Github link:- https://github.com/yogaprasath/FDA-JCOMP-WEATHER_FORCASTING_SYSTEM.git

FALL SEMESTER 2022-2023

A J-Component Project Report

on

Weather Forecasting

to be submitted in partial fulfilling of the requirements for the course

on

Fundamentals of Data Analytics– (CSC3005)

(E1+TE1)

By

DHARMASEELAN M   20BCS0115

YOGA PRASATH M   20BCS0162

LOKESH G   20BCS0144

SURYA M   20BCS0133

TABLE OF CONTENTS

ABSTRACT

Weather forecasting is the application of science and technology to predict the state of the atmosphere for a given location. Ancient weather forecasting methods usually relied on observed patterns of events, also termed pattern recognition. For example, it might be observed that if the sunset was particularly red, the following day often brought fair weather. However, not all of these predictions prove reliable.

Here this system will predict weather based on parameters such as temperature, humidity and wind. User will enter current temperature; humidity and wind, System will take this parameter and will predict weather(rainfall in inches) from previous data in database(dataset). The role of the admin is to add previous weather data in database, so that system will calculate weather(estimated rainfall in inches) based on these data. Weather forecasting system takes parameters such as temperature, humidity, and wind and will forecast weather based on previous record therefore this prediction will prove reliable. This system can be used in Air Traffic, Marine, Agriculture, Forestry, Military, and Navy etc.

INTRODUCTION

The remarkable improvement in the quality of weather forecasts is one of the great successes of environmental science in the 20th century, which continues at a sustained pace at the beginning of the 21st century (see Figure 1 and Bauer et al, 2015). This is due to the progress of numerical prediction systems and the increasing number and variety of observations of the state of the atmosphere and related media (ocean, soils, vegetation, cryosphere), including observations from Earth observation satellites. The rapid development of supercomputers has been one of the keys to this success, which has also required significant scientific work.

Each country in the world has a National Meteorological Service (NMS), whose mission is to make regular observations of the atmosphere and to issue forecasts for government, industry and the public. But only the most advanced countries have Numerical Weather Prediction (NWP) centres, whose products are also distributed to other countries, in exchange for their observations, within the framework of the World Meteorological Organization .

Among the main NWP centres outside Europe are those in the United States, Canada, Japan, Korea, China, Russia, Australia, India, Morocco, South Africa and Brazil. In Europe, only France, the United Kingdom and Germany make numerical forecasts for the entire globe, while the other countries have NWP centres covering only regional areas. The European countries have also come together in a “super-centre”, which is responsible for providing them with medium-range numerical forecasts.

REVIEW-1

Analysis of existing system

GLOBAL FORECAST SYSTEM

The Global Forecast System (GFS) is a global numerical weather prediction system containing a global computer model and variational analysis run by the United States' National Weather Service (NWS).

History

The Global Forecast System (GFS) has been in NWS operations since 1980 and is continuously improved by the Environmental Modeling Center (www.emc.ncep.noaa.gov) whose mission is to maintain, enhance and transition-to-operations advanced numerical guidance systems for the Nation's weather/water/climate enterprise and the global community for the protection of life/property and the enhancement of the economy.

Operation

The mathematical model is run four times a day, and produces forecasts for up to 16 days in advance, but with decreased spatial resolution after 10 days. The forecast skill generally decreases with time (as with any numerical weather prediction model) and for longer term forecasts, only the larger scales retain significant accuracy. It is one of the predominant synoptic scale medium-range models in general use.

Application

As with most works of the U.S. government, GFS data is not copyrighted and is available for free in the public domain under provisions of U.S. law. Because of this, the model serves as the basis for the forecasts of numerous private, commercial, and foreign weather companies.

Principles

The GFS model is a FV3 model with an approximate horizontal resolution of 13 km for the days 0-16 days. In the vertical, the model is divided into 127 layers and extends to the mesopause (roughly ~80 km), and temporally, it produces forecast output every hour for the first 120 hours,[1] three hourly through day 10 and 12 hourly through day 16. The output from the GFS is also used to produce model output statistics.

Accuracy

By 2015, the GFS model had fallen behind the accuracy of other global weather models.[2][3] This was most notable in the GFS model incorrectly predicting Hurricane Sandy turning out to sea until four days before landfall, while the European Centre for Medium-Range Weather Forecasts' model predicted landfall correctly at 7 days. Much of this was suggested to be due to limits in computational resources within the National Weather Service. In response, the NWS purchased new supercomputers, increasing processing power from 776 teraflops to 5.78 petaflops.[4][5][6] As of the 12z run on 19 July 2017, the GFS model has been upgraded. Unlike the recently-upgraded ECMWF, the new GFS behaves a bit differently in the tropics and in other regions compared to the previous version.[7] This version accounts more accurately for variables such as the Madden–Julian oscillation and the Saharan Air Layer. In 2018, the processing power was increased again to 8.4 petaflops,[8] The agency also tested a potential replacement model with different mechanics, the flow-following, finite-volume icosahedral model (FIM), in the early 2010s; it abandoned that model around 2016, after it did not show substantial improvement over the GFS.

In 2019, A major upgrade was held for the GFS, converting it from the GSM, (Global Spectral Model) to the new FV3 dycore. Horizontal and vertical resolution remained the same but this set the foundation for what's now known as the UFS (Unified Forecast System)

On March 22, 2021, the NOAA upgraded the GFS model, coupling it with the WaveWatch III global wave model, which will increase the GFS's resolution from 64 to 127 vertical levels, while extending the WaveWatch III forecasting window from 10 to 16 days. This left some meteorologists hopeful that the GFSv16 upgrade would be enough to close the accuracy gap with the ECMWF's model, which was considered to be the most accurate global weather model at the time.[9][10]

The Future of the GFS

With the initial operational implementation of FV3GFS now accomplished, NOAA's Environmental Modeling Center (EMC) global modeling focus has turned towards development of the next GFS (v16) upgrade, which will include doubled vertical resolution (64 to 127 layers), more advanced physics, data assimilation system upgrades, and coupling to a

NCEP's Global Wave Model using the Unified Forecast System (UFS) community model. GFSv16 was implemented on March 22, 2021.[14]

On 23 September 2020, the first global UFS application at NCEP was implemented in the Global Ensemble Forecast System (GEFS v12). The components of this upgrade include:

- Use of the FV3 global model (same version as GFS v15) as the atmospheric component of GEFS
- Increase in horizontal resolution to ~25 km
- Forecast length increased from 10 to 16 days
- Increased from 21 to 31 members
- Coupling the GEFS atmospheric component to the NCEP Global Wave model
- Run a 32nd member to 5 days (GEFS-Aero) for aerosol prediction, inline aerosol representation based on GOCART (GSD-Chem).

This implementation is the first global-scale coupled system at NCEP, and replaces the previous standalone Global Wave Ensemble and the NEMS GFS Aerosol Component (NGAC) systems. More details can be found at the EMC Model Evaluation Group's GEFS v12 web site, the EMC GEFS web page, and the EMC GEFS-Aerosol web page

Objectives :

The Global Data-processing and Forecasting System (GDPFS) facilitates cooperation and exchange of information in operational meteorology and related field. By contributing to capacity development amongst all Members, especially developing and least developed countries, the Programme supports:

- weather forecasting and warning services, contributing to Disaster Risk Reduction (DRR) and to socio-economic sectors such as agriculture and food security, aviation, marine safety and transportation, and so on;
- climate prediction and the production of climate services, contributing to the Global Framework for Climate Services (GFCS); and
- specific applications through the provision of specialized products

Benefits :

The report estimates that highly weather-sensitive sectors such as agriculture, energy, transport and construction and disaster risk management can benefit by over US $160 billion per year from potential improvements in weather forecasting capabilities that would be within reach given our current state of scientific knowledge and our technology.

A new report published by the World Bank, produced in collaboration with the World Meteorological Organization and the Met Office (UK), estimates improving the collection and international exchange of surface-based observational data will deliver additional socioeconomic benefits worth more than US $5 billion a year.

Challenges :

Forecasting ultimately is a three-step process. These include:

1) Observing

2) Forecasting

3) Communicating.

When a forecaster comes on duty, their first order of business is to become familiar with what is currently happening in the weather. This includes looking at satellite imagery, surface data, precipitation reports, and getting a briefing from other forecasters on duty. The next order of business is to project weather changes into the future to derive a forecast. Short range forecasting, looking a few hours into the future, typically depends on closely observing how weather systems are currently evolving and tracking, and projecting their movement into the future based on what is understood about dynamics of the atmosphere. Forecasts out beyond a day rely more on numerical weather modeling. Numerical forecast models compile data from surface observations, weather balloons and satellite imagery to create a computer-generated simulation of the weather into the future. The model simulations use dynamic equations that express how the atmosphere will respond to changes in temperature, pressure and humidity over time. Forecasters deal with many forecast models that are run several times a day and must decide which ones to rely on based on how well they seem to be handling current weather, how realistic their output is, and how consistent the forecast models are being from one run to the next. The forecaster may even decide that no model can be relied on at that time.

Once the forecasters have created a forecast that they are reasonably confident in, their next role is to deliver that forecast in a way that people can understand and appropriately respond to. The mission of the National Weather Service is to communicate forecasts in a way that helps save lives and protect property. When the potential exists for a weather event that may impact our forecast area, a weather watch will be issued such as a winter storm watch, a high wind watch or a thunderstorm watch. When a watch is issued, those who may be impacted are advised to stay closely tuned to updated forecasts for potential warnings. A weather warning means the warned event is expected and people in the affected area should prepare. National Weather Service warnings often include "call to action" statements which alert the public on how to prepare for the expected weather.

Requirement gathering

Platform requirements:

Hardware/Software	Hardware / Software element	Specification /version
Hardware	Processor	i3
	RAM	2GB
	Hard Disk	250GB
Software	OS	Windows
	Python IDE	Jupyter notebook, pandas
	Neural Prophet	Neuralprophet0.3.2

Data Set

The dataset is a public weather dataset available on Kaggle.

weatherAUS.csv

LITERATURE REVIEW

Daily Weather Forecasting using Artificial Neural Network

ABSTRACT

Daily Weather forecasting is used for multiple reasons in multiple areas like agriculture, energy supply, transportations, etc. Accuracy of weather conditions shown in forecast reports is very necessary. In this paper, the review is conducted to investigate a better approach for forecasting which compares many techniques such as Artificial Neural Network, Ensemble Neural Network, Backpropagation Network, Radial Basis Function Network, General Regression Neural Network, Genetic Algorithm, Multilayer Perceptron, Fuzzy clustering, etc. which are used for different types of forecasting.

Among which neural network with the backpropagation algorithm performs prediction with minimal error. Neural network is a complex network which is self-adaptive in nature. It learns by itself using the training data and generates some intelligent patterns which are useful for forecasting the weather. This paper reviews various techniques and focuses mainly on neural network with back propagation technique for daily weather forecasting. The technique uses 28 input parameters to forecast the daily weather in terms of temperature, rainfall, humidity, cloud condition, and weather of the day

INTRODUCTION

Weather forecasting is a process of identifying and predicting to a certain accuracy the climatic conditions using multiple technologies. Many of the live systems rely on weather conditions to make necessary adjustments in their systems. Forecasting helps to take necessary measures to prevent damage to life and property to a large extent. Quantitative forecast like temperature, humidity and rainfall are important in agriculture area, as well as to traders within commodity markets. Temperature forecasts are used by utility companies to estimate demand over coming days. Since outdoor activities are severely restricted by heavy rain, snow and the chill; forecasts can be used to plan activities around these events, and to plan ahead and survive them .

Nowadays multiple computing techniques are available which can be used for forecasting enhancing its accuracy. Different categories of forecasting methods are Naive approach, Judgmental methods, Quantitative and Qualitative method, Causal or econometric forecasting methods, Time series methods, Artificial intelligence methods, etc. The weather forecast reports needs some intelligent computing which can read the nonlinear data and generate some rules and patterns to study and train from the observed data to predict the weather in future. Use of ANN will give results which are more accurate. Here, the error may or may not reduce completely. But the accuracy will improve as compared to previous forecasts. The weather forecasting is live forecasting where output of the model may be required for daily weather guide or weekly or monthly weather plans. Thus, the accuracy of the result is a very important aspect in this forecasting. Multiple issues are discussed which can be considered to get the accurate results. In Section two, a reviews multiple literature on weather forecasting. Section three introduces different terms about the neural network. Section four proposes a neural network model with all the specifications for forecasting weather with a high degree of accuracy.

ARTIFICIAL NEURAL NETWORK FOR WEATHER FORECASTING

In Forecasting it is intuitive that accuracy is very important .The input parameters for a weather forecasting model are different types of data need different types of methods; and need to be handled accordingly. Statistical methods are usually associated with linear data whereas Artificial Intelligence methods are associated with nonlinear data [13]. Different learning models based on Artificial Intelligence are genetic algorithms, neuro-fuzzy logic and neural networks. Among which neural networks is preferred for time series forecasting for applications such as “stock index forecasting” in financial markets or “fault detection” in machine maintenance [14]. Weather forecasting can be done International Journal of Computer Applications (0975 – 8887) Volume 121 – No.22, July 2015 11 more accurately using ANN. Because daily weather data has multiple parameters representing temperature, humidity, rainfall amount, cloud distance and size, wind speed and direction, etc. All these parameters are not linear, but they need to be processed together to determine temperature, rainfall, humidity or weather status for the next day.

Such type of applications needs the models which are complex in nature and can produce the required result by generating the patterns on its own by performing self-learning using the training data given to the model. To develop an ANN model for weather forecasting, selection of region for input data and parameters is necessary. The input data is to be taken from a specific area on which the model is trained and tested so that the model is able to generate accurate results. The number of input data given to model also helps to improve accuracy of the model by giving the results with a high degree of similarity between predicted and actual output data. The available data may be noisy thus, data should be cleaned. Similarly, it has to be normalized because, all the parameters are of different units and normalization will help the input and output parameters to correlate with each other [6]. The data should be divided in training and testing samples in proper proportion so that the results can be predicted, tested

CONCLUSION AND DISCUSSION

Different methods for weather forecasting are reviewed. ANN with backpropogation is recommended for weather forecasting. ANN with backpropagation uses an iterative process of training where, it repeatedly compares the observed output with targeted output and calculates the error. This error is used to readjust the values of weights and bias to get an even better output. Hence this method tries to minimize the error. Thus, Artificial Neural network with Backpropagation algorithm seems to be most appropriate method for forecasting weather accurately.

The weather Forecasting has a big challenge of predicting the accurate results which are used in many real time systems like electricity departments, airports, tourism centers, etc. The difficulty of this forecasting is the complex nature of parameters. Each parameter has a different set of ranges of values. This issue is addressed by ANN. It accepts all complex parameters as input and generates the intelligent patterns while training and it uses the same patterns to generate the forecasts. The Artificial Neural Network model proposed in this paper indicates all the parameters for input and output, training and testing data set, number of hidden layers and neurons in each hidden layer, weight, bias, learning rate and activation function. The Mean Squared Error between predicted output and the actual output is used to check accuracy.

PROBLEM IDENTIFICATION

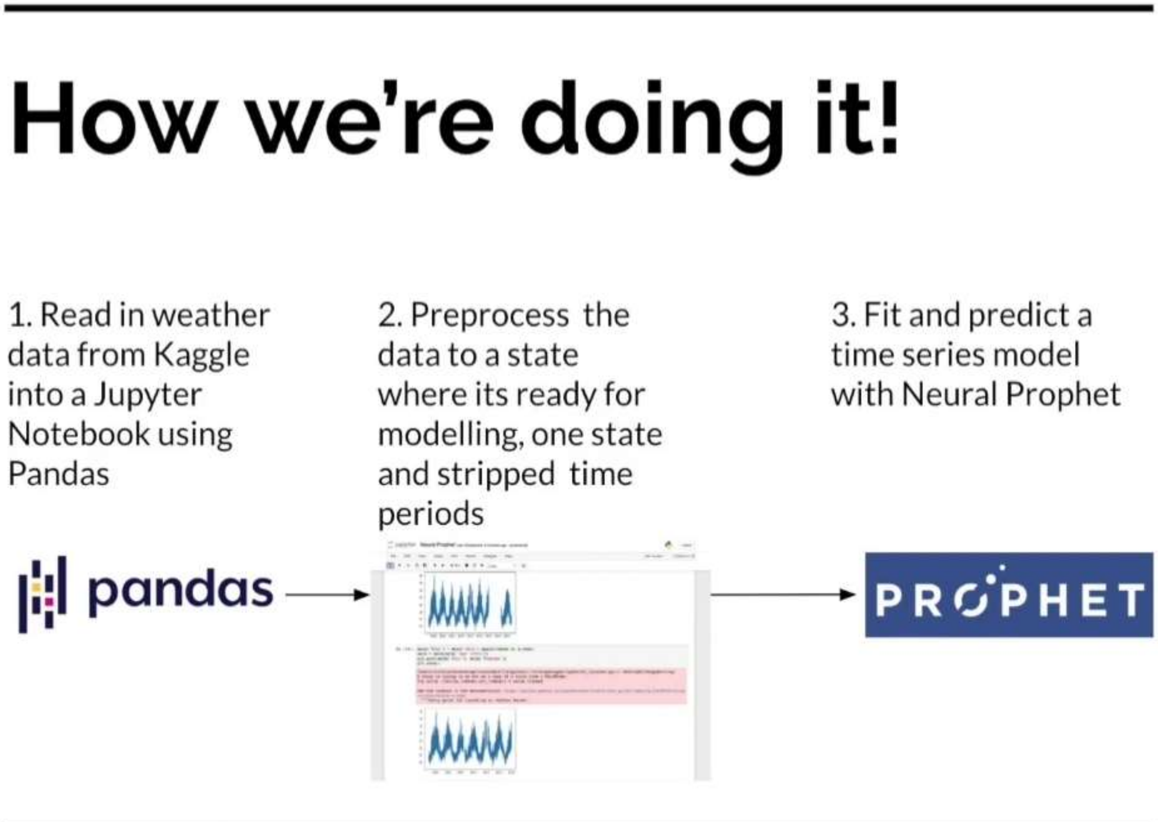
- Weather forecasting is a complex and often challenging skill that involves observing and processing vast amounts of data.
- Weather systems can range from small, short lived thunderstorms only a few miles in diameter that last a couple hours to large scale rain and snow storms up to a thousand miles in diameter and lasting for days.
- Forecasting ultimately is a three step process. These include:

  1) Observing

  2) Forecasting

  3) Communicating.

DEVELOPMENT



What's Covered

1. Preprocessing a weather dataset from Kaggle using Pandas

2. Training a time series forecasting model to predict temperature using Neural Prophet

3. Forecasting temperature into the future using the trained model

USED:

Jupyter notebook:

- The Jupyter Notebook is an open source web application that you can use to create and share documents that contain live code, equations, visualizations, and text.
- Jupyter Notebook is maintained by the people at Project Jupyter.
- Jupyter Notebooks are a spin-off project from the Ipython project, which used to have an Ipython Notebook project itself.
- The name, Jupyter, comes from the core supported programming languages that it supports: Julia, Python, and R. Jupyter ships with the Ipython kernel, which allows you to write your programs in Python, but there are currently over 100 other kernels that you can also use.

KAGGLE:

- Kaggle is an online community platform for data scientists and machine learning enthusiasts. Kaggle allows users to collaborate with other users, find and publish datasets, use GPU integrated notebooks, andcompete with other data scientists to solve data science challenges.
- The aim of this online platform (founded in 2010 by Anthony Goldbloom and Jeremy Howard and acquired by Google in 2017) is to help professionals and learners reach their goals in

their data science journey with the powerful tools and resources it provides.

- As of today (2021), there are over 8 million registered users on Kaggle.

Pandas:

- Pandas is a Python package providing fast, flexible, and expressive data structures designed to make working with "relational" or "labeled" data both easy and intuitive. It aims to be the fundamental high-level building block for doing practical, real-world data analysis in Python.
  Additionally, it has the broader goal of becoming the most powerful and flexible open source data analysis/manipulation tool available in any language. It is already well on its way toward this goal.

NeuralProphet:

- NeuralProphet, an evolution of the Prophet algorithm created by Facebook, is a time series prediction algorithm.
- Originally, the will of Facebook with Prophet is to provide an easily workable, tweakable, and explainable tool to predict time series. However, one problem persisted: poor performance. To clear up the issue, NeuralProphet was created.

REVIEW 3

EXECUTION AND EVALUATION

Open Jupyter notebook

Import dependencies

```
import pandas as pd
from neuralprophet import NeuralProphet
from matplotlib import pyplot as plt
import pickle
df = pd.read_csv('weatherAUS.csv')
df.head()
```

```
In [3]: df = pd.read_csv('weatherAUS.csv')
        df.head()
```

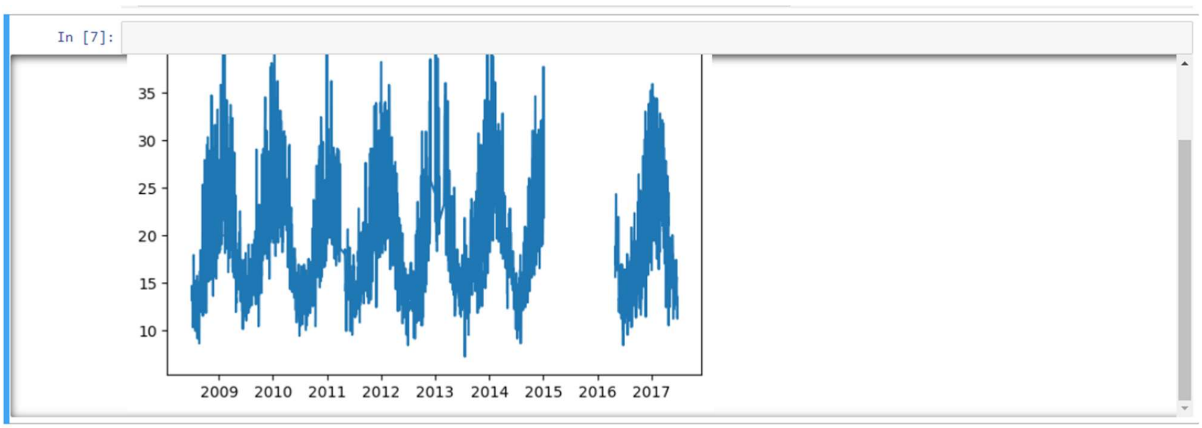
Out[3]:

	Date	Location	MinTemp	MaxTemp	Rainfall	Evaporation	Sunshine	WindGustDir	WindGustSpeed	WindDir9am	...	Humidity9am	Humidity3pm	Pressure9
0	2008-12-01	Albury	13.4	22.9	0.6	NaN	NaN	W	44.0	W	...	71.0	22.0	100
1	2008-12-02	Albury	7.4	25.1	0.0	NaN	NaN	WNW	44.0	NNW	...	44.0	25.0	101
2	2008-12-03	Albury	12.9	25.7	0.0	NaN	NaN	WSW	46.0	W	...	38.0	30.0	100
3	2008-12-04	Albury	9.2	28.0	0.0	NaN	NaN	NE	24.0	SE	...	45.0	16.0	101
4	2008-12-05	Albury	17.5	32.3	1.0	NaN	NaN	W	41.0	ENE	...	82.0	33.0	101

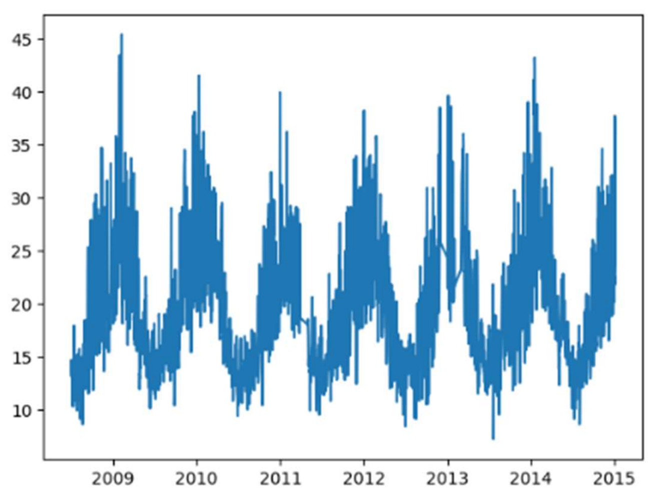
5 rows × 23 columns

Read in data and process dates

```
melb = df[df['Location']=='Melbourne']
melb['Date'] = pd.to_datetime(melb['Date'])
plt.plot(melb['Date'], melb['Temp3pm'])
plt.show()
```




```
melb['Year'] = melb['Date'].apply(lambda x: x.year)
melb = melb[melb['Year']<=2015]
plt.plot(melb['Date'], melb['Temp3pm'])
plt.show()
```



```
data = melb[['Date', 'Temp3pm']]
data.dropna(inplace=True)
data.columns = ['ds', 'y']
data.head()
```

```
rsus-a-copy
  data.dropna(inplace=True)
```

Out[11]:

	ds	y
67200	2008-07-01	14.6
67201	2008-07-02	13.7
67202	2008-07-03	13.9
67203	2008-07-04	13.1
67204	2008-07-05	14.6

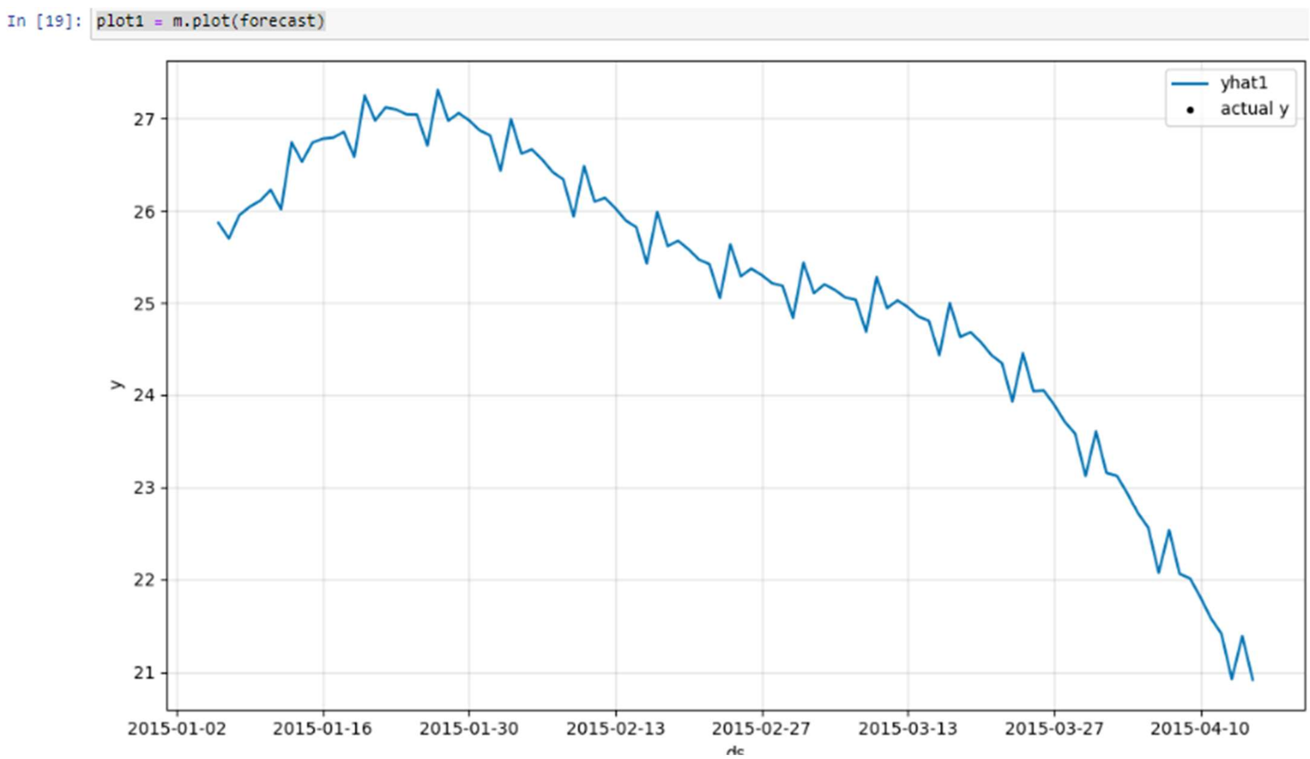
Train model

```
m = NeuralProphet(epochs=1000)

model = m.fit(data, freq='D')
```

```
In [16]: m = NeuralProphet(epochs=1000)

In [17]: model = m.fit(data, freq='D')

INFO - (NP.df_utils._infer_frequency) - Major frequency D corresponds to 99.694% of the data.
INFO - (NP.df_utils._infer_frequency) - Defined frequency is equal to major frequency - D
INFO - (NP.config.init_data_params) - Setting normalization to global as only one dataframe provided for training.
INFO - (NP.utils.set_auto_seasonalities) - Disabling daily seasonality. Run NeuralProphet with daily_seasonality=True to overri
de this.
INFO - (NP.config.set_auto_batch_epoch) - Auto-set batch_size to 32
  0%|          | 0/134 [00:00<?, ?it/s]
INFO - (NP.utils_torch.lr_range_test) - lr-range-test results: steep: 7.71E-02, min: 7.98E-01
  0%|          | 0/134 [00:00<?, ?it/s]
INFO - (NP.utils_torch.lr_range_test) - lr-range-test results: steep: 7.71E-02, min: 1.09E+00
INFO - (NP.forecaster._init_train_loader) - lr-range-test selected learning rate: 8.41E-02
Epoch[1000/1000]: 100%|█| 1000/1000 [06:51<00:00,  2.43it/s, SmoothL1Loss=0.0145, MAE=3.02, RMSE=3.92, Loss=0.0108, Reg
```

Forecast

```
future = m.make_future_dataframe(data, periods=100)

forecast = m.predict(future)

 plot1 = m.plot(forecast)
```

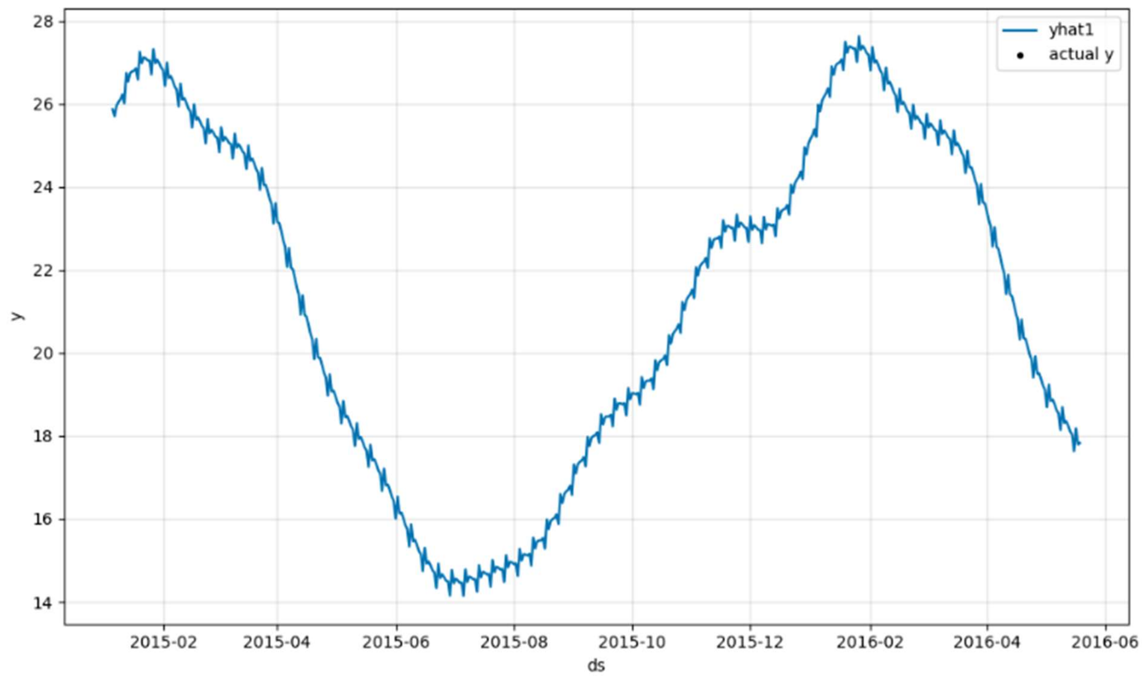


Save model

```
with open('saved_model.pkl', "wb") as f:
pickle.dump(m, f)
del m
with open('saved_model.pkl', "rb") as f:
 m = pickle.load(f)
future = m.make_future_dataframe(data, periods=500)
```

```
forecast = m.predict(future)
forecast.head()
plot1 = m.plot(forecast)
```

```
In [25]: plot1 = m.plot(forecast)
```



CONCLUSION

Weather forecasts still have their boundaries regardless of the modern era and progressed strategies to expect the weather. Weather forecasting is complicated and no longer usually correct, especially for days further within the future, because the weather can be chaotic and unpredictable. If weather patterns are noticeably solid, the patience technique of forecasting gives an extraordinarily beneficial approach to are expecting the climate for tomorrow. Weather statement techniques have stepped forward, and there were technological improvements in predicting the climate nowadays. Despite this fundamental scientific and technical development, many challenges continue to be concerning long-time period climate predictability. The accuracy of character climate forecasts varies substantially.