27.163	31.929	36.959	41.758	36.682	33.617	34.556	33.344	26.183	22.465

Figure 4.1: MET Dataset of the Patiala city

4.2 Snow and Rain Transition Dataset

The dataset is provided by US researchers in collaboration with universities of the US. This detailed hydrometeorological data from the Reynolds Creek (in southwestern Idaho) is consist of water years from 2004 to 2014. Here, the water years are called as the number of years when the water flowing from the mountains to the plain areas. This dataset records air temperature, relative humidity and snow depth from all eleven different locations of the region in the watershed while, solar radiation, precipitation, wind speed and direction, and soil moisture and temperature are available at selected stations.

The motive of this research is to understand the effects of temperature by the climate on a higher elevation. The climate is going warms, the land area of the ice is

decreasing and the rate of the flowing water from the mountains have altered their times. So, to understand this rain-to-snow transition zone this study is very helpful. The dataset of observed study is illustrated in Table 4.2 and the information of the dataset is illustrated in Figure 4.2.

Table 4.2: Snow and rain transition dataset

Variable	Definition	Units	Duration
Date_time	Date and Time	10/26/2012 12:00 to 9/30/2014 23:00	10/26/2012 12:00 to 9/30/2014 23:00
WY	Describe a time period of 12 months for which precipitation totals are measured. (1 October through 30 September)	2013 and 2014 only	-
ppt_a	percentage of precipitation that is in the form of rain, snow, or a mixture of rain and snow.	mm	-
T_g_**	Soil temp. At **cm depth.	°C	Installed at 8 of the 11 sites at 5,20,50,75,100 (cms), depths.
S_m_*	Soil moisture at **cm depth.	m ³ .m ⁻³	Installed at 8 of the 11 sites at 5,20,50,75,100 (cms), depths.
T_a	Air Temperature above the 3m ground surface	°C	-
RH	Relative Humidity above 3m of the ground surface	NA	0.01 to 1
e_a	Water vapor pressure above 3m pressure	NA	-
T_d	Dew point Temperature (It is the temperature at which the air is saturated)	°C	-0.5 °C is considered all snow and above +0.5 °C is considered all rain.

1	Date_time	WY	ppt_a	perc_snow	T_g_5	T_g_20	T_g_50	T_g_75	T_g_100	s_m_5	s_m_20	s_m_50	s_m_75	s_m_100	T_a	RH	e_a	T_d
2	10/26/2012 12:00	2013	0	1	1	2.6	6.9	8.4	9.5	0.17	0.062	0.082	0.072	0.077	3.8	0.44	353	-6.5
3	10/26/2012 13:00	2013	0	1	1	2.7	6.9	8.4	9.5	0.169	0.066	0.086	0.076	0.081	5.3	0.42	374	-5.8
4	10/26/2012 14:00	2013	0	1	1.1	2.6	6.9	8.4	9.5	0.173	0.062	0.082	0.072	0.077	6.4	0.41	394	-5.2
5	10/26/2012 15:00	2013	0	1	1.1	2.7	6.7	8.4	9.5	0.17	0.072	0.092	0.082	0.087	5.9	0.45	418	-4.5
6	10/26/2012 16:00	2013	0	1	1	2.6	6.7	8.4	9.3	0.172	0.062	0.082	0.072	0.077	5.3	0.5	445	-3.8
7	10/26/2012 17:00	2013	0	1	1.1	2.6	6.7	8.4	9.3	0.174	0.066	0.086	0.076	0.081	5.7	0.51	467	-3.2
8	10/26/2012 18:00	2013	0	1	1.1	2.7	6.7	8.4	9.3	0.173	0.062	0.082	0.072	0.077	5.5	0.55	497	-2.5
9	10/26/2012 19:00	2013	0	1	1	2.6	6.7	8.3	9.3	0.174	0.066	0.086	0.076	0.081	5.2	0.59	522	-1.9
10	10/26/2012 20:00	2013	0	1	1	2.7	6.7	8.3	9.3	0.176	0.062	0.082	0.072	0.077	3.3	0.61	472	-3.1
11	10/26/2012 21:00	2013	0	1	1.1	2.7	6.7	8.3	9.3	0.177	0.062	0.082	0.072	0.077	2.2	0.66	473	-3.1
12	10/26/2012 22:00	2013	0	1	1	2.6	6.6	8.2	9.3	0.178	0.066	0.086	0.076	0.081	3.6	0.67	530	-1.7
13	10/26/2012 23:00	2013	0	1	1	2.6	6.6	8.3	9.5	0.178	0.068	0.088	0.078	0.083	3.5	0.68	534	-1.6
14	10/27/2012 0:00	2013	0	1	0.9	2.6	6.6	8.2	9.3	0.178	0.062	0.082	0.072	0.077	3.8	0.67	537	-1.6
15	10/27/2012 1:00	2013	0	1	0.9	2.7	6.6	8.2	9.2	0.178	0.066	0.086	0.076	0.081	3.7	0.66	526	-1.8
16	10/27/2012 2:00	2013	0	1	0.9	2.6	6.6	8.2	9.2	0.178	0.066	0.086	0.076	0.081	3.6	0.64	506	-2.3
17	10/27/2012 3:00	2013	0	1	0.8	2.5	6.7	8.3	9.2	0.177	0.066	0.086	0.076	0.081	3.7	0.6	478	-2.9
18	10/27/2012 4:00	2013	0	1	0.8	2.5	6.6	8.2	9.2	0.176	0.066	0.086	0.076	0.081	3	0.65	493	-2.6
19	10/27/2012 5:00	2013	0	1	0.8	2.5	6.6	8.2	9.2	0.177	0.066	0.086	0.076	0.081	3.5	0.63	495	-2.5
20	10/27/2012 6:00	2013	0	1	0.8	2.6	6.6	8.2	9.3	0.176	0.062	0.082	0.072	0.077	4	0.55	447	-3.7
21	10/27/2012 7:00	2013	0	1	0.7	2.5	6.6	8.2	9.2	0.174	0.066	0.086	0.076	0.081	4	0.52	423	-4.4
22	10/27/2012 8:00	2013	0	1	0.8	2.6	6.5	8.2	9.2	0.175	0.066	0.086	0.076	0.081	4.5	0.48	404	-4.9
23	10/27/2012 9:00	2013	0	1	0.8	2.4	6.5	8.2	9.2	0.175	0.062	0.082	0.072	0.077	7.3	0.39	399	-5.1
24	10/27/2012 10:00	2013	0	1	0.9	2.5	6.5	8.2	9.2	0.173	0.066	0.086	0.076	0.081	6.8	0.39	385	-5.5
25	10/27/2012 11:00	2013	0	1	1	2.5	6.6	8.1	9.2	0.173	0.068	0.088	0.078	0.083	7.4	0.4	412	-4.7
26	10/27/2012 12:00	2013	0	1	1.1	2.5	6.5	8.1	9.3	0.173	0.062	0.082	0.072	0.077	7.7	0.38	399	-5.1
27	10/27/2012 13:00	2013	0	1	1.1	2.4	6.5	8.1	9.3	0.179	0.062	0.082	0.072	0.077	9.1	0.32	370	-6

Figure 4.2: Snow and rain transition dataset

4.3 Michigan University dataset

This is the raw dataset of the harvested crops such as corn, soy, and wheat. The reading values are recorded using GPS Yield monitor, it collects several attributes such as crop flow, total yield value, and moisture of the soil. This dataset is a part of the project of Michigan University.

The aim of this study is to examine the basic ecological relationships in field-crop ecosystems in the US. Researchers address the impact of cropping system management with the use of a special green environment. A recommended green environment is set up for the crop and the behavior of the crop and the soil is navigated. The schema of the dataset is depicted in Table 4.3. The values of this dataset are depicted in Figure 4.3. It shows all the information of this dataset such as different parameters with their numeric values like longitudes and latitude of the sensors, duration of the recordings, moisture value and elevation level of the farm field.

Table 4.3: Michigan University dataset

Variable	Definition	Units	Duration
longitude		Degree	-
Latitude		Degree	-
crop_flow_lb_s	Flow rate of grain passing the sensor	lb/s	-
datetime	Date and time of the sampling	2013-07-24 19:20:04+00	1996-10-29 0:12:06:00 To 2013-07-24 19:20:04:00
Duration	Number of seconds since the last reading	Seconds	1 or 2 Secs
distance_in	Length of the sampled area	Inches	-
swth_wdth_in	Width of sampled area (swath)	Inches	-
moisture	Moisture of grain	NA	-
Status	Header status (up or down)	digit	0 or 1 only
pass_num	A number that is incremented on each pass of the field	number	-
serial_number	The serial number of yield monitoring device (WSN)	number	-
Field	User-assigned designation for sampled field	Predefined units	-
Dataset	User-assigned designation for load within a field	Predefined unit	-
Product	User-assigned designation for the crop being sampled	Predefined units	-
elevation_ft	The altitude of the GPS system	foot	-

1	longitude	latitude	crop_	datetime	duration	distance	swth_wdth_in	moisture	status	pass_num	serial_	field	dataset	product	elevation_ft
2	-85.37282	42.40828	3.29	2013-07-24 19:20:04+00	1	56	180	6.6	0	141	5648	F1: Lysimeter	L1:	Wheat (Sft Rd Wtr)	953.5
3	-85.37284	42.40829	2.89	2013-07-24 19:20:03+00	1	47	180	6.6	0	141	5648	F1: Lysimeter	L1:	Wheat (Sft Rd Wtr)	953.5
4	-85.37285	42.4083	2.89	2013-07-24 19:20:02+00	1	38	180	6.6	0	141	5648	F1: Lysimeter	L1:	Wheat (Sft Rd Wtr)	953.5
5	-85.37286	42.4083	3.69	2013-07-24 19:20:01+00	1	18	180	6.6	0	141	5648	F1: Lysimeter	L1:	Wheat (Sft Rd Wtr)	953.5
6	-85.37286	42.4083	3.39	2013-07-24 19:20:00+00	1	39	180	6.6	0	141	5648	F1: Lysimeter	L1:	Wheat (Sft Rd Wtr)	953.5
7	-85.37285	42.40829	3.99	2013-07-24 19:19:59+00	1	61	180	6.6	0	141	5648	F1: Lysimeter	L1:	Wheat (Sft Rd Wtr)	953.5
8	-85.37284	42.40828	6.39	2013-07-24 19:19:58+00	1	67	180	6.6	0	141	5648	F1: Lysimeter	L1:	Wheat (Sft Rd Wtr)	953.5
9	-85.37284	42.40826	5.89	2013-07-24 19:19:57+00	1	56	180	6.6	0	141	5648	F1: Lysimeter	L1:	Wheat (Sft Rd Wtr)	953.5
10	-85.37284	42.40825	5.59	2013-07-24 19:19:56+00	1	19	180	6.6	0	141	5648	F1: Lysimeter	L1:	Wheat (Sft Rd Wtr)	953.17
11	-85.37284	42.40825	4.49	2013-07-24 19:19:55+00	1	10	180	6.6	0	141	5648	F1: Lysimeter	L1:	Wheat (Sft Rd Wtr)	953.5
12	-85.37284	42.40825	3.49	2013-07-24 19:19:54+00	1	41	180	6.6	0	141	5648	F1: Lysimeter	L1:	Wheat (Sft Rd Wtr)	953.5
13	-85.37284	42.40826	2.49	2013-07-24 19:19:53+00	1	62	180	6.6	1	141	5648	F1: Lysimeter	L1:	Wheat (Sft Rd Wtr)	953.5
14	-85.37284	42.40828	1.89	2013-07-24 19:19:52+00	1	75	180	6.6	1	141	5648	F1: Lysimeter	L1:	Wheat (Sft Rd Wtr)	953.5
15	-85.37283	42.40829	3.39	2013-07-24 19:19:51+00	1	82	180	6.6	1	141	5648	F1: Lysimeter	L1:	Wheat (Sft Rd Wtr)	953.5
16	-85.37283	42.40831	3.29	2013-07-24 19:19:50+00	1	89	180	6.6	1	141	5648	F1: Lysimeter	L1:	Wheat (Sft Rd Wtr)	953.5
17	-85.37283	42.40833	3.29	2013-07-24 19:19:49+00	1	90	180	6.6	1	141	5648	F1: Lysimeter	L1:	Wheat (Sft Rd Wtr)	953.92
18	-85.37283	42.40835	4.39	2013-07-24 19:19:48+00	1	90	180	6.6	1	141	5648	F1: Lysimeter	L1:	Wheat (Sft Rd Wtr)	953.92
19	-85.37283	42.40838	3.39	2013-07-24 19:19:47+00	1	90	180	6.6	1	141	5648	F1: Lysimeter	L1:	Wheat (Sft Rd Wtr)	953.92
20	-85.37283	42.4084	3.89	2013-07-24 19:19:46+00	1	89	180	6.6	1	141	5648	F1: Lysimeter	L1:	Wheat (Sft Rd Wtr)	953.92
21	-85.37283	42.40842	3.39	2013-07-24 19:19:45+00	1	98	180	6.6	1	141	5648	F1: Lysimeter	L1:	Wheat (Sft Rd Wtr)	953.92
22	-85.37283	42.40844	3.39	2013-07-24 19:19:44+00	1	103	180	6.6	1	141	5648	F1: Lysimeter	L1:	Wheat (Sft Rd Wtr)	953.92
23	-85.37283	42.40847	2.99	2013-07-24 19:19:43+00	1	111	180	6.6	1	141	5648	F1: Lysimeter	L1:	Wheat (Sft Rd Wtr)	954.33

Figure 4.3: Michigan University dataset

4.4 IMD Pune Soil-based Dataset

The IMD dataset of Pune is owned by the Indian meteorological department (IMD), Government of India. This department has service of different type of meteorological units such as weather forecasting, climate services, agricultural data, and monsoon reports. The aim of IMD is to maintain the records of information on the several environmental factors and preserve them for future usages. In this dataset, the environmental temperature, soil temperature at different depth of the earth such as 5cm, 15cm and 30 cm, relative humidity, wind direction, rain, and the evapotranspiration are discussed. A schema of this dataset is clearly depicted in Table 4.4 and the information of all attributes is clearly depicted in Figure 4.4. The Name of the variable used by the sensor is in the first column, its definition is in the second column and lastly, the units used for the measurement of the value is depicted in the third column.

Table 4.4: IMD Pune soil-based dataset

Name	Definition	Units
IND	The index number of the station	-
YEAR	The year of the recording value	y
WK	Week number (1 -52)	w
DT	Date	d
MN	Month	m
DBT1	Dry Bulb Temperature	Degree per hour
WBT1	Wet bulb temperature device 1	Degree per hour
DBT2	Dry bulb temperature device 2	Degree per hour
WBT2	Wet bulb temperature device 2	Degree per hour
MAX	Maximum Temperature	°C
MIN	Minimum Temperature	°C
GMIN	Grass minimum temperature	°C
TMP1	Soil temperature at 5 cm depth per hour (device/region 1)	Degree per hour
STMP2	Soil temperature at 15cm depth per hour (device/region 1)	Degree per hour
STMP3	Soil temperature at 30 cm depth per hour (Device/region 1)	Degree per hour
STMP4	Soil temperature at 5 cm depth per hour (device/region 2)	Degree per hour
STMP5	Soil temperature at 15 cm depth per hour (device/region 2)	Degree per hour
STMP6	Soil temperature at 30 cm depth per hour (device/region 2)	Degree per hour
VP1	Vapour Pressure (Device /region 1)	mm (Hg)
VP2	Vapour Pressure (Device / region 2)	mm (Hg)
RH1	Relative humidity (Device/ region 1)	% per hour
RH2	Relative humidity (Device/ region 2)	% per hour

D1 and D2	Wind Direction in 16 points of the compass 00 Calm, 02 N-NE, 05 NE, 07 E-NE, 09 E, 11 E-SE, 14 SE, 16 S-SE, 18 S, 20 S-SW, 23 SW, 25 W-SW, 27 W, 29 W-NW, 32 NW, 34 N-NW, 36 N, 99 VARIABLE	-
SSH	Sunshine of the previous day	Hours
DRF	Rain in the last 24 hours (cycle begins at 08:30)	mm
EVP1	Evaporation (Device / region 1) cycle begins at 08:30	mm
EVP2	Evaporation (Device / region 2) cycle begins at 08:30	mm

1	IND	YEAR	WK	DT	MN	DBT1	WBT1	DBT2	WBT2	MAX	MIN	GMIN	STMP1	STMP2	STMP3	STMP4	STMP5	STMP6	VP1	VP2	RH1	RH2	MWSP	D1	D2	SSH	DRF	EVP1
2	J94	2002	1	1	1	6.8	6.8	12	10.4	13.2	6.6	5	9.5	10.4	13.6	16.5	13.5	13	7.4	8.5	100	81	3.8	0	32	2.5	0	0.5
3	J94	2002	1	2	1	4.8	4.8	13.2	11.6	12.6	4.6	4	8.4	9.4	13	19	15	13	6.5	9.3	100	82	3.2	0	32	0	0	0.7
4	J94	2002	1	3	1	4.6	4.6	15.8	12.4	15	4.4	1.8	8	9	13	19.4	15.5	13.4	6.4	8.7	100	65	1.5	0	32	5.5	0	0.5
5	J94	2002	1	4	1	5	5	16.8	13.8	17.4	4.2	1	8	9	13	18	15	13	6.5	10	100	70	1.7	0	32	6.2	0	0.6
6	J94	2002	1	5	1	2.8	2.8	19.4	14.2	18.6	2.6	-0.4	5.5	7.5	12.5	20	15.5	12.5	5.6	9	100	53	1	0	32	5.5	0	0.5
7	J94	2002	1	6	1	3.8	3.8	18.8	14	20	3.5	-0.4	6.4	8.8	12.6	19	16.5	12.6	6	8.5	100	56	1.2	0	0	6.4	0	0.7
8	J94	2002	1	7	1	4.2	4.2	19.4	13.6	21.2	4	-0.6	5.2	7.2	12.4	19.5	15.5	12.5	6.2	9.2	100	49	0.9	0	32	5.2	0	0.9
9	J94	2002	2	8	1	5.8	5.8	19.2	15	20	5	0.6	5	7	12	20	15.5	12	6.9	9.6	100	61	1.8	0	32	8	0	1
10	J94	2002	2	9	1	5	5	18.6	14.6	19.8	4.5	1.4	8	9	13.5	20.5	16	12.5	6.5	10	100	62	2.1	0	32	8.2	0	1
11	J94	2002	2	10	1	6.8	6.8	19.2	15	19.4	6.2	2	7.6	9	12.5	21.5	17	12.5	7.4	10	100	61	2	0	32	6	0	1
12	J94	2002	2	11	1	6.6	6.6	20	13.8	19.8	6.2	0	5.6	7.6	13	22	17.4	13.4	7.3	8.1	100	46	3.1	0	32	7.3	0	1.2
13	J94	2002	2	12	1	7.4	7.4	21.2	16	20.3	7	0	5.7	8	13	24	18	13	7.7	11	100	55	3.6	0	32	8.8	0	1.3
14	J94	2002	2	13	1	12.6	11.2	18.2	15.6	22.2	9.8	5.2	10.4	11	13	20	15	13	9.1	12	84	75	4.2	14	18	8.5	0	2
15	J94	2002	2	14	1	15.8	13.4	19	15	18.6	15	12.8	13.2	13.4	14.2	20.5	15	15	10	10	75	63	11.7	14	14	0	0	2.2
16	J94	2002	3	15	1	14.4	12	17.6	14.8	19.5	14	12	12.2	12.7	14.2	16.5	15	14	9.1	11	74	72	11.9	14	14	12	0	4
17	J94	2002	3	16	1	13.6	12.8	15	14	18	13	10	12.8	13	15	15	14.5	14.5	11	11	91	89	9	14	0	9	6	1.4
18	J94	2002	3	17	1	9.4	9.4	16.8	15	16.6	9.4	5.5	11.2	11.6	14.2	21	16.5	14.5	9	12	100	82	2.3	0	32	2.3	0	0.7
19	J94	2002	3	18	1	8.6	8.6	17	14	22.2	8.5	5	11	12	14.4	22	18	14.4	8.4	10	100	70	1.9	0	32	1.9	0	0.2
20	J94	2002	3	19	1	5.8	5.2	17.6	13	17.8	5.5	1	7.4	9.5	14	20.5	17.4	16	6.4	8.4	91	56	1.8	0	32	1.8	0	1.7
21	J94	2002	3	20	1	4.4	4.2	18	10.8	18.8	4.4	-2.2	5.2	8	13.6	24	19	14	6.1	5.4	97	35	0.6	0	32	0.6	0	1.4
22	J94	2002	3	21	1	4.2	3.8	16.8	12	19	3.8	-1	5	7.6	13.4	23.6	18	13.6	5.8	7.6	93	53	2.2	0	32	2.2	0	1.7
23	J94	2002	4	22	1	5.6	5.4	17.6	13	18	5.4	0.4	6.4	8.4	13.4	23	18	13.5	6.6	8.4	97	56	2.5	0	32	8.5	0	1.6

Figure 4.4: IMD Pune soil-based dataset

4.5 IMD Pune Weather-based Dataset

There is also another dataset from the IMD, Pune. It has environment-based parameters. It includes serial order of sensors such as a timestamp of each sensor, temperature records, rainfall, evaporation, rainfall, weather type (like drizzle, snow, thunderstorm). This dataset has a term squell i.e., a sudden increase in the wind speed

and lasts for the minutes, this winds highly associated with the active weather, such as thunderstorms, rain showers, and heavy snow. The whole architecture of the dataset is depicted in Table 4.5. The difference between these two datasets of the IMD, Pune is their parameters, the soil based dataset works on the soil properties such as soil temperature at different depths, vapour pressure and the relative humidity while another dataset of the IMD named as environment based dataset consists of environmental parameters such as rainfall, wind, precipitation, and type of weather.

Table 4.5: IMD Pune weather-based dataset

Name	Definition	Units
INDEX	The index number of the station	-
YEAR	Year of the recording date	Y
MN	Month	M
DT	Date	d
MAX	Maximum Temperature	°C
MIN	Minimum Temperature	°C
R/F	Rainfall (Cycle start at 08:30)	mm
EVP DR	Evaporation (Cycle starts at 08:30)	mm
DR NRF	Duration of rainfall	hours: minutes
SSH	Duration of the sunshine	hours
SQ	Total number of squell (Strom / Violent Wind)	number
RA	Weather day with rain	-
DZ	Weather day with drizzle	-
SN	Weather day with snow	-
SL	Weather day with sleet	-
HA	Weather day with hail	-
TH	Weather day with thunder	-
DS	Weather day with dust storm	-

FG	Weather day with fog	-
GA	Weather day with gale	-
T	Type of weather 0 – Lightning, 1 - Haze, 2 – Mist, 3 - Sand/dust storm, 4 – Fog, 5 – Drizzle, 6 - Rain/Showers, 7 – Squall, 8 – Gale, 9 - Thunderstorm, J - Hailstorm, K - Dust fog, L - Line squall, M - Ground frost, N – Dew, O - Snow or sleet	Predefined
G	TIME OF COMMENCEMENT OF WEATHER 1 - 0001 to 0300 hrs IST 2 - 0301 to 0600 hrs IST 3 - 0601 to 0900 hrs IST 4 - 0901 to 1200 hrs IST 5 - 1201 to 1500 hrs IST 6 - 1501 to 1800 hrs IST 7 - 1801 to 2100 hrs IST 8 - 2101 to 2400 hrs IST	Predefined

1	INDEX	YEAR	MN	DT	MAX	MIN	R/F	EVP	DR	DRNRF	SSH	SQ	DZ	SN	SL	HA	TH	DS	FG	GA	T	G
2	6202	1992	1	1	16.6	4.6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2
3	6202	1992	1	2	17	5.4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	3
4	6202	1992	1	3	17.2	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1
5	6202	1992	1	4	13.1	6.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
6	6202	1992	1	5	18.4	7.8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2
7	6202	1992	1	6	19.5	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2
8	6202	1992	1	7	20.6	11	0	0	0	0	0	0	0	0	0	0	0	0	1	0	4	3
9	6202	1992	1	8	19.9	11	0	0	0	0	1	0	0	0	0	0	0	0	1	0	4	7
10	6202	1992	1	9	19.6	12	0	0	0	0	0	1	0	0	0	0	0	0	1	0	4	1
11	6202	1992	1	10	15	5.5	0	0	20	1	2	0	0	0	0	1	0	0	0	9	5	
12	6202	1992	1	11	18.9	11	6.2	0	0	1	0	0	0	0	0	0	1	0	1	0	9	1
13	6202	1992	1	12	11.3	5.2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1
14	6202	1992	1	13	16.6	7.2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	3
15	6202	1992	1	14	16.6	4.6	0	0	0	0	0	0	0	0	0	0	0	0	1	0	4	2
16	6202	1992	1	15	17.1	4.1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	4	1
17	6202	1992	1	16	16.6	3.7	0	0	0	0	0	0	0	0	0	0	0	0	1	0	4	1
18	6202	1992	1	17	19.2	6.6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
19	6202	1992	1	18	20.2	6.7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	3
20	6202	1992	1	19	18.5	3.9	0	0	0	0	0	0	0	0	0	0	0	0	1	0	4	2
21	6202	1992	1	20	20.1	4.3	0	0	0	0	0	0	0	0	0	0	0	0	1	0	4	3
22	6202	1992	1	21	20.4		0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	5
23	6202	1992	1	22	21	5.7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1

Figure 4.5 IMD Pune weather-based dataset

4.6 IIT-K Dataset

This dataset is acquired by the Indian Institute of Technology, Kanpur. Dr. Rajib Maity uses this dataset in the research of soil moisture to correlate the relation of the soil moisture content at multiple depths [22]. The researcher uses the linear model in his experiment. This dataset is composed of the reading of the four different depths such as 10m, 25m, 50m, and the 80m in earth surface from 16 June 2011 16:30 to 15 October 2014 15:51. Initially, this dataset is composed of four different files. Each depth has its own data file. A schema of the raw dataset is depicted in Table 4.6. The date column has the date of the month when the data is recorded and the time has information in the hour format that store the timestamp of the reading. Network and the Station represent the location of the sensor. Depth level depicts the depth level of the soil, that records the readings. The last attribute is the soil moisture that is the important variable, it includes all the readings value of the soil moisture.

Table 4.6: IIT-K Dataset

Name	Definition	Units	Example
Date	Date of the recording	mm/dd/yy	11/22/2011
Time	The recording time of the Day	hh: mm	23:00
Network	Network server	NA	IIT_KANPUR
Station	Sensor Name	NA	IITK_Airstrip
Depth	Depth level of the soil	cm	0.8 m
Soil Moisture	Reading value of the moisture	m/m ³	0.249

The IIT-K dataset is used in this study as experimentation, to increase the efficiency of the model this raw data is processed. The valuable readings of all the depths are merged to single CSV file with respect to the date and time of the readings. The Date and the Time column of the raw dataset are merged to Date_Time and the reading information of soil moisture of all depths are added to the next column of the dataset

such as readings of 10, 25m, 50m, and 80m respectively. The final dataset is described in Table 4.7. The readings of the moisture values are depicted in Figure 4.1. It has the soil moisture readings at certain depths of 10m, 25m, 50m, and 80m from 11-Nov-2011 to 22-Nov-2012. It has four different lines such as green, blue, yellow, and brown, each representing different depth level of 10m, 25m, 50m, and the 80m. The processed dataset is depicted in Figure 4.6, that illustrates the soil moisture at all different depths such as 10 m, 25 m, 50m, and the 80m.

Table 4.7: IIT-K Processed dataset

Date_Time	Soil moisture at 10 m	Soil moisture at 25 m	Soil moisture at 50 m	Soil moisture at 80 m
-----------	-----------------------	-----------------------	-----------------------	-----------------------

1	date_time	Soil moisture at 10m	Soil moisture at 25m	Soil moisture at 50 m	Soil moisture at 80m
2	11/22/2011 1:00	9.2	12.9	18.7	24.9
3	11/22/2011 2:00	9.2	12.9	18.7	24.9
4	11/22/2011 3:00	9.2	12.9	18.7	24.9
5	11/22/2011 4:00	9.2	12.9	18.7	24.4
6	11/22/2011 5:00	9.2	12.6	18.7	24.4
7	11/22/2011 6:00	8.9	12.2	18.3	24.4
8	11/22/2011 7:00	8.6	11.9	17.9	24
9	11/22/2011 8:00	8.9	11.9	17.4	24
10	11/22/2011 9:00	8.9	11.9	17.4	24
11	11/22/2011 10:00	9.2	12.2	17.4	24
12	11/22/2011 11:00	9.2	12.2	17.9	24
13	11/22/2011 12:00	9.2	12.6	17.9	24
14	11/22/2011 13:00	9.6	12.6	18.3	24.4
15	11/22/2011 14:00	9.6	12.9	18.3	24.4
16	11/22/2011 15:00	9.6	12.9	18.3	24.4
17	11/22/2011 16:00	9.6	12.9	18.7	24.4
18	11/22/2011 17:00	9.6	12.9	18.7	24.4
19	11/22/2011 18:00	9.6	12.9	18.7	24.4
20	11/22/2011 19:00	9.6	12.9	18.7	24.4
21	11/22/2011 20:00	9.6	12.9	18.7	24.4
22	11/22/2011 21:00	9.6	12.9	18.7	24.9
23	11/22/2011 22:00	9.6	12.9	18.7	24.9

Figure 4.6: IIT-K processed Dataset

Table 4.8: Summary of all datasets

Sr no.	Dataset	Feature	Records	Duration	URL
1	MET Dataset (2007)	Min Temperature, Max Temperature, Average Temperature, Precipitation, Vapour Pressure, Wet Day Frequency, Evapotranspiration, Cloud Cover	1352 X 8	1983-2002	http://www.india-waterportal.org/met_data/
2	Snow and rain transition dataset. (2018)	Date, time, water year, soil temperature, soil moisture, atmospheric temperature, relative humidity, precipitation.	16909 X 18	26 Oct 2012 to 30 Sep 2014	bit.ly/2S7Q7Be
3	Michigan university dataset. (2017)	Longitude, latitude, yield production, DateTime, duration, field size moisture, status, counter, WSN number, field size, dataset, product, elevation.	767827 X 15	29 Oct 1996 to 24 July 2013	https://lter.kbs.msu.edu/datatables/80
4	IMD Pune soil based dataset. (2016)	Index, week date, month, atmospheric temperature, soil temperature, vapor pressure, relative humidity, evaporation.	5375 X 28	1 Jan 2002 to 31 Dec 2016	Acquired by mail https://drive.google.com/drive/folders/1sCzSPDkd21-t0ScuY_xjKj3MJKDK9YPC?usp=sharing
5	IMD Pune weather based dataset.	Index, year, month, date, max temp., Min. temp., relative humidity, evapotranspiration, type of weather, time of	10101 X 20	1 Jan 1992 to 31 Dec 2008	Acquired by mail https://drive.google.com/drive/folders/1sCzSPDkd21-

	(2016)	weather change.			t0ScuY_xjKj3MJ kDK9YPC?usp=s haring
6	IIT K Dataset. (2016)	Soil moisture at 10 m, soil moisture at 25 m, soil moisture at 50m, soil moisture at 80 m.	345600 X 4	16 June 2011 to 15 Oct 2014	https://www.geo.tuwien.ac.at/insitu/data_viewer/#

As discussed above, it is founding that different datasets have different parameters. Several datasets are computed for this experiment. Initially, the MET dataset of the Patiala, Punjab is acquired from the IMD website. It is evaluated to monitor the temperature as a forecasting unit. This MET dataset does not play any huge role in the precision agriculture-based datasets. because it lacks several other parameters like humidity, moisture, *etc* and it also has multistep in the reading values. Afterward, the snow and rain dataset is used to forecasts the soil moisture. This dataset is as expected as required for this study, but this has reading values of the higher elevated region instead of the values of the plain. This has information about hilly areas that may result with the inaccurate evaluation of the model.

A University of Michigan is also working on the crop recommender system. This study uses soil moisture with the sensor data and other farm-related parameters such as field size, field evaluation, and yield production. The drawback of this dataset is that it includes information about the area of the land and field instead of the environmental attributes. Furthermore, there is a dataset from the IMD is acquired. The soil-based IMD dataset has the soil-based parameters such as soil temperature, vapor pressure, and the humidity from 2002 to 2016 while the other IMD dataset named as IMD Pune weather dataset consists of weather attributes. It has relative humidity, type of weather condition such as hazy, shiny, foggy, *etc.* from 1992 to 2008. This dataset has deficiencies such as lots of unmatched date and time values for combining both the dataset and a high number of missing values.

In the end, a dataset of the soil moisture of different depth like 10m, 25 m, 50 m, and the 80 m is acquired. As discussed above, this dataset is merged to single CSV for better computation of the model. It has only the soil moisture values with date and time of the moisture readings.

Models for Building Soil Moisture Prediction System

This chapter discusses the proposed technique used to forecasts the soil moisture. The study involves the time series based Statistical and deep learning model for performing the experimentation. The soil moisture dataset of multi-depth is used to build the model and the soil moisture is predicted using ARIMA and LSTM model of the time series.

5.1 Time Series

In this section time series analysis-based studies are discussed. Several researchers use this methodology for the early prediction of the information. It uses the historical observation for the forecasting of the real-time data such as weather forecasting, stock price prediction and diesel price prediction are the daily life example that is mainly used by the data analyst. In the time series analysis, there are different types of the model. These several types of time series models are discussed in this section.

5.1.1 Working of the Time series models

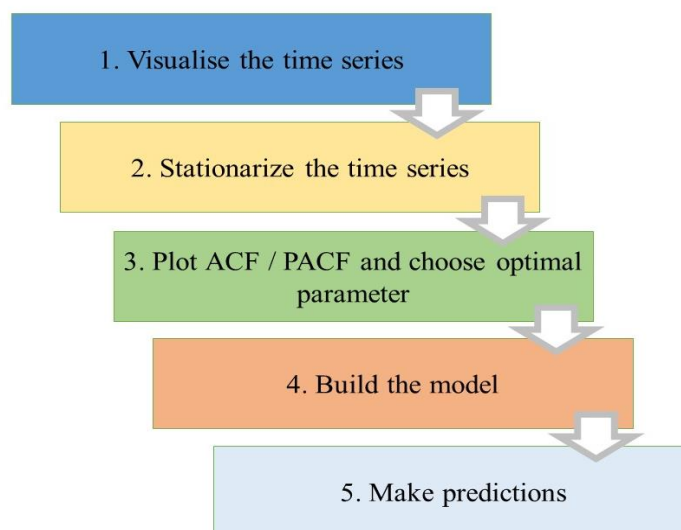


Figure 5.1: Flowchart of the time series

The working of the time series is depicted in Figure 5.1. It summarizes the timeline of the time series modeling. Initially, the time series dataset is analyzed whether it is

stationary or not. This dataset is converted to the stationary dataset from the non-stationary dataset using lagging or differencing order of the dataset to increase the computation level and the accuracy. The ACF/ PACF graphs are plotted to determine the parameter values such as Auto-regressive (p) or Moving Average (q) of the different model used in model building. The model is built using the recommended parameter and the new values are predicted from the model. This above whole process is simply discussed in this chapter.

In time series the data is converted to stationary using two methods such as differencing and the transformation and the more about the ACF and the PACF describe as below:

- **Differencing**

Time series datasets may consist of seasonality and trends, it is required to remove these trends before the modeling of the dataset and make the time series based data to stationary, It has a difference between the data points. Differencing is mainly performed by subtracting the previous observation from the current observation. Assumes a dataset having an original time series is:

$X_1, X_2, X_3, \dots, X_n$

The series with a degree of differencing is equated in the equation (5.1):

$$\begin{aligned}
 \text{If } d=0: x_t &= x_t \\
 \text{If } d=1: x_t &= (x_t - (x_{t-1})) \\
 \text{If } d=2: x_t &= (x_t - (x_{t-1})) - ((x_{t-1}) - (x_{t-2}))
 \end{aligned}
 \tag{5.1}$$

Table 5.1: Differencing order of the sample observations

nth term	Differencing
X1	X2 - X1
X2	X3 - X2
Xn	Xn - X(n-1)

After taking the difference of 1 degree in Table 5.1, plot the series and see if there is any improvement in the Autocorrelation Functions (AFC) curve. If not, then try a

second or even a third order differencing. Remember, the more the degree of difference, the more complicated analysis is becoming.

- **Transformation**

Another method of making time series data to stationary is transforming the variables. The Log transform is probably the most commonly used transformation. The Time series having an exponential distribution can be built to the linear by using the logarithm of the dataset. This is termed as a log transform.

5.1.2 Autocorrelation and Partial Autocorrelation Functions (ACF and PACF)

To determine a proper model for a given time series data, it is necessary to carry out the ACF and PACF analysis. These statistical measures reflect how the observations in a time series are related to each other. For modeling and forecasting purpose it is often useful to plot the ACF and PACF against consecutive time lags.

Autocorrelation Function: The function used to measure the correlation of a Time Series with lags of itself. This is a significant metric because it shows if the previous states (lagged observations) of the time series has an influence on the current state. In the autocorrelation chart, it is used to determine if the time series is stationary or not. A stationary time series will have the autocorrelation fall to zero quickly but for a non-stationary series, it drops gradually. An example of the autocorrelation of the stationary dataset is shown in Figure 5.2.

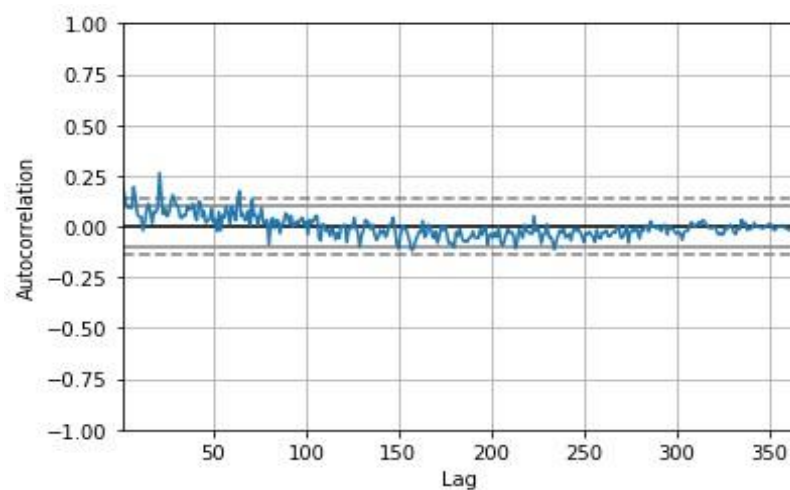


Figure 5.2: Example of Autocorrelation

Partial Autocorrelation is the correlation of the time series with a lag of itself, with the linear dependence of all the lags between them removed. An example of Partial-Autocorrelation is shown in Figure 5.3.

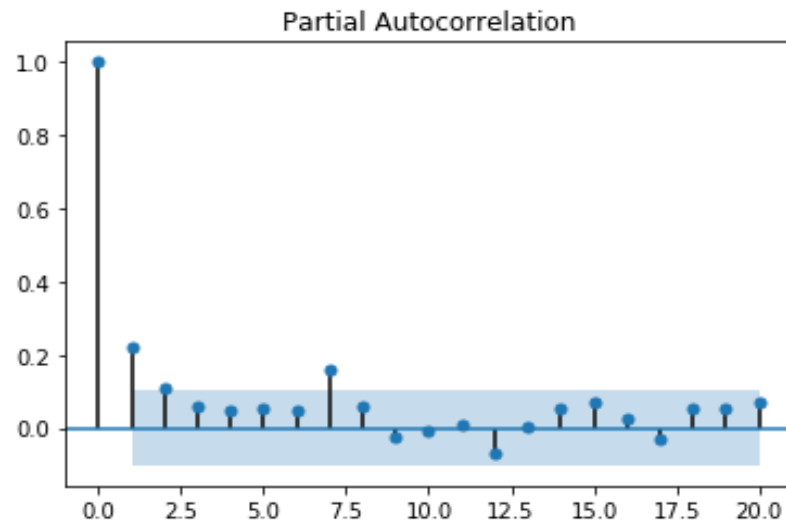


Figure 5.3: Example of partial autocorrelation

Apart from this, these functions are required to build a model. The graphical representation of these functions is analyzed and the recommended values are used as a parameter for a model like ACF function is used as Automatic Regression.

5.2 Models based on Time Series Analysis

There are various types of model in Time series. These models use ACF and the PACF values to fit the data observations. These time series models are described as follow.

5.2.1 Automatic Regressive (AR)

Autoregression is a time series model that uses previous time steps of the data observations as a new input to the data and a regression equation is used to predict the next time step. The automatic regressive is the simplified term that provides accurate results to forecasts a range of data. It is a linear regression term that is applied to the dataset and uses the linear combination of the input data, based on which the output values are determined. The prediction of the data is represented by equation (5.2).

$$\hat{y} = B_0 + B_1 * x_i \quad (5.2)$$

where the \hat{y} is the term to be predicted, B_0 and B_1 are coefficients terms. These coefficients terms are founded by optimizing the model using the training dataset, and x_i is an input value at the i^{th} term. This technique is mostly used with the lag variables. The lag variable is defined as a dataset that takes input variable as observations of the previous time steps. As an instance, the next observations of the dataset is predicted using two previous observations of the dataset and the regression model is used to predict the next observation, An equation (5.3) depicts the working of the autoregressive model.

$$x(t+1) = B_0 + B_1 * x(t-1) + B_2 * x(t-2) \quad (5.3)$$

This regression model is also called as the autoregression (regression of self) when it uses data from the same input variable at previous time steps.

5.2.2 Moving Average (MA)

The Moving Average is a strategy used in the time series to make trending information of the dataset. It includes the average value of the n consecutive terms, where n is a value used to include the number of transactions of the last n terms, that represents as a window's width. It depicts the number of n observation is used to calculate the moving average value. The “moving” part of the moving average refers to the ongoing window width of the data that increments the element for the next n observations by one transaction and decrements the last observation by one.

The moving average method is mostly used to forecasts long-term trends. The two types of Moving Average methods such as a Centered Moving Average and the Trailing Moving Average are discussed below.

5.2.2.1 Centered Moving Average

The Centered Moving Average (CMA) has a centered range instead of the consecutive observations of the data. If one chooses n observation as a dataset, then it has half of the n observation before and after n in the window width. Equation (5.4) is used as an instance to compute the centered Moving Average of three observations.

$$CMA(t) = \text{mean}(\text{obs}(t-1), \text{obs}(t), \text{obs}(t+1)) \quad (5.4)$$

This method generally used to remove the trend and seasonality of the time series dataset. It is rarely used to forecast the observation in the time series analysis.

5.2.2.2 Trailing Moving Average

The trailing moving average (TMA) have an average range of consecutive observations of the data. The chosen window width of n value is used to computes the average of n next observations. An example of this method is represented by equation (5.5).

$$TMA(t) = \text{mean}(\text{obs}(t-2), \text{obs}(t-1), \text{obs}(t)) \quad (5.5)$$

The trailing moving average uses the historical observation to forecast the data on basis of the time series analysis.

In the Moving Average, it has a problem in the predictions that it has an extent of the averages which eliminates the oscillatory fluctuations of the observations. This method seems to become a linear in trend and do not provides the expected outcome of all observations. Hence, this is not an efficient model used for forecasting future trend, i.e., the main objective of the time series analysis.

5.2.3 ARMA

The autoregressive-moving average (ARMA) model is the basic model to analyze the phenomenon of the time series analysis. It is a univariate model of the time series, it consists of two time series model such as Auto-Regression (p) and the Moving Average (q) to forecast the data. The p is the autoregressive term and the q is the value of the moving average. The notation for the ARMA model is represented by the equation (5.6), where p is used to observe the autoregressive value for a data items, similarly, the q is used to compute the moving average of the same dataset. The $e(t)$ is the noise that mainly remains to zero.

$$x(t) = p(t-1) + q(t-1) + e(t) \quad (5.6)$$

However, this ARMA model assumes that the data is in a stationary form, but this case is not ever. In real time datasets, the trends and periodicity exist, so it is required to remove these effects and make the dataset stationary. The differencing method is

used for removal of trends and it may be once, twice or three times, until the series is at the best in its stationary (removing trends or periodicities) of data. This differencing method is added to the ARIMA model.

5.2.4 ARIMA

An ARIMA is an Autoregressive Integrated Moving Average. It is the generalized version of the Autoregressive Moving Average (ARMA) model with an addition of the integration notation. It is one of a statistical model to analyze and forecasts the time series dataset. The ARIMA model is composed of three parameters such as Autoregressive (p), differencing (d), and the Moving Average (q). The Autoregressive model is the number of observations used to compute the autoregressive value of the model, differencing is the number of times that a dataset must be differenced from the original by a lag value, it is also called as the degree of difference. Moving average is the size of the window of the moving average to compute the value of the moving average, it is also called the order of moving average.

This model uses the linear regression model to constructs the ACF and PACF to compute the value for the p and q and is also used to make the dataset stationary by changing the degree of the differencing that helps to remove the trends and the seasonal structure of the data. In ARIMA if a parameter has a zero (0) value, that means it is not required to be used in the model building as shown in equation (5.7).

$$\Delta Y_t = p_i \Delta y_{t-1} + d_i \Delta y_{t-1} + q_i \Delta y_{t-1} \quad (5.7)$$

The equation (5.6) elaborates the mathematically working of the ARIMA model, here p_i includes all the Autoregressive values, d_i has a degree of differencing that manipulate the dataset by a lag order, while the q_i computes the Moving average for the model [23].

5.2.4.1 Working of the ARIMA Model

An Autoregressive Integrated Moving Average model is a regressive and the average based model that depends on a unique variable to forecast the data. The first step in ARIMA modeling is to acquire a dataset for an experiment and pre-process the data using feature selection techniques of machine learning. Afterward, the differencing order of the dataset is computed to make a dataset in stationary form. The

"Stationarity" implies that the dataset remains at a constant level over time. It has a notation of integration, which means differencing the raw data with a lag order. The stationary of the data can be avail by using differencing of the data observations. This allows the model to be independent of the other observations, removes the trend and stabilize the means value of the time-series-based dataset.

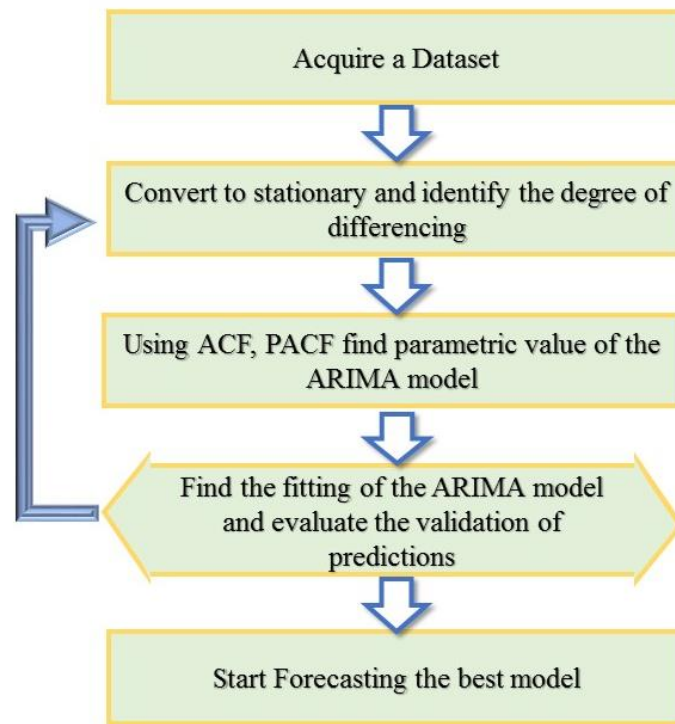


Figure 5.4: Flowchart of ARIMA

After converting a non-stationary data to the stationary data then it is required to fetch the (p, q) parametric values for the ARIMA model. These values can be acquired using the ACF function for the p , and the PACF for the q . where p is used for the Autoregression model and the q is used for the Moving Average model. These (p, q) values are in the range of integers datatype. Like as a sample model the ACF and PACF values are observed in the range 1 to 3 in a dataset, Hence the researcher must use all these parametric values one by one for the model. As an example, the ARIMA $(0, 1, 0)$, ARIMA $(1,2,1)$, ARIMA $(1,1,2)$ are used to build the model using the parametric values and these can be changed for the better performance of the model. These different parametric values help the model to forecasts the best results of observations. A graphical representation of the model is illustrated in Figure 5.4 [38].

If the series has positive autocorrelations in a dataset with a high number of lags (say, 10 or more), then it resulted with a higher order of differencing and if the series has negative autocorrelation then it does not have a higher order of differencing or the series may be over differenced. Moreover, the ARIMA is a univariate model that uses one attribute of the dataset at a single time to build the model for each attribute of the dataset respectively. It cost a system with a high degree of performance in computing the results and analyzing the dataset.

5.2.5 LSTM

The LSTM is the Long Short-Term Memory networks. It is the neural network based model used to predict the value based on the time series analysis. It is based on the Recurrent Neural Network (RNN) that takes a new input every time (one review at the time) and does some processing for each input. The processing of the data is the same for all the inputs and form of a chain of repeating modules of the neural network to the dataset. It has four elements, three gates such as Input gate, Forget gate and the Output gate and the one cell state. This model is briefly explained below.

In Artificial Neural Networks, LSTM is a special kind of neural network that uses short term and long-term memories. The architecture of an LSTM cell is shown in Figure 5.5. This model uses gates to flow data from the input state to the output state. There are three gates (input gate, the output gate, and the forget gate) and one cell state through which the information passes and operations are performed on the model. The new data of present time t is passed to LSTM through Input gate (i). The Forget gate (f) calculates the proportion of previous data to be kept. It receives some input data using a sigmoid function and calculates how much of the current data should be kept. The Output gate (o) is used to obtained results using another sigmoid function [24]. The cell state preserves the data of the previously hidden layer and passes it to the next layer for the processing. In LSTM, each gate has its own activation function. There are three sigmoid functions and two \tanh functions which are used for modeling of data. The h_{t-1} is the output of the previous hidden layer and h_t is the output of the current hidden layer. The X_t is the current input to the model, C_{t-1} is cell state of previous time t and C_t is the current cell state of the model, O_t is the output of the model. These defined states of LSTM are represented by equation (5.8), equation (5.9) and equation (5.10) [35, 39].

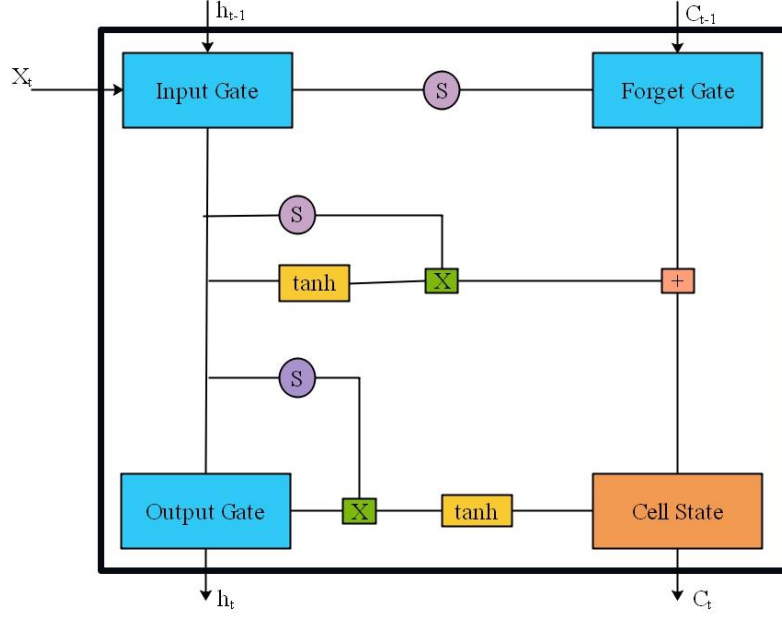


Figure 5.5: Architecture of an LSTM [26]

$$i_t = \sigma (W_{xi}X_t + W_{hi}h_{t-1} + W_{ci}C_{t-1} + b_i) \quad (5.8)$$

$$f_t = \sigma (W_{xf}X_t + W_{hf}h_{t-1} + W_{cf}C_{t-1} + b_f) \quad (5.9)$$

$$c_t = f_t C_{t-1} + i_t \tanh (W_{hc}h_{t-1} + W_{xc}X_t + b_c) \quad (5.10)$$

These above equations representing a function of the input gate (i_t), forget gate (f_t) and the cell state. There is a logistics sigmoid function which is represented as σ . W_{xi} , W_{hi} , and W_{ci} are the weight matrix of the input gate. W_{xf} , W_{hf} , W_{cf} are weight matrix of the forget gate. In Cell state $f_t C_{t-1}$ is the vector multiplication of forget gate and the previous cell state, the $i_t \tanh$ is the tanh function applied on input state (i_t) and W_{hc} , W_{xc} are weight matrix for hidden state and cell state.

$$h_t = o_t \tanh(c_t) \quad (5.12)$$

In the equation (5.12), a \tanh function is performed on the cell state and the resulting information is passed to the forget gate to perform an operation on next t state. The next inputs of the dataset are passed to the LSTM network at each timestamp t for processing of the model as denoted in the equations above. Once the last input of the dataset being processed, the final output for the whole dataset is returned by the network. Hence, this process is iterative for the whole dataset and the model is built using historical dataset [27].

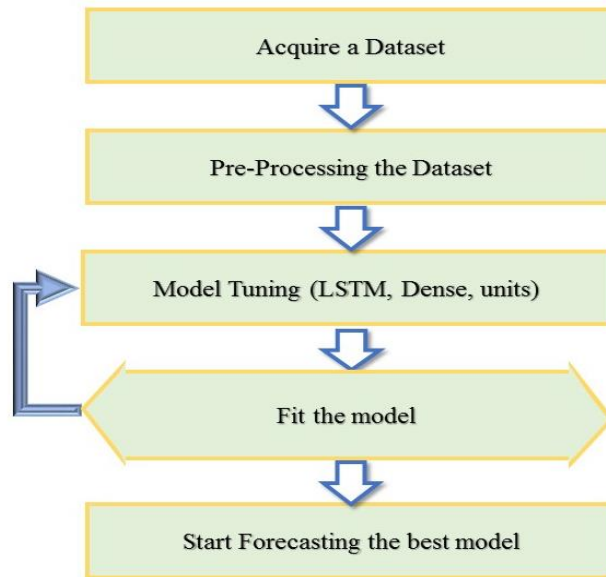


Figure 5.6: Flowchart of the LSTM

5.2.5.1 Working of the LSTM model

An LSTM is a Neural Network based model. It uses the deep learning approaches such as Recurrent Neural Network (RNN) to forecasts the observations. The first step in LSTM is to acquire dataset for the experiment and then pre-process the data using feature selection technique of Machine Learning. After acquiring the dataset, it has been scaled on the scale of 0 to 1 integer for the efficient performance of the model. Further, the scaled dataset has been reshaped to the three-dimensional format to observe the unique variance of the data.

This manipulated dataset is passed to the LSTM to build the model for the forecasting of the observations. The keras library is used for the LSTM, it has several different parameters such as number of LSTM and Dense layers, neurons and the epochs to be used for the model building of this proposed study. The LSTM is tuned with different values of these parameters to achieves the best accuracy of the model. This tuned model is fitted with the dataset of the soil moisture and the observations of the soil moisture are forecasted. The graphical representation of this model is depicted in Figure 5.6. It depicts the whole process of the LSTM model of this study.

5.3 Validation Phase

Validation of the model is very crucial to compute the performance of the model. The validation is done using different statistical models of the performance and evaluation

metrics such as MAE (Mean Absolute Error), MAPE (Mean Absolute Percentage Error), MSE (Mean Squared Error), RMSE (Root Mean Squared Error) are used [28, 40,41]. The reason for computing several metrics is to compare this study to another forecasting model. The Lower value of the error depicts the best model for the experiment. In the experiment, the testing data set is used as a validated dataset and the outputs from the model are used as the predicted dataset.

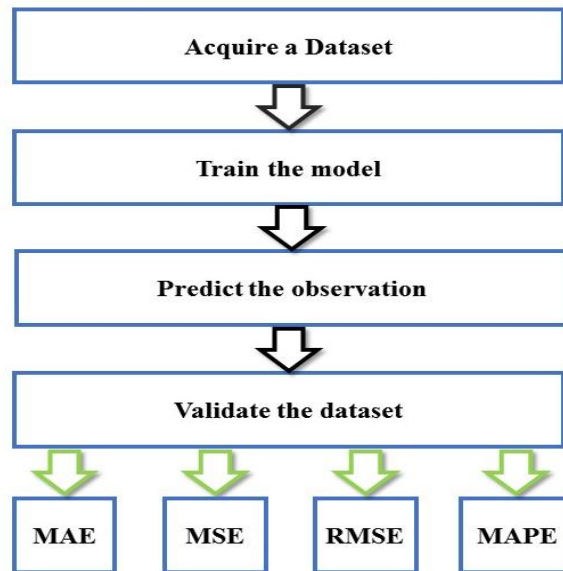


Figure 5.7: Process to perform Validations

$$\mathbf{Ab\ Err} = X_i - Y_i \quad (5.13)$$

In the above equation, i is the reading at t time. Equation (5.13) depicts the absolute error. The *AB err* is the absolute error, it is the difference between the actual data (X_i) and the predicted data (Y_i).

$$\mathbf{MAE} = \frac{\sum_{i=1}^n |Absolute\ Error|}{n} \quad (5.14)$$

Equation (5.14) is the mean absolute error. The MAE is calculated as the average of the absolute error. There are pipes on both ends of the error that makes the nature of this result always positive. The n value in this equation represents the total number of the readings. It calculates the mean average error value of all observations.

$$\mathbf{MAPE} = \frac{\sum_{i=1}^n \frac{Absolute\ Error_i}{Actual_i}}{Actual_i} * 100 \quad (5.15)$$

Equation (5.15) is the error rate in percentage form. The MAPE of each time t is calculated as an absolute error and divided by the time t . The resultant value is further multiplied by 100. The whole dataset values are compiled to depict the percentage error rate of a dataset.

$$\mathbf{MSE} = \frac{\sum_{i=1}^n (\text{Absolute Error})^2}{n} \quad (5.16)$$

Equation (5.16) equated the MSE. MSE is the minimum square error. It is the square mean of the absolute error. The absolute error of all observations of a dataset are simply square and divided by the total number of observations (n).

$$\mathbf{RMSE} = \sqrt{\frac{\sum_{i=1}^n (\text{Absolute Error})}{n}} \quad (5.17)$$

The last equation (5.17) is the root means square of an error value. RMSE is the mean of the square root of differences between actual readings and predicted readings. The results of the MSE error values of all observations are rooted with square root and divided by the total number of the dataset rows.

5.4 Chapter summary

In this chapter, the discussion about the working of the time series model is carried out. Different models of the time series are explored and introduced. The AR and the MA are the basic model of the time series, through which the ARMA and the ARIMA are generalized. The AR uses the autocorrelation (p) value and the MA uses the moving average (q) value of the dataset. In ARMA both these models are combined to increases the computation level and a degree of differencing order is added to the ARIMA model. These above models are statistical models that forecast the soil moisture values. In the view of advanced computing, the deep learning based approach known as LSTM is used in this study to forecasts the soil moisture. The LSTM model consists of neurons, a dense layer and a set of epochs to build the model for the forecasting of the soil moisture. The LSTM model uses a particular percent ratio of the trained observation to build the model. The predicted values of the ARIMA and the LSTM models are compared to the validated dataset. Several performance metrics such as MAE, MAPE, MSE and the RMSE are used to compute the error rate of the model. Hence, the working phenomenon of the times series model and the validation strategy used in the experiment is discussed in this chapter. Moreover, the implementation phases of the proposed study are discussed in the next chapter.

Implementation of Models for Soil Moisture Prediction System

This chapter highlight the implementation of the soil moisture forecasting model. It focuses on the implementation of the proposed work and the recommendation for this experiment. The proposed study uses Artificial Intelligence (AI) based approaches that follow the machine learning and deep learning models to predict the observations of moisture.

6.1 System Components

Several components of software and hardware for this experiment are used, each plays a vital role in the implementation of the models. This proposed study is implemented on a 64-bit based computer system. The processor system is Intel-based i3 with 2.40 GHz speed. The recommended RAM size should be 4 GB or more, the more RAM in size increases efficiency and fasten the results. The minimum requirement of the hard disk size is 320 GB.

Similarly, in the software configuration of this study. Windows 7.0 is used for the implementation of this study. The Anaconda Navigator is used to create a virtual environment for the system, it is like a (Graphical User Interface) GUI platform, that consists the several tools used in the AI such as Spyder, R studio, Pycharm and many more. The Spyder IDE is used to perform computation on the dataset and Python language is used as computation language in the Spyder. The installed Spyder is 3.3.2 version and the Python is 3.6.8. MS Office is used to traverse, process and modify the dataset. MS Excel is used to read and write of the CSV file and MS Word is used to creates the documentation of the proposed study. Notepad++ is used to view the .py files if there is an offline mode of the Anaconda Navigator.

This study uses different libraries to implement the experiment. The following is the list of libraries that are used in this study.

- **pandas:** This package offers the computations on numerical observations. It has several types of function such as plotting ACF /ACF, import/export dataset.
- **os:** This library has functions based on an operating system such as directory-based services.

- **numpy**: It includes the operations on the arrays and the vectors. The multi-dimensional arrays with a large collection of datasets are operated with this package.
- **matplotlib**: This library package is used to plot the graphs, images.
- **sklearn**: It has packages for the such as scaling, performance measuring.
- **statsmodels**: It includes the package for the Statistical operations such as modeling of the ARIMA.
- **keras**: This library has deep learning models such as LSTM, Dense, Dropout. It is based on the Tensorflow.

6.2 Implementation of the ARIMA model

This project is composed of two models i.e., ARIMA and the LSTM. The flow of the ARIMA model is depicted in Figure 6.1 and the LSTM model is discussed in the latter section. The flowchart of the ARIMA includes importing of the dataset, model building, model evaluation using performance metrics such as RMSE, MAE, after the successful evaluation of the model the results are obtained and can be used as important observation of the soil moisture. The results with the predicted observations are stored in the CSV format to the local disk. Figure 6.1 depicts the working of the ARIMA model.

The ARIMA model is a time-series based model. The initial process of the experiment is to pre-process the dataset for working of the model. The raw dataset of the soil moisture with all its depth is converted to the processed data, the unrealistic values are removed from the dataset and all important data values are stored to the single CSV file. The second step is to activate the virtual environment for the ARIMA model. The virtual environment is the collection of required libraries for any type of experiment and model evaluation. It is created using the CMD (Command Prompt) or the anaconda navigator. This virtual environment named as ARIMA consists of statistical packages such as Pyramid and the Statsmodel. The Spyder software is used and it must be called from the virtual environment window. The required libraries are imported using the Spyder IDE for the experimentation.

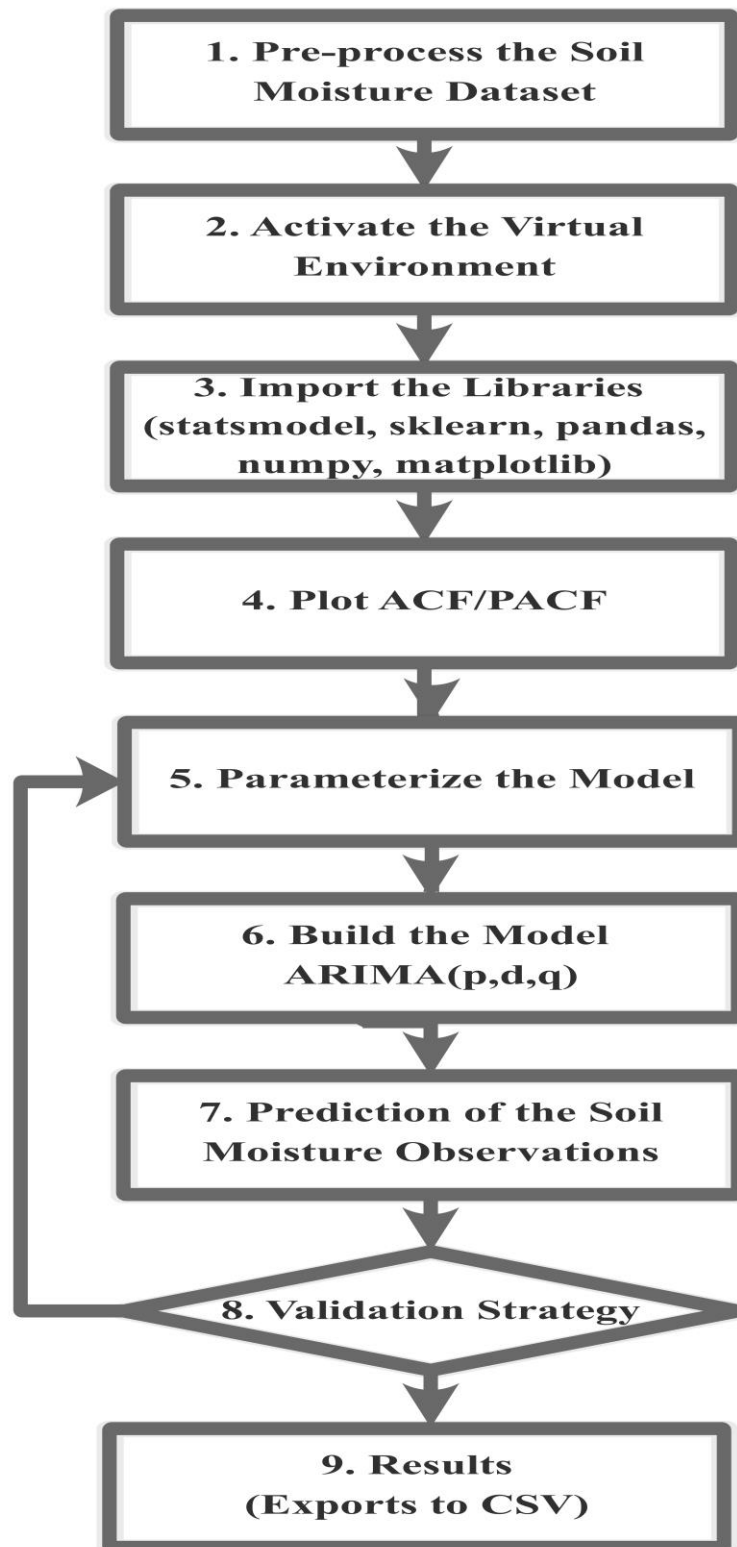


Figure 6.1: Workflow of the ARIMA model

After importing all required libraries, it is required to compute the ACF and the PACF values for the model. In the third step these both values can be acquired using the graphical representation of the dataset. This ACF and the PACF values are simply

plotted using the sklearn and the matplotlib libraries. In the further step, the required parameter for the model such as p from ACF plots and q from PACF plots are configured and these are discussed in the next section. In the next step, these recommended values are used to build the model and this process is called as parametrizing the model. Several iterations of multi-values are parametrized to build the best model. In case of soil moisture at 10 m the parameterized p value is 0 to 2, d value is 1 or 2 and the q value is 0 to 3, the permutations on this fetched information are applied to build the model.

In the last step, these predicted values are retrieved from the built model which are used for the validation of the data to test the error rate of the observations. There are four different performance metrics such as MAE, MAPE, MSE, and RMSE. The lower value in the error are rate is computed as the best accuracy of the model. If the resulted error rate is high then it is suggested to parametrize the model and repeat the previous steps to build the model. The resulted soil moisture values are stored to the local drive in the format of the CSV file.

6.3 Parametrize the ARIMA model

The parametric values of the dataset have been computed for the different level of the soil moisture as follow:

Soil moisture at 10 m

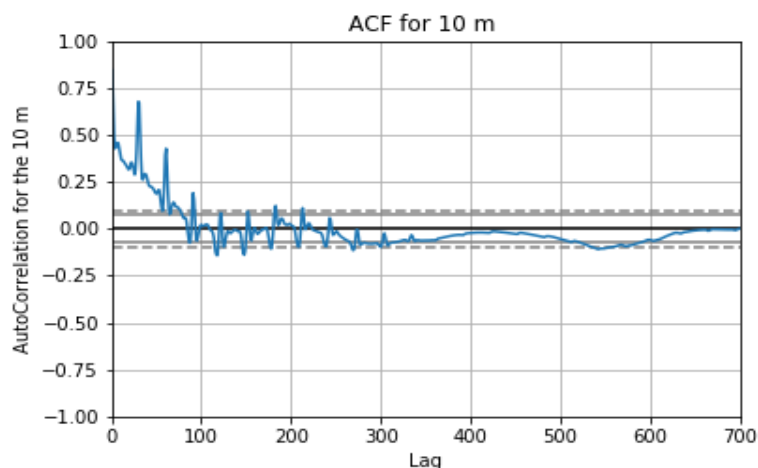


Figure 6.2: Autocorrelation plot of the soil moisture at 10m

In Figure 6.2, the ACF plot of the soil moisture at 10 m is depicted. It is shown that the recommended ACF values are 0, 1 and 2 for the soil moisture at 10 m to build the model.

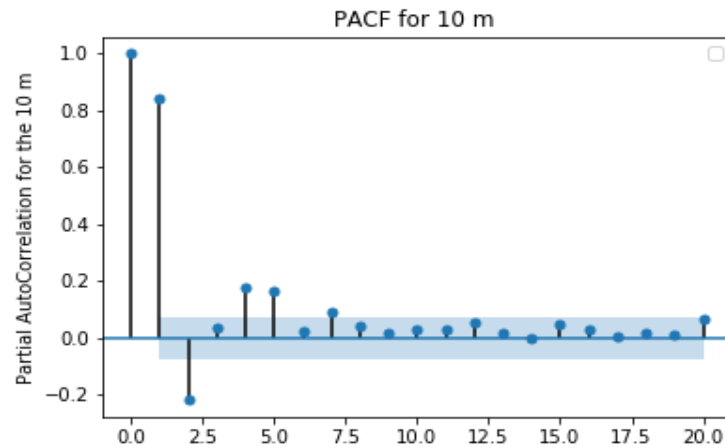


Figure 6.3: Partial-Autocorrelation plot of the soil moisture at 10m

The above Figure 6.3, the PACF plot of the soil moisture at 10 m is depicted. The observed PACF values are 0,1,2 and 3 for the soil moisture at 10 m, In PACF as the number of lags increases, it counts as a recommended observation for a dataset.

Soil moisture at 25 m

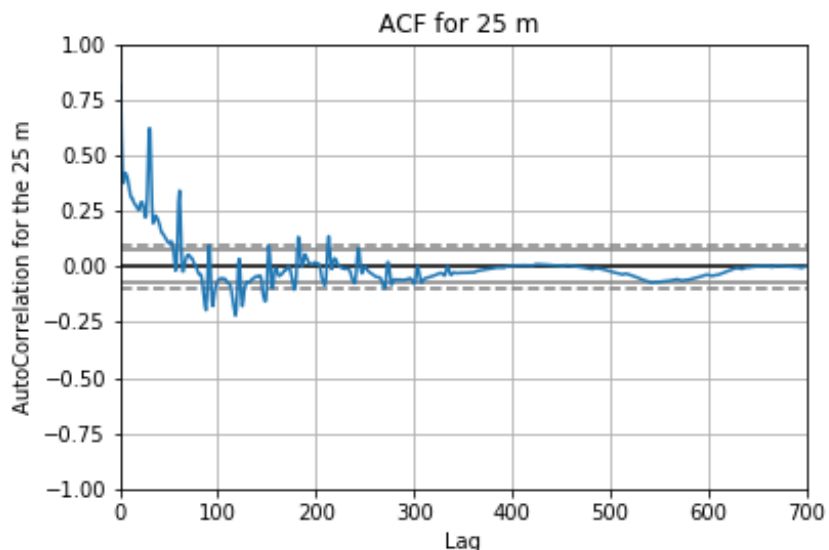


Figure 6.4: Autocorrelation plot of the soil moisture at 25 m

In Figure 6.4, the ACF plot of the soil moisture at 25 m is drawn. It determines that the AR value for the ARIMA model of 25 m depth of the soil moisture must be 0 or 1.

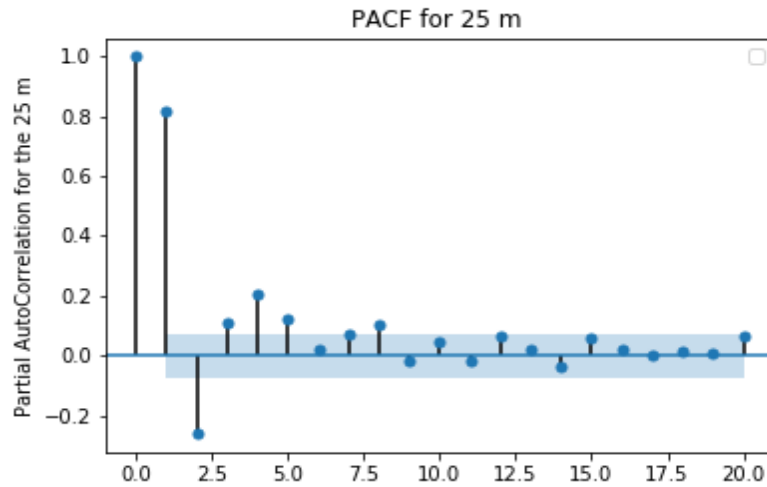


Figure 6.5: Partial-Autocorrelation plot of the soil moisture at 25m

In Figure 6.5, the PACF plot of the soil moisture at 25 m is illustrated. The observed value for the PACF or MA(q) is 0,1,2 and to build the model for the soil moisture at 25 m.

Soil moisture at 50 m

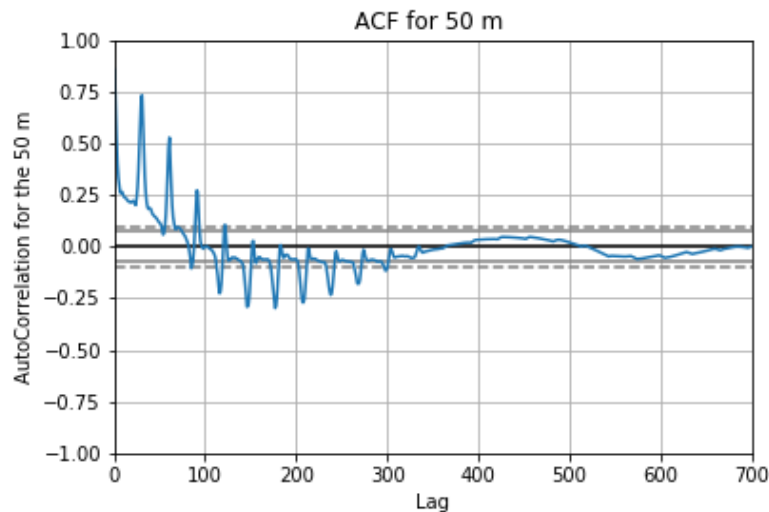


Figure 6.6: Autocorrelation plot of the soil moisture at 50 m

The ACF diagram of the soil moisture at 50 m is illustrated in Figure 6.6. It shows that the ACF values are 1 and 2 for the soil moisture at 50 m.

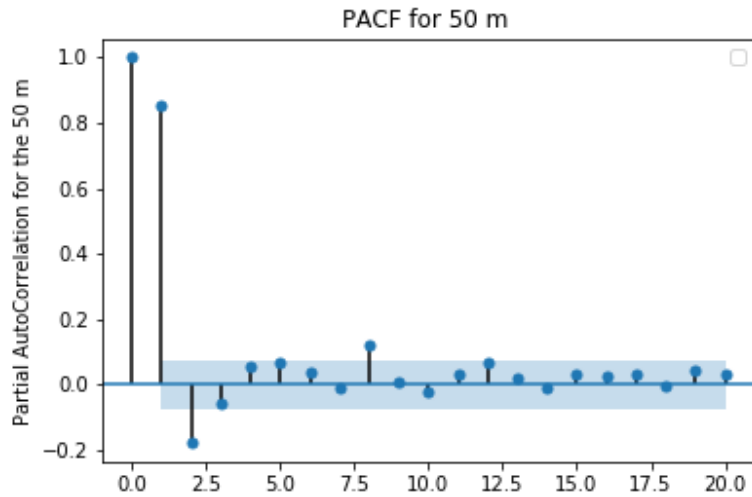


Figure 6.7: Partial-Autocorrelation plot of the soil moisture at 50 m

Figure 6.7 shows the graphical representation of the PACF for the soil moisture at 50 m. It depicts that the recommended PACF values for ARIMA model of soil moisture at 50 m depth are 0,1 and 2.

Soil moisture at 80 m

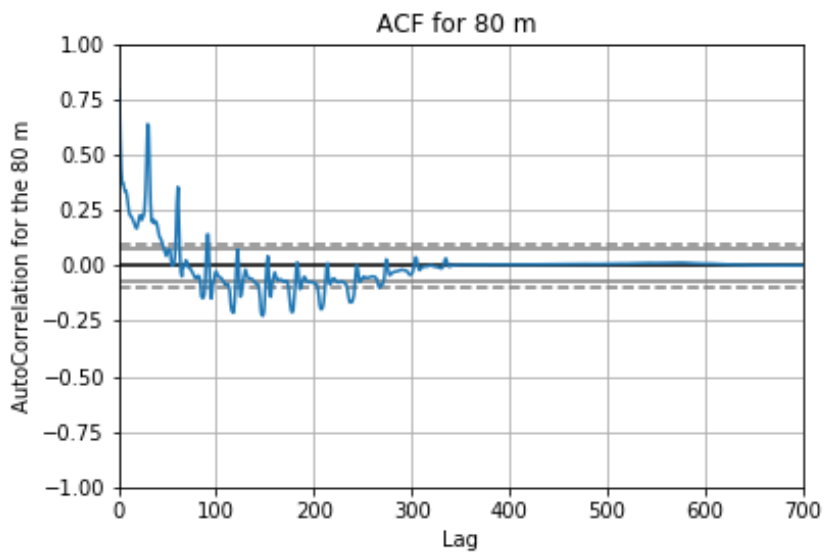


Figure 6.8: Autocorrelation plot of the soil moisture at 80 m

The ACF plot for the soil moisture at 80 m is depicted in Figure 6.8. The fetched values for the AR(p) are 0 and 1 for the soil moisture at 80 m.

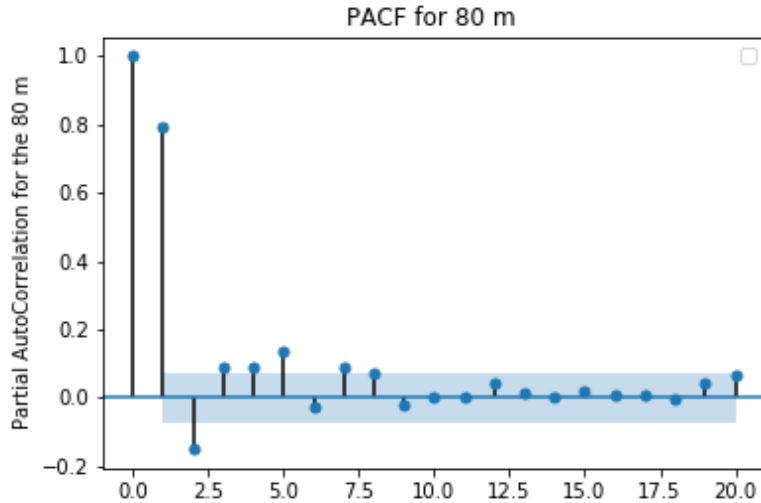


Figure 6.9: Partial-Autocorrelation plot of the soil moisture at 80 m

The PACF (q) plot of the soil moisture at 80 m is depicted in Figure 6.9. The observed values from the PACF plot are 0, 1 and 2 for the soil moisture at 80 m. These values are input to the ARIMA model with the training dataset and used to build the model. The above figures have different correlation and partial autocorrelation, for better understanding, it is illustrated in Table 6.1.

Table 6.1: Parametric values of ACF and PACF for this experiment

Name	Autocorrelations		Partial Autocorrelations	
	Min	Max	Min	Max
Soil moisture at 10 m	0	2	0	3
Soil moisture at 25 m	0	1	0	3
Soil moisture at 50 m	0	2	0	2
Soil moisture at 80 m	0	1	0	2

In the model, the ARIMA has 3 parameters as depicted in the above section. It has “ p ”, “ q ”, and “ d ”. The p is the auto-correlation value for the dataset, q is the moving average value and the last d is the lag value that depicts the number of gaps allowed for the dataset to make dataset stationery. This model is also trained with different

parameters by changing the values of its parameter. The last and best values used for the model building is depicted in Table 6.2. In the end, these parameters depict better performance than others. Hence, these are used for the model building.

Table 6.2: Values of p, q, and d used in ARIMA

Caption	p	d	q
Soil moisture at 10 m	1	2	1
Soil moisture at 25 m	1	1	1
Soil moisture at 50 m	1	1	2
Soil moisture at 80 m	1	2	1

An ARIMA model for each depth layer is trained separately with different parameters using training data of the acquired dataset and the predictions from the model are processed. These predictions are inversely transformed to the normal dataset for the actual values and saved into a new variable. In the final step, the predicted values have been compared with the testing dataset to compute the error rate of the model. The whole process of this experiment is elaborated in Figure 2 and the results of this study are depicted below.

6.4 Implementation of the LSTM model

The LSTM is a deep learning based approach. It is an advanced methodology used in the field of computer science. In the case of the LSTM model, the initial step of the model is the same as the ARIMA model to convert the raw data to the processed data for the best utilization of the system resources. The nonrequired values are removed and the required attributes are merged to the single file in the format of CSV. The second step is the installation of the TensorFlow. The LSTM Model of the TensorFlow is installed using the virtual environment of the disk. The required packages such as theano, tensorflow and the keras are installed in this virtual environment. The third step is the processing of the dataset. The merged dataset is scaled to the 0 to 1 float values. The scaled dataset is used by the LSTM model to predict the soil moisture.

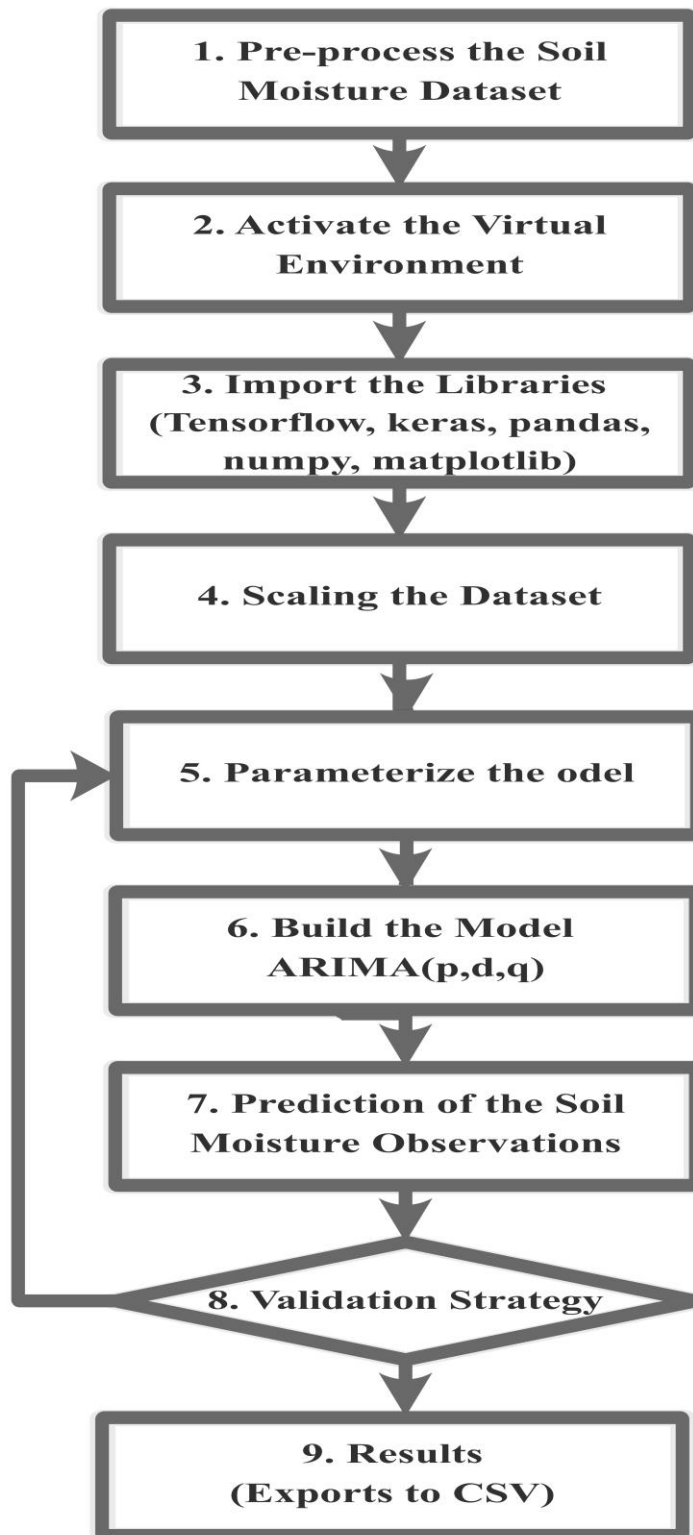


Figure 6.10: Workflow of the LSTM model

The further step is the parameterizing the LSTM model. It is parameterized with different sets of iterations, several experiments are computed to find an optimal value for the LSTM parameters. This parameterizing of the LSTM model is discussed in the

next section. In the next step, the best values of the parameter are used to build the model and predict the observations of the soil moisture. These predicted values are further compared with the testing dataset in the last step to compute the error rate of the model. The MAE, MAPE, MSE and the RMSE are used as performance metrics to compute the error rate. The lower value of the error rate is computed as the best value and the higher error rate is called as insufficient of the model and it required to re-parametrizing of its parameters. The observation with the best model is saved to the local disk of the computer for the future use in the CSV format. A workflow of the LSTM model is depicted in Figure 6.10.

6.5 Parametrize the LSTM model

The implementation of the experiment requires a different parameter used in LSTM to perform the operation. An LSTM model has layers such as the LSTM layer and the Dense layer. It also has a Dropout rate and the number of epochs required to fit the model. These important parameters for this experiment are computed with the number of experiments on this study and tuned with the parameter having the maximum accuracy. These several tuned parameters are summarized in Table 6.4.

Table 6.3: Parameter for the LSTM model

Name	Unit(s)
LSTM layer	1
Neurons in LSTM layer	50
Dropout Rate	0.20
Dense Layer	2
Epochs	125

The LSTM layer is the important layer that used to acquire the model from the keras and perform operations on the dataset. It consists of Tensorflow packages as a backend for the implementations and it also has a number of neurons to performs the operations. The dropout rate is constant for the whole dataset, it is used to remove the noise created during the computation of the data values. This 0.20 dropout rate is the success rate for this experiment. The dense layer is termed as the hidden layer that

follows the Forget gate to perform the operations. This experiment has two number of dense layer. The Epochs are the number of iterations for the dataset. At each epoch, the random observations are fetched by the LSTM model and used to make a strong and successful model. After the implementations of the model, it is required to fetch the results of the soil moisture based observations.

6.6 Chapter summary

In summary, this chapter describes the implementation of both models ARIMA and the LSTM. In the initial section, the implementation of the model is carried out. The ACF and the PACF values are computed for the ARIMA model. Table 6.2 depicts the parametric values of the ARIMA model. different depth level of the soil moisture based dataset uses different values of p , d , q , and an ARIMA model is built with the best value for the forecasting of the soil moisture.

Apart from this, the LSTM modelbuilding is discussed in the chapter. The LSTM model is trained with a different iteration of the parametric value. The best observed values are depicted in Table 6.4. It has one LSTM layer with 50 neurons, two dense layers, and epochs of 125 units to build the best LSTM model. The results of this study are discussed in the next chapter.

7.1 Evaluation of results for ARIMA model

This model is implemented in the Python 3.6.8, and Spyder IDE is used to build the model. The updated model is built in the i3 processor, 4 GB RAM, and 32-bit Windows Operating System. Initially, the dataset is imported from the local disk in CSV format. The *statsmodel* library is used to decompose the data into seasonal and monthly import the ARIMA model and the pyramid library is used to import the ARIMA model. It is further split into training (70) and testing dataset (30), where training dataset has the readings taken from 11-Nov-2011 till 11-Nov-2012, and the values are recorded from 8-Oct-2012 to 22-Nov-2012 are used for the testing of the forecasting model.

An ARIMA model for each depth layer is trained separately with different parameters using training data of the acquired dataset. These predicted values have been compared with the testing dataset to compute the error rate of the model. the results of these experiments are analyzed below.

The results of the ARIMA model are illustrated in Figures from 7.1 to 7.4. These results are obtained using the matplotlib of the python. These Figures have two lines to compare the datasets. One line is green in color and it describes the actual values of the dataset, similarly, the red line in the graph illustrates the predicted values of the dataset. These values are predicted by the model. The range of the values in all depths of the soil moisture ranges from 25 m/m³ to 45 m/m³.

There are four different depths of the soil moisture and each has its own reading values. The first graph in Figures depicts the output of the soil moisture at 10m. This compares the actual and predicted values and give a minimum absolute percentage error is 0.38. Similarly, the next graphs depict the soil moisture forecasted and the actual reading at different depths such as 25 m, 50 m, and 80 m respectively. Each depth has its own error rate between actual and the predicted values and a tabular representation of all error rate are depicted in Table 7.1.

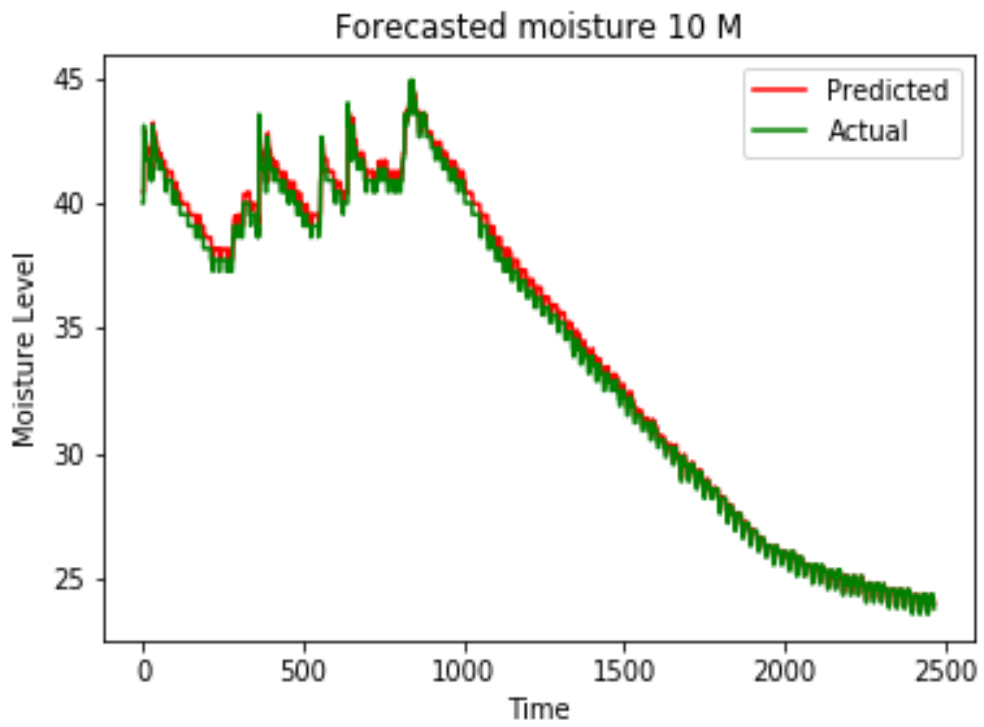


Figure 7.1: Actual and Predicted soil moisture at 10 m

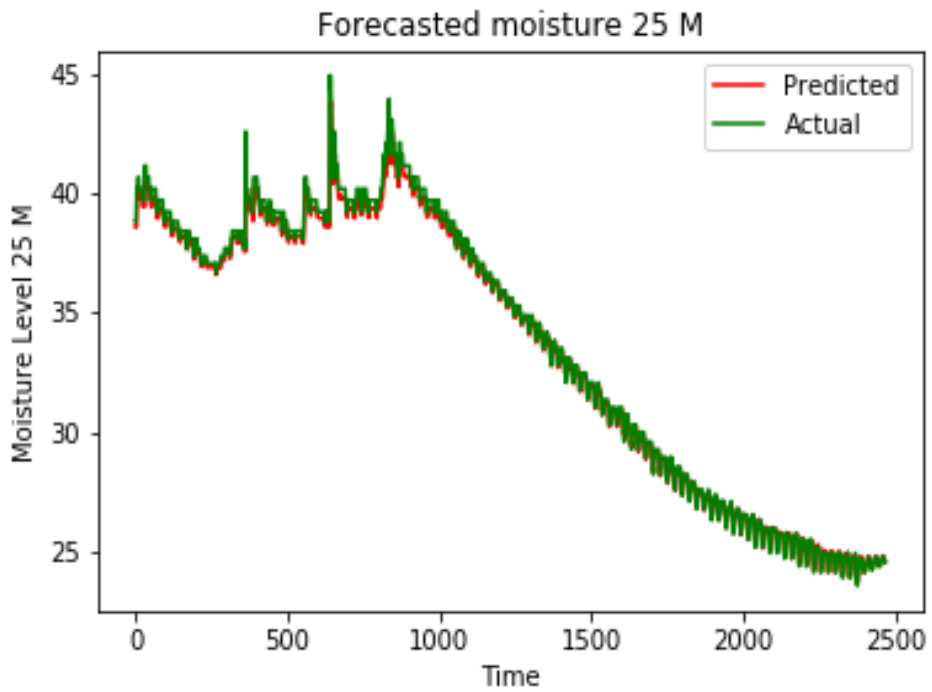


Figure 7.2: Actual and Predicted soil moisture at 25 m

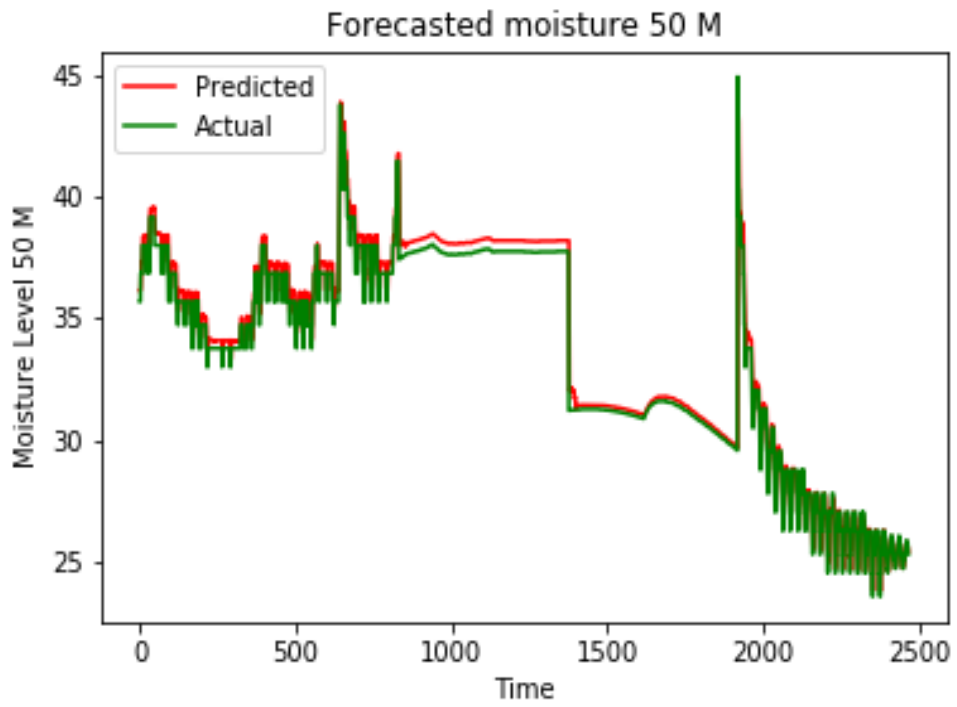


Figure 7.3: Actual and Predicted soil moisture at 50 m

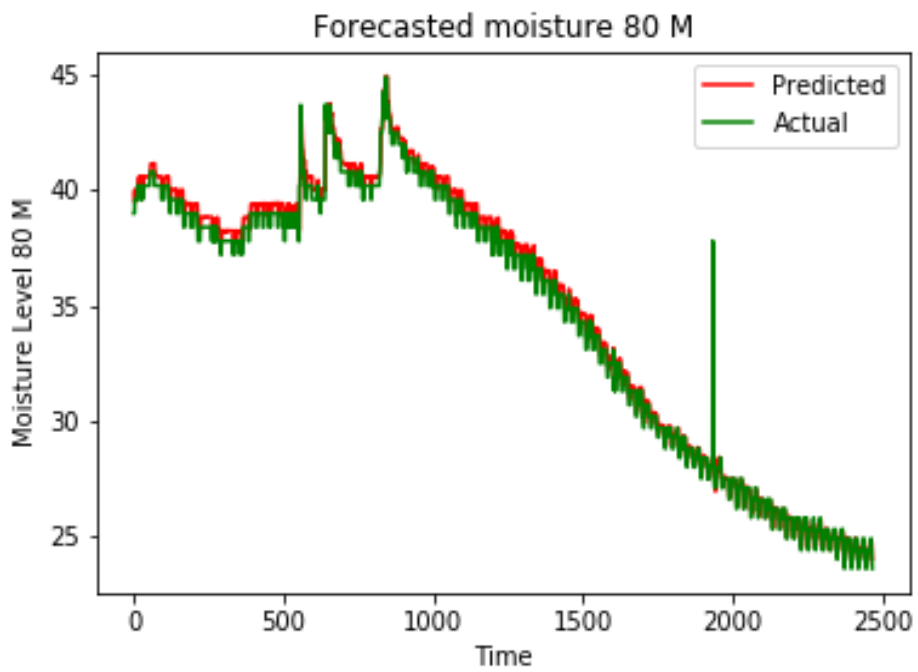


Figure 7.4: Actual and Predicted soil moisture at 80 m

Table 7.1: Performance metrics evaluation of ARIMA

Depth Level	MAE	MAPE	MSE	RMSE
Soil moisture at 10 m	0.13	0.29	0.08	0.28
Soil moisture at 25 m	0.16	0.49	0.19	0.43
Soil moisture at 50 m	0.17	0.53	0.30	0.54
Soil moisture at 80 m	0.16	0.47	0.21	0.45
Soil moisture of all depths*	0.16	0.44	0.19	0.35

Table 7.1 depicts the different error rate of the soil moisture using the ARIMA model. The overall error rate is computed using the mean values of all different depths. The MAE is 0.16, MAPE is 0.44 and the MSE is 0.19. This error rate is measured in the lowest terms. The lowest value of the error illustrates the best efficiency of the model and it also helps to compare this model to another model.

7.2 Evaluation of results for LSTM model

The forecasting model is implemented in Python 3.6.5, and Spyder IDE is used for the development of the model on a system with *i3* processor, *4GB* RAM, and *64-bit* Ubuntu Operating system. The dataset is imported as same as in the ARIMA model. The required libraries and directories which have been used for the model building such as *numpy*, *pandas*, and *sklearn* that are used for the arrays and dataset processing, *matplotlib* for graph plotting. *Tensorflow* and *Keras* libraries are used for model building are imported. The fetched dataset is processed using *MinMaxScaler* function on the scale from 0 to 1 float values. After scaling the acquired dataset, it is further split into training (70) and testing dataset (30), where training dataset has the readings taken from 11-Nov-2011 till 11-Nov-2012, and the values recorded from 8-Oct-2012 to 22-Nov-2012 are used for the testing of the forecasting model.

In the model, the *Keras* library uses *Tensorflow* as a backend to perform the operations. The *keras* has the *LSTM*, it uses 50 units of neurons of the LSTM layer and has two dense layers with 1 neuron and an epoch value of 125 is used for model fitting.

This model is also trained with different parameters by changing its number of neurons, layers, and epochs. In the end, these parameters depict better performance than others. Hence, these are used for the model building.

An LSTM model is trained using training data of the acquired dataset and the predictions from the model are processed. These predictions are inversely transformed to the normal dataset for the actual values and saved into a new variable. In the final step, the predicted values have been compared with the testing dataset to compute the error rate of the model.

The result of the LSTM model is computed using the graphical representation of the dataset. The actual and the predicted values of the soil moisture at different depths are illustrated in Figure 7.5 to 7.8. The figure illustrates the soil moisture at 10 m depth. The moisture values it predicts remains in the range of 15 to 40 m^3/m^3 . The minimum absolute error is 0.11. This low error rate depicts that actual and predicted values have less change in their values. Moreover, graphical representation of other soil moisture depths such as 25 m, 50 m, 80 m are represented in the below figures and the error rate of all depth is summarized in Table 7.3.

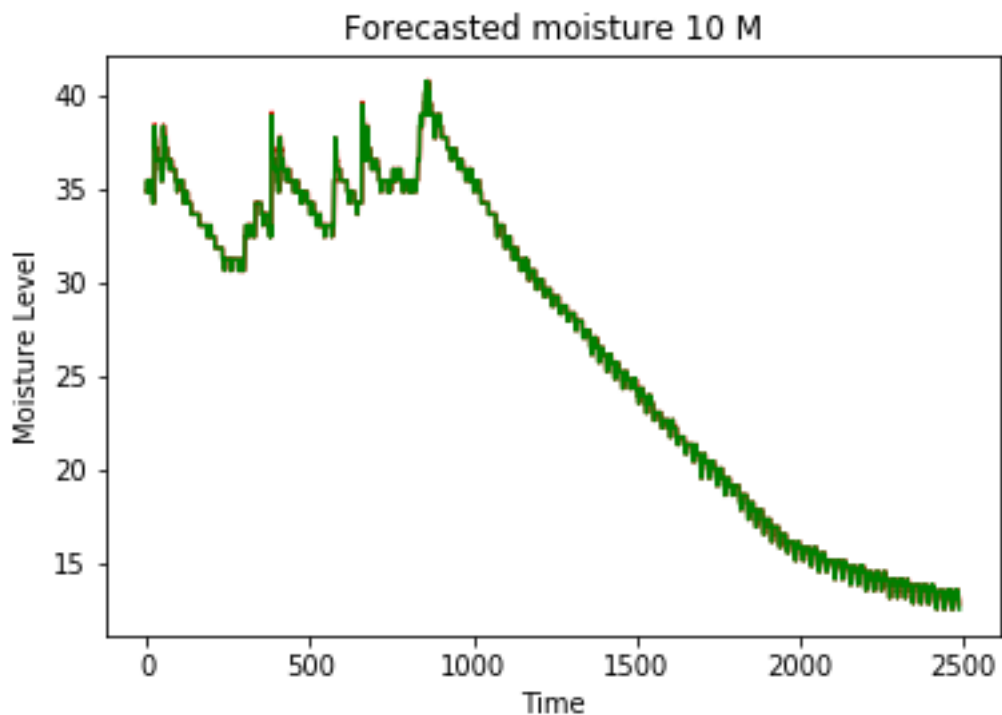


Figure 7.5: Actual and Predicted soil moisture at 10 m.

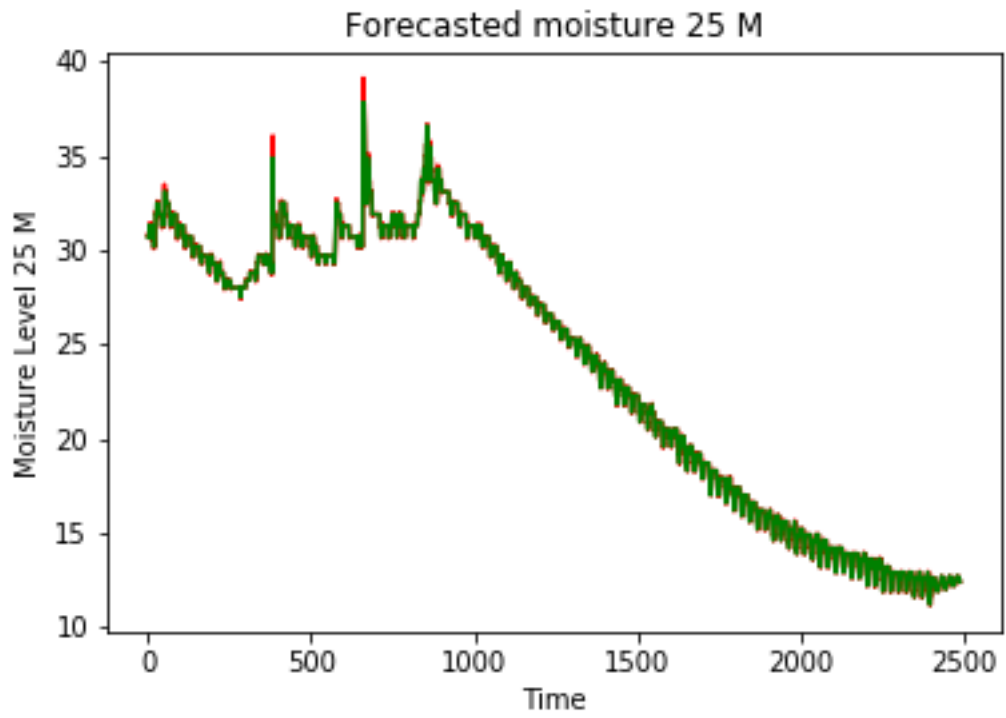


Figure 7.6: Actual and Predicted soil moisture at 25 m

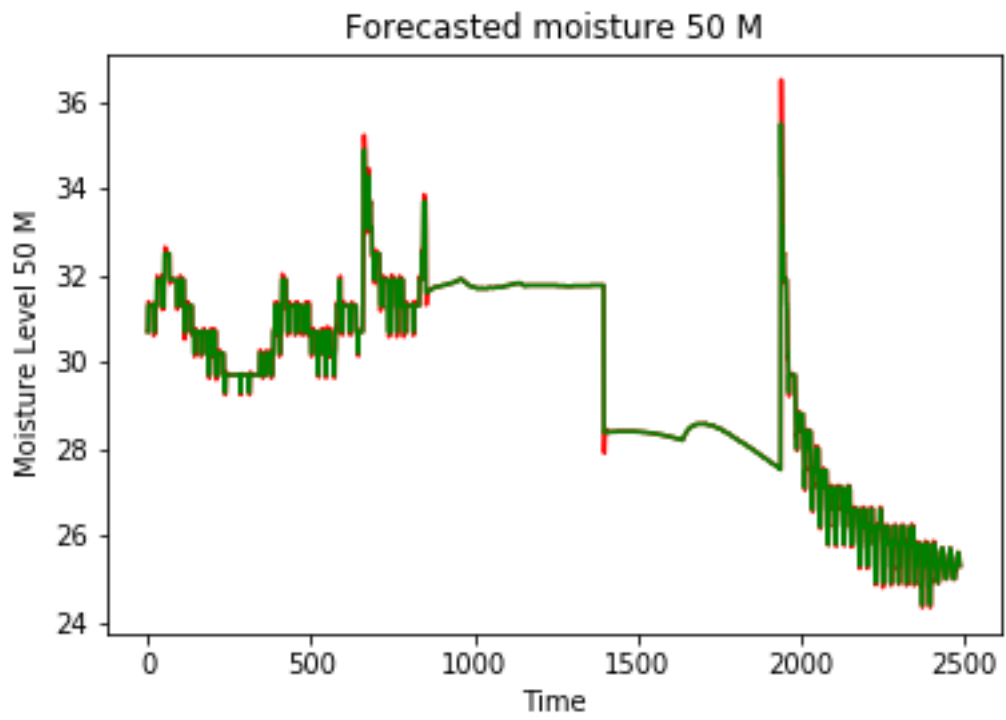


Figure 7.7: Actual and Predicted soil moisture at 50 m

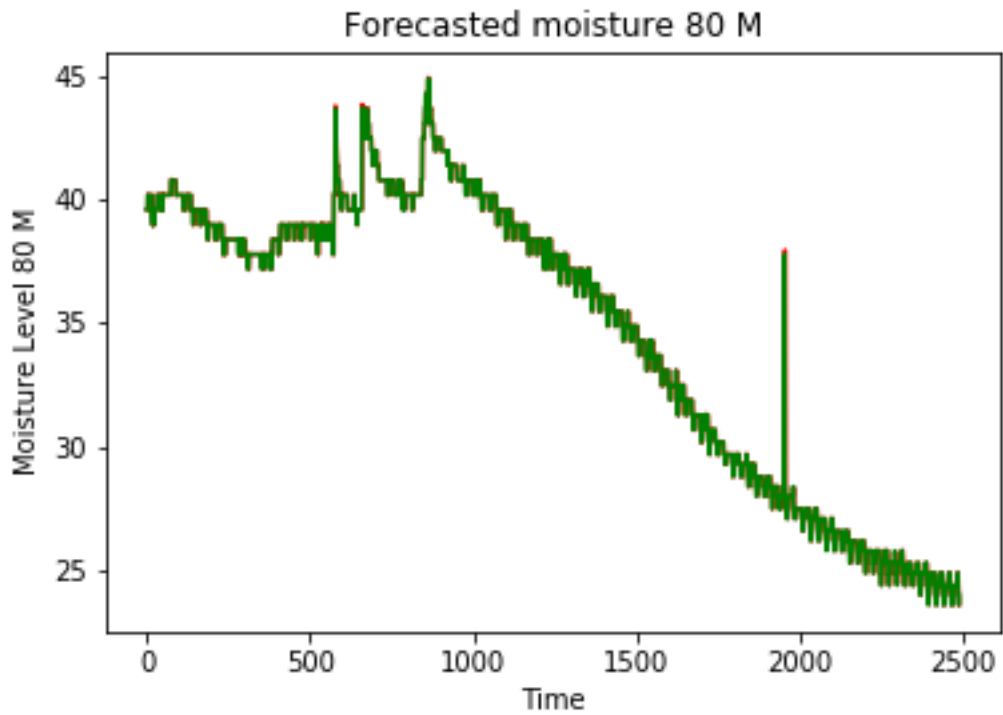


Figure 7.8: Actual and Predicted soil moisture at 80 m

Table 7.2: Performance metrics evaluation of LSTM

Depth Level	MAE	MAPE	MSE	RMSE
Soil moisture at 10 m	0.08	0.24	0.05	0.22
Soil moisture at 25 m	0.10	0.29	0.06	0.25
Soil moisture at 50 m	0.13	0.40	0.23	0.47
Soil moisture at 80 m	0.12	0.36	0.14	0.38
Soil moisture of all depths*	0.11	0.32	0.12	0.33

Table 7.2 depicts the error rate of the soil moisture at different depths using different performance metrics such as MAE, MAPE, MSE, and RMSE. The soil moisture of all depths is computed using the mean of all error rate at different depths of the soil

moisture. MAE is 0.12 for all depth level similarly the MAPE i.e., the percentage measurement of the data is 0.46. and the MSE of all depth is 0.12.

7.3 Comparison analysis of the ARIMA and LSTM

The results of the ARIMA and the LSTM models are acquired with the best accuracy of its tuned parameters. Table 7.1 depicts the error rate of the soil moisture with ARIMA model using the different performance metrics such as MAE, MSE, MAPE, and the RMSE, similarly, Table 7.2 discusses the value of the error rate using the same performance metrics for the LSTM model. It is lucid that the LSTM model has a low error rate as compared to the ARIMA model. On comparing the two tables of both the models, it has been observed that the MAE is 0.11 for the LSTM model, while the ARIMA has an error rate of 0.16 in MAE. Secondly, the MAPE is 0.32 of the LSTM model and 0.44 for the ARIMA model, which depicts that the LSTM model performs better over the ARIMA model. Moreover, the MSE of the LSTM model is 0.12 and the ARIMA has 0.19 as an error rate. The lower the MSE in error rate depict the best accuracy of the model.

It can be concluded that the results obtained during the experiment the ARIMA model is not as efficient as the LSTM model. The LSTM is attaining the low score in error rate of performance metrics, which depicts that this model is better to analyze and forecast the data based on the multi-depth data of the soil moisture. A tabular representation of the mean error rate of the soil moisture observation is depicted in Table 7.3.

Table 7.3: Comparison of the error rates of ARIMA and LSTM

Model	MAE	MAPE	MSE	RMSE
ARIMA	0.16	0.44	0.19	0.35
LSTM	0.11	0.32	0.12	0.33

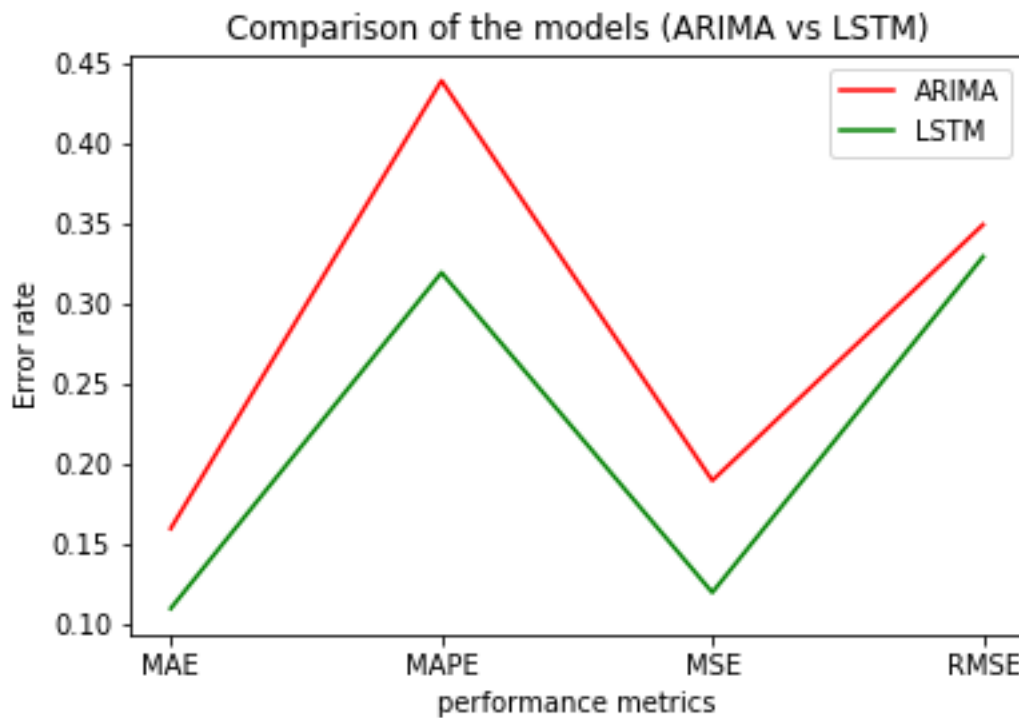


Figure 7.9: Graphical comparison of ARIMA and the LSTM

The graphical representation of all performance metrics is depicted in Figure 7.9. The green line is the performance of the LSTM model and the red line is the ARIMA model's performance. The error rate in low value depicts the best efficiency of the model. Hence, this graphical representation shows that ARIMA has high in error rate and the LSTM is well efficient in its performance.

7.4 Chapter Summary

In short, this chapter analyses and discuss the output of the models. The actual and predicted results of the observation are illustrated by graphical representations. The error rate of both models are summarised in Table 7.1 and Table 7.2 for ARIMA and the LSTM respectively. A comparison representation of both models is depicted in Table 7.3 and Figure 7.9 that help to understand the analysis of this study. The Conclusion and the future scope of this study are discussed in the next chapter.

Initially, the discussions of the Precision Agriculture and its importance, and the time series analysis are carried out. Latterly, the literature review and its several existing studies and experiments are discussed. Further, the problem statement and the objective of the proposed study is discussed. Thereafter, the relevant datasets based on precision agriculture are discussed. In the later chapters, the working strategy of the proposed study, Implementation and the evaluation of the results are analyzed. In this chapter, the conclusion and the future scope of the research work of this proposed study are presented.

8.1 Work Accomplished

After the implementation of this system, the work accomplished during this proposed study are discussed as follow.

- Import the best suitable dataset of the soil moisture based on the multi-depth of the soil.
- The ARIMA and the LSTM models are parameterized with different values.
- The models are built with the improved value of the parameters.
- Observations are forecasts using these models, where training dataset is used as a source of observations.
- Compute the error rate of forecasted observation using different performance metrics such as MAE, MAPE, MSE, and the RMSE, the testing dataset is used as a validated dataset. It is used to compare the results of forecasted observation with the testing dataset.
- Compare the results of both models and analyze the best performance of them.

8.2 Conclusion

The proposed research work presented a comparative study of two different models. The ARIMA and the LSTM models are used in this study to compute the comparison on a dataset based on the attributes of the soil moisture. The dataset of the soil moisture is acquired from the IIT Kanpur's agricultural site. It has the values of soil moisture at four different depths of the soil. The performance analysis of this study is

computed using MAE, MAPE, MSE, and RMSE as performance metrics to compute the accuracy of the model.

The purpose of this research is to suggest the best model forecast the soil moisture. In this experiment, the LSTM model performs better as comparing different performance metrics such as MAPE, MAE, MSE, and RMSE to the ARIMA model. The error rate of the LSTM model is lower than the error rate of the ARIMA model. Hence, in this study, the LSTM model performs better. The MAE for the ARIMA and the LSTM are 0.16 and 0.11 and the MAPE of both models are 0.44 and the 0.32 respectively.

8.3 Social Benefits

In the era of modern tools and technology, the equipment of the agriculture sector are improving, that led to an increase in the productivity of the crop. This study is benefitted to the farmers to estimates the soil moisture of a depth. The farmer can acquire the moisture level from the system that helps him to attain the required resources for the crop and utilize them in the best efforts to improves the quality and increase the quantity and health of the crop and the soil.

8.4 Future Work

The soil moisture based study is beneficial to understand the relations of the soil moisture with multi-depth. This proposed experiment can be extended to the user-defined data for the Smart Farming team at Thapar Institute of Engineering, Patiala campus. There are several other parameters related to the precision agriculture (such as soil temperature and environment variables like weather, precipitation, humidity and the temperature of the environment) can be added to improve the effectiveness of this experiment. This team is work on the soil moisture prototype to increase the fertility of the soil. The fetched data can be trained with the LSTM model and the forecasted observations are obtained from the model. The forecasted observations help to increases the health of the crop by providing the level of the water in the soil.

References

- [1] Zhu, Y., Song, J. and Dong, F., 2011. Applications of wireless sensor network in the agriculture environment monitoring. *Procedia Engineering*, 16, pp.608-614.
- [2] Mat, I., Kassim, M.R.M. and Harun, A.N., 2015, November. Precision agriculture applications using wireless moisture sensor network. In *2015 IEEE 12th Malaysia International Conference on Communications (MICC)* (pp. 18-23). IEEE.
- [3] Tan, Li. "Cloud-based decision support and automation for precision agriculture in orchards." *IFAC-PapersOnLine* 49, no. 16 (2016): 330-335.
- [4] Zhang, P., Zhang, Q., Liu, F., Li, J., Cao, N. and Song, C., 2017, July. The construction of the integration of water and fertilizer smart water saving irrigation system based on big data. In *2017 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC)* (Vol. 2, pp. 392-397). IEEE.
- [5] Sreeram, K., Kumar, R.S., Bhagavath, S.V., Muthumeenakshi, K. and Radha, S., 2017, March. Smart farming—A prototype for field monitoring and automation in agriculture. In *2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)* (pp. 2189-2193). IEEE.
- [6] Ip, R.H., Ang, L.M., Seng, K.P., Broster, J.C. and Pratley, J.E., 2018. Big data and machine learning for crop protection. *Computers and Electronics in Agriculture*, 151, pp.376-383.
- [7] Lenka, S.K. and Mohapatra, A.G., 2015, December. Gradient descent with momentum based neural network pattern classification for the prediction of soil moisture content in precision agriculture. In *2015 IEEE International Symposium on Nanoelectronic and Information Systems* (pp. 63-66). IEEE.

- [8] Mohapatra, A.G. and Lenka, S.K., 2016. Neural network pattern classification and weather dependent fuzzy logic model for irrigation control in WSN based precision agriculture. *Procedia Computer Science*, 78, pp.499-506.
- [9] Ponce-Guevara, K.L., Palacios-Echeverría, J.A., Maya-Olalla, E., Domínguez-Limaico, H.M., Suárez-Zambrano, L.E., Rosero-Montalvo, P.D., Peluffo-Ordóñez, D.H. and Alvarado-Pérez, J.C., 2017, October. GreenFarm-DM: A tool for analyzing vegetable crops data from a greenhouse using data mining techniques (First trial). In *2017 IEEE Second Ecuador Technical Chapters Meeting (ETCM)* (pp. 1-6). IEEE.
- [10] Han, J., Mao, K., Xu, T., Guo, J., Zuo, Z. and Gao, C., 2018. A soil moisture estimation framework based on the CART algorithm and its application in China. *Journal of hydrology*, 563, pp.65-75.
- [11] Priya, R., Ramesh, D. and Khosla, E., 2018, September. Crop Prediction on the Region Belts of India: A Naïve Bayes MapReduce Precision Agricultural Model. In *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)* (pp. 99-104). IEEE.
- [12] Sobayo, R., Wu, H.H., Ray, R. and Qian, L., 2018, April. Integration of Convolutional Neural Network and Thermal Images into Soil Moisture Estimation. In *2018 1st International Conference on Data Intelligence and Security (ICDIS)* (pp. 207-210). IEEE.
- [13] Prasad, R., Deo, R.C., Li, Y. and Maraseni, T., 2019. Weekly soil moisture forecasting with multivariate sequential, ensemble empirical mode decomposition and Boruta-random forest hybridizer algorithm approach. *Catena*, 177, pp.149-166.
- [14] Soni, K., Kapoor, S., Parmar, K.S. and Kaskaoutis, D.G., 2014. Statistical analysis of aerosols over the Gangetic–Himalayan region using ARIMA model based on long-term MODIS observations. *Atmospheric research*, 149, pp.174-192.

- [15] Bari, S.H., Rahman, M.T., Hussain, M.M. and Ray, S., 2015. Forecasting monthly precipitation in Sylhet city using ARIMA model. *Civil and Environmental Research*, 7(1), pp.69-77.
- [16] Taki, M., Ajabshirchi, Y., Ranjbar, S.F., Rohani, A. and Matloobi, M., 2016. Heat transfer and MLP neural network models to predict inside environment variables and energy lost in a semi-solar greenhouse. *Energy and Buildings*, 110, pp.314-329.
- [17] Babu, C.N. and Sure, P., 2016. Partitioning and interpolation based hybrid ARIMA–ANN model for time series forecasting. *Sādhanā*, 41(7), pp.695-706.
- [18] Zhang, J., Zhu, Y., Zhang, X., Ye, M. and Yang, J., 2018. Developing a Long Short-Term Memory (LSTM) based model for predicting water table depth in agricultural areas. *Journal of hydrology*, 561, pp.918-929.
- [19] Zhang, L., Lin, J., Qiu, R., Hu, X., Zhang, H., Chen, Q., Tan, H., Lin, D. and Wang, J., 2018. Trend analysis and forecast of PM_{2.5} in Fuzhou, China using the ARIMA model. *Ecological Indicators*, 95, pp.702-710.
- [20] Qing, X. and Niu, Y., 2018. Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM. *Energy*, 148, pp.461-468.
- [21] Bendre, M. and Manthalkar, R., 2019. Time series decomposition and predictive analytics using MapReduce framework. *Expert Systems with Applications*, 116, pp.108-120.
- [22] Pal, M., Maity, R. and Dey, S., 2016. Statistical Modelling of vertical soil moisture profile: coupling of memory and forcing. *Water resources management*, 30(6), pp.1973-1986.
- [23] Haiges, R., Wang, Y.D., Ghoshray, A. and Roskilly, A.P., 2017. Forecasting electricity generation capacity in Malaysia: an auto regressive integrated moving average approach. *Energy Procedia*, 105, pp.3471-3478
- [24] Li, Y. and Cao, H., 2018. Prediction for tourism flow based on LSTM neural network. *Procedia Computer Science*, 129, pp.277-283.

- [25] Bouktif, S., Fiaz, A., Ouni, A. and Serhani, M., 2018. Optimal deep learning lstm model for electric load forecasting using feature selection and genetic algorithm: Comparison with machine learning approaches. *Energies*, 11(7), p.1636.
- [26] Hochreiter, S. and Schmidhuber, J., 1997. Long short-term memory. *Neural computation*, 9(8), pp.1735-1780.
- [27] Sharma, A. and Rani, R., 2017. An optimized framework for cancer classification using deep learning and genetic algorithm. *Journal of Medical Imaging and Health Informatics*, 7(8), pp.1851-1856.
- [28] Fischer, T. and Krauss, C., 2018. Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, 270(2), pp.654-669.
- [29] <https://study.com/academy/lesson/bronze-age-farms-farming-tools.html>
- [30] <https://www.worldatlas.com/articles/did-the-british-agricultural-revolution-lead-to-the-industrial-revolution.html>
- [31] <https://study.com/academy/lesson/the-bronze-age-inventions-tools-technology.html>
- [32] https://www.crestcapital.com/tax/history_of_agriculture
- [33] <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2610166/>
- [34] <https://regenerationinternational.org/2017/02/24/what-is-regenerative-agriculture/>
- [35] <https://whatis.techtarget.com/definition/precision-agriculture-precision-farming>
- [36] <http://www.madhavuniversity.edu.in/precision-agriculture.html>
- [37] <https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/>
- [38] https://www.slideshare.net/21_venkat/arima-26196965
- [39] <https://www.cse.iitb.ac.in/~cs725/notes/lecture-slides/2016b/lecture-23-annotated.pdf>
- [40] https://en.wikipedia.org/wiki/Mean_absolute_percentage_error
- [41] <https://www.dataquest.io/blog/understanding-regression-error-metrics>

List of Publications

- [1] S Singh, P. Kumar, S. Kaur. “Forecasting Soil Moisture based on Evaluation of Time Series Analysis” Springer International Conference on Computing, Power and Communication technologies (GUCON), NCR New Delhi, India, September 2019. [Accepted]
- [2] S Singh, S. Kaur. “Forecasting Soil Moisture based on Time Series Analysis” Asian-Australasian Conference on Precision Agriculture, Punjab Agriculture University, Ludhiana, October 2019. [Accepted]

Plagiarism Report

agri

ORIGINALITY REPORT

14%

SIMILARITY INDEX

10%

INTERNET SOURCES

8%

PUBLICATIONS

0%

STUDENT PAPERS

PRIMARY SOURCES

1	dspace.thapar.edu:8080 Internet Source	1%
2	machinelearningmastery.com Internet Source	1%
3	www.jamasoftware.com Internet Source	1%
4	Susha Lekshmi, S.U., D.N. Singh, and Maryam S. Baghini. "A Critical Review of Soil Moisture Measurement", Measurement, 2014. Publication	1%
5	lter.kbs.msu.edu Internet Source	1%
6	r-statistics.co Internet Source	1%
7	epdf.tips Internet Source	<1%
8	es.scribd.com Internet Source	<1%