RAINFALL PREDICATION USING MACHINE LEARNING

K.SUPARNA[1], PEDAPATI.SUBRAHMANYAM [2]

[1] **Assistant Professor MSc (CS), DEPT,** Dantuluri Narayana Raju College, **Bhimavaram, Andhra Pradesh**
Email id:- suparnakalidindi@gmail.com
[2]**PG Student of MSc Computer Science,** Dantuluri Narayana Raju College, **Bhimavaram, Andhra Pradesh**
Email id :- pedapatisubramanyam707@gmail.com

ABSTRACT

Rainfall prediction is one of the challenging and uncertain tasks which has a significant impact on human society. Timely and accurate predictions can help to proactively reduce human and financial loss. This study presents a set of experiments which involve the use of prevalent machine learning techniques to build models to predict whether it is going to rain tomorrow or not based on weather data for that particular day in major cities of Australia. This comparative study is conducted concentrating on three aspects: modeling inputs, modeling methods, and pre-processing techniques. The results provide a comparison of various evaluation metrics of these machine learning techniques and their reliability to predict the rainfall by analyzing the weather data.

1 INTRODUCTION

India's welfare is agriculture. The achievement of agriculture is dependent on rainfall. It also helps with water resources. Rainfall information in the past helped farmers better manage their crops, leading to economic growth in the country. Prediction of precipitation is beneficial to prevent flooding that saves people's lives and property. Fluctuation in the timing of precipitation and its amount makes forecasting of rainfall a problem for meteorological scientists. Forecasting is one of the utmost challenges for researchers from a variety of fields, such as weather data mining, environmental machine learning, functional hydrology, and numerical forecasting, to create a predictive model for accurate rainfall. In these problems, a common question is how to infer the past predictions and make use of future predictions. A variety of sub-processes are typically composed of the substantial process in rainfall. It is at times not promising to predict the precipitation correctly by on its global system. Climate forecasting stands out for all countries around the globe in all the benefits and services provided by the meteorological department. The job is very complicated because it needs specific numbers, and all signals are intimated without any assurance. Accurate precipitation forecasting has been an important issue in hydrological science as early notice of stern weather can help avoid natural disaster injuries and damage if prompt and accurate forecasts are made. The theory of the modular model and the integrati2on of different models has recently gained more interest in rainfall forecasting to address this challenge. A huge range of rainfall prediction methodologies is available in India. In India, there are two primary methods of forecasting rainfall. Regression, Artificial Neural Network (ANN), Decision Tree algorithm, Fuzzy logic and team process of data handling are the majority frequently used computational methods used for weather forecasting the basic goal is to follow information rules and relationships while gaining intangible and potentially expensive knowledge. Artificial NN is a promising part of this wide field

Rainfall prediction remains a serious concern and has attracted the attention of governments, industries, risk management entities, as well as the scientific community. Rainfall is a climatic factor that affects many human activities like agricultural production, construction, power generation, forestry and tourism, among others [1]. To this extent, rainfall prediction is essential since this variable is the one with the highest correlation with adverse natural events such as landslides, flooding, mass movements and avalanches. These incidents have affected society for years [2]. Therefore, having an appropriate approach for rainfall prediction makes it possible to take preventive and mitigation measures for these natural phenomena. To solve this uncertainty, we used various machine learning techniques and models to make accurate and timely predictions. These papers aim to provide end to end machine learning life cycle right from Data preprocessing to implementing models to evaluating them. Data Preprocessing steps include imputing missing values, feature transformation, encoding categorical features, feature scaling and feature selection. We implemented models such as Logistic Regression, Decision Tree, K Nearest Neighbour, Rule-based and Ensembles. For evaluation purpose.

2. LITERATURE SURVEY AND RELATED WORK

Climate Change and Human Health - Risks and Responses

The long-term good health of populations depends on the continued stability and functioning of the biosphere's ecological and physical systems, often referred to as life-support systems. We ignore this long-established historical truth at our peril: yet it is all too easy to overlook this dependency, particularly at a time when the human species is becoming increasingly urbanized and distanced from these natural systems. The world's climate system is an integral part of this complex of life-supporting processes, one of many large natural systems that are now coming under pressure from the increasing weight of human numbers and economic activities.

By inadvertently increasing the concentration of energy-trapping gases in the lower atmosphere, human actions have begun to amplify Earth's natural greenhouse effect. The primary challenge facing the world community is to achieve sufficient reduction in greenhouse gas emissions to avoid dangerous interference in the climate system. National governments, via the UN Framework Convention on Climate Change (UNFCC), are committed in principle to seeking this outcome. In practice, it is proving difficult to find a politically acceptable course of action—often because of apprehensions about possible short-term economic consequences.

This volume seeks to describe the context and process of global climate change, its actual or likely impacts on health, and how human societies should respond, via both adaptation strategies to lessen impacts and collective action to reduce greenhouse gas emissions. As shown later, much of the resultant risk to human populations and the ecosystems upon which they depend comes from the projected extremely rapid rate of change in climatic conditions. Indeed, the prospect of such change has stimulated a great deal of new scientific research over the past decade, much of which is elucidating the complex ecological disturbances that can impact on human well-being and health—as in the following example.

The US Global Change Research Program (Alaska Regional Assessment Group) recently documented how the various effects of climate change on aquatic ecosystems can interact and ripple through trophic levels in unpredictable ways. For example, warming in the Arctic region has reduced the amount of sea ice, impairing survival rates for walrus and seal pups that spend part of their life cycle on the ice. With fewer seal pups, sea otters have become the alternative food source for whales. Sea otters feed on sea urchins, and with fewer sea otters' sea urchin populations are expanding and consuming more of the kelp that provides breeding grounds for fish. Fewer fish exacerbate the declines in walrus and seal populations. Overall, there is less food available for the Yupik Eskimos of the Arctic who rely on all of these species.

Global climate change is thus a significant addition to the spectrum of environmental health hazards faced by humankind. The global scale makes for unfamiliarity—although most of its health impacts comprise increases (or decreases) in familiar effects of climatic variation on human biology and health. Traditional environmental health concerns long have been focused on toxicological or microbiological risks to health from local environmental exposures. However, in the early years of the twenty-first century, as the burgeoning human impact on the environment continues to alter the planet's geological, biological and ecological systems, a range of larger-scale environmental hazards to human health has emerged. In addition to global climate change, these include: the health risks posed by stratospheric ozone depletion; loss of biodiversity; stresses on terrestrial and ocean food-producing systems; changes in hydrological systems and the supplies of freshwater; and the global dissemination of persistent organic pollutants.

Climate change and stratospheric ozone depletion are the best known of these various global environmental changes. Human societies, however, have had a long experience of the vicissitudes of climate: climatic cycles have left great imprints and scars on the history of humankind. Civilizations such as those of ancient Egypt, Mesopotamia, the Mayans, the Vikings in Greenland and European populations during the four centuries of Little Ice Age, all have both benefited and suffered from nature's great climatic cycles. Historical analyses also reveal widespread disasters, social disruption and disease outbreaks in response to the more acute, inter-annual, quasi-periodic ENSO (El Niño Southern Oscillation) cycle (1). The depletion of soil fertility and freshwater supplies, and the mismanagement of water catchment basins via excessive deforestation, also have contributed to the decline of various regional populations over the millennia.

Geomorphology, natural hazards, vulnerability and prevention of natural disasters in developing countries

The significance of the prevention of natural disasters is made evident by the commemoration of the International Decade for Natural Disaster Reduction (IDNDR). This paper focuses on the role of geomorphology in the prevention of natural disasters in developing countries, where their impact has devastating consequences. Concepts such as natural hazards, natural disasters and vulnerability have a broad range of definitions; however, the most significant elements are associated with the vulnerability concept. The latter is further explored and considered as a key factor in understanding the occurrence of natural disasters, and consequently, in developing and applying adequate strategies for prevention. Terms such as natural and human vulnerabilities are introduced and explained as target aspects to be considered in the reduction of vulnerability and for prevention and mitigation of natural disasters. The importance of the incorporation not only of geomorphological research, but also of geomorphologists in risk assessment and management programs in the poorest countries is emphasized.

Atmospheric and climatic hazards: Improved monitoring and prediction for disaster mitigation

The last few years have seen enormous damage and loss of life from climate and weather phenomena. The most damaging events have included the severe 1997/98 El Niño (with its near-global impacts), Hurricane Mitch, and floods in China in mid-1998. What have we learnt regarding the causes, variability, and predictability, of these phenomena? Can we predict the occurrence of these extreme events, and thereby mitigate their damage? This paper reviews what we have learnt in the last decade or so regarding the predictability of these climate and weather extremes. The view starts with the largest (El Niño) scales and works towards the scale of individual thunderstorms. It focuses on the practical outcomes of our improved knowledge regarding decreasing the impact of natural disasters, rather than describing in detail the scientific knowledge underlying these outcomes. The paper concludes with a discussion of some of the factors that still restrict our ability to mitigate the deleterious effects of atmospheric and climatic hazards.

Exploratory Data Analysis: the best way to Start a Data Science Project

Exploratory Data Analysis is a set of techniques that were developed by Tukey, John Wilder in 1970. The philosophy behind this approach was to examine the data before building a model. John Tukey encouraged statisticians to explore the data, and possibly formulate hypotheses that could lead to new data collection and experiments. Today Data scientists and analysts spend most of their time in Data Wrangling and Exploratory Data Analysis also known as EDA. But what is this EDA and why it is so important? This article explains what EDA is and how to apply EDA techniques to a dataset.

3 EXISTING SYSTEM

Rainfall prediction is important as heavy rainfall can lead to many disasters. Predictions help people to take preventive measures and moreover the prediction should be accurate. There are two types of prediction short term rainfall prediction and long-term rainfall. Prediction, mostly short-term prediction can give us accurate results. The main challenge is to build a model for long term rainfall prediction. Heavy precipitation prediction could be a major drawback for the earth science department because it is closely associated with the economy and lifetime of humans.

Disadvantages:

We can just do it by having the historical data analysis of rainfall and can predict the rainfall for future seasons. We can apply many techniques like classification, regression according to the requirements and, we can calculate the error between the actual and prediction and the accuracy. Different techniques produce different accuracies, so it is important to choose the right algorithm and model it according to the requirements

4 PROPOSED WORK AND ALGORITHM

It's a cause for natural disasters like flood and drought that square measure encountered by individuals across the world each year. Accuracy of rainfall statement has nice importance for countries like India whose economy is basically dependent on agriculture. The dynamic nature of atmosphere applied mathematics techniques fail to provide sensible accuracy for precipitation statement. The prediction of precipitation using machine learning techniques may use regression. The intention of this project is to offer non-experts' easy access to the techniques and approaches utilized in the sector of precipitation prediction and provide a comparative study among the various machine learning techniques..

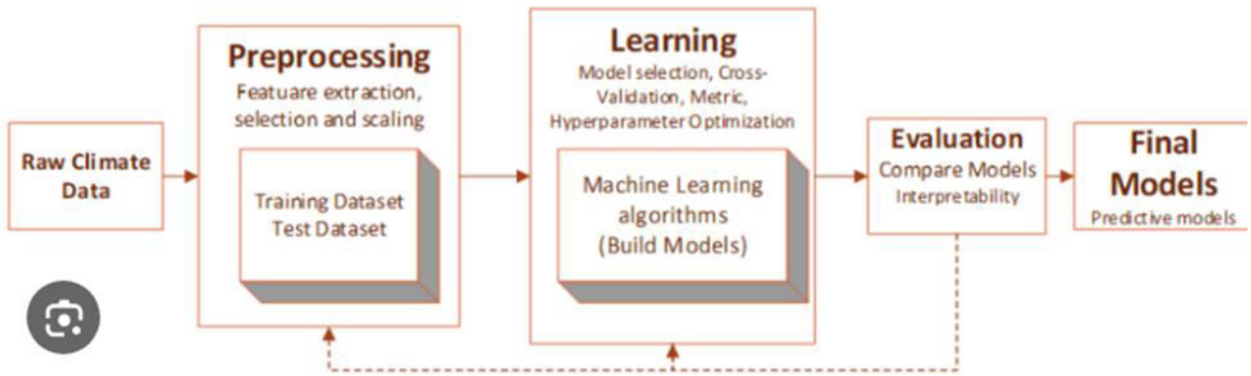

FIG 1: SYSTEM ARCHITECTURE

5 METHODOLOGIES

MODULES

1. Add Product Details

To build project I used some sample products image to train product identification models

2. Train Model

In this Module screen train model generated with 100% accuracy and now show product to web cam.

3. Add/Remove Product from basket

To allow application to identify product image and then show in text area and if we again show same product then application will remove from text area

6 RESULTS AND DISCUSSION SCREENSHOTS

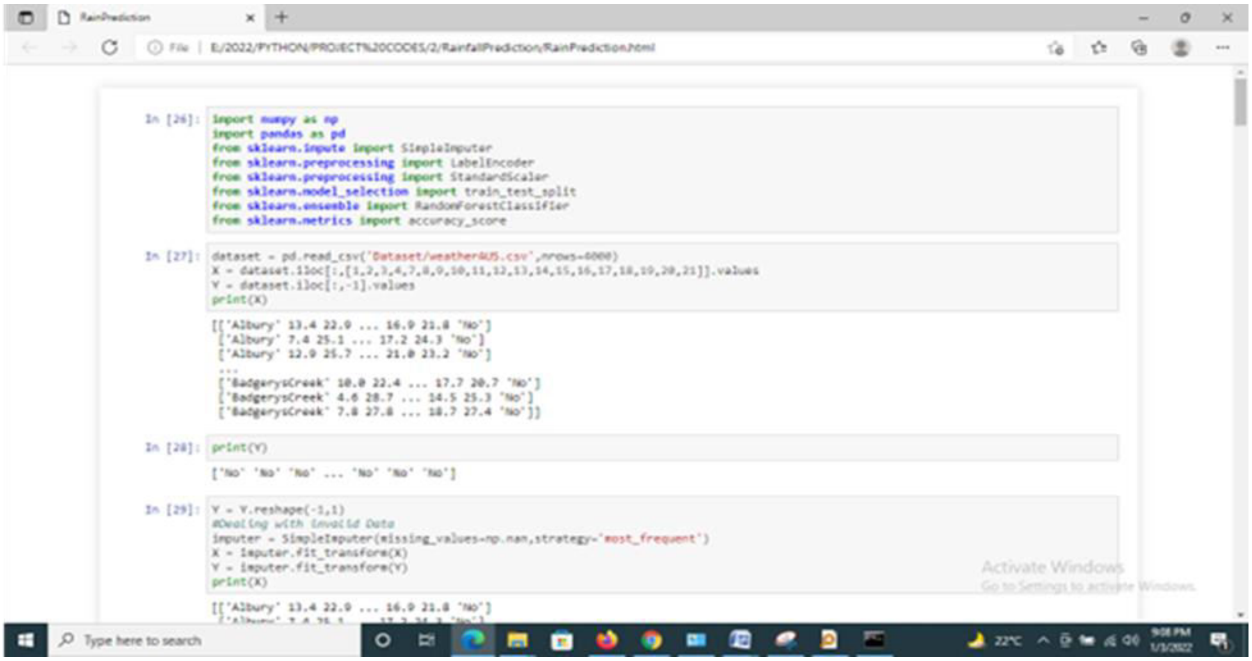

Fig 1: Package 1

Fig 2:- Analysis

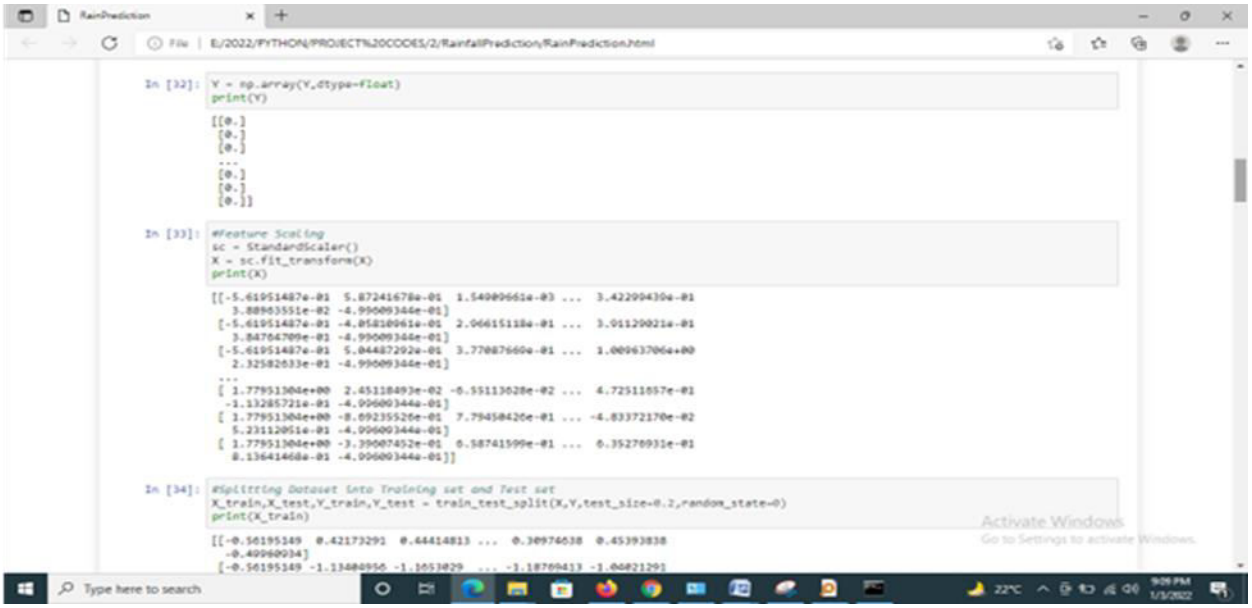

Fig 3:- Training

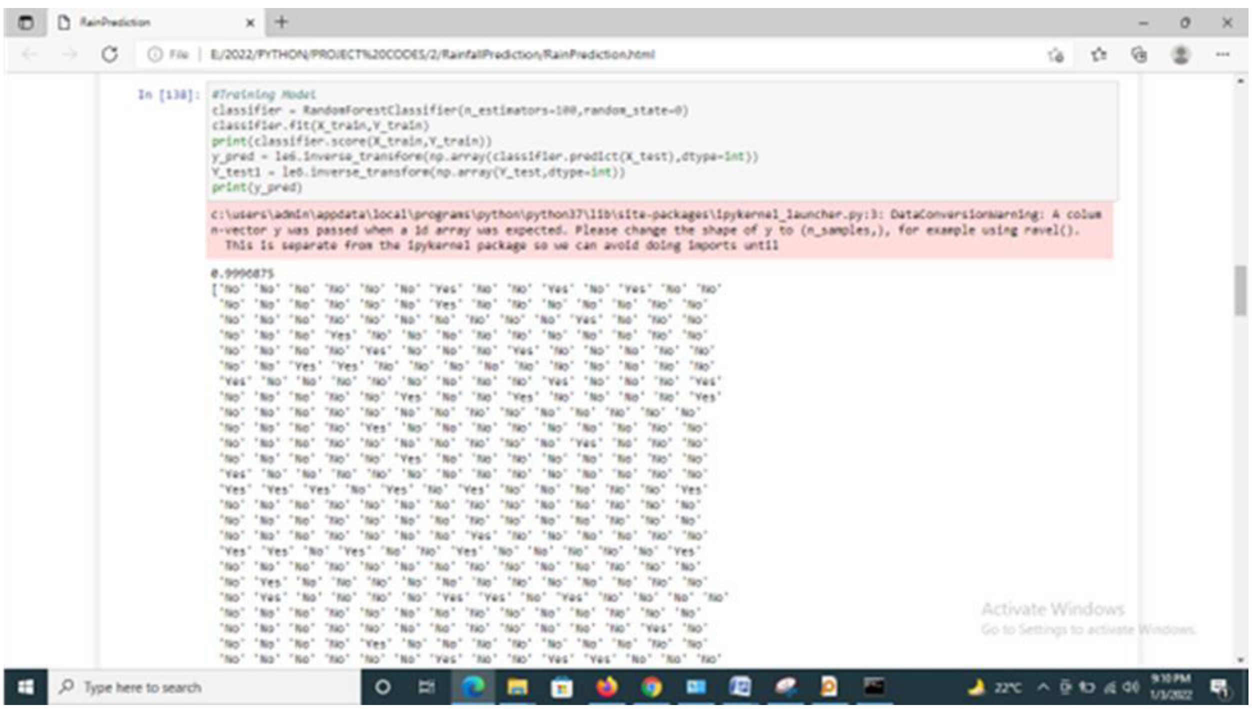

Fig 4:- Algorithm

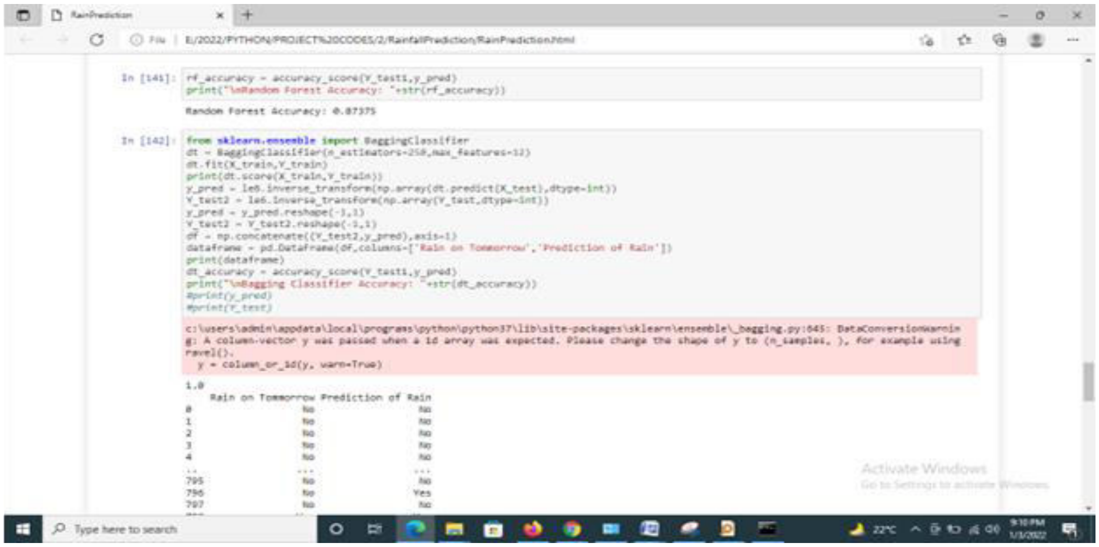

Fig 5: Random Forest

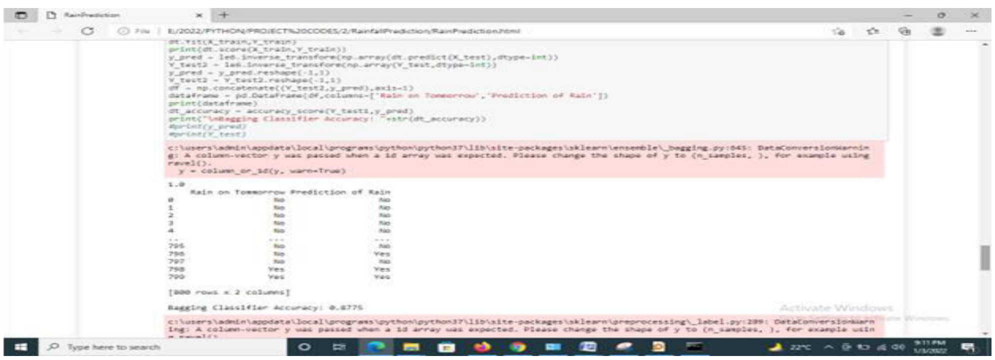

Fig 6: Bagging

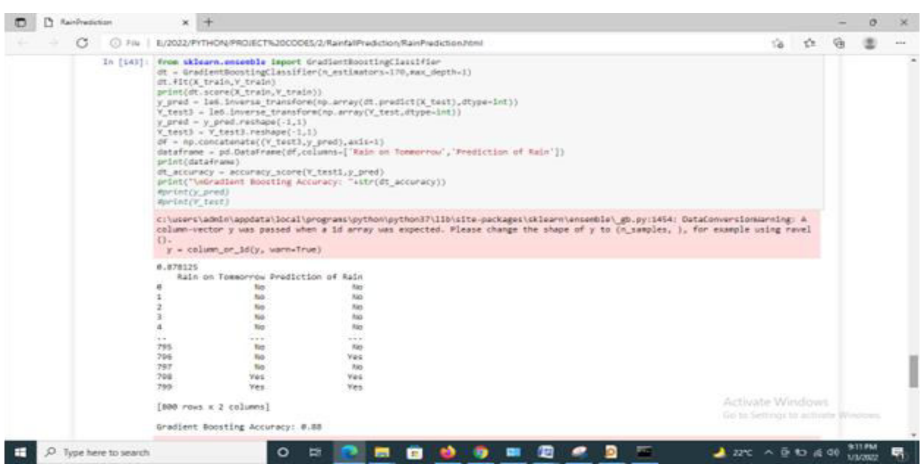

Fig 7: Gradient Boosting

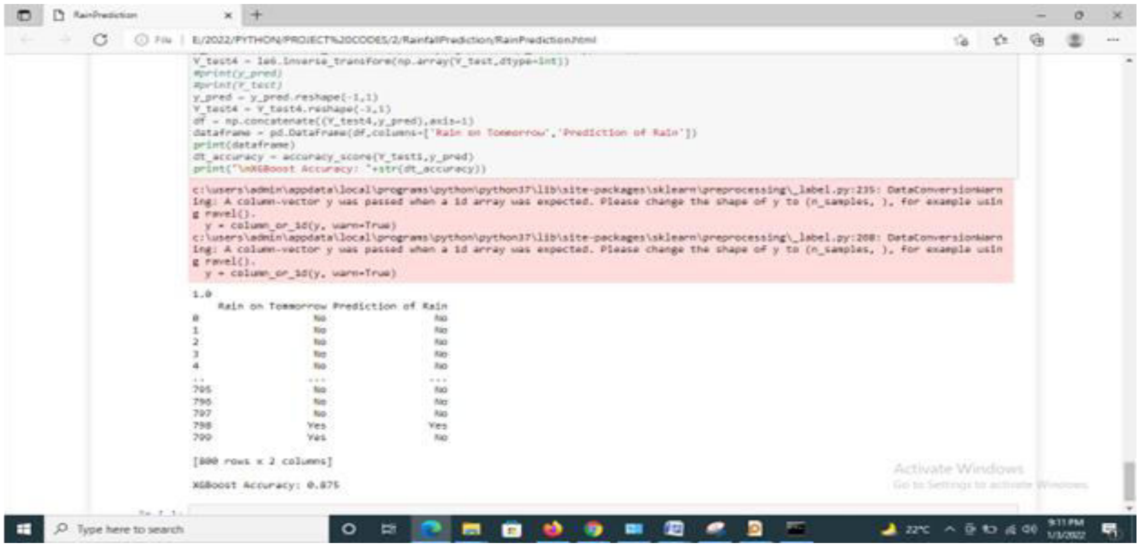

Fig 8 : Xg Boost

6.CONCLUSION AND FUTURE SCOPE

In this paper, we explored and applied several preprocessing steps and learned there impact on the overall performance of our classifiers. We also carried a comparative study of all the classifiers with different input data and observed how the input data can affect the model predictions.

We can conclude that Australian weather is uncertain and there is no such correlation among rainfall and the respective region and time. We figured certain patterns and relationships among data which helped in determining important features. Refer to the appendix section. As we have a huge amount of data, we can apply Deep Learning models such as Multilayer Perceptron, Convolutional Neural Network, and others.

It would be great to perform a comparative study between the Machine learning classifiers and Deep learning models.

7 REFERENCES

1. World Health Organization: Climate Change and Human Health: Risks and Responses. World Health Organization, January 2003

2. Alcntara-Ayala, I.: Geomorphology, natural hazards, vulnerability and prevention of natural disasters in developing countries. Geomorphology 47(24), 107124 (2002)

3. Nicholls, N.: Atmospheric and climatic hazards: Improved monitoring and prediction for disaster mitigation. Natural Hazards 23(23), 137155 (2001)

4. [Online] InDataLabs, Exploratory Data Analysis: the Best way to Start a Data Science Project. Available: https://medium.com/@InDataLabs/ why-start-a-data-science-project-with-exploratory-data-analysis-f90c0efcbe49

5.[Online] Pandas Documentation. Available: https://pandas.pydata.org/ pandas-docs/stable/reference/api/pandas.get_dummies.html

6.[Online] Sckit-Learn Documentation Available: https://scikit-learn.org/ stable/modules/generated/sklearn.feature_extraction.FeatureHasher. html

7.[Online] Sckit-Learn Documentation Available: https://scikit-learn.org/ stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html

8.[Online] Sckit Learn Documentation Available: https://scikit-learn.org/ stable/modules/generated/sklearn.feature_selection.SelectKBest.html

9. [Online] Raheel Shaikh, Feature Selection Techniques in Machine Learning with Python Available: https://towardsdatascience.com/ feature-selection-techniques-in-machine-learning-with-python-f24e7da3f36e

10.[Online] Imbalanced-learn Documentation Available: https://imbalanced-learn. readthedocs.io/en/stable/introduction.html

11. V. Veeralakshmi and D. Ramyachitra, Ripple Down Rule learner (RIDOR) Classifier for IRIS Dataset. Issues, vol 1, p. 79-85.

12. [Online] Aditya Mishra, Metrics to Evaluate your Machine Learning Algorithm Available: https://towardsdatascience.com/ metrics-to-evaluate-your-machine-learning-algorithm