Weather Forecasting using Deep Learning

by

Sheikh Abdul Fahad
18101462
Takowa Islam Trina Akand
17101333
Md. Reaj Uddin Ahammed Raju
17101385
Ranjita Yadav
18201202

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science

Department of Computer Science and Engineering
Brac University
September 2022

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Sheikh Abdul Fahad	Takowa Islam Trina Akand
18101462	17101333
Md. Reaj Uddin Ahammed Raju	Ranjita Yadav
17101385	18201202

Approval

The thesis/project titled "Weather Forecasting using Deep Learning" submitted by

1. Sheikh Abdul Fahad (18101333)
2. Takowa Islam Trina Akand (17101333)
3. Md. Reaj Uddin Ahammed Raju (17101385)
4. Ranjita Yadav (18201202)

Of Summer, 2022 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on September 20, 2022.

Examining Committee:

Supervisor:
(Member)

____

Tanvir Ahmed

Lecturer
Department of Computer Science and Engineering
Brac University

Program Coordinator:
(Member)

____

Md. Golam Rabiul Alam, PhD

Professor
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

____

Sadia Hamid Kazi

Chairperson and Associate Professor
Department of Computer Science and Engineering
Brac University

Abstract

Global human life is significantly impacted by weather forecasts. It takes a lot of computing resources to solve mathematical equations that are based on climatic circumstances. As a result, during the past ten years, deep learning algorithms have been integrated with enormous amounts of weather observation data. At the moment, a lot of data is being consumed. So, we can increase the accuracy of weather forecasts by combining this enormous amount of data with deep learning methods. In this paper, we implement benchmark datasets for autoencoder and linear regression. We are using z500 dataset, temp 2m dataset and t850 data set. As training the linear regression on the full data will take a lot of memory which is why we took every 5th time step that almost give the same result. Using a linear regression approach and an auto encoder model, we trained and obtained day-level predictions using the ERA5 reanalysis dataset (Hersbach et al., 2020) to determine the accuracy of the test data and training data.

Keywords: Deep learning; Weather forecasting;linear regression; auto encoder; Prediction; de-noise; Linear Regression Analysis

Acknowledgement

Firstly, all praise to the Great Allah for whom our thesis have been completed without any major interruption.
Secondly, to our supervisor Tanvir Ahmed sir for his kind support and advice in our work. He helped us whenever we needed help.
And finally to our parents without their throughout support it may not be possible. With their kind support and prayer we are now on the verge of our graduation.

Table of Contents

Declaration **i**

Approval **ii**

Abstract **iii**

Dedication **iv**

Acknowledgment **iv**

Table of Contents **v**

List of Figures **vii**

List of Tables **ix**

Nomenclature **x**

1 Introduction **1**

2 Literature Review **4**
- 2.1 Agrawal et al. (1980) . . . 5
- 2.2 F. Mekanik and M. A. Imteaz (2013) . . . 6
- 2.3 Scher (2018) and Messori (2019) . . . 6
- 2.4 Reddy and Babu (2017) . . . 7
- 2.5 Weyn et al. (2019) . . . 7

3 Methodology **8**
- 3.1 Data set . . . 8
- 3.2 Linear regression . . . 11
- 3.3 Auto Encoder Architecture . . . 12
- 3.4 ARIMA And SARIMA Model Architecture . . . 14
- 3.5 Facebook's Prophet Model Architecture . . . 17
- 3.6 LSTM and Bi-LSTM Model Architecture . . . 18

4 Findings **22**
- 4.1 Training Data set . . . 22
- 4.2 Findings of Linear regression . . . 24
- 4.3 Findings of Auto Encoders . . . 26
- 4.4 Findings of Auto Encoders vs linear regression . . . 28

4.5 Findings of ARIMA and SARIMA Model . . . 30
4.6 Findings of Facebook Prophet Model . . . 32
4.7 Findings of LSTM and Bi-LSTM Model . . . 34
4.8 Comparison of RMSE among the used models . . . 35

5 Summary, Conclusion and Discussion **36**

Bibliography **40**

List of Figures

3.1 Data file . . . 9
3.2 Numerical Data plotted of temperature at 2 meter . . . 9
3.3 Data Pre-processing . . . 9
3.4 Linear regression . . . 11
3.5 Predicting Data with Linear regression . . . 12
3.6 auto encoder architecture . . . 13
3.7 Predicting data using auto encoder architecture . . . 13
3.8 ARIMA model architecture . . . 14
3.9 SARIMA model architecture . . . 15
3.10 Data's seasonality and Trend . . . 15
3.11 Stationy And Null Hyphothesis Test . . . 16
3.12 ACF plot . . . 16
3.13 PACF plot . . . 17
3.14 Facebook's Prophet Model Architecture . . . 18
3.15 General LSTM Model Architecture . . . 19
3.16 LSTM Model Architecture of our testing . . . 19
3.17 General Bi-LSTM Model Architecture . . . 20
3.18 Bi-LSTM Model Architecture of our testing . . . 20

4.1 Geo potential 500 Training data . . . 23
4.2 Total Precipitation Training data . . . 23
4.3 2m temperature Training data . . . 23
4.4 2m temperature Numerical Training data(Daily-interval) . . . 23
4.5 2m temperature Numerical Training data(Monthly-interval) . . . 24
4.6 Weather forecast of Geo-potential 500 for 7 days using Linear regression . . . 25
4.7 Weather forecast of Total Precipitation for 7 days using Linear regression . . . 25
4.8 Weather forecast of 2m temperature for 7 days using Linear regression . . . 25
4.9 Weather forecast of Geo-potential 500 for 7 days using Auto Encoder . . . 26
4.10 Weather forecast of Total Precipitation for 7 days using Auto Encoder . . . 27
4.11 Weather forecast of 2m temperature for 7 days using Auto Encoder . . . 27
4.12 Weather forecast of Geo-potential 500 for 7 days using auto encoder vs linear regression 28
4.13 Weather forecast of Total Precipitation for 7 days using auto encoder vs linear regression 29
4.14 Weather forecast of 2m temperature for 7 days using auto encoder vs linear regression 29
4.15 Data's seasonality and trend . . . 30
4.16 ACF . . . 30
4.17 PACF . . . 31
4.18 Prediction from ARIMA model . . . 31
4.19 Prediction from SARIMA model . . . 32

4.20 Prediction from Facebook Prophet model 32
4.21 Data's seasonality and trend 33
4.22 Training Data 34
4.23 Prediction from LSTM model vs actual data 34
4.24 Prediction from Bi-LSTM model vs actual data 35
4.25 RMSE comparison 35

List of Tables

3.1 List of variables contained in the benchmark dataset.[23] 10

4.1 Chart of train mse, test mse and predicted value of each data set from linear regression 24

4.2 Chart of train mse, test mse and predicted value of each data set from auto encoder . 26

Nomenclature

The next list describes several symbols & abbreviation that will be later used within the body of the document

ai Artificial Intelligence

CNN Convolutional neural network

lat Latitude

lon Longitude

lr Linear regression

$t2m$ Temperature at 2 meter

tp Total Precipitation

$z500$ Geo-potential at 500 millibars

Chapter 1

Introduction

Weather forecasting is the atmospheric conditions prediction of a specific location and time to indicate the weather conditions which are likely to occur which includes humidity, temperature, rainfall, wind speed and dew points. Since the 19th century and for millennia people have made several efforts to forecast the weather formally and informally. This forecasting happens by securing quantitative data of the present atmospheric conditions at a given place to analyze how the weather pattern changes over time. Instruments like outdoor thermometers (air temperature), barometers (air pressure), anemometers (wind speed and direction), radar and hygrometers (humidity) are used in collecting data of present and previous weather conditions, by tracking motion of clouds and airs. Weather forecasting has been an important science in meteorology.It is also one of the most scientifically and technologically challenging problems in the world. There are diversions of ways to use it. People's lives depend on weather forecasting as the accurate forecasting allows communities to assemble better over the impacts of climate change. The key importance of weather forecasting is to protect human lives and property, improve health, safety, and economic prosperity. Temperature and precipitation based forecasting are crucial for agriculture as well as to the farmers as it allows them to track when it's best to plant or help them protect their crops and to traders within commodity markets. Temperature based forecasting is not important for farming but also used by the utility companies to evaluate demands for the future days. Weather forecasting is also important for daily uses of people since it affects what to wear on the day depending on the forecast prediction. It is used to plan outdoor activities as these activities can be severely curtailed by heavy rain, wind chill or snows and to protect or to plan and survive ahead.

There are various methods to predict the forecasting and these methods mostly depend on what a forecaster like in selecting terms, depending on the experiences and knowledges of the meteorologist, availability on the amount of information regarding the matter to the forecaster, the level of difficulty and the degree of accuracy that the forecast situation presents and . Based on these characteristics, there are four methods created for forecasting weather conditions. There might be much information and data available but not all these data are used for the prediction.

Simplest way to predict the atmospheric condition in a method is used as a Climatological forecast. Climatology is the broad study of climate and the differentiation over time. Climatologists pointed out three major aspects of climate. The very first aspect is the weather condition that determines the standard pattern throughout the world in different regions. Secondly climatologists are trying to acknowledge the correspondence between different aspects of weather such as temperature and sunlight. The last and third aspect of climate that the climate scientists investigate is the changes of weather conditions over time. This forecasting method requires averaging weather statistics accumulated over many years to make the forecast. Current atmospheric conditions are not needed for forecasting since it uses information about the region's climate on a seasonal basis. This forecasting method only works well when the weather conditions are similar to that expected for the chosen time of year. But if the pattern is not the same or unusual for the given time of year, the climatology method will often fail miserably.

Another of the simplest methods is the Persistence Method; that is used for producing a forecast. Persistence forecast is the opposite of climatological forecast method because in this method it only uses the current weather conditions and also used often as a standard of comparison in measuring the degree of skill of forecasts prepared by other methods, especially for very short projections. This forecast is used when we decide what to wear on that day, basically it depends on the daily atmospheric condition. This method reassures that the patterns at the time of the forecast will stay unchangeable. For an example of this method, imagine it is 55 degrees outside with a bright sunny day, so it will be bright and sunny with 55 degrees outside tomorrow with the persistence method prediction. On the other hand if it says it is gonna be raining with three inches of rain falling today, the persistence method would predict the same condition of weather for tomorrow as well. This method is perfect when the atmospheric conditions substitute are trivial and presents on the weather maps move very slowly. For example this method works well in places like Southern California where the weather conditions stay unchangeable significantly from day to day basis. This persistence forecast might break down and not come in handy if weather conditions change significantly. This is considered an unskilled forecast because it does not need any special training or grounding to happen. Because it takes the day to day, present weather and predicts that future coming day or the next day weather will exactly be the same.

Analog forecasting is a simple yet powerful method to generate the forecast with a touch of complication. This method basically works and involves the examination of the current day weather conditions and recalling the exact same day weather pattern from the past. More like doing an analogy. Examined records then predicted that the weather forecast will act as same as the previous situation under those conditions as it did previously. It can be quite difficult as it is almost impossible to find a perfect analog in real life and virtually. Numerous weather patterns seldom align themselves with their earlier times in the same regions. Even a nuance between present time with the analog can lead to very different results. Also as the time passes, more weather datas on weather conditions are recorded for finding a "good match" by chance.

And the last and most advanced method for weather forecasting is Numerical forecast which is also known as numerical weather prediction (NWP). It is one of the advanced methods where the methodologist produces weather forecasts through large computer simulations by using the fundamental laws governing the flow of the atmosphere. The daily weather pattern or conditions in the United States forecasting happens with the methodology of NWP (Numerical Weather Prediction), alongside a supercomputer at the National Oceanic and Atmospheric Administration (NOAA) in Washington, DC. Almost all the countries' weather forecasting is based on NWP since it works as the key guidance for their operational weather prediction. For this forecasting, there are three classical physical laws being used in the terms of mathematical equations. Then by solving these equations with the high-performance supercomputers, they predict the weather forecast in the most accurate ways. It is also called an ensemble prediction system as through this method methodologists produce not only a single forecast or two but multiple forecasts. In fact, through this forecasting system , they are trying to represent all the uncertainties that exist in the atmosphere of our planet by doing the NWP (Numerical Weather Prediction).

As we all know, there are numerous weather forecasting models in use today that are extremely accurate, but over time, the accuracy significantly decreases. For example, if we use data from a month, the accuracy will be highly precise, but if we use data from a year, the accuracy will decline. Because the data required for weather forecasting is so large, it is difficult to implement and frequently crashes while doing so. As a result, for such large data, we want powerful technology that can assemble the data quickly.

Nowadays there are numerous models that can generate accurate forecasts, but as time goes on, such predictions get less accurate, so the purpose of our study is to ascertain how long the data is true. With time, the prediction's precision decreases.

In our research we have done two different parts. In the first part we restricted our data set among geopotential 500, total precipitation and temperature at 2m though we had a massive collection of data. In our findings we have only computed the raw prediction and focused on its accuracy of various points. we have not gone with any kind of interpretation of how the climate will behave like if it is going to rain or any other stuff. But at the end of the first part we were not satisfied yet. Hence we decided to try some other models with numerical values of temperature at 2 meter only of a particular latitude and longitude. We have tried to find out the accuracy of theses models in our research.

Chapter 2

Literature Review

Synopsis of the previous work In this part of section, we are going to give a short description of the existing studies on predicting the weather forecast which based on dataset, model and accuracy.

In order to anticipate solar irradiance, a variety of artificial intelligence (AI) techniques have been developed. These methods comprise statistical and machine learning (ML) methods, numerical prediction, and prediction based on images. Time-series data include data on solar radiation or data that continuously change over time. Because they are well-known, simple to compute, and produce accurate estimates of solar irradiance, linear prediction techniques were frequently utilized in the past. Traditional forecasting methods include the autoregressive with exogenous inputs (ARX), autoregressive moving average (ARMA), autoregressive integrated moving average (ARIMA), autoregressive combined moving average with exogenous inputs (ARIMAX), [15][7], and autoregressive moving average with exogenous inputs (ARMA). The ARIMA and ARMA models were presented by Belmahdi et al.[29] to estimate the parameters of the global solar radiation. In these simulations, the only parameter considered was solar radiation.Since the models assume linearity in the data and are therefore unable to capture complex nonlinear patterns, no geographic or meteorological characteristics were included in the model training process. In order to anticipate grid- connected solar efficiency, Ferlito et al. compared eleven data-driven online and offline models. Gensler et al. reduced the dimensionality of historical data using an automated encoder, and they predicted sun irradiance using LSTM. Data on solar irradiation were pre-processed using multi-level wavelet decomposition by Zhen et al. to boost forecast accuracy. Another Zhen study created a brand-new daily model for predicting solar irradiance using a time-section fusion pattern and mutual iterative optimization[22]

Yagi, etc. Using radiation data gathered from satellites in various places, we assessed 68 machine learning models.[28]Multilayer perceptron (MLP) models exhibit the best performance, according to studies. This study's design took into account the artificial neural network (ANN) model utilized in day-ahead prediction. ANN layers use nonlinear transformations to examine data and can identify intricate data structures. It is suitable for complicated variable forecasting with time series since it can mimic noisy systems from data. When dependencies and data need to be tracked as they move between data time stages, these are all excellent solutions to design issues. To calculate sun irradiance, a deep recurrent neural network can be employed, which will simplify the model and make feature extraction easier. The suggested method performs better than conventional feedforward ANNs and SVMs in terms of performance. We are able to determine the sequential nature of data node associations while storing data energies 2022, 15, and 2226 3 of 12 using a simple recurrent neural network (RNN) design. RNNs, on the other hand, are vulnerable to gradients that explode and fade. RNN expansions include the Long Short-Term Memory (LSTM), Gated Recurrent Unit

(GRU), bidirectional LSTM networks that replace memory cells, and gating algorithms that regulate the information flow in networks of conventional perceptron designs. It was finished. In order to predict solar radiation for the following day, the time series forecasting technique LSTM was created. When compared to the other prediction methods included in this investigation, the performance of the LSTM model was superior. proposed a way to forecast the hourly sun radiation for the following day using meteorological information. RNN models include conventional memory-based models and attention-based models. Along with other memory-based models, there are bidirectional RNN, LSTM, and GRU. Self-aware generative adversarial networks, attention-directed LSTMs, and multi-head LSTMs are a few examples of attention-based models. [4] A system of mathematical equations used to describe the physical characteristics and dynamic behavior of the atmosphere is known as a physics technique. There are frequently very brief and very lengthy applications. Most of the time, these techniques rely on satellite imagery, sky photography, and numerical weather prediction (NPW).They are categorized as global or mesoscale physical approaches depending on the simulated atmospheric extent (global or restricted). The power output of a solar system should only be predicted using mesoscale models with resolutions of 16 to 50 km .[4][21]

Due of the numerous factors that might influence how well predictive systems operate, comparing them can generally be challenging. There are numerous variables, including the accessibility of historical information and forecasts for the weather, the time span and resolution, the climate, the location, and the installation requirements. When employing statistical approaches, improved performance and reduced computing costs are also dependent on proper data preparation (for example, eliminating nighttime sampling when no power is being generated). Despite being qualitative rather than quantitative, the review's findings do shed light on the efficacy of various strategies. Recent reviews include comparisons that take into account both the contributions of other writers and statistical shortcomings. Because experimental settings and procedures vary, comparisons cannot be made quantitatively

To estimate solar irradiance, memory-based RNN models are most frequently utilized in the literature. The study did not, however, incorporate attention-based RNN models. This research proposes an attention-based Bi-LSTM method for predicting solar irradiance. He trained the model using real meteorological data from three places in Kuwait: Al-Wafer, Kia, and Abdaly. To assess the proposed attention-based biprecision LSTM, we employ trustworthy statistical methods such mean squared error (RMSE), mean squared error (MSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). Compare it to other things and evaluate it. existing design.[26].

2.1 Agrawal et al. (1980)

In the paper, the author used time series regression models to demonstrate how the yield of rice in the Raipur quarter might be predicted based on daily data and rainfall characteristics. According to Kou and Sun (1993), the Tanshui region of Taiwan had storms and other exceptionally strong rainfall that caused an intervention pattern to last an average of 10 days before flowing into the sewer and causing confusion. To simulate daily, periodic, and periodic overflows at eight acceptable peaks, Chiew et al. (1993)[3] compared six reduced runoff modeling approaches. They came to the conclusion that the time-series technique is in line with credible projections of continuous and recurrent income in watershed water centers. A different strategy is used by Langu (1993), who searches for significant changes within different storm statistical corridors by employing statistical analysis to track changes in storm and runoff patterns. This innovation was first described by Box and Jenkins (1994) in the early 1970s. In this author's method,[1] a type of statistics known as univariate Box-Jenkins

(UBJ) ARIMA modeling was developed for univariate instances. a determinant of precipitation the terminology used in this sentence evolves with time. When utilizing any parameter, statistical model, or new fortune teller model, as set by the author to characterize each time series, this time series data is employed (Chatfield 1994; Montgomery and Lynnwood 1996)[6]. The notes attracted the interest of people who were curious as well (Sivakumar 2001 Sivakumar et al.1999; Men 2004).[5] Since seasonal weather change is a complex phenomenon that involves many specialized fields of knowledge that must be addressed in the field of meteorology, all presumptions must be taken into account with original and global climate variables. The current state of nature and the causes of seasonal weather change are still unknown. The probabilistic rush model has been developed by the world. This is mostly used to forecast natural disasters brought on by rapid climate change and to offer advice on how to address them using climatic factors. Turkish technical sciences relevant study by Marmara University College of Technology (2010) Weather at Goztepe, Istanbul, Turkey,[10] using statistical and neuro fuzzy network models for vaccination fuzzy conclusion system[11] and machine that is network-based the efficiency of ARIMA and ANFIS was evaluated using a retrogression moving average model; different models were used for different exercises. The criteria used to evaluate a test dataset was Additionally, the performance of the ARIMA and ANFIS models for soothsaying are estimated and compared using the calculated performance score. a thorough overview. The Arima and Anfi models are taken into consideration by Mahmudur Rahman, A.H.M. Saiful Islam, Saha Yasser Maknoon Nadvi, and Rashedul M. Rahman (2013)[14]. Describe how ARIMA models are more effective. Taking into account the dynamic properties of rainfall, As an example, the minimum temperature, maximum temperature, moisture content, and air pressure need to be compared using performance metrics like root, R-square, and total squared errors (SSE) ARIMA delivers superior outcomes to competing systems. modeling techniques akin to ANFIS [2]

2.2 F. Mekanik and M. A. Imteaz (2013)

The researchers found that these significant and complex climate variables also affect rainfall in Australia. However, there have not been many attempts to ascertain the combined impact of these factors on rainfall in order to improve knowledge and forecasting systems.[8] Given that rainfall is a complicated meteorological event, linear techniques might not be able to accurately describe its characteristics.Using artificial neural networks,[9][12] this study seeks to identify a nonlinear link between the Victorian rainfall and the localized lagged indices (ANN). It was found that ANN modeling, as opposed to linear techniques, is able to produce greater correlations when employing the lagged indices to anticipate spring rainfall. For the three case study stations in Victoria, Australia—Horsham, Melbourne, and Orbost—using these indices in an ANN model boosted the model correlation to 99[13]

2.3 Scher (2018) and Messori (2019)

In a simplified reality scenario, these two experiments tackled the problem of data-driven weather forecasting. The "reality" consisted of lengthy runs of sped-up general circulation models (GMCs). The model fields were forecast by neural networks many days in advance. One of the characteristics of the CNNs used in the neural network architecture is the encoder-decoder system. The immediate 3D model fields from one timestep are fed into them, and the same model fields are output at a later time. Scher trained an entirely new network up to 14 days' worth of lead-time for each. Scher and Messori (2019)[24][20][27] created longer forecasts frequently after only receiving instruction for 1-day forecasts. Unexpectedly, direct prediction networks fared better than iterative networks when they were taught to make predictions up to five days in advance. The forecasts were evaluated using the anomalous percentage of the correlation of 800 hPa and Z500 temperature as well as the root mean squared error. In 2018 Scher demonstrated[20] very high forecasting skill using a relatively simple GCM devoid of a hydrological cycle. Additionally, they were able to use the network to produce steady "climate" runs, which are lengthy collections of successive forecasts. Scher and Messori (2019b) made use of a number of more intricate and realistic GCMs. Although the data-driven model produced reasonably accurate short-term forecasts, it was unable to provide consistent and accurate "climate" runs. Regarding neural-network architectures, they demonstrated that the same structures, which were tuned for simpler GCMs, also function on more complicated GCMs and have some predictive ability on single-level reanalysis data.[20][24]

2.4 Reddy and Babu (2017)

The authors wanted to examine several methodologies, evaluate their effectiveness, and pinpoint the issues with weather forecasting. In their study they looked at short term, medium term and long term. They were able to compare a variety of methods and models by examining multiple papers and comparing the data sets and hypotheses, the time period of the investigation, input parameters, methodology, findings, and recommendations, including the MapReduce model, linear regression method, random forest, support vector machine neural network, REP tree and bagging algorithm, Naive Bayes, wavelet artificial neural network and nearest neighbors modeling. This survey's taxonomy was weak, and it lacked a section on research methodology and an analysis of the findings. It was not a systematic review in this paper. [18]

2.5 Weyn et al. (2019)

This study uses deep CNNs to forecast the 700-300 hPa and Z500 thicknesses from reanalysis at 6-hourly time intervals. The information is from the 2.5-degree horizontal resolution reanalysis starting from 1979 till 2010 at Climate Forecast System[10], using a North American hemisphere crop. The authors explored using a Convolutional network with encoder-decoders, as those utilized by Messori (2019)[27][24] and Scher, as well as long and short period memory hidden layer (2018).To produce frequent forecasts inputs the outputs of the model are used just like Scher and Messori (2019) did. Using two input time steps that are separated by six hours and two anticipated output time steps, according to the authors, outperformed using just one step. They outperform a climatology benchmark with shorter lead times, up to 120 hours, with their best CNN forecast., while up to fourteen days during longer lead periods, it seems to asymptote toward persistence projections. The first stages towards data-driven forecasting are outlined in these three strategies. The variations among the suggested strategies already draw attention to the necessity of a standard benchmark case to measure prediction accuracy. [27]

Chapter 3

Methodology

In this section we will get to know about the methods we are going to follow for our research regarding weather forecasting. And also the dataset we have collected and the details of that and also the process of that in detail.
For our research we are using deep learning method for the forecasting process. Here we have already collected a bunch of data from weather benchmarks and numerical data which we will use for our training and testing purpose. And then from the mentioned below models, we will get our findings of research. There are many established methods and models to create weather forecasting using deep learning with various accuracy rates of their own. In our research we will know how data usages and data portion selection may have an impact on the accuracy of the weather forecasting of days ahead.

3.1 Data set

For our testing and training purpose we have used two different types of data sets, one of which was reanalysed and gives the best possible guess of the atmospheric state at any time and combination of forecasting model with the available data observation. The data set includes hourly data from 1979 to 2018 on a 0.25° latitude longitude grid(721×1440 grid points) with 37 vertical levels. As the main raw data is quite massive in size to compute our project we choose a smaller portion of the 5.625° data set. This has 32x64 grid points. The regridding was done with the xesmf Python package (Zhuang, 2019) using a bilinear interpolation. The reason behind choosing a shorter data set is, since high resolutions are still hard to compute for deep learning models because of GPU memory resource consumption and constraint of its I/O speed. The processed data (table 1) are available at https://mediatum.ub.tum.de/1524895 (Rasp et al., 2020).[23] The data set is splitted yearly in NetCDF format. The whole dataset for 5.625° is almost 250GB in size. And the second dataset is a CSV file which carries only temperature at 2 meter of latitude 23.7803 degree and longitude of 90.4071 degree. This dataset was collected from https://power.larc.nasa.gov/data-access-viewer/WER — Data Access Viewer (nasa.gov). The dataset we collected contains only temperature at 2 meter from january 1st 2000 to 31st december of 2020 and this was daily data.

Figure 3.1: Data file

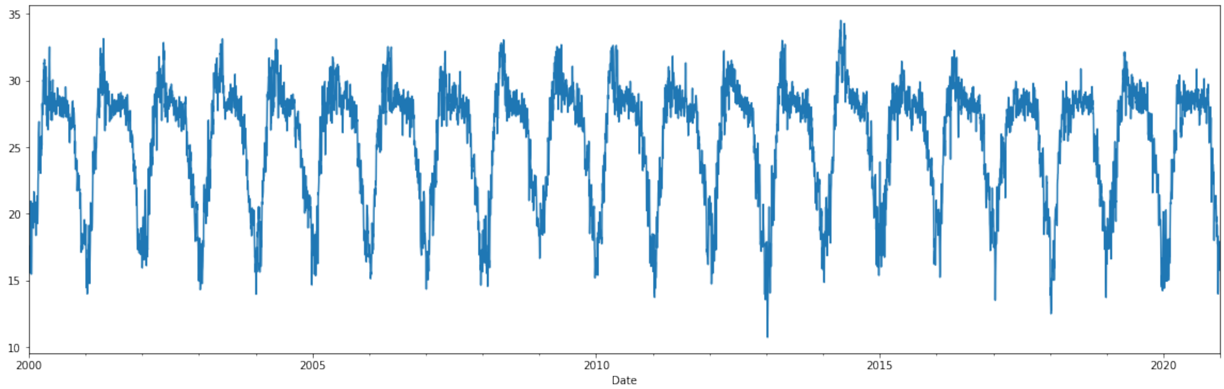

Figure 3.2: Numerical Data plotted of temperature at 2 meter

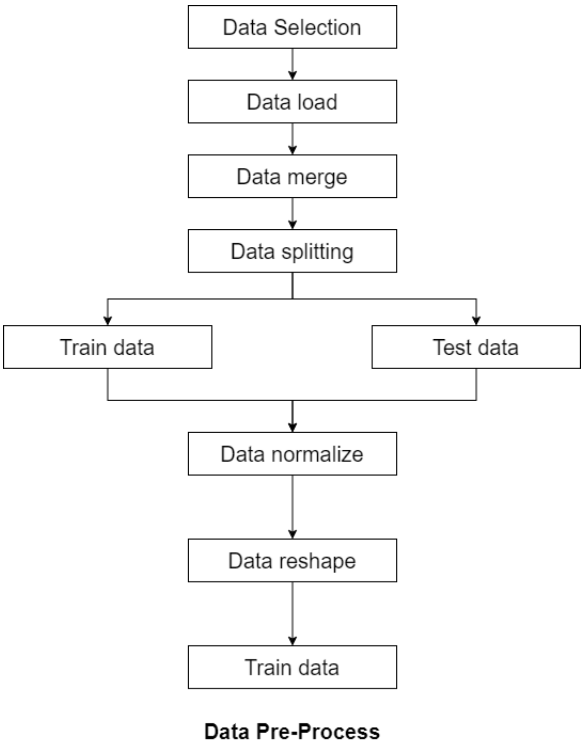

Figure 3.3: Data Pre-processing

Long name	Short name	Description	Unit	Levels
geopotential	z	Proportional to the height of a pressure level	[m$^2 s^{-2}$]	13 levels
temperature	t	Temperature	[K]	13 levels
specific_humidity	q	Mixing ratio of water vapor	[kgkg^{-1}]	13 levels
relative_humidity	r	Humidity relative to saturation	[%]	13 levels
u_component_of_wind	u	Wind in x/longitude-direction	[m s^{-1}]	13 levels
v_component_of_wind	v	Wind in y/latitude direction	[m s^{-1}]	13 levels
vorticity	vo	Relative horizontal vorticity	[1 s^{-1}]	13 levels
potential_vorticity	pv	Potential vorticity	[Km$^{-2} kg^{-1} s^{-1}$]	13 levels
2m_temperature	t2m	Temperature at 2 m height above surface	[K]	Single level
10m_u_component_of_wind	u10	Wind in x/longitude-direction at 10 m height	[m s^{-1}]	Single level
10m_v_component_of_wind	v10	Wind in y/latitude-direction at 10 m height	[m s^{-1}]	Single level
total_cloud_cover	tcc	Fractional cloud cover	(0-1)	Single level
total_precipitation	tp	Hourly precipitation	[m]	Single level
toa_incident_solar_radiation	tisr	Accumulated hourly incident solar radiation	[Jm^{-2}]	Single level
constants		**File containing time-invariant fields**		
land_binary_mask	lsm	Land-sea binary mask	(0/1)	Single level
soil_type	slt	Soil-type categories	see text	Single level
orography	orography	Height of surface	[m]	Single level
latitude	lat2d	2D field with latitude at every grid point	[°]	Single level
longitude	lon2d	D field with longitude at every grid point	[°]	Single level

Table 3.1: List of variables contained in the benchmark dataset.[23]

3.2 Linear regression

Linear regression is a technique of data analysis which is a model used to predict the accuracy of the predictions from the known or related data of same usages. It helps to plot dependent or unknown variables as a linear equation. The Main idea of linear regression create a set of predictor variables do a good job in predicting an dependent variable and also which variables in particular are significant predictors of the outcome variable, and in what way do they-indicated by the magnitude and sign of the beta estimation-impact the outcome variable.These regression estimates are used to explain the relationship between one dependent variable and one or more independent variables. The simplest form of the regression equation with one dependent and one independent variable is defined by the formula y = c + b*x, where y = estimated dependent variable score, c = constant, b = regression coefficient, and x = score on the independent variable.

If we look at the figure here we will find that the linear regression model plots a predicted value which is scattered in the plot and finds a straight line among them like the figure 3.3

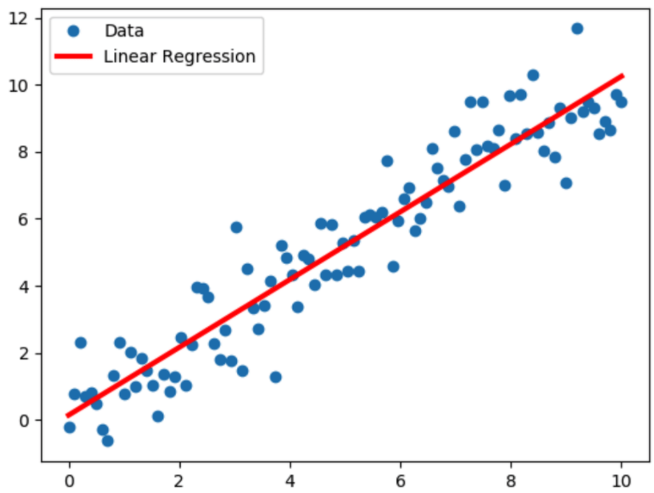

Figure 3.4: Linear regression

In our research, we splitted our dataset into two portions. One is for data training and another one for testing purposes. In the testing set, we choose data from 2012 to 2016 for temperature 2m, geopotential at 500 millibars and total precipitation. All of the datas are of 5.625° and after selecting the data we merged the data of them we normalized our data sets, reshaped them. The process of that is likely the figure 3.4

In our linear regression training model we had variables based on our test data which we directed from the data training method. Then we trained the data again and calculated the RMSE for both test data and training data as well. Then a linear regression prediction function is used to draw the value for each data set of temperature 2m, geopotential at 500 millibars and total precipitation. We tested this for several days. Which results we are going to see in the findings section.

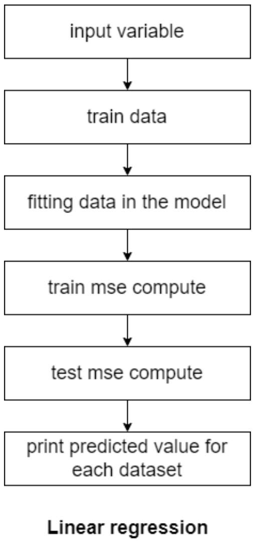

Figure 3.5: Predicting Data with Linear regression

3.3 Auto Encoder Architecture

An Architecture which is a special kind of neural network that has been trained to create duplicate of the input and generates an output which is known as autoencoder. It has a secret layer into that which represents the code that correspond to the input. The network can be thought of as having two components: one of them is decoder that creates a rebuilding (r=g) and an encoder function

$$h = f(x)$$

This building's architecture is seen in figure. An autoencoder is not particularly beneficial if it learns to set

$$g(f(x)) = x$$

everywhere. Autoencoders are instead made to be unable to improve their copying. Typically, these restrictions only accept them to copy its input nearly and the input that can closely mimics the training data. Since the model should decide which parts of the input must be duplicated, it repeatedly picks up useful data attributes. The concepts of an encoder and a decoder have been extended in modern autoencoders from deterministic functions of stochastic mappings, Pencoder(h $\mid$ x) and Pencoder(x $\mid$ h). For decades, the concept of autoencoders has been present in the development of neural networks (LeCun, 1987; Bourlard and Kamp, 1988; Hinton and Zemel,1994). Autoencoders have customarily been employed for feature learning or dimensionality lessening. Autoencoders have recently gained prominence in generative modeling due to theoretical linkages between them and latent variable models, as we will see in chapter 20. Autoencoders can be regarded of as a specific instance of feedforward networks and can be trained by utilizing the similar methods, generally using minibatch gradient descent after back-propagation-derived gradients. Recirculation (Hinton and McClelland, 1988), a researching approach created on evaluating the initiations of the network on the model input to the activations on the recreated input, can also be used to train autoencoders, in contrast to generic feedforward networks. Although it is infrequently utilized for machine learning applications, recirculation is seen to be more biologically plausible than back-propagation.

Using the data we had collected and the methods we used in our research, we have processed our data for input and trained our data and implemented denoiser. After then it reshaped the denoised data

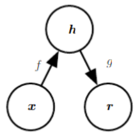

Figure 3.6: auto encoder architecture

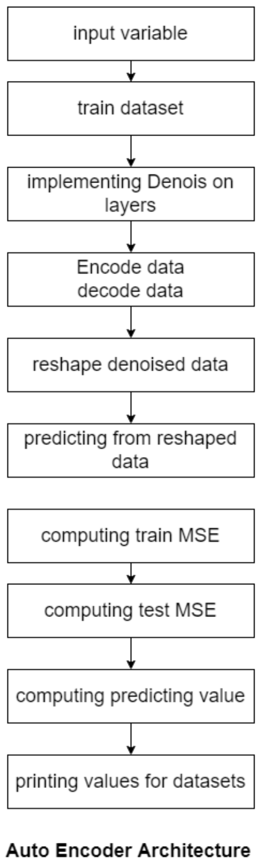

Figure 3.7: Predicting data using auto encoder architecture

and computes the prediction of test MSE from train MSE and from them computes the predicted value. which is also shown in the figure 3.6[17]

3.4 ARIMA And SARIMA Model Architecture

ARIMA and SARIMA models are well known models for time series forecasting. ARIMA is mainly a combination of three parts AR, I and MA which stands for Auto Regression, Integration, Moving Average. And same as ARIMA, SARIMA includes seasonal effects with that. These models create a linear equation describing and also forecast time series data based on the data set that is provided. ARIMA has been used since the 1970s for analyzing time series data. To use these models effectively we need to consider and look for three things. Firstly, if there is any known seasonality in the dataset. Secondly, if there are a lot of outliers in the dataset. And lastly, if the variation of the data about mean is inconsistent or not[15][7].

The ARIMA model is always represented by p, d and q. Where p represents the AR terms, q for MA and d for I. For that we write (P,D,Q) for ARIMA. But for SARIMA we write it (P,D,Q,s). Where s stands for seasonality.

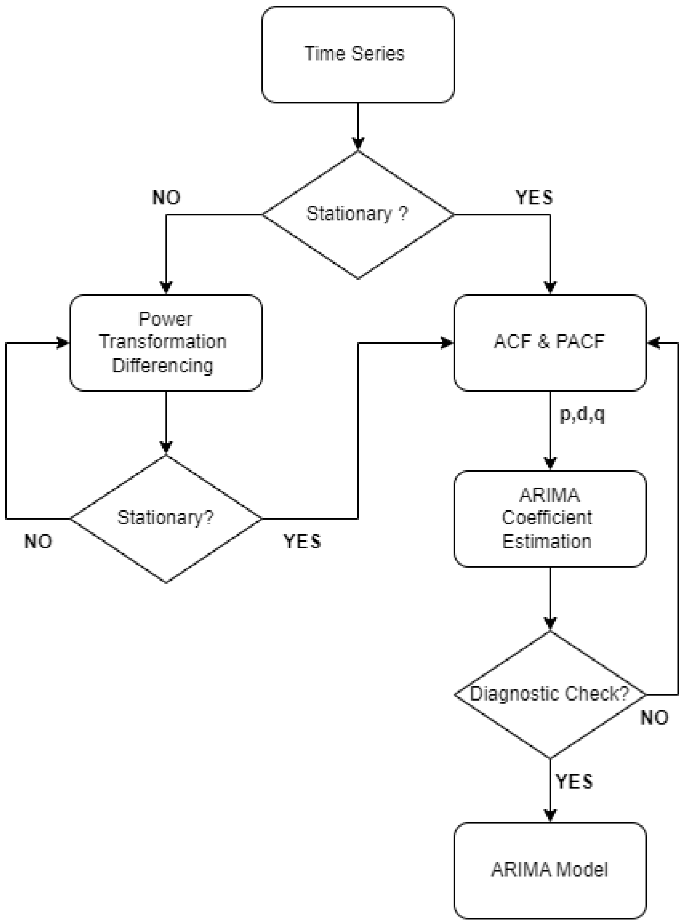

Figure 3.8: ARIMA model architecture

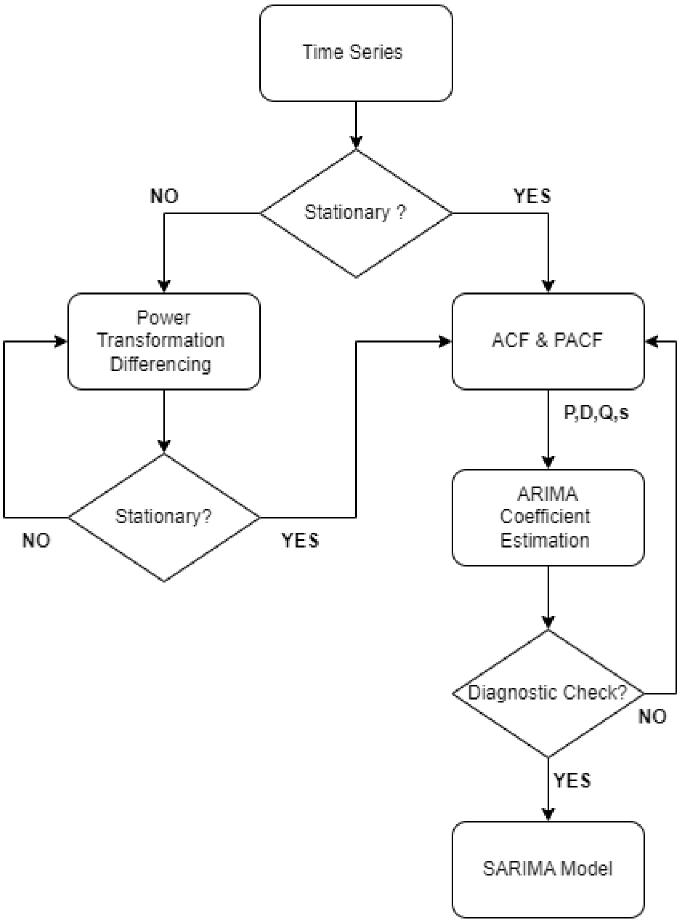

Figure 3.9: SARIMA model architecture

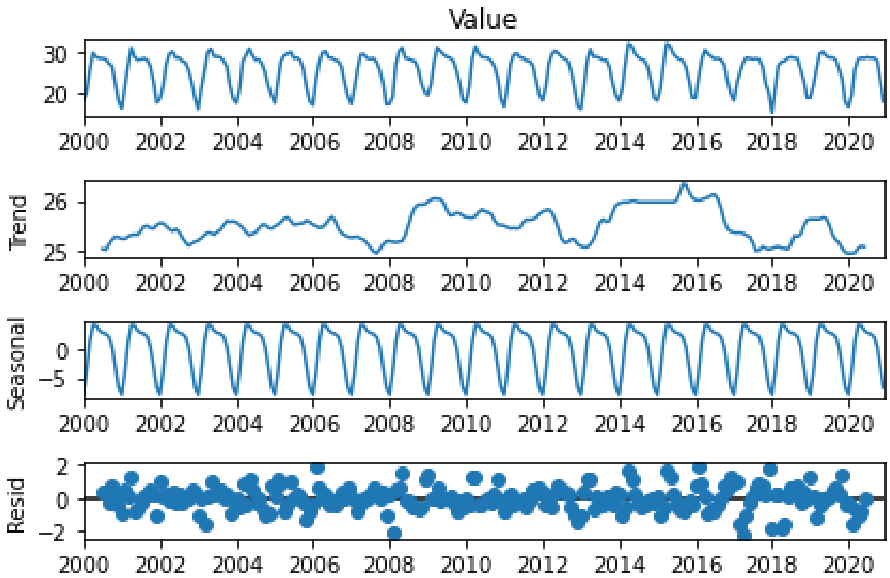

Figure 3.10: Data's seasonality and Trend

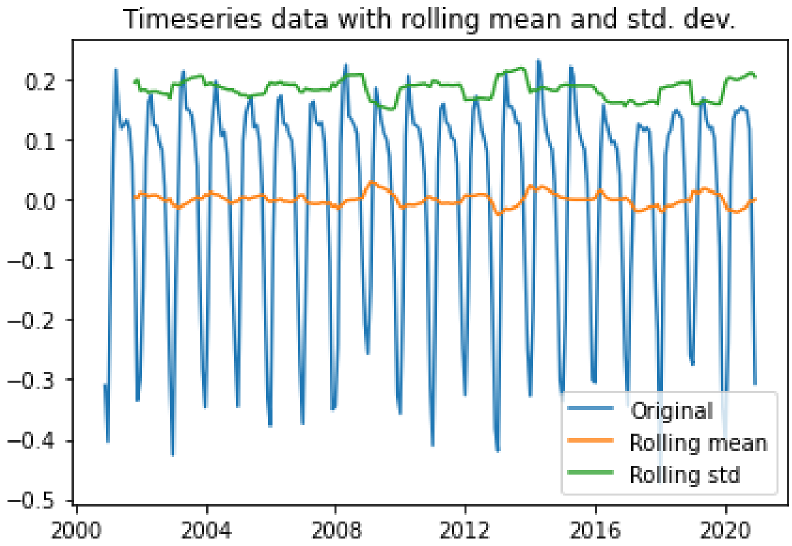

Figure 3.11: Stationy And Null Hyphothesis Test

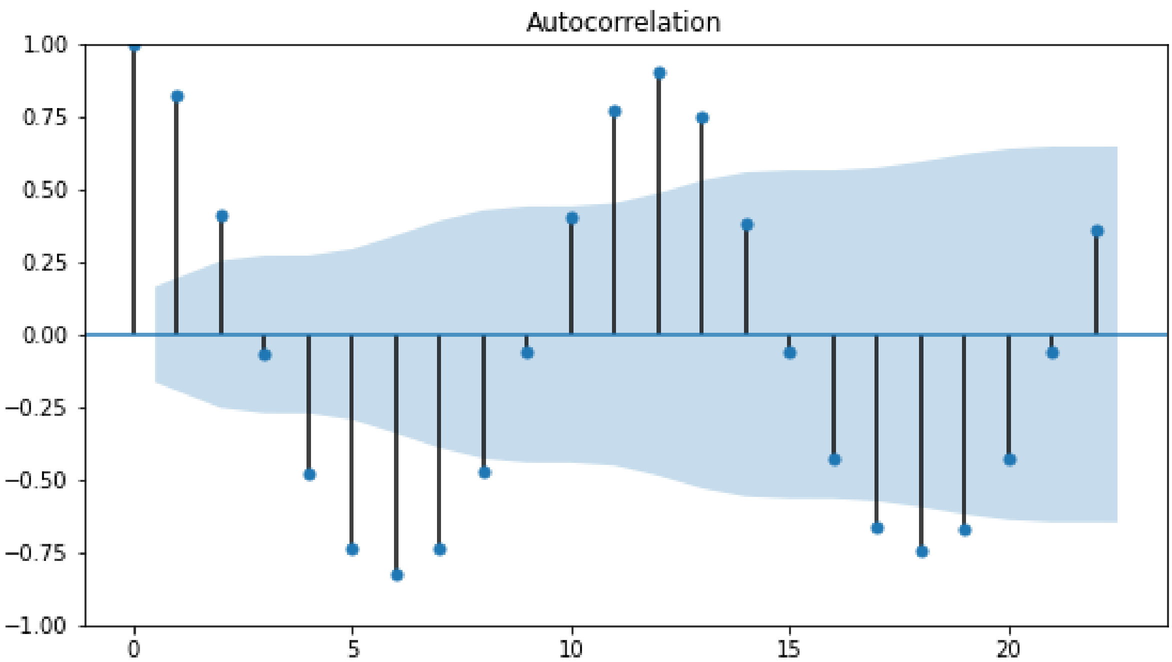

Figure 3.12: ACF plot

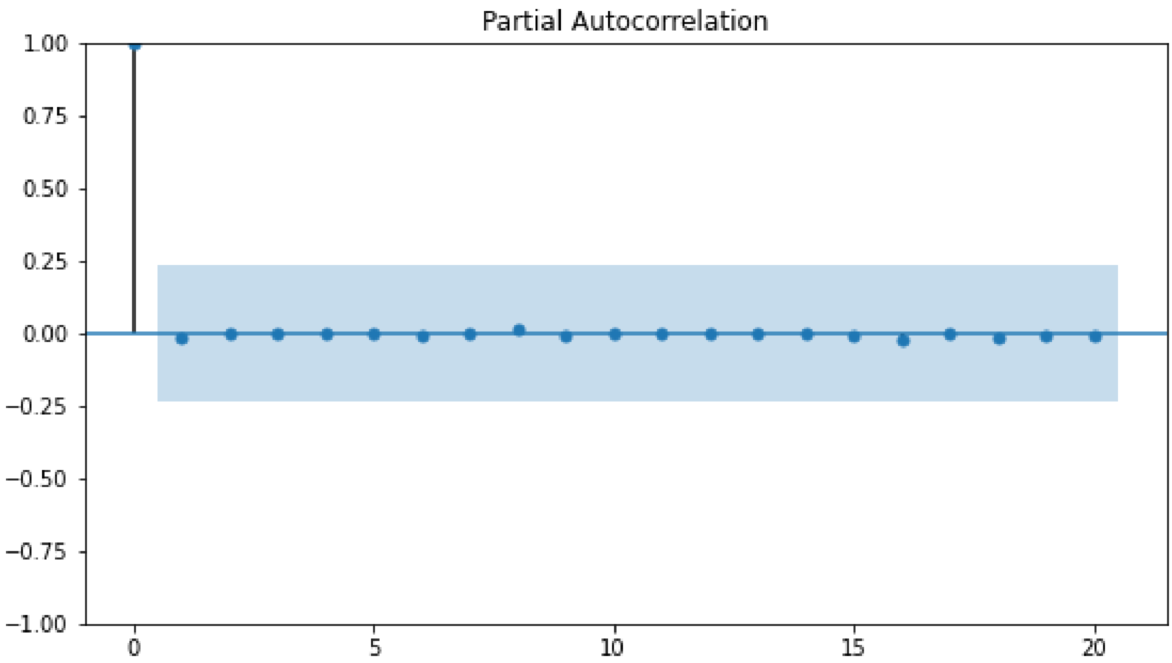

Figure 3.13: PACF plot

In our use case we took the data set and preprocessed that and after that we tested if the data was seasonal or not which we get from watching the trend and seasonality test. In our testing we found that our data has a trend of seasonality in it. After that we tested stationarity of the data by visual analysis and by null-hypothesis testing. For visual analysis we used rolling standard deviation and rolling average. For null hypothesis testing we have used the adfuller testing. Here we get Mackinson's P value which is less than 0.05 which is rejecting the null hypothesis and the data is stationary and also seasonal. But to remove the seasonality we logged scale the data subtracting the moving average we almost get a constant value the rolling mean close to 0 but the data is having a character of seasonality. Before fitting the data in the model we need the value of p and q from ACF where PACF plots where it gets to 0. We found 3 for P and 1 for Q. and we splitted the data for train and test and fitted into the model and for SARIMA we included the s Value of 12 which indicates the seasonality of the data

3.5 Facebook's Prophet Model Architecture

The Prophet model by facebook was introduced in 2017. The model is fast and accurate in terms of forecasting. This model allows adjustment of parameters and also customized seasonality which basically helps to improve the forecast. It can also handle the outliers as well as the other data issues by itself.

It is an additive regression model that works based on four main ingredients. First of all this model identifies the changes in the trends from looking into the changing points of the data. Secondly, using the fourier series it creates a model of the yearly components. Then using dummy variables, it generates a model of weekly components. Lastly, a list of important holiday lists given by the user if needed. Here linear or logistic growth curve for modeling non-periodic changes in time series is mentioned as g(t). Secondly , periodic changes or seasonality is mentioned by s(t). Thirdly, the effects of holidays (user-provided) with irregular schedules is mentioned by h(t). Lastly, e(t) indicates an error term that accounts for any unusual changes not accommodated by the model.[19][31].

$$y(t) = g(t) + s(t) + h(t) + E_t$$

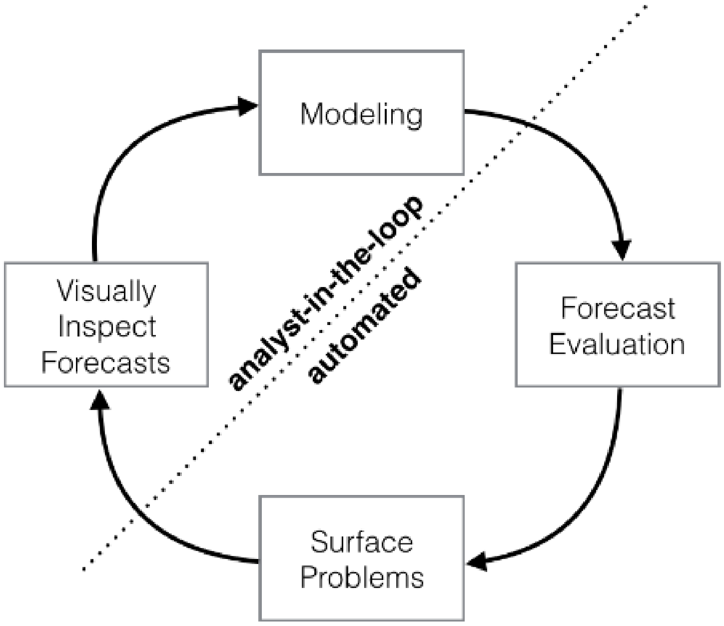

Figure 3.14: Facebook's Prophet Model Architecture

First of all we decomposed the data to check its trends, seasonality and residual component. In our testing we found that our data has a trend of seasonality in it. To fit the data to model we need to reshape the columns as ds and y.. After that we put the model with interval width and daily seasonality marked on as our data-set is seasonal. And for prediction we need to set the period for how much the model is required to model and frequency as well. As we don't have any holiday here to mention we are not required to do anything about it. From the forecast model we find that there is part call yhat which also show the uncertain interval like a range indulging the main forecast.

3.6 LSTM and Bi-LSTM Model Architecture

LSTM stands for long short term memory. It contains memory blocks of recurrent hidden layers. These types of memory cells can store temporal states of the network which are self connected with gates to control the flow of any data. Every memory block has an input as well output gate. These gates respectively control the flow. There is a forget-gate added to the memory cells. It discards the data that is not required in the model[25].

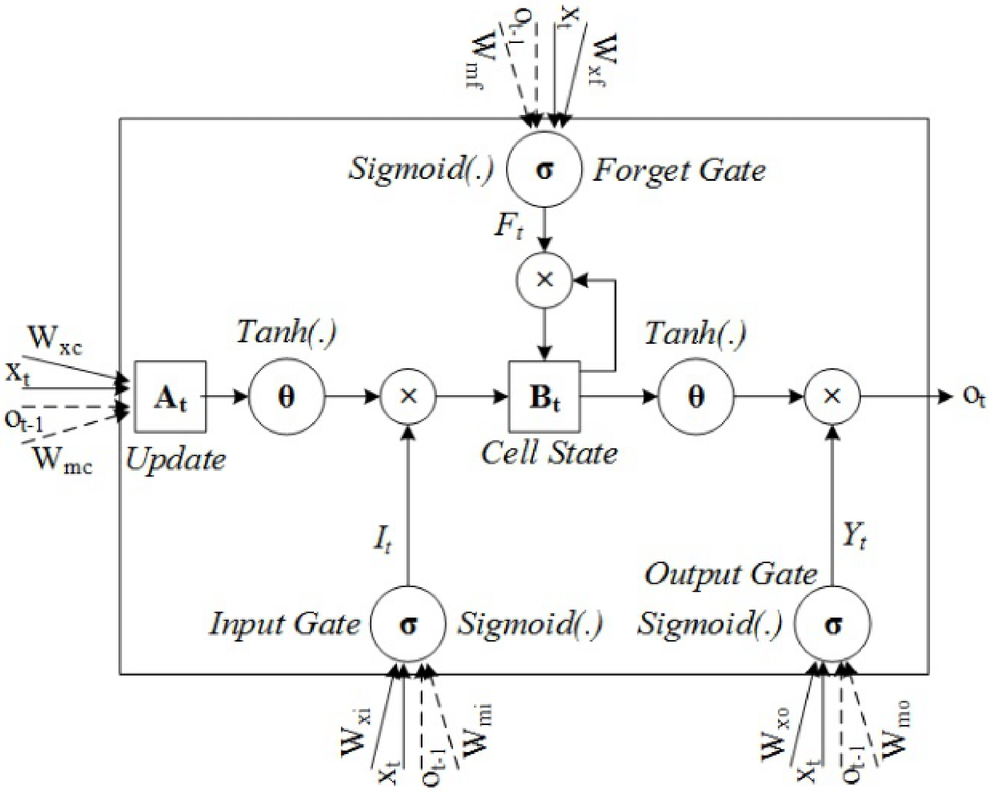

Figure 3.15: General LSTM Model Architecture

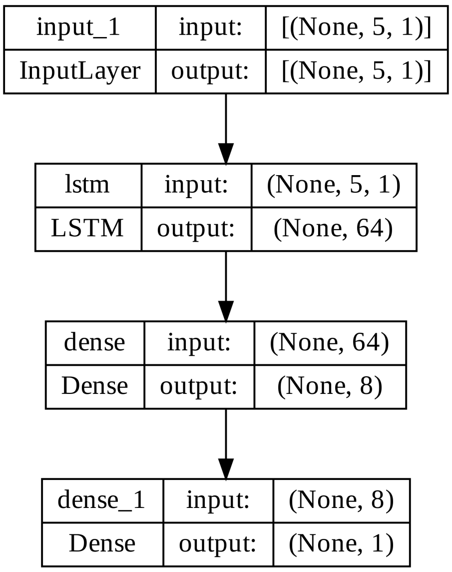

Figure 3.16: LSTM Model Architecture of our testing

Bi-LSTM stands for bidirectional long short-term memory. In bi-lstm it doesn't just train a single model but 2 models. The first one gets the sequential pattern of the input. Then the second one learns the reverse of the sequential pattern. Both of these mechanisms are combined into one then. The default merging policy is the concatenation. There are others like sum, multiplication and averaging[16][30].

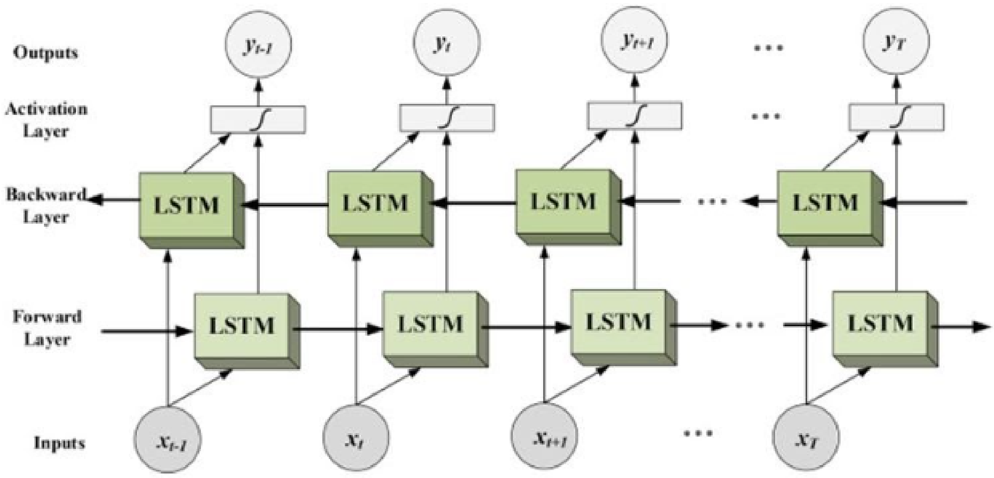

Figure 3.17: General Bi-LSTM Model Architecture

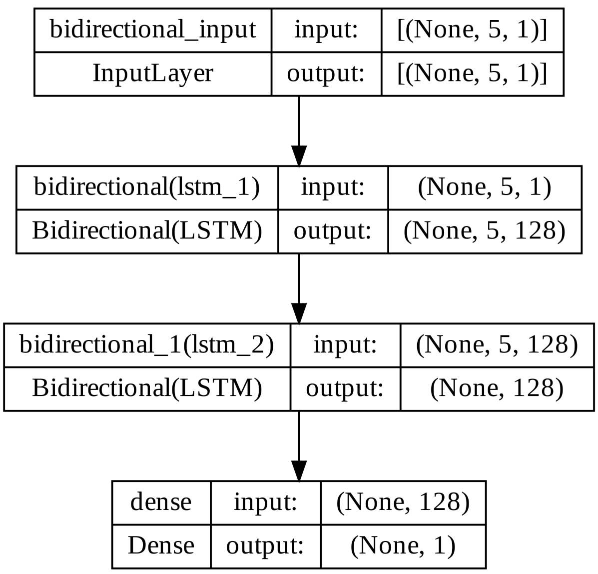

Figure 3.18: Bi-LSTM Model Architecture of our testing

In our testing we splitted our data-set for training and testing and for validation. We took a sequential model, we added input layer, dense and LSTM for LSTM and for Bi-LSTM, we added bidirectional layer into the Bi-LSTM. Here we added an optimiser called Adam for learning rate, we calculated loss function to get the rmse to check the accuracy. We fitted the model with train and validation data with 100 epochs. Then we compared the predicted data with actual to test the accuracy. In our testing the models gave us a pretty good outcome.

Chapter 4

Findings

Here in our research, we have used two different data set. For the first one we used data of ERA5 reanalysis data-set (Hersbach et al., 2020)[23] for training and testing using linear regression method and an auto encoder model to train and get prediction for days to get the test data and training data accuracy.

And for the second one we used the numerical data of temperature at 2 meter of latitude 23.7803 degree and longitude of 90.4071 degree. This data-set was collected from https://power.larc.nasa.gov/data-access-viewer/WER — Data Access Viewer (nasa.gov). The data-set we collected contains only temperature at 2 meter from January 1 2000 to December 31 of 2020 and this was daily data.
Here are some findings and charts and figures shown below to get a better understanding of our research and its purpose.

4.1 Training Data set

For the first testing, we selected from 2012 to 2016 for training data purposes. Though we wanted to take more data for training purpose but due to the hardware limitation we had to stick with shorter training period with lower resolution of data which is of 5.625° which are in following figure 4.1-4.3

And for the second part, we took numerical data of temperature at 2 meter from January 1 2000 to December 31 2020. For ARIMA, SARIMA we used monthly interval data due to longer training periods and for the rest we used daily interval, which are in following figure 4.4-4.5

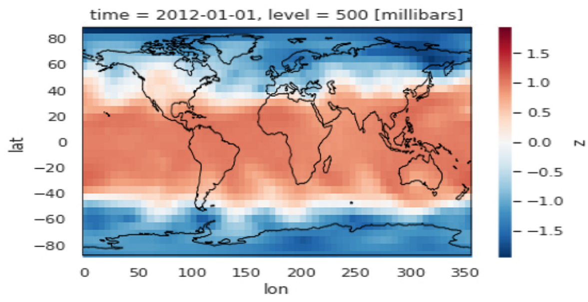

Figure 4.1: Geo potential 500 Training data

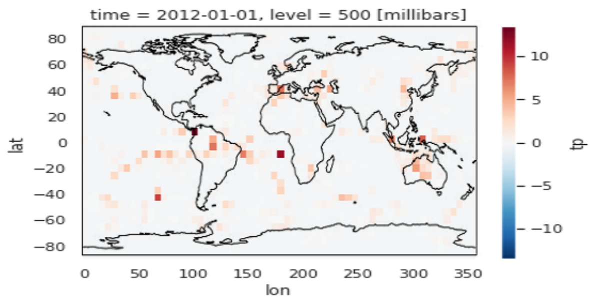

Figure 4.2: Total Precipitation Training data

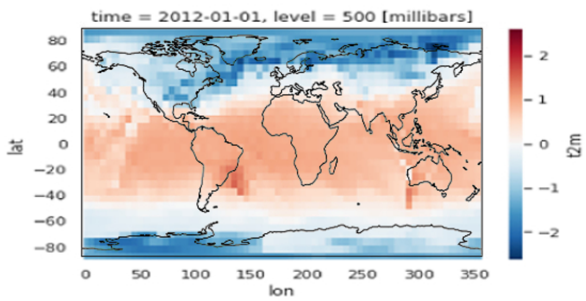

Figure 4.3: 2m temperature Training data

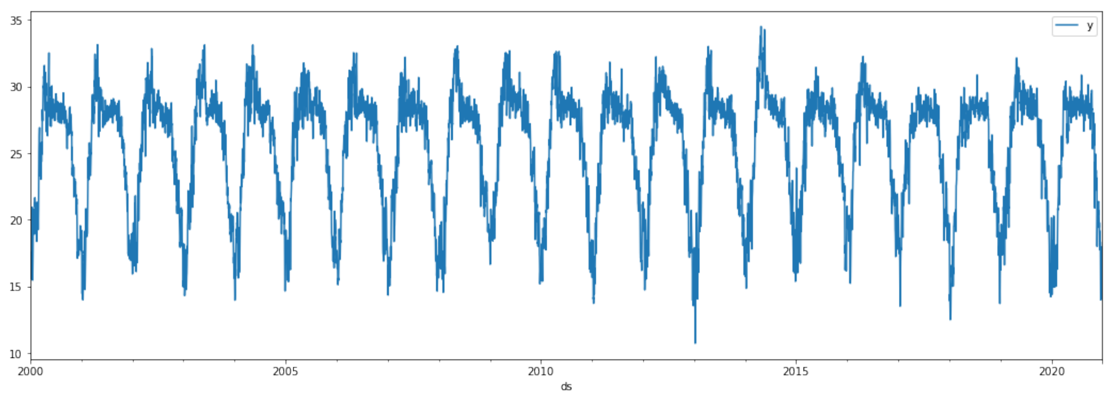

Figure 4.4: 2m temperature Numerical Training data(Daily-interval)

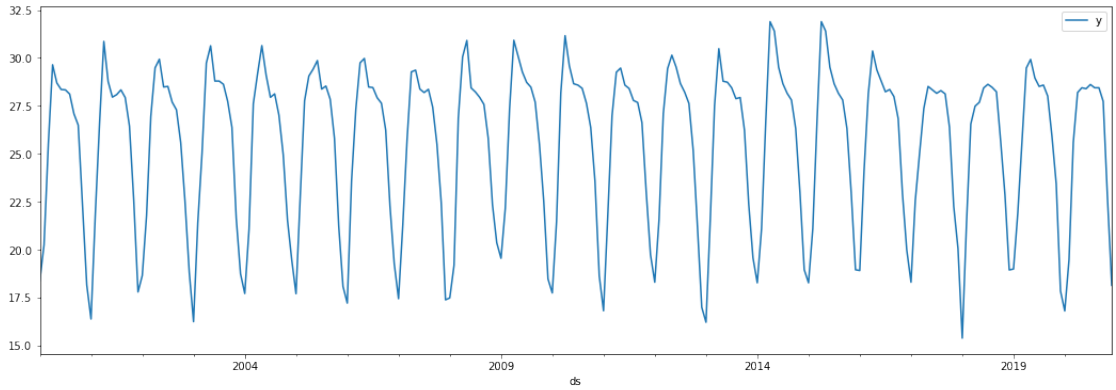

Figure 4.5: 2m temperature Numerical Training data(Monthly-interval)

4.2 Findings of Linear regression

We have plotted few predictions of linear regression that we found from our testing and the Value is also mentioned below in the table 4.1 and figure 4.4-4.6

Days	z500 train mse/ test mse/ z500	tp train mse/test mse/ tp	t2m train mse/ test mse/ t2m
Day 1	0.008/0.315/524.835	0.625/1.491/0.002	0.004/0.016/2.221
Day 2	0.015/0.067/770.811	0.643/1.547/0.002	0.005/0.026/2.794
Day 3	0.019/0.091/896.0.78	0.647/1.566/0.002	0.006/0.032/3.059
Day 4	0.022/0.106/967.916	0.648/1.578/0.002	0.006/0.035/3.218
Day 5	0.023/0.116/1011.174	0.657/1.582/0.002	0.006/0.038/3.322
Day 6	0.024/0.122/1036.479	0.650/1.590/0.02	0.007/0.039/3.388
Day 7	0.025/0.126/1053.522	0.650/1.589/0.002	0.007/0.041/3.429

Table 4.1: Chart of train mse, test mse and predicted value of each data set from linear regression

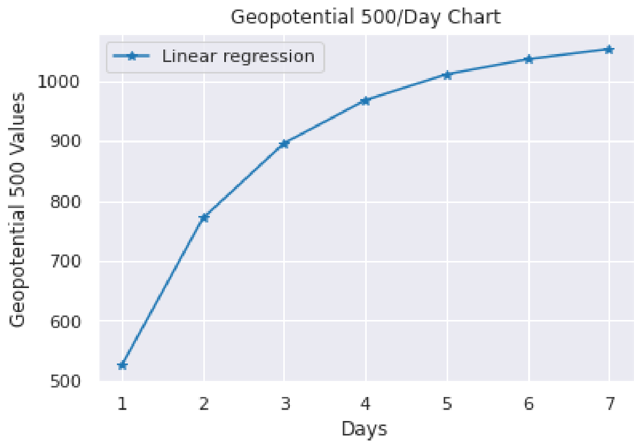

Figure 4.6: Weather forecast of Geo-potential 500 for 7 days using Linear regression

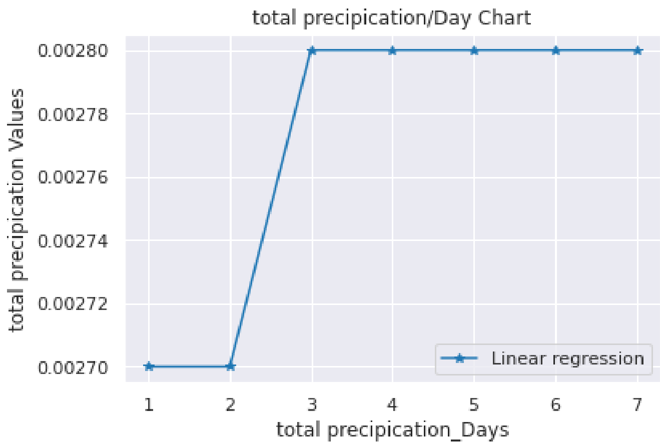

Figure 4.7: Weather forecast of Total Precipitation for 7 days using Linear regression

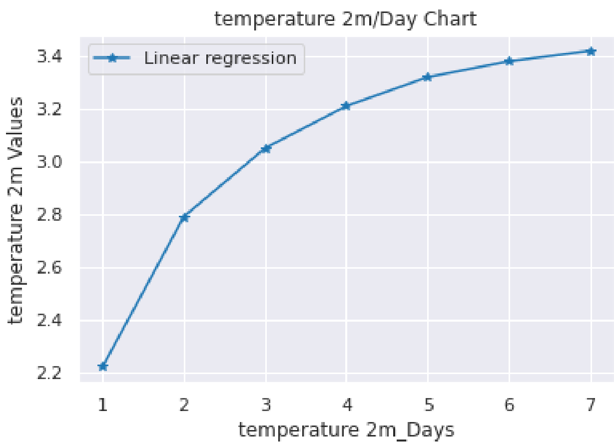

Figure 4.8: Weather forecast of 2m temperature for 7 days using Linear regression

Days	z500 train mse/ test mse/ z500	tp train mse/test mse/ tp	t2m train mse/ test mse/ t2m
Day 1	0.557/0.545/1702.549	1.139/1.154/0.002	0.634/0.615/9.553
Day 2	0.562/0.550/1722.374	1.139/1.154/0.002	0.634/0.616/9.567
Day 3	0.565/0.554/1738.986	1.139/1.154/0.002	0.635/0.616/9.573
Day 4	0.568/0.557/1748.746	1.139/1.155/0.002	0.635/0.617/9.576
Day 5	0.568/0.588/1751.762	1.139/1.155/0.002	0.635/0.617/9.578
Day 6	0.569/0.559/1755.514	1.139/1.155/0.002	0.635/0.617/9.580
Day 7	0.569/0.559/1756.067	1.139/1.155/0.002	0.635/0.617/9,575

Table 4.2: Chart of train mse, test mse and predicted value of each data set from auto encoder

4.3 Findings of Auto Encoders

We have plotted few predictions of Auto Encoders that we found from our testing and the Value is also mentioned below in the table 4.2 and figure 4.7-4.9

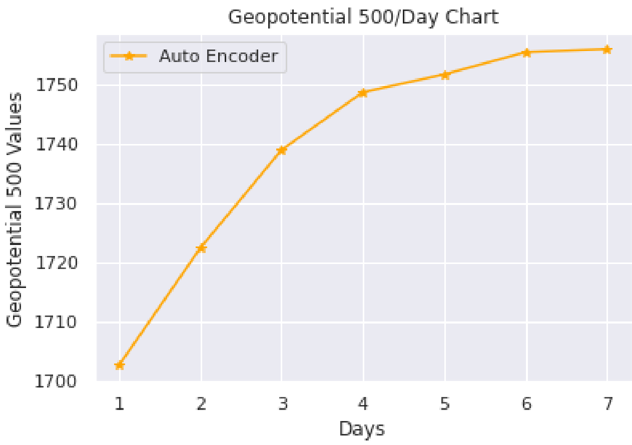

Figure 4.9: Weather forecast of Geo-potential 500 for 7 days using Auto Encoder

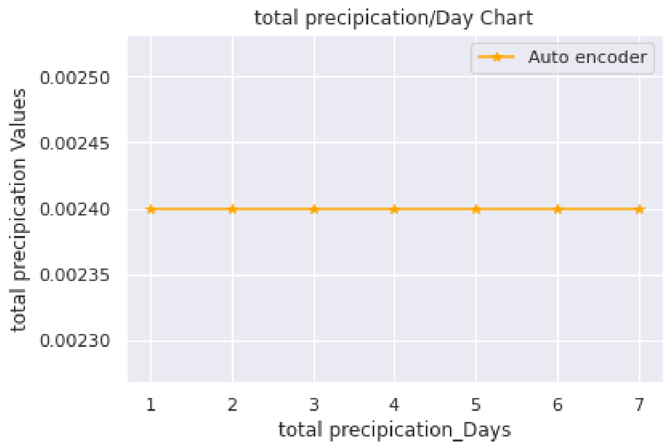

Figure 4.10: Weather forecast of Total Precipitation for 7 days using Auto Encoder

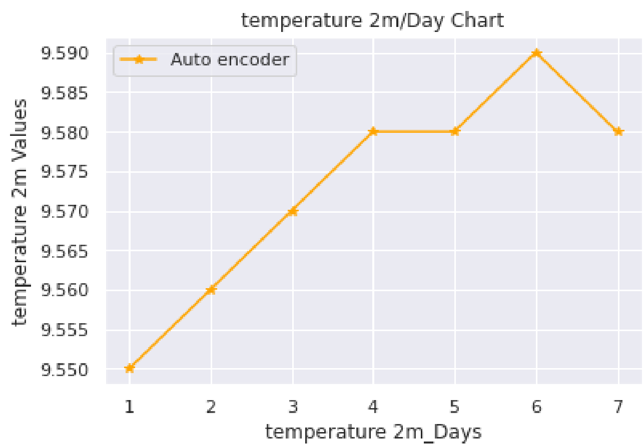

Figure 4.11: Weather forecast of 2m temperature for 7 days using Auto Encoder

4.4 Findings of Auto Encoders vs linear regression

We have plotted few comparison between predictions of Auto Encoder and linear regression model that we found from our testing and the Value is also shown below in figure

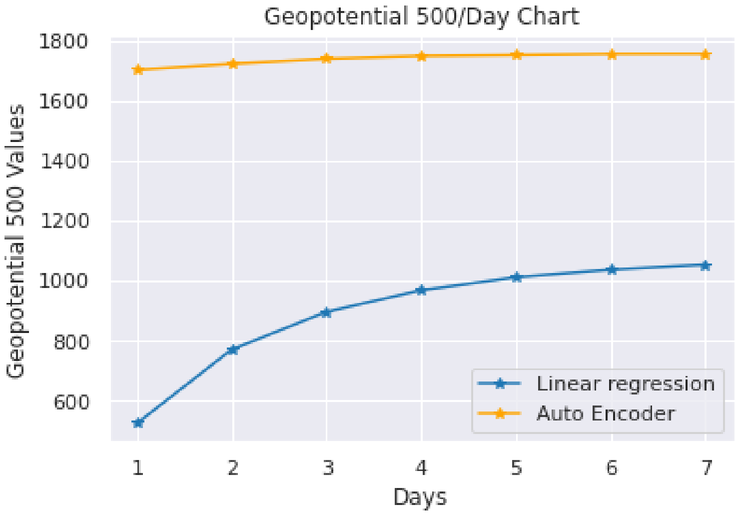

Figure 4.12: Weather forecast of Geo-potential 500 for 7 days using auto encoder vs linear regression

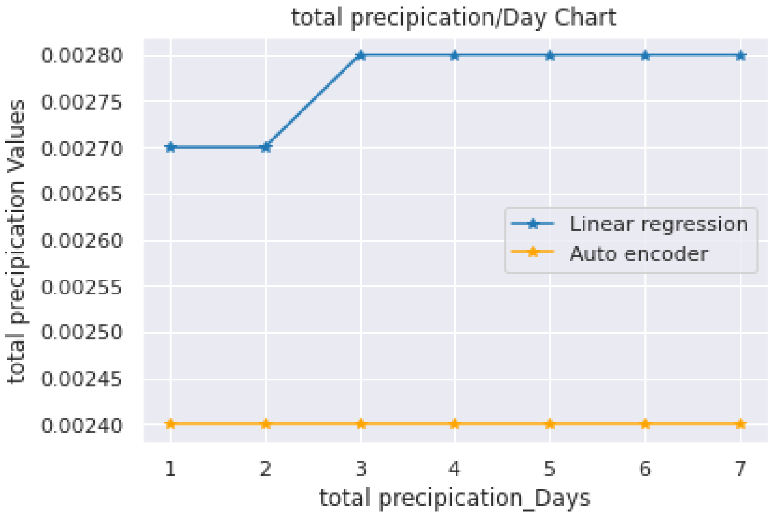

Figure 4.13: Weather forecast of Total Precipitation for 7 days using auto encoder vs linear regression

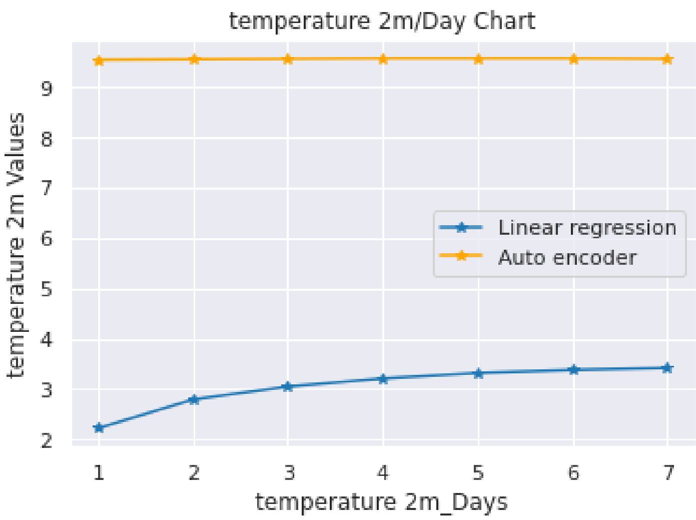

Figure 4.14: Weather forecast of 2m temperature for 7 days using auto encoder vs linear regression

4.5 Findings of ARIMA and SARIMA Model

In our testing we have trained our ARIMA and SARIMA models using the numerical data set of temperature at 2 meter and found these as the results.

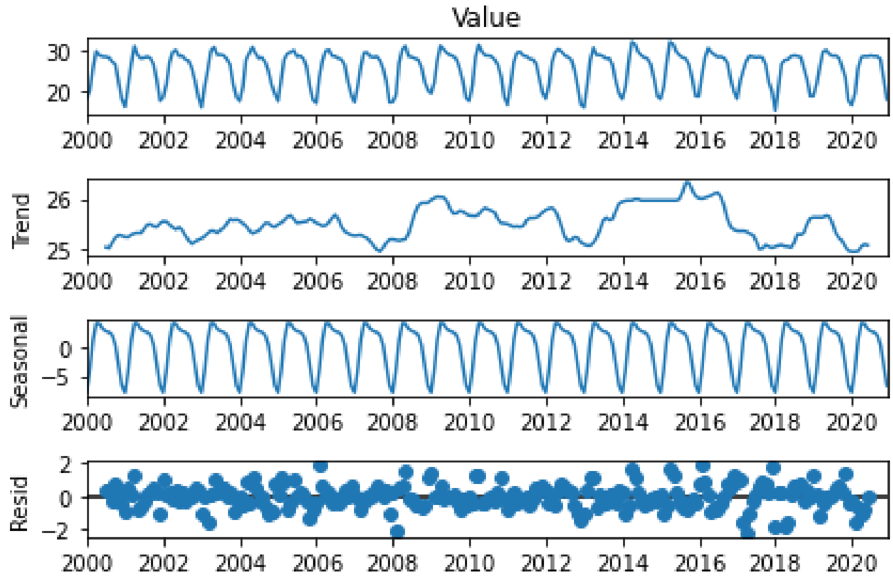

Figure 4.15: Data's seasonality and trend

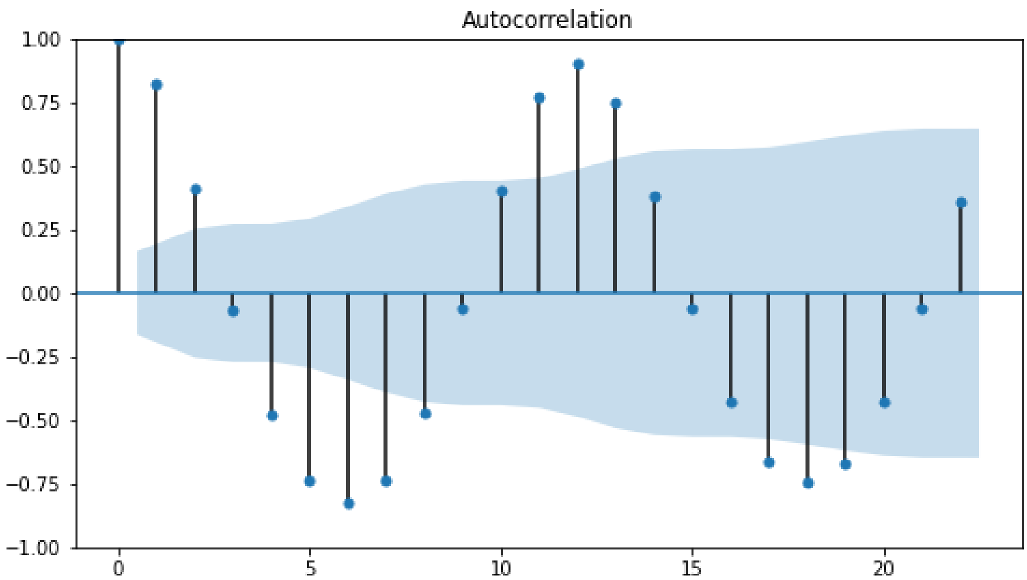

Figure 4.16: ACF

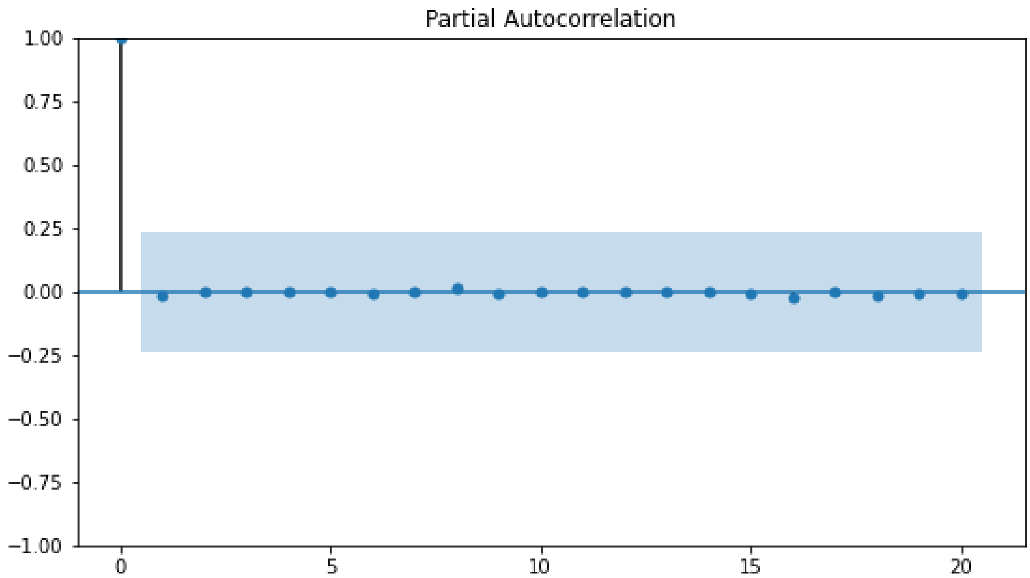

Figure 4.17: PACF

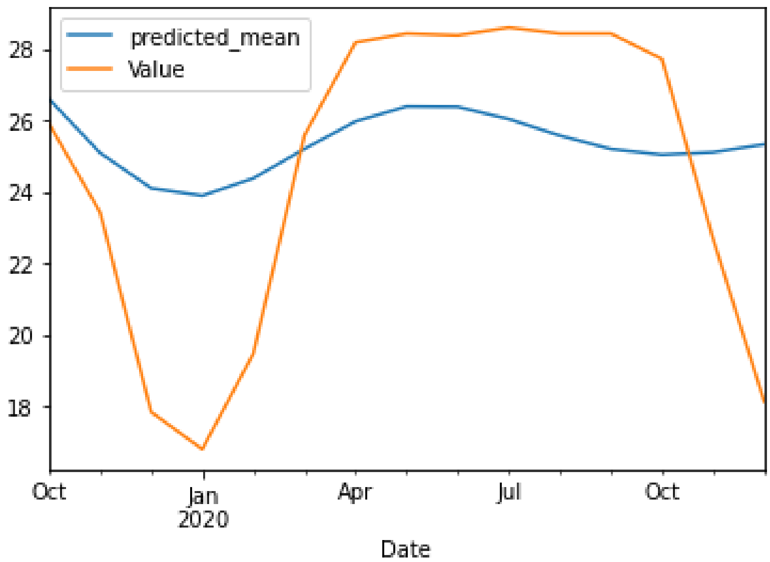

Figure 4.18: Prediction from ARIMA model

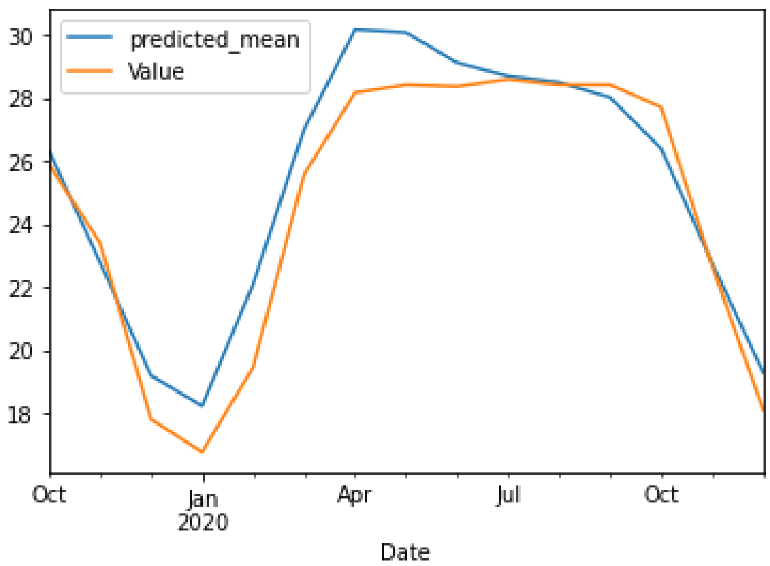

Figure 4.19: Prediction from SARIMA model

4.6 Findings of Facebook Prophet Model

In our testing we have trained our Facebook Prophet model using the numerical data set of temperature at 2 meter and found that as the results.

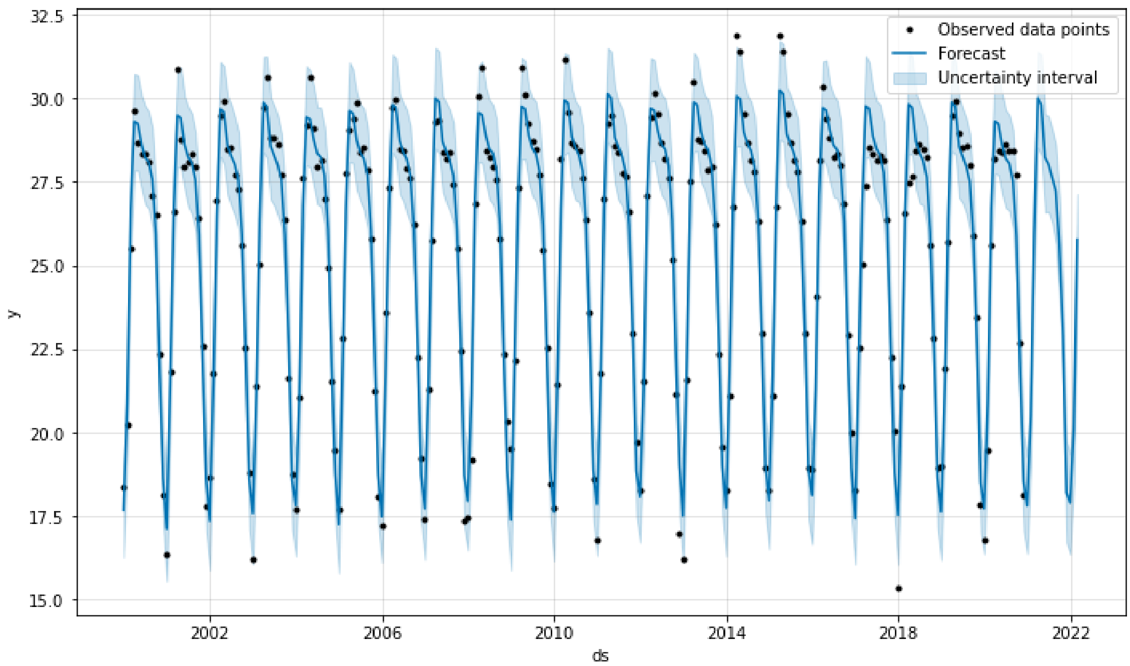

Figure 4.20: Prediction from Facebook Prophet model

in the Figure 4.18 the black dots are the actual data and the blue line that is passing across the dots is the prediction from the model.

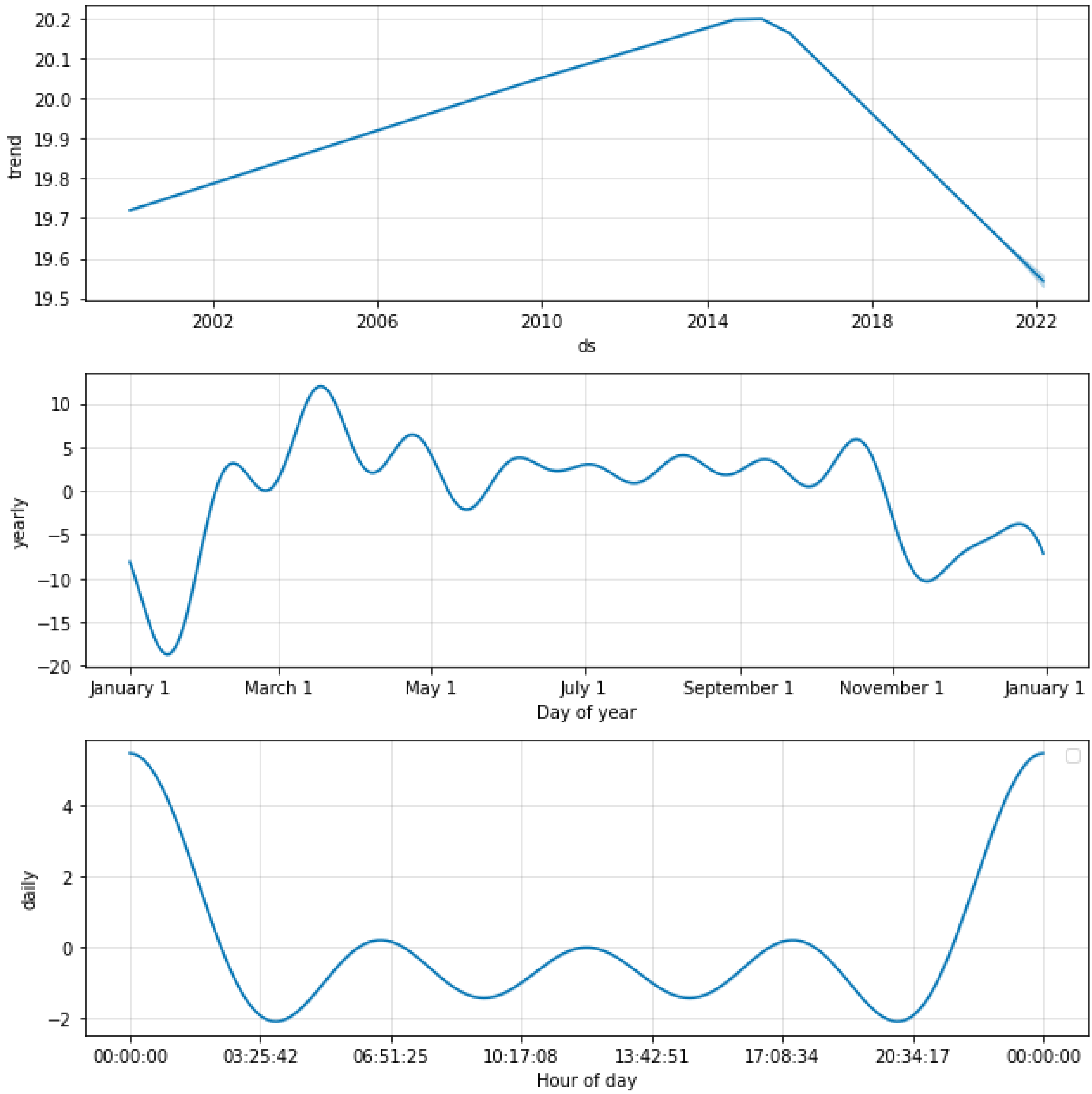

Figure 4.21: Data's seasonality and trend

4.7 Findings of LSTM and Bi-LSTM Model

In our testing we have trained both LSTM and Bi-LSTM models using the numerical data set of temperature at 2 meter and found these as the results.

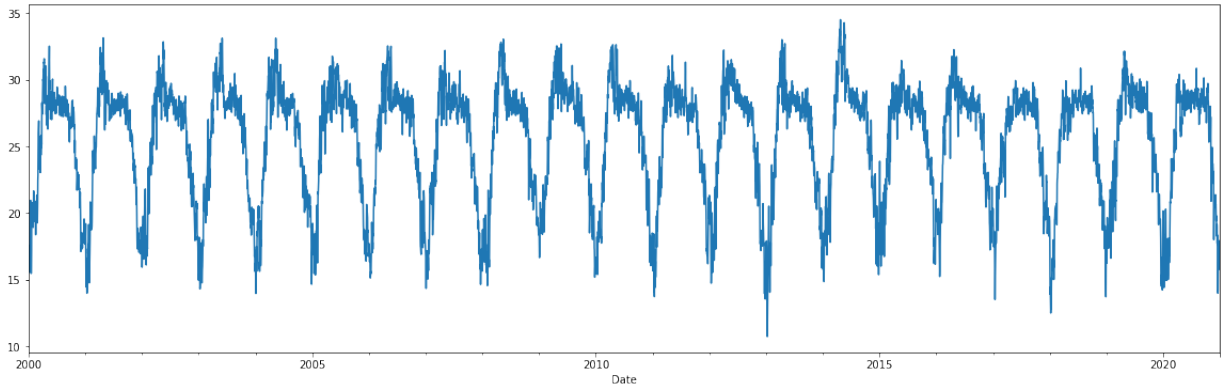

Figure 4.22: Training Data

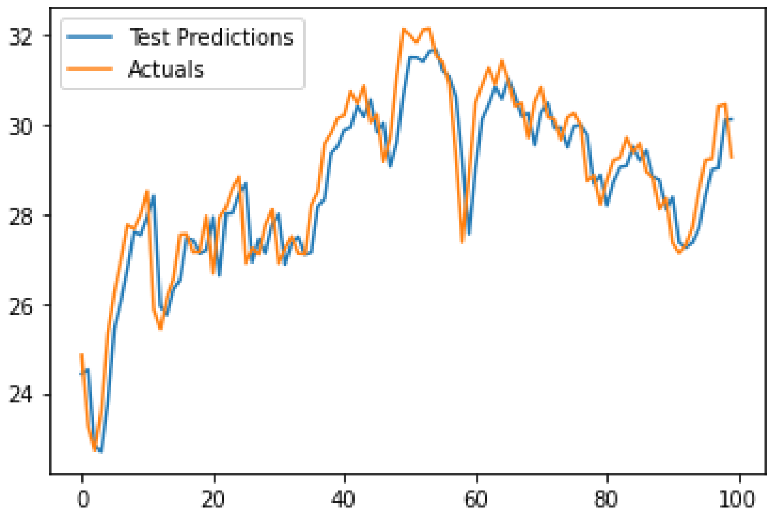

Figure 4.23: Prediction from LSTM model vs actual data

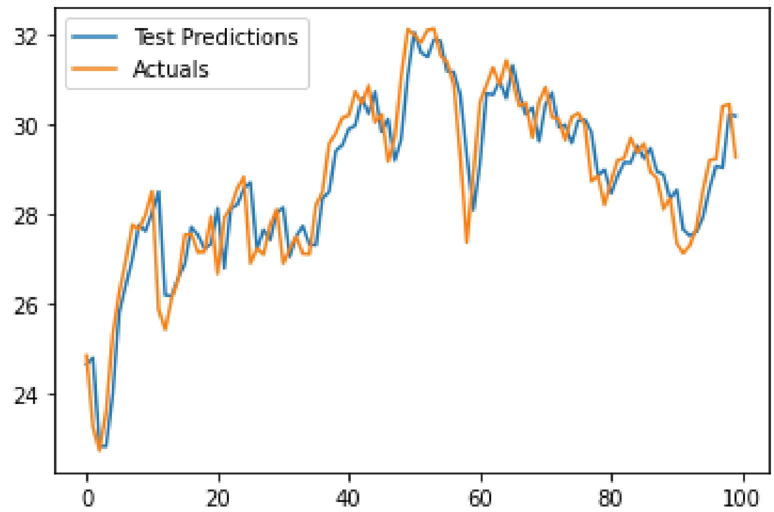

Figure 4.24: Prediction from Bi-LSTM model vs actual data

4.8 Comparison of RMSE among the used models

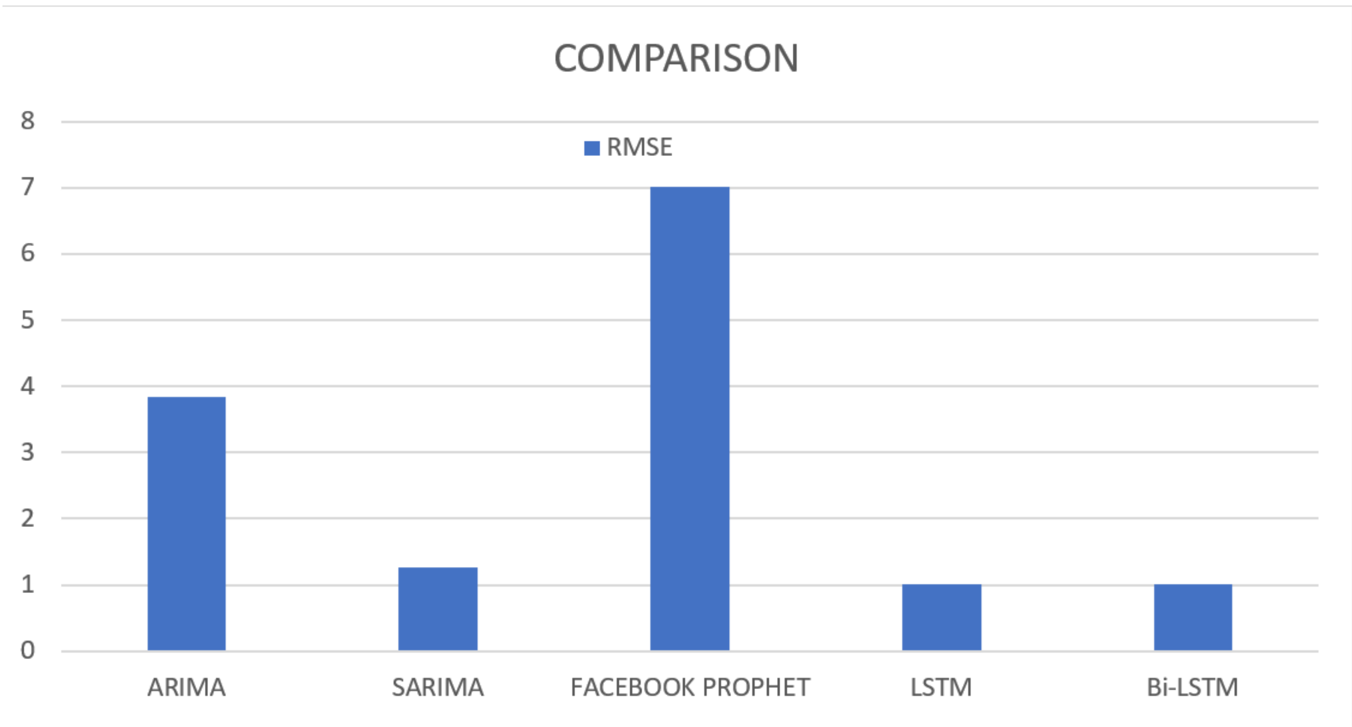

Figure 4.25: RMSE comparison

Chapter 5

Summary, Conclusion and Discussion

The reanalyzed data sets that provide the best estimate of the atmospheric condition at any given time and the integration of the forecasting model with the available data observation are utilized in this work for testing and training purposes. In order to improve prediction accuracy, we used this data to reshape and normalize the data for training. We discovered that the precision of the data declines as time passes. Here, we employed three sets of data: the z500 dataset, the temp 2m dataset, and the t850 dataset.

Our first goal was to determine how long the data is true since, as we all know, there are many models that can make correct predictions, but as time goes on, those predictions become less accurate. The prediction's accuracy declines as time goes on. Since we learned a lot from the literature review, we read those publications and acquired the data that can aid us in our research. These connected works assisted us in obtaining the knowledge we required. For the forecasting procedure in our research, deep learning is being used. Here, we've already gathered a ton of meteorological benchmark data that we used for training and testing. Deep learning may be used to construct a variety of weather forecasting techniques and models, each with a different level of accuracy. In our study, we demonstrated how data utilization and data chunk choice could affect how accurately the weather will be predicted in the coming days. A data analysis technique called linear regression uses a model to forecast the accuracy of predictions from known or similar data that are used in the same way. Plotting dependent or ambiguous variables as a linear equation is useful. We divided our dataset into two parts for our study. One is used for testing data, and the other is for testing. We selected data for temperature at two meters, geopotential at 500 millibars, and total precipitation for the testing set from 2012 to 2016. After picking the data, we combined all of the data of 5.625°, normalized our data sets, and reshaped them. We used test data-based variables in our linear regression training model that were controlled by data training. After that, we trained the data once more and determined the RMSE for both the test and training sets of data. The value for each data set of temperature at 2 meters, geopotential at 500 millibars, and total precipitation is then calculated using a linear regression prediction function. We put this to the test for a few days. Which outcomes will be presented in the findings section. An unsupervised neural network called an auto encoder is used to train it and then copy the taught data to its output. Our auto encoder model comprises three components. First, a denoise model is applied to variable data from training sets and input via an auto encoder. The denoise model then begins to reshape the data, using the Keras sequential library to do so precisely to the desired size. Finally, it provides the data to the prediction and accuracy model, which is continually trained over a period of time to predict the coordination's and qualities. Additionally, it calculates the train and test MSE and outputs an MSE result along with a projected value for each data set. In this study, we used the ERA5 reanalysis dataset for training and testing[23] using the linear regression approach and an auto encoder model to train and get prediction for days to get the test data and training data accuracy.

But at the end of the first part we were not yet satisfied with the task. Hence we tried ARIMA, SARIMA, LSTM, Bi-LSTM and Prophet model to find out the accuracy of these model's using the second data set we collected and in our time series prediction models we have found from the chart of RMSE comparison that Bi-LSTM model gives us the best accurate prediction among all these models that we have tested in our research. As Bi-LSTM uses both forward and reverse directional input, the model gave us the best possible outcome among these models we have tested.

According to the study, the accuracy of the prediction reduces steadily as time passes. For instance, if we use data from a single month, the accuracy of the prediction is good, but if we use data from a full year, the accuracy of the prediction declines significantly. We encountered many problems while working on the project, such as the datasets being so vast that they were difficult to deal with and causing multiple crashes. It takes considerably better hardware to work with this kind of data. In order to advance this research, we are considering incorporating picture data into the GAN model to produce a prediction model. Tensorflow's GAN model will be implemented.

Bibliography

[1] D. J. Bartholomew, *Time series analysis forecasting and control.* 1971.

[2] R. Agrawal, J. RC, J. MP, *et al.*, "Forecasting of rice yield using climatic variables," 1980.

[3] L.-J. Lin, "Scaling up reinforcement learning for robot control," in *Proc. 10th Conf. on Machine Learning*, 1993, pp. 182–189.

[4] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, ISSN: 0899-7667. DOI: 10.1162/neco.1997.9.8.1735. eprint: https://direct.mit.edu/neco/article-pdf/9/8/1735/813796/neco.1997.9.8.1735.pdf. [Online]. Available: https://doi.org/10.1162/neco.1997.9.8.1735.

[5] B. Sivakumar, "Rainfall dynamics at different temporal scales: A chaotic perspective," *Hydrology and Earth System Sciences*, vol. 5, no. 4, pp. 645–652, 2001.

[6] C. Chatfield, *The analysis of time series: an introduction.* Chapman and hall/CRC, 2003.

[7] P. Zhang, "Zhang, g.p.: Time series forecasting using a hybrid arima and neural network model. neurocomputing 50, 159-175," *Neurocomputing*, vol. 50, pp. 159–175, Jan. 2003. DOI: 10.1016/S0925-2312(01)00702-0.

[8] P. Guhathakurta, "Long-range monsoon rainfall prediction of 2005 for the districts and sub-division kerala with artificial neural network," *Current Science*, vol. 90, no. 6, pp. 773–779, 2006.

[9] C. Hamzaçebi, "Improving artificial neural networks' performance in seasonal time series forecasting," *Information Sciences*, vol. 178, no. 23, pp. 4550–4559, 2008.

[10] T. Mehmat, "Weather forecasting using anfis and arima models, a case study for istanbul environmental research," *Engineering and Management*, vol. 1, no. 51, p. 5, 2010.

[11] H. Saima, J. Jaafar, S. Belhaouari, and T. Jillani, "Arima based interval type-2 fuzzy model for forecasting," *International Journal of Computer Applications*, vol. 28, no. 3, pp. 17–21, 2011.

[12] K. Abhishek, A. Kumar, R. Ranjan, and S. Kumar, "A rainfall prediction model using artificial neural network," in *2012 IEEE Control and System Graduate Research Colloquium*, IEEE, 2012, pp. 82–87.

[13] F. Mekanik, M. Imteaz, S. Gato-Trinidad, and A. Elmahdi, "Multiple regression and artificial neural network for long-term rainfall forecasting using large scale climate modes," *Journal of Hydrology*, vol. 503, pp. 11–21, 2013.

[14] M. Rahman, A. S. Islam, S. Y. M. Nadvi, and R. M. Rahman, "Comparative study of anfis and arima model for weather forecasting in dhaka," in *2013 international conference on informatics, electronics and vision (ICIEV)*, IEEE, 2013, pp. 1–6.

[15] Y. Li, Y. Su, and L. Shu, "An armax model for forecasting the power output of a grid connected photovoltaic system," *Renewable Energy*, vol. 66, pp. 78–89, 2014.

[16] H. Sak, A. W. Senior, and F. Beaufays, "Long short-term memory recurrent neural network architectures for large scale acoustic modeling," in *INTERSPEECH*, 2014.

[17] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, http://www.deeplearningbook.org.

[18] P. C. Reddy and A. S. Babu, "Survey on weather prediction using big data analystics," in *2017 Second international conference on electrical, computer and communication technologies (ICECCT)*, IEEE, 2017, pp. 1–6.

[19] S. Taylor and B. Letham, *Forecasting at scale*, Sep. 2017. DOI: 10.7287/peerj.preprints.3190v2.

[20] S. Scher, "Toward data-driven weather and climate forecasting: Approximating a simple general circulation model with deep learning," *Geophysical Research Letters*, vol. 45, no. 22, pp. 12–616, 2018.

[21] S. Srivastava and S. Lessmann, "A comparative study of lstm neural networks in forecasting day-ahead global horizontal irradiance with satellite data," *Solar Energy*, vol. 162, pp. 232–247, 2018, ISSN: 0038-092X. DOI: https://doi.org/10.1016/j.solener.2018.01.005. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0038092X18300173.

[22] Z. Zhen, X. Wan, Z. Wang, F. Wang, H. Ren, and Z. Mi, "Multi-level wavelet decomposition based day-ahead solar irradiance forecasting," in *2018 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, 2018, pp. 1–5. DOI: 10.1109/ISGT.2018.8403376.

[23] S. Rasp, "Coupled online learning as a way to tackle instabilities and biases in neural network parameterizations," *arXiv preprint arXiv:1907.01351*, 2019.

[24] S. Scher and G. Messori, "Weather and climate forecasting with neural networks: Using general circulation models (gcms) with different complexity as a study ground," *Geoscientific Model Development*, vol. 12, no. 7, pp. 2797–2809, 2019. DOI: 10.5194/gmd-12-2797-2019. [Online]. Available: https://gmd.copernicus.org/articles/12/2797/2019/.

[25] S. Siami Namini, N. Tavakoli, and A. Siami Namin, "The performance of lstm and bilstm in forecasting time series," Dec. 2019, pp. 3285–3292. DOI: 10.1109/BigData47090.2019.9005997.

[26] Y. Wang, Y. Shen, S. Mao, X. Chen, and H. Zou, "Lasso and lstm integrated temporal model for short-term solar intensity forecasting," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 2933–2944, 2019. DOI: 10.1109/JIOT.2018.2877510.

[27] J. A. Weyn, D. R. Durran, and R. Caruana, "Can machines learn to predict weather? using deep learning to predict gridded 500-hpa geopotential height from historical weather data," *Journal of Advances in Modeling Earth Systems*, vol. 11, no. 8, pp. 2680–2693, 2019.

[28] G. M. Yagli, D. Yang, and D. Srinivasan, "Automatic hourly solar forecasting using machine learning models," *Renewable and Sustainable Energy Reviews*, vol. 105, pp. 487–498, 2019, ISSN: 1364-0321. DOI: https://doi.org/10.1016/j.rser.2019.02.006. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1364032119300905.

[29] B. Belmahdi, M. Louzazni, and A. E. Bouardi, "One month-ahead forecasting of mean daily global solar radiation using time series models," *Optik*, vol. 219, p. 165 207, 2020.

[30] M. J. J. Gul, M. H. Firmansyah, S. Rho, and A. Paul, "Bi-lstm-lstm based time series electricity consumption forecast for south korea," in *Advances in Artificial Intelligence and Applied Cognitive Computing*, H. R. Arabnia, K. Ferens, D. de la Fuente, E. B. Kozerenko, J. A. Olivas Varela, and F. G. Tinetti, Eds., Cham: Springer International Publishing, 2021, pp. 897–902, ISBN: 978-3-030-70296-0.

[31] M. Khayyat, K. Laabidi, N. Almalki, and M. Al-Zahrani, “Time series facebook prophet model and python for covid-19 outbreak prediction,” *Cmc -Tech Science Press-*, vol. 67, pp. 3781–3793, Mar. 2021. DOI: 10.32604/cmc.2021.014918.