



2CSOE78: Scientific Programming

Special Assignment Report

on

Weather Forecasting

Submitted By:

Dhyan Thakkar (20BEC027)
Karan Singh Choudhary (20BEC051)

[B. Tech. (ECE), Semester - V]

Submitted to:

Prof. Usha Patel

Weather Forecasting

(Condition Classification and TimeSeries Forecasting)

Aim:

To Create Machine Learning Models to Classify Current Day Conditions, and Predict Next 15 Days forecast With LSTM Network.

About Dataset:

Dataset Source: Kaggle <https://www.kaggle.com/datasets/akashram/indian-summer-over-the-years>

Indian Summers (April to June) are extremely Hot and Humid. Surprisingly, (or not so), 2022 is hotter, not just in the usual cities across the country, but also in places that normally are known for warm temperatures.

File 1 - The dataset has daily temperatures, dew, humidity, wind speed and more for the top 15 populous cities of India for the period between 2012 to 2021.

File 2 - Metadata about the data present in file 1

File 3 - The dataset has daily temperatures, dew, humidity, wind speed and more for the top 15 populous cities of India for the period between 2007 to 2011

All 3 are CSV files.

Empty values (or null values) are used within the data set to indicate an absence of data, such as missing weather information or unknown data. They are not used to indicate a zero value. For example, an unknown humidity value will be marked as empty or null. A zero amount of humidity will be indicated by the value zero.

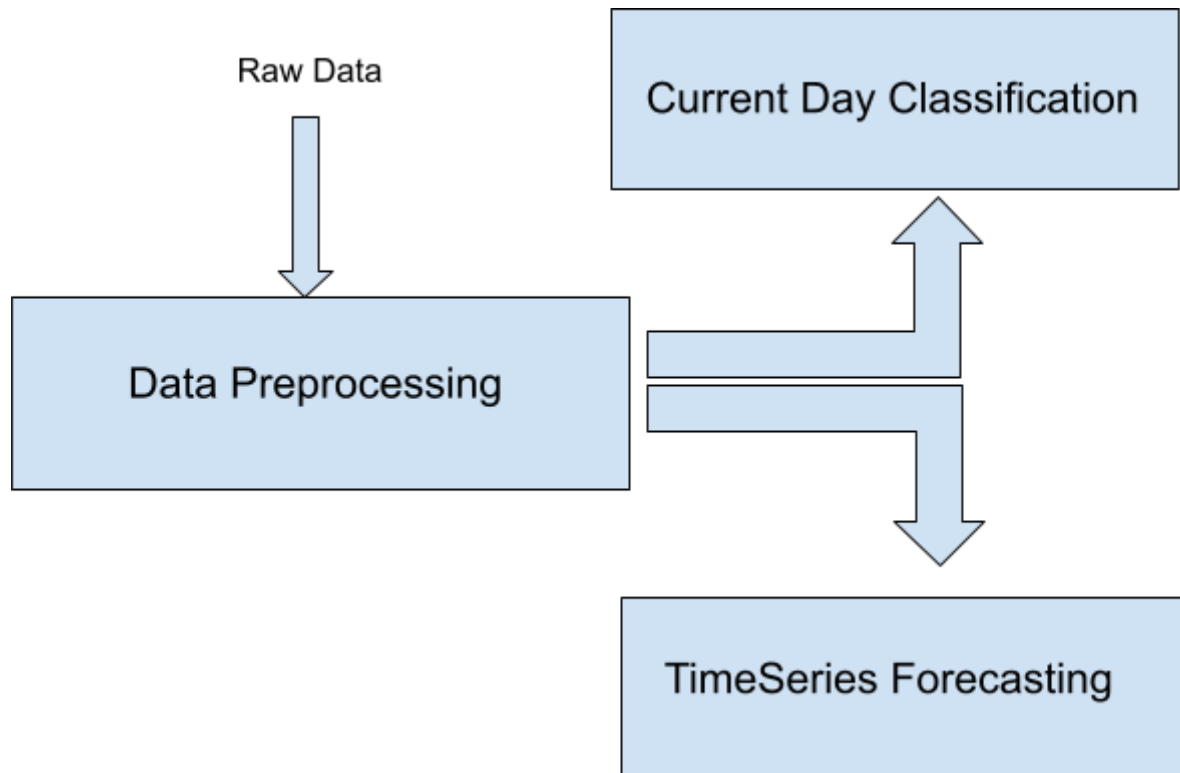
Tasks

1. Classification of conditions for Daily Temperature.
2. Timeseries Forecasting to predict daily temperature..

Libraries Used:

Numpy
Pandas
Matplotlib
Seaborn
MissingNo
Keras

Data Pipeline (Flow):



Data Preprocessing:

- Data from 2011 to 2021 was added.
- NaN celled rows are dropped.
- Object Data Type ["Date", "city", "sunrise", "sunset"] are processed
 - City and Conditions are Ordinally Encoded
 - Date, Sunrise time, and sunset time are parsed as int and presented in different columns.
- Transient Features ['temp', 'dew', 'humidity', 'windspeed', 'sealevelpressure', 'moonphase', 'Year', 'Month', 'Date_split', 'winddir'] are separated for the LSTM Model.

Classification data Processing

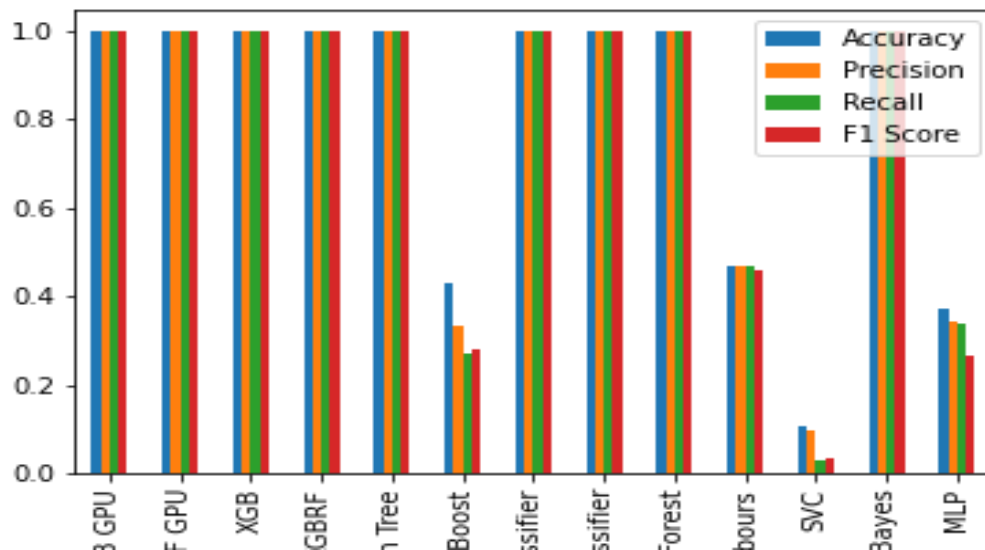
- Various Classification Algorithms like ['XGB GPU', 'XGBRF GPU', 'XGB', 'XGBRF', 'Desicion Tree', 'AdaBoost', 'Bagging Classifier', 'Extra Trees Classifier', 'Random Forest', 'KNeighbours', 'SVC', 'Naive Bayes', 'MLP'] are used.

Transient Data Processing

- Discrete Non Uniform Data is converted to Uniform Data.
- Split_sequences is used to prepare the data in
train_data =(Sample,LookBack,In_features) and
test_data = (Sample,Forecast_Window, Out_features)
- Data is Scaled using MinMaxScalar to accelerate Computing and have Faster Training.
- Trends Of Winter and Summer are Learnt by the model.

Results:

Categorical Classification



	Accuracy	Precision	Recall	F1 Score
XGB GPU	1.0	1.0	1.0	1.0
XGBRF GPU	0.999361	0.99914	0.999579	0.999356
XGB	0.999681	0.99957	0.999789	0.999679
XGBRF	0.999361	0.99914	0.999579	0.999356
Desicion Tree	1.0	1.0	1.0	1.0
AdaBoost	0.43213	0.333333	0.269732	0.280198
Bagging Classifier	1.0	1.0	1.0	1.0
Extra Trees Classifier	1.0	1.0	1.0	1.0
Random Forest	1.0	1.0	1.0	1.0

KNeighbours	0.468221	0.468959	0.468436	0.458542
SVC	0.105078	0.098617	0.028555	0.035896
Naive Bayes	0.999361	0.99938	0.999034	0.999202
MLP	0.370169	0.342348	0.336855	0.26756

Time Series Forecasting:

