# Climate Time Series Analysis

**CIVE7100 35340 Time Ser/Geospatial Data Sci**

**SEC V30 Spring 2022**

**Made by -**
**Soumyo Dey**

# Introduction

Delhi's climate is intense in both directions. As a result, the primary goal of this project is to aid in the comprehension of underlying trends or systemic patterns across time. Weather forecasting is important because it assists in predicting future climatic changes. Using data visualizations, users can see seasonal trends and dig deeper into why these trends occur.

Climate time series analysis is particularly difficult because it is very hard to predict the future weather conditions accurately. A number of factors make it difficult to obtain accurate results from climate time series. Some of these factors are:

- Even with a perfect model and increased observations, there are numerous unpredictable atmospheric fluctuations.
- Uncertainties or errors in data can amplify as the models run realizations to try and predict the weather conditions.

# Steps

1. Data Obtainment
2. Data Cleaning
3. Plot different types of graphs,plots and heatmaps to visualize the data.
4. Choice of Models
5. Final result comparison

# Dataset

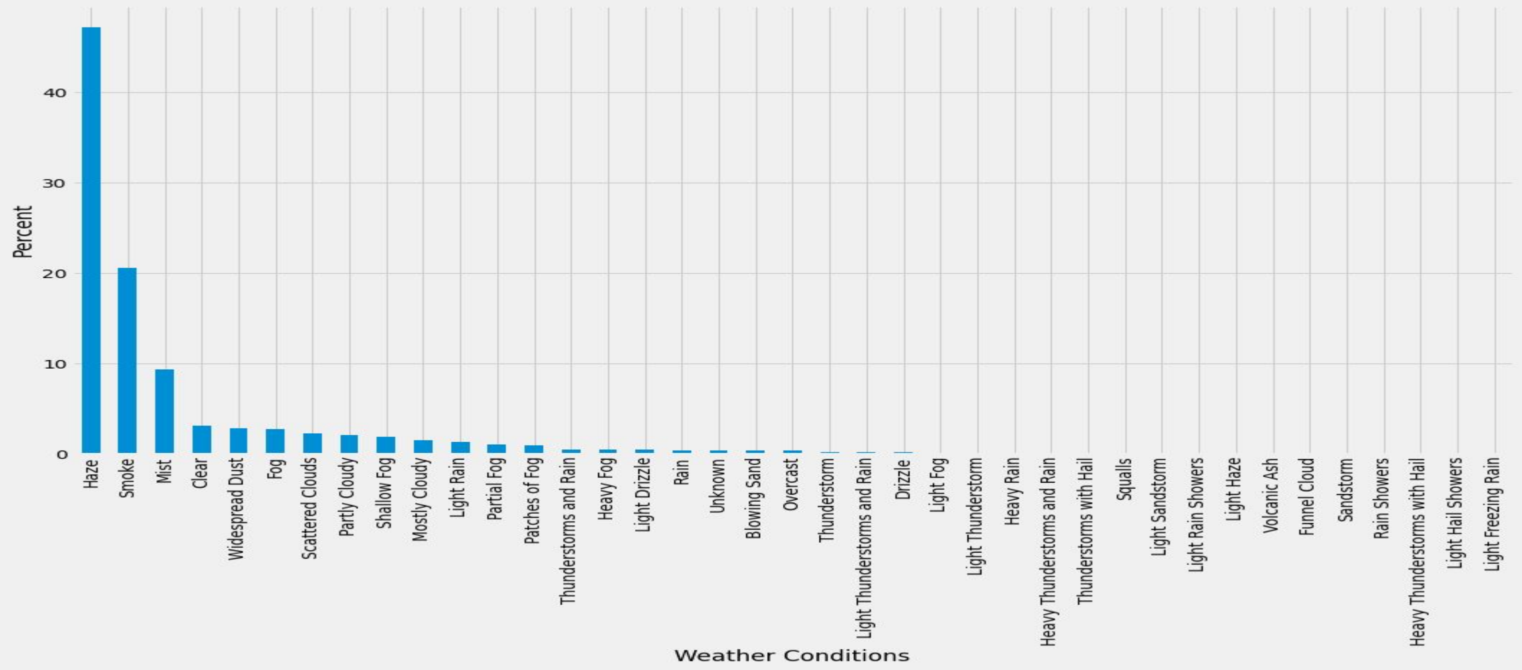
| datetime_utc | _conds | _dewptm | _fog | _hail | _heatindexm | _hum | _precipm | _pressurem | _rain | _snow | _tempm | _thunder | _tornado | _vism | _wdird | _wdire |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1996-11-01 11:00:00 | Smoke | 9.0 | 0 | 0 | NaN | 27.0 | NaN | 1010.0 | 0 | 0 | 30.0 | 0 | 0 | 5.0 | 280.0 | West |
| 1996-11-01 12:00:00 | Smoke | 10.0 | 0 | 0 | NaN | 32.0 | NaN | -9999.0 | 0 | 0 | 28.0 | 0 | 0 | NaN | 0.0 | North |
| 1996-11-01 13:00:00 | Smoke | 11.0 | 0 | 0 | NaN | 44.0 | NaN | -9999.0 | 0 | 0 | 24.0 | 0 | 0 | NaN | 0.0 | North |
| 1996-11-01 14:00:00 | Smoke | 10.0 | 0 | 0 | NaN | 41.0 | NaN | 1010.0 | 0 | 0 | 24.0 | 0 | 0 | 2.0 | 0.0 | North |

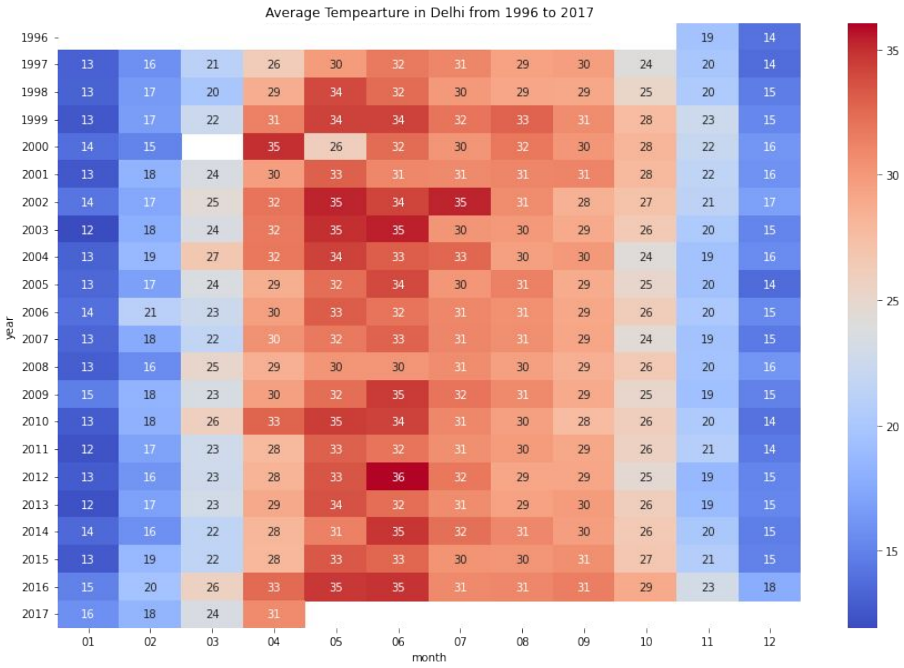
**This dataset contains hourly weather data collected in the city of Delhi from the period of 21 years (from 1996 to 2017). Link - [https://www.kaggle.com/datasets/mahirkukreja/delhi-weather-data](https://www.kaggle.com/datasets/mahirkukreja/delhi-weather-data) . There are 20 columns and 100991 rows.**

# Example of cleaning of data

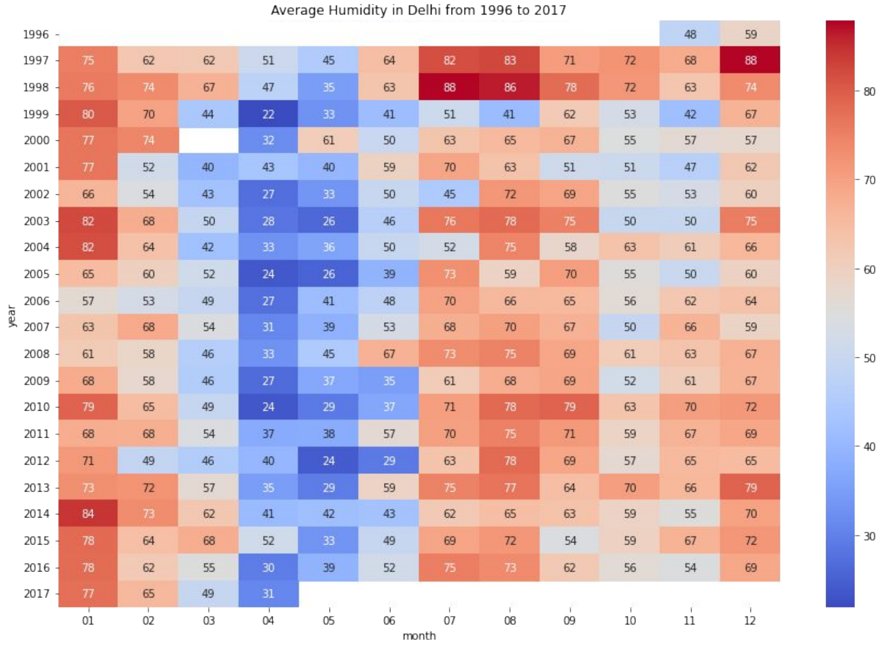
| | humidity | temprature |
|---|---|---|
| **count** | 100990.000000 | 100990.000000 |
| **mean** | 57.957422 | 25.438222 |
| **std** | 23.821218 | 8.487994 |
| **min** | 4.000000 | 1.000000 |
| **25%** | 39.000000 | 19.000000 |
| **50%** | 59.000000 | 27.000000 |
| **75%** | 78.000000 | 32.000000 |
| **max** | 243.000000 | 90.000000 |

# Visualization






Average Temperature




Average Humidity

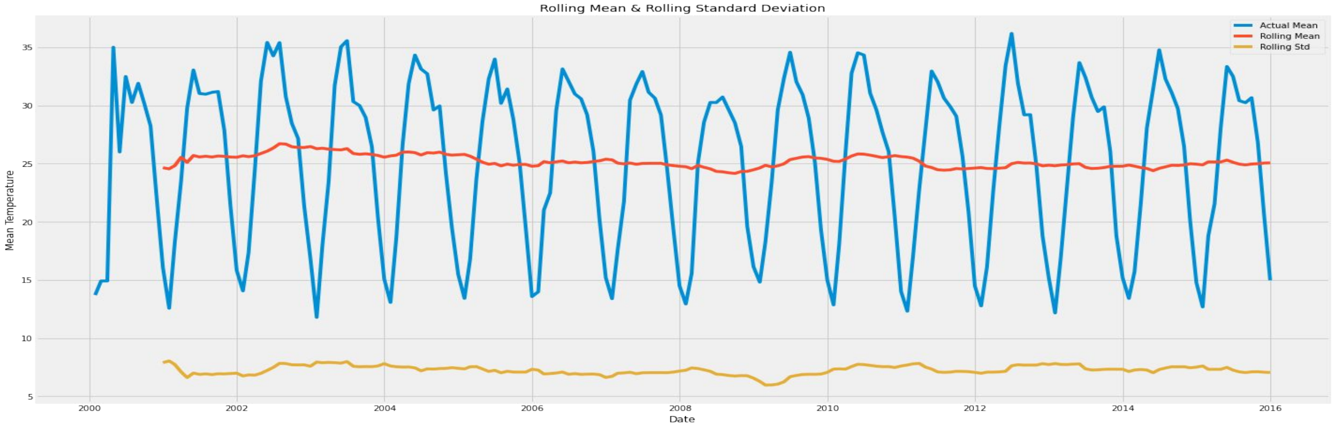
# Choice of Models

1. **ARIMA** - **Autoregressive integrated moving average (ARIMA) models predict future values based on past values. ARIMA makes use of lagged moving averages to smooth time series data.**

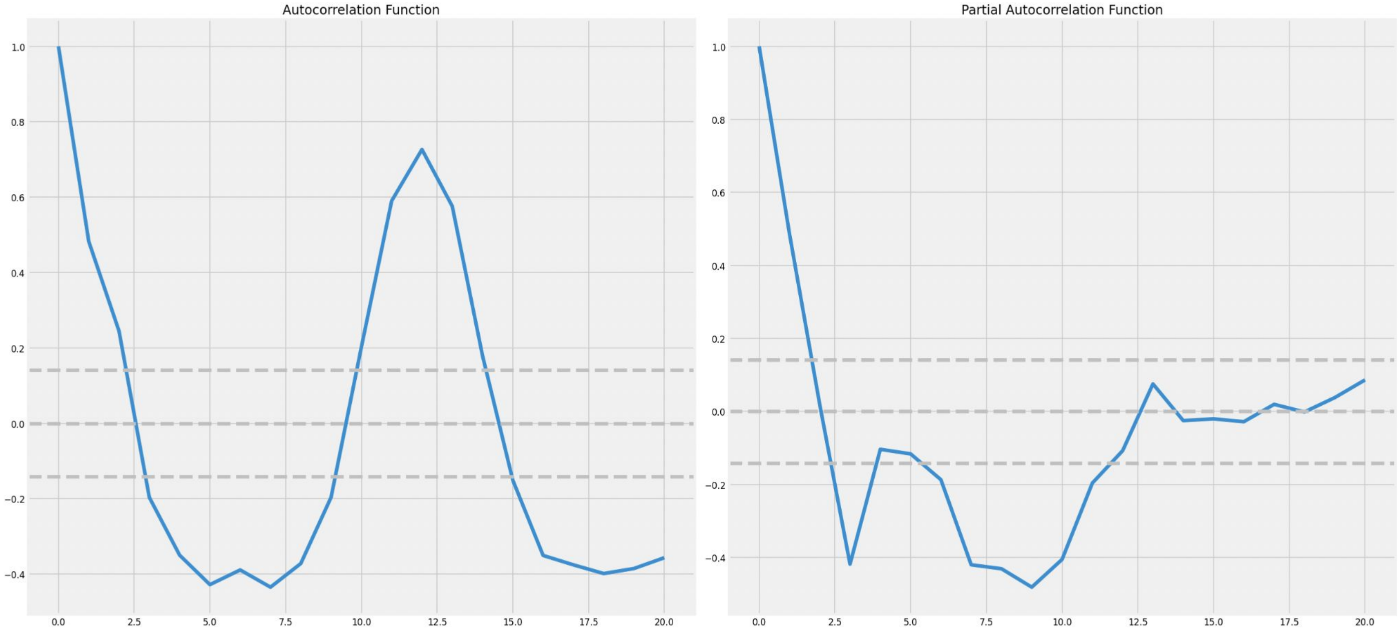
2. **LSTM** - **LSTM has feedback connections. Therefore, it can predict values for point data and can predict sequential data like weather.**

# ARIMA

## 1. Stationarity



## 2. Autocorrelation and Partial autocorrelation



# The Model Summary

```
                              ARMA Model Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  192
Model:                     ARMA(2, 2)   Log Likelihood                -454.355
Method:                       css-mle   S.D. of innovations              2.552
Date:                Mon, 25 Apr 2022   AIC                            920.709
Time:                        00:27:58   BIC                            940.254
Sample:                             0   HQIC                           928.625

==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const         25.1917      0.119    211.046      0.000      24.958      25.426
ar.L1.y        1.6785      0.024     69.835      0.000       1.631       1.726
ar.L2.y       -0.9519      0.023    -41.164      0.000      -0.997      -0.907
ma.L1.y       -0.9726      0.098     -9.919      0.000      -1.165      -0.780
ma.L2.y        0.1453      0.090      1.618      0.107      -0.031       0.321
                                    Roots
=============================================================================
                  Real          Imaginary           Modulus         Frequency
-----------------------------------------------------------------------------
AR.1            0.8816            -0.5227j            1.0250           -0.0852
AR.2            0.8816            +0.5227j            1.0250            0.0852
MA.1            1.2685            +0.0000j            1.2685            0.0000
MA.2            5.4264            +0.0000j            5.4264            0.0000
-----------------------------------------------------------------------------
```

# Final results

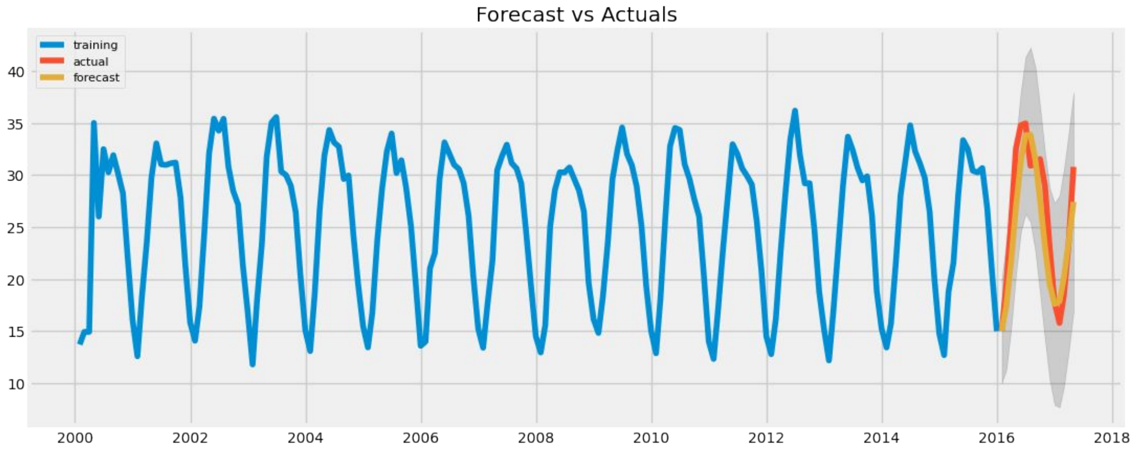
1. **ARIMA**

**Prediction of temperature -**

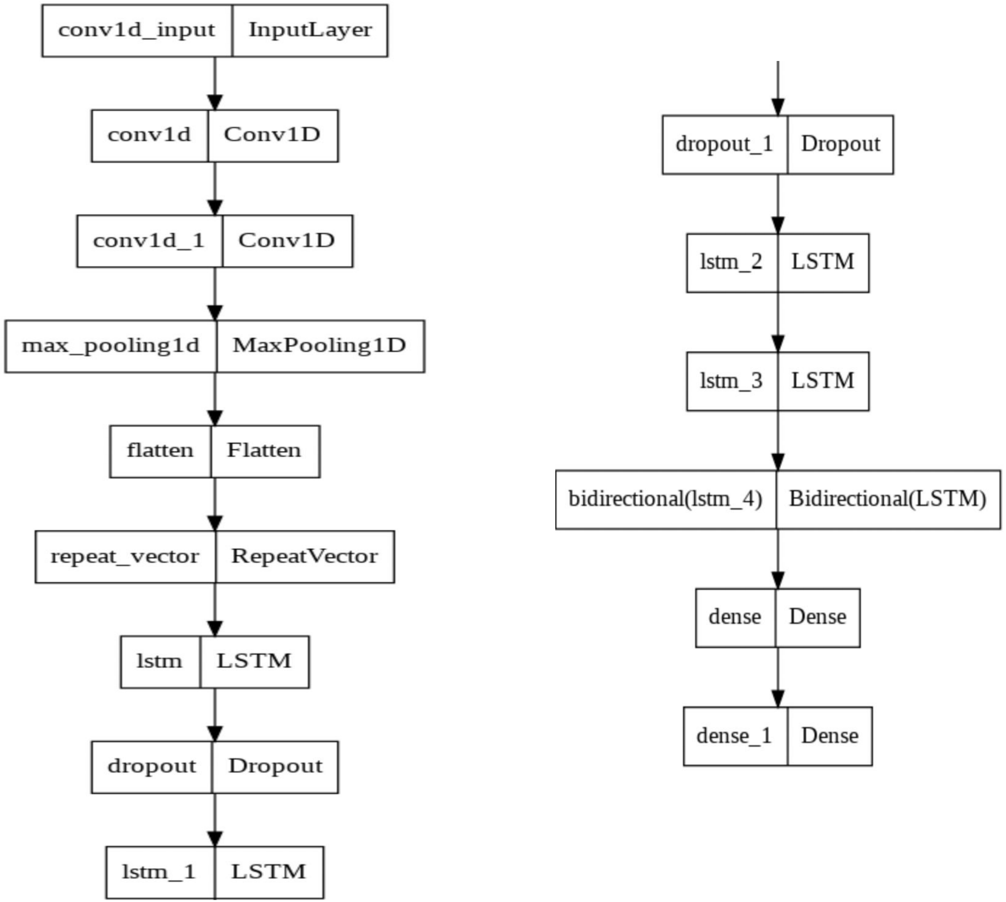
**Mean of temperature - 25.438222**

**RMSE - 3.1057635856013266**

**Error - 12.209043484%**



# LSTM



conv1d_input | InputLayer
conv1d | Conv1D
conv1d_1 | Conv1D
max_pooling1d | MaxPooling1D
flatten | Flatten
repeat_vector | RepeatVector
lstm | LSTM
dropout | Dropout
lstm_1 | LSTM
dropout_1 | Dropout
lstm_2 | LSTM
lstm_3 | LSTM
bidirectional(lstm_4) | Bidirectional(LSTM)
dense | Dense
dense_1 | Dense

# Final Results

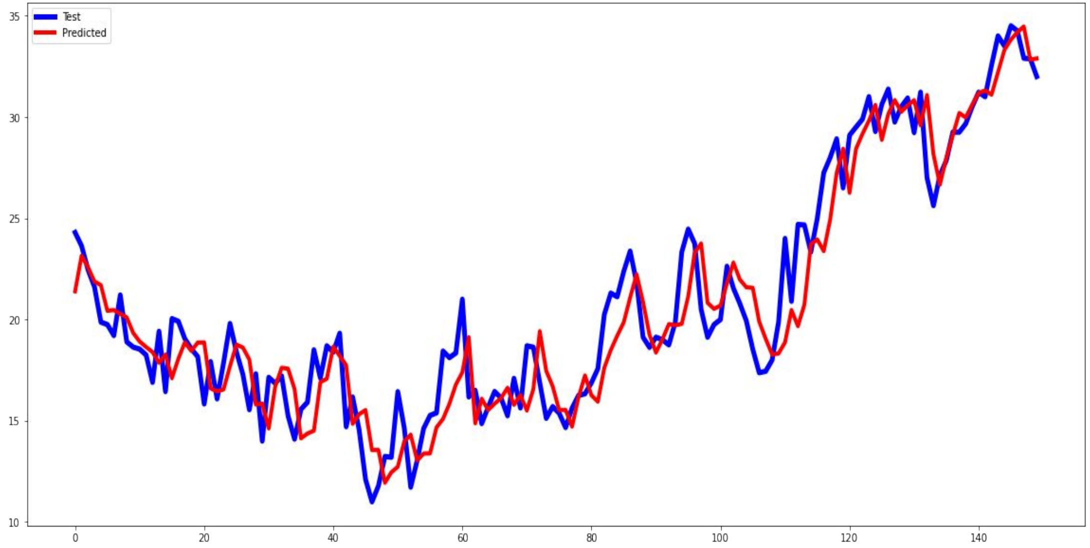
## 2. <u>LSTM</u>

**Prediction of temperature -**

**Mean of temperature - 25.438222**

**RMSE - 1.847351078574147**

**Error - 7.26210346%**



# Conclusion

**We can easily observe from the above results that LSTM outperforms ARIMA on the dataset.**

**This could be due to the model's limitation, which is that ARIMA can only assess the linear portion. However, the non-linear element of the data may not be white noise, which means that the ARIMA model may miss some information. LSTM is a type of RNN and deep learning application that is designed to learn temporal patterns, capture non-linear dependencies, and preserve useable memory for a longer period of time, resulting in superior results in scenarios when the dataset is large.**

# Citations


1. Comparison of ARIMA and LSTM for prediction of hemorrhagic fever at different time scales in China - Rui Zhang,Hejia Song,Qiulan Chen,Yu Wang,Songwang Wang ,Yonghong Li ( January 14 2022 )
2. Data Science for Weather Prediction – The Prerequisite to all Natural Disasters
3. How to Develop LSTM Models for Time Series Forecasting - Jason Brownlee ( November 14 2018)
4. How to Create an ARIMA Model for Time Series Forecasting in Python - Jason Brownlee (January 9 2017)

# Big Picture

Accurate weather prediction can help in various fields -

- To help people take proper precautions to secure themselves and their families in case of unwanted occurrences.
- Organizations can work better with the help of accurate weather predictions and it helps to deliver visual forecasts by various methods that most companies prefer.
- Weather forecasting highly benefits the agriculture sector for buying/selling livestock. It also assists the farmers to decide when to plant crops, pastures, and when to irrigate.
- It provides the business with valuable information that the business can use to make decisions about future business strategies.

Accurate weather forecasting can provide information to people and organizations that can utilize it to reduce weather-related losses and improve societal advantages such as life and property protection, public health and safety, and economic prosperity and quality of life.


Thank You