**A PROJECT REPORT ON**

**RAIN PREDICTION USING RANDOM FOREST ALGORITHM**

**Submitted in partial fulfillment for the requirement of the award of Training in**

**Data Analytics, Machine Learning and AI using Python**

*Submitted By:*

Name: Pranav Jadhav

Year Of Study: B.Tech 1st year

College: Rajiv Gandhi Institute of Petroleum Technology

**Mentor: Bandenawaz Bagwan**

**ACKNOWLEDGEMENT**

My sincere gratitude and thanks towards my project paper guide Bandenawaz Bagwan. It was only with his backing and support that I could complete the report. He provided me all sorts of help and corrected me if ever seemed to make mistakes. I have no such words to express my gratitude. I acknowledge my sincere gratitude to the two weeks training in which I learn various new things. In this training I learn about Python modules in Data Science and Machine Learning which gave my start in my new journey.

**INTRODUCTION**

In this modern world where technology have risen to such a level that without using it, we can’t even stay a single minute. This has proven very beneficial for human kind but at the same time it has disturbed the natural flow of nature. Like uneven rain. Because of technology we have disturbed the cycles of rainy seasons which led to uneven rain on occasional days. It’s like It seem to be hot at one instance and then rain at that very moment. It made difficult to figure out our travelling as who knows when will it start raining and flourish our plains or daily routine work. So, it has become very important to know whether it will rain today or tomorrow so that we can plain our activities according to those conditions. To get thorough this I have made a Machine Learning model which can predict whether there will be rain tomorrow or not. By this prediction result, people can plain there routines.

**Problem Statement**

Make prediction of weather condition whether it will rain tomorrow or not.

**Technology and Concepts**

**Data Science:**

In Data Science, first the concept is properly visualized in the form of graphs and charts to get clear idea about the problem statement. It is commonly known as Exploratory Data Analysis. This gives clear idea about the dependent and independent variables which is required in building Machine Learning model. This also tells whether which model we should use to predict the required result. Exploring data is very important as it gives us strong foundation whether the prediction, we have made through Machine Learning model is correct or not. Exploratory Data Analysis also helps to visualize data which intern gives some prediction about future values.

**Machine Learning:**

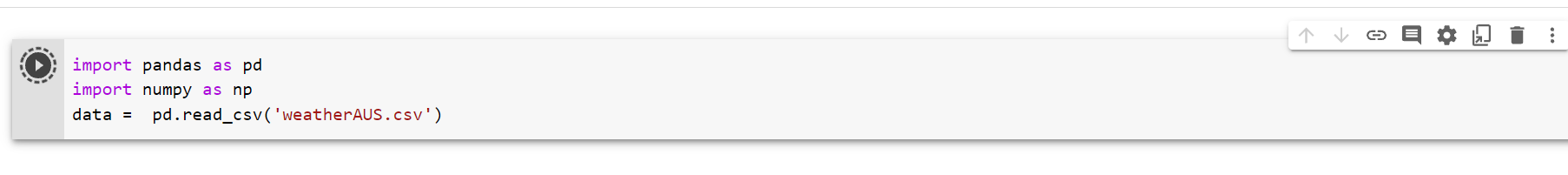
The main thing which actually predicts the result is done by Machine Learning model. It is usually done after analysis of data. Machine Learning is a strong tool which has emerged as a powerful tool when it comes in predicting real world problems. It is of two types: 1)Supervised Learning

2)Unsupervised Learning

Both are useful according to problems. Here in this project, I have used Random Forest algorithm which comes under Supervised Learning. Supervised Learning have various algorithms like Linear Regression, Logistic Regression and few more too which are applicable according to conditions. In the dataset through which I predicted whether it will rain today and tomorrow, to deal with NaN values that is of which data is not available, I have use Simple Imputer which I import from Sklearn Library. To deal with strings values I have use Label Encoder as Machine Learning only deals with numbers. In the entire project I have used various libraries like pandas, numpy, sklearn, Matplotlib and Seaborn.

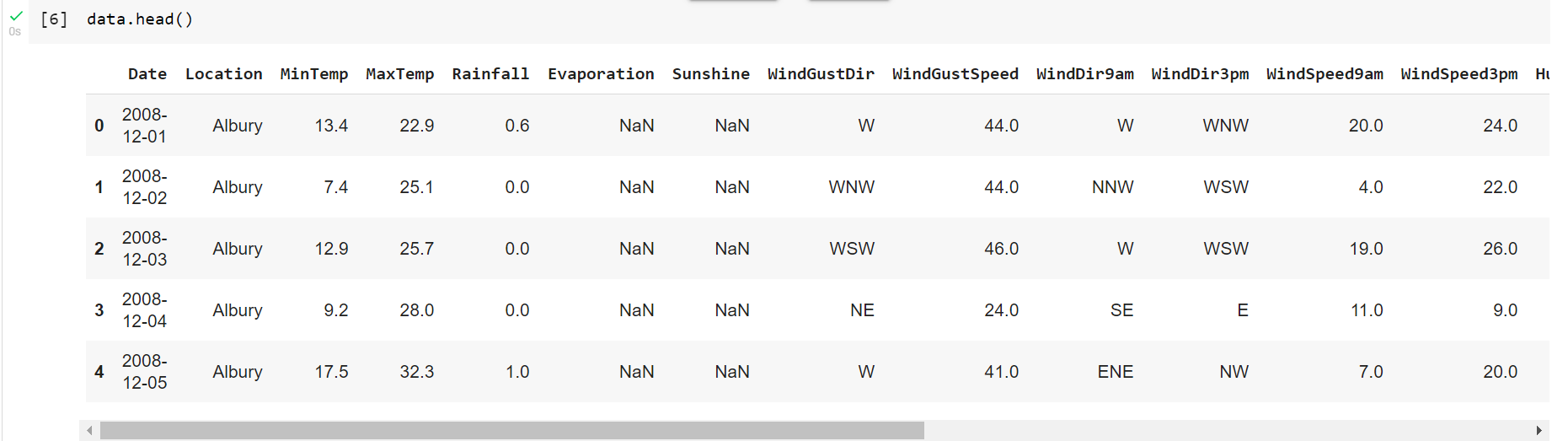
**METHODOLOGY**

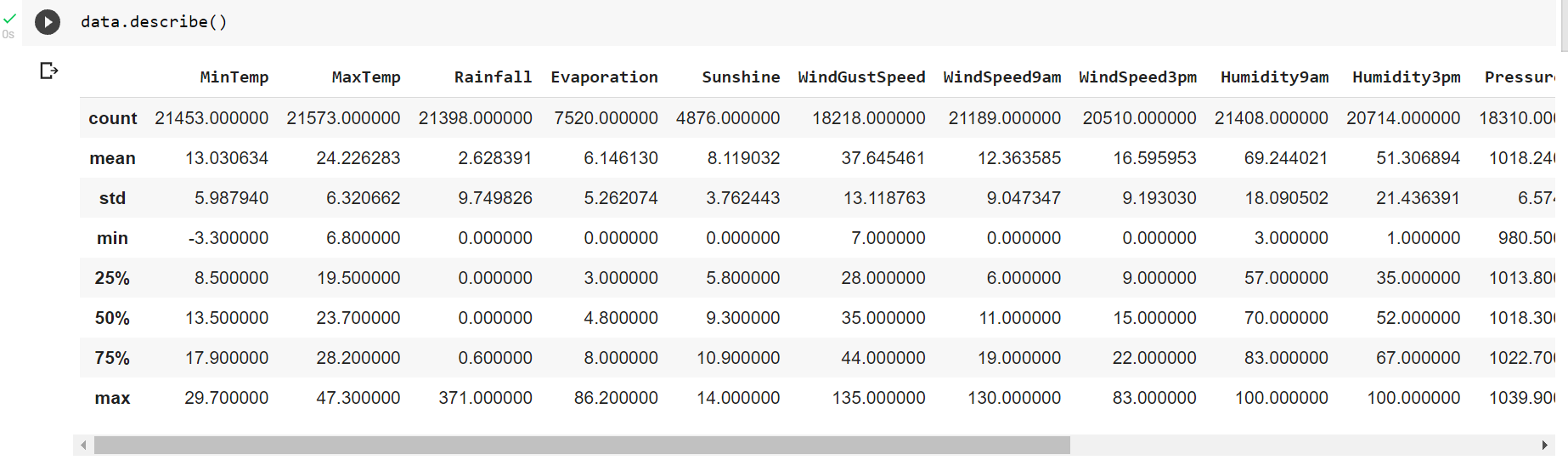
1. **Import data**

Here I have imported the weather file in Google colab using pandas.

1. **Perform Exploratory Data Analysis**

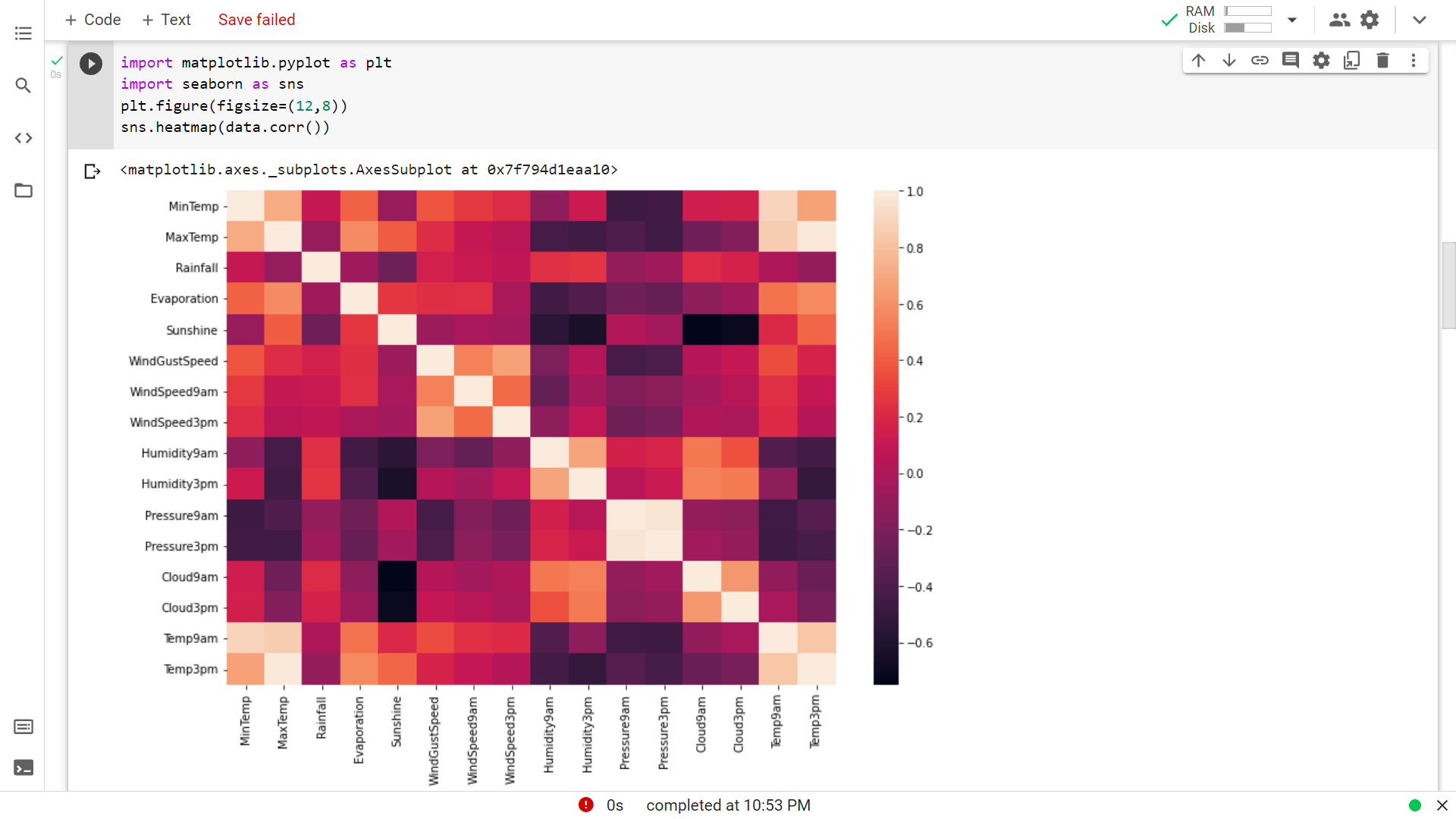
**2.1)** By using head function I have checked what is inside the file name weatherAUS.csv



**2.2)** By using describe function I checked the count, mean, standard deviation, quantile and max-min

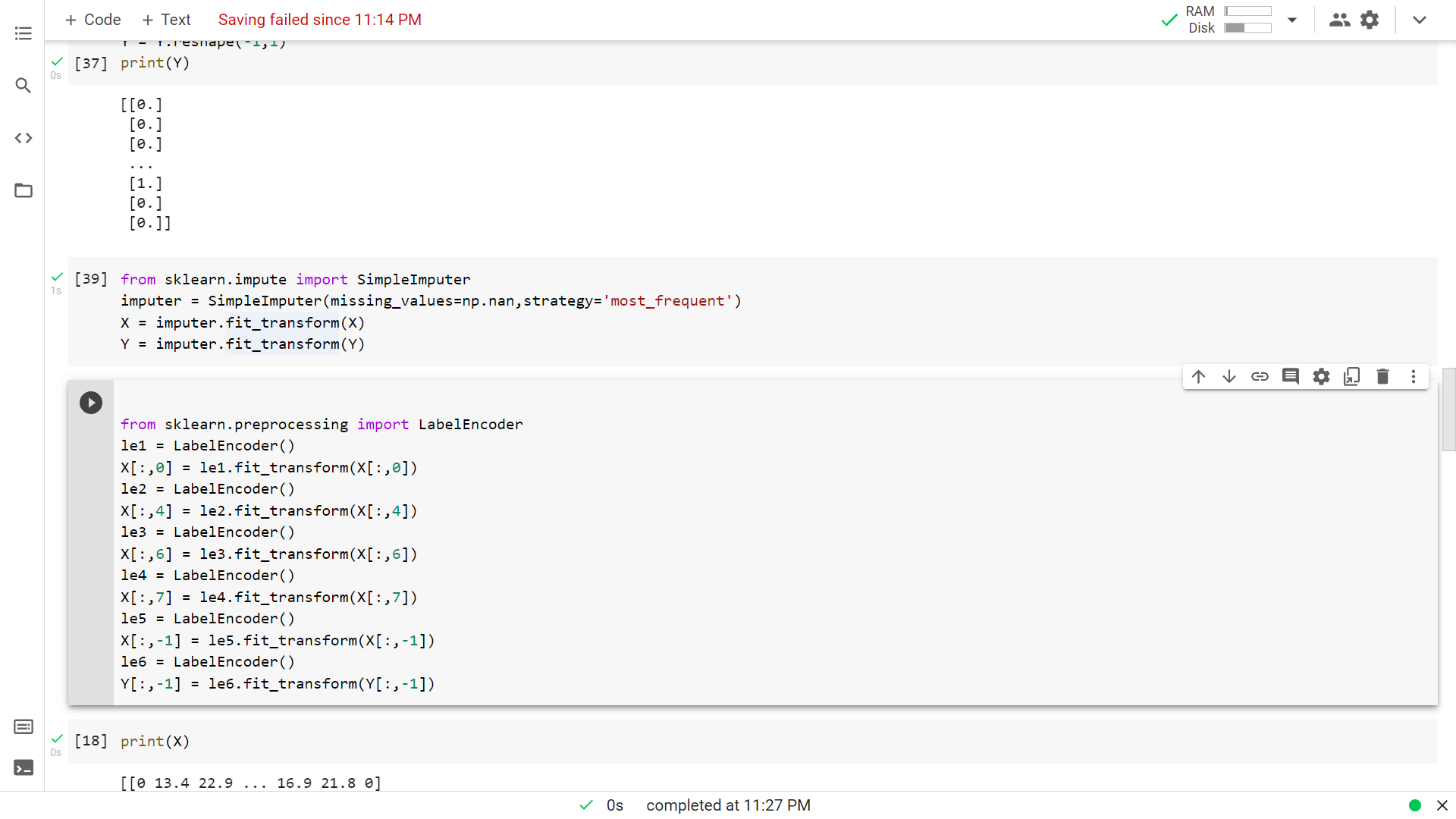
**2.3)** To check how many null values are present I use info function. ****

**2.4)** To visualize date in the graph, I have use heatmap to see relation which quantity affect the rain prediction most. The lighter the color is the more it affects the rain prediction.

****

**2.5)** Separating dependent and independent variables

1. **Dealing with invalid Data i.e NaN values**

****To deal with NaN values I have use SimpleImputer module from sklearn. I have replaced the missing values i.e NaN values with the most frequent values. I could have even replaced it with mean values but here most frequent seems best**.**

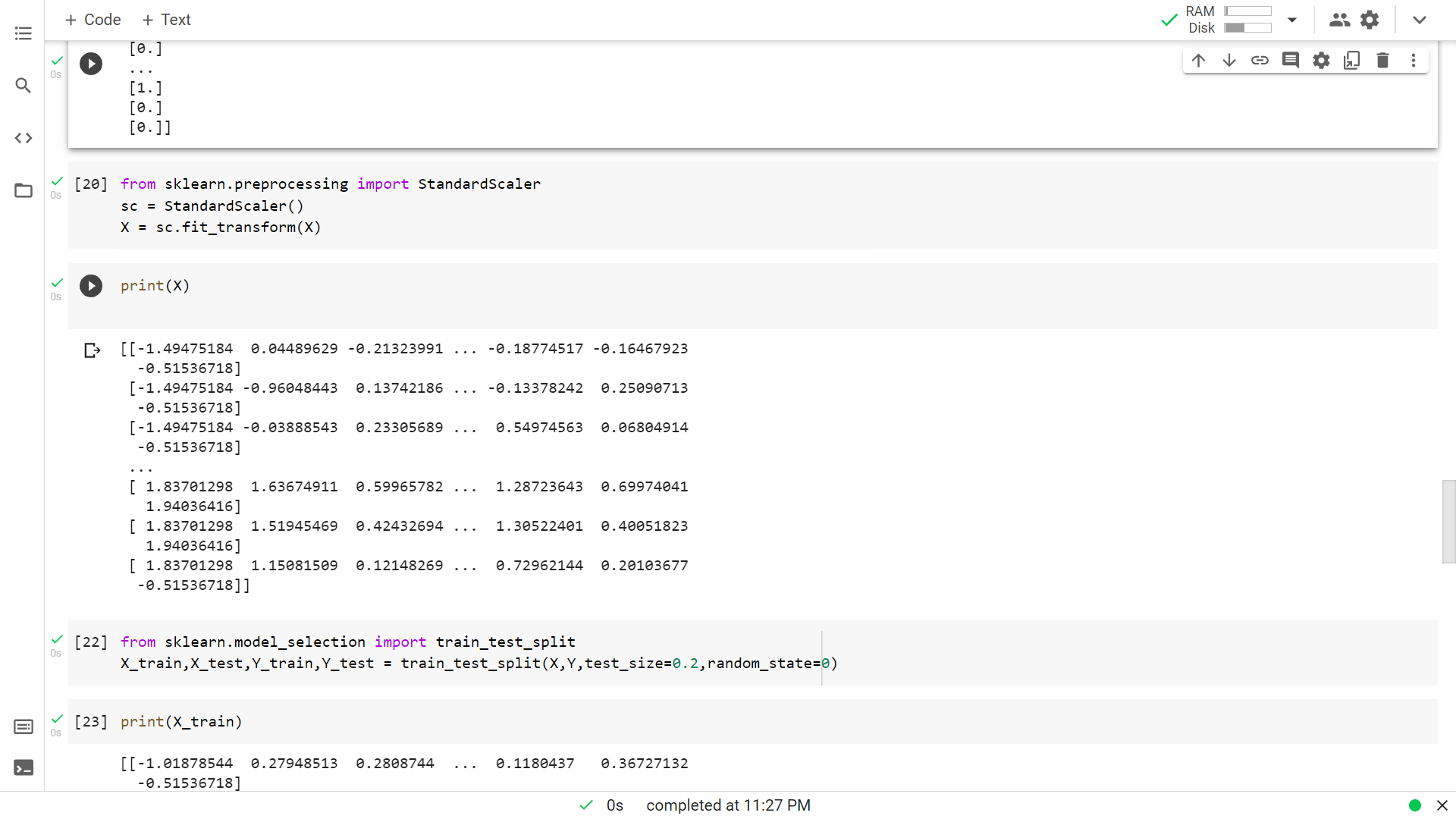
1. **Encoding DataSet**

Machine Learning model doesn’t deal with string values so to convert String values to numbers, I have used LabelEncoder from Sklearn. I could have even used One Hot Encoding but here LabelEncoder seems best as it doesn’t create multiple columns.

****

1. **Feature Scaling**

As the values are too much scattered so to make them fall in a range between -3 to 3, so that it can be visualize easily I have used Standard Scaler from Sklearn.

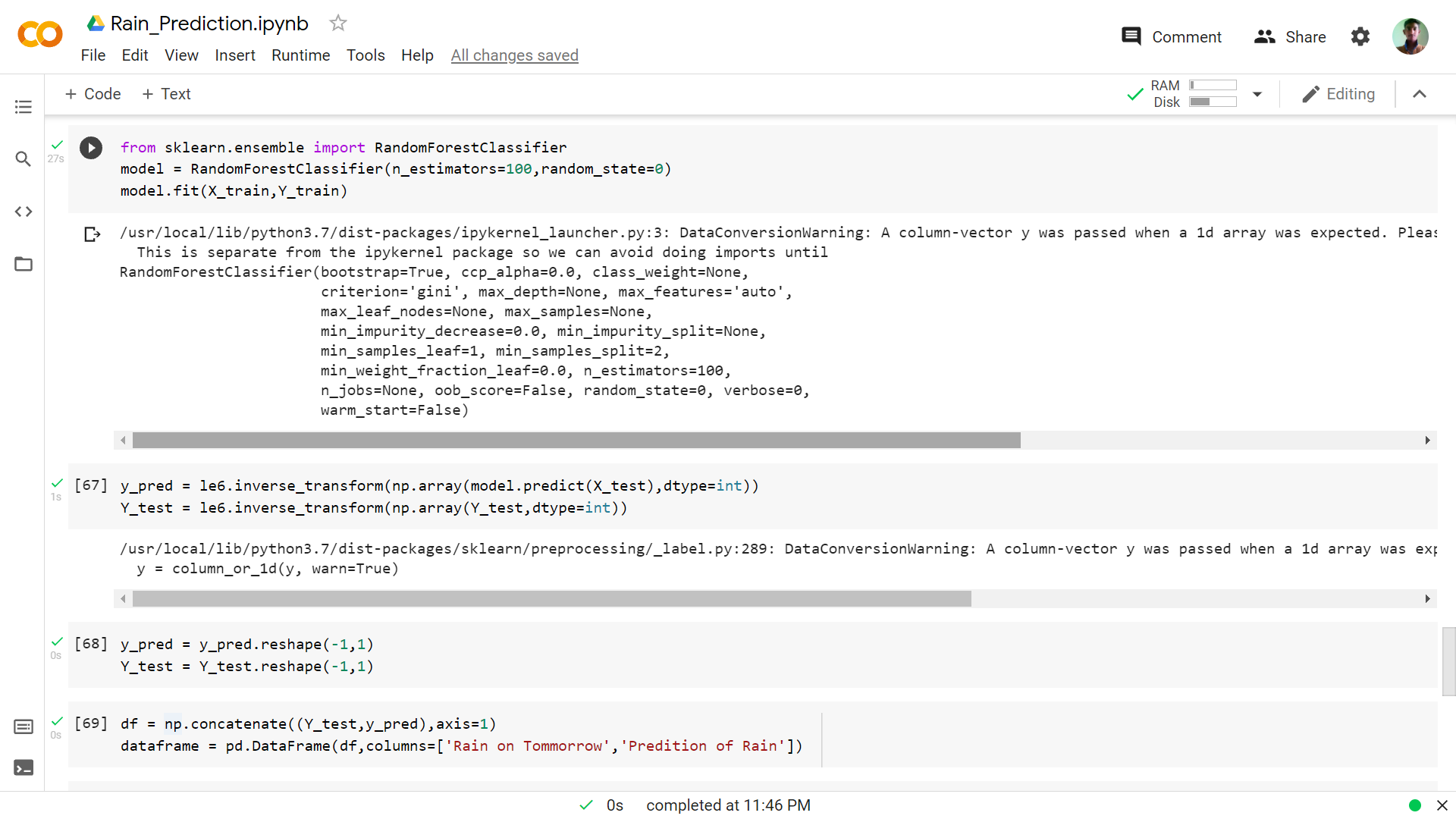
****

1. **Splitting Dataset into Training set and Test set**

Here, I started building my Machine Learning model. I have split my train and test fields with test size of 20% and train size of 80%.

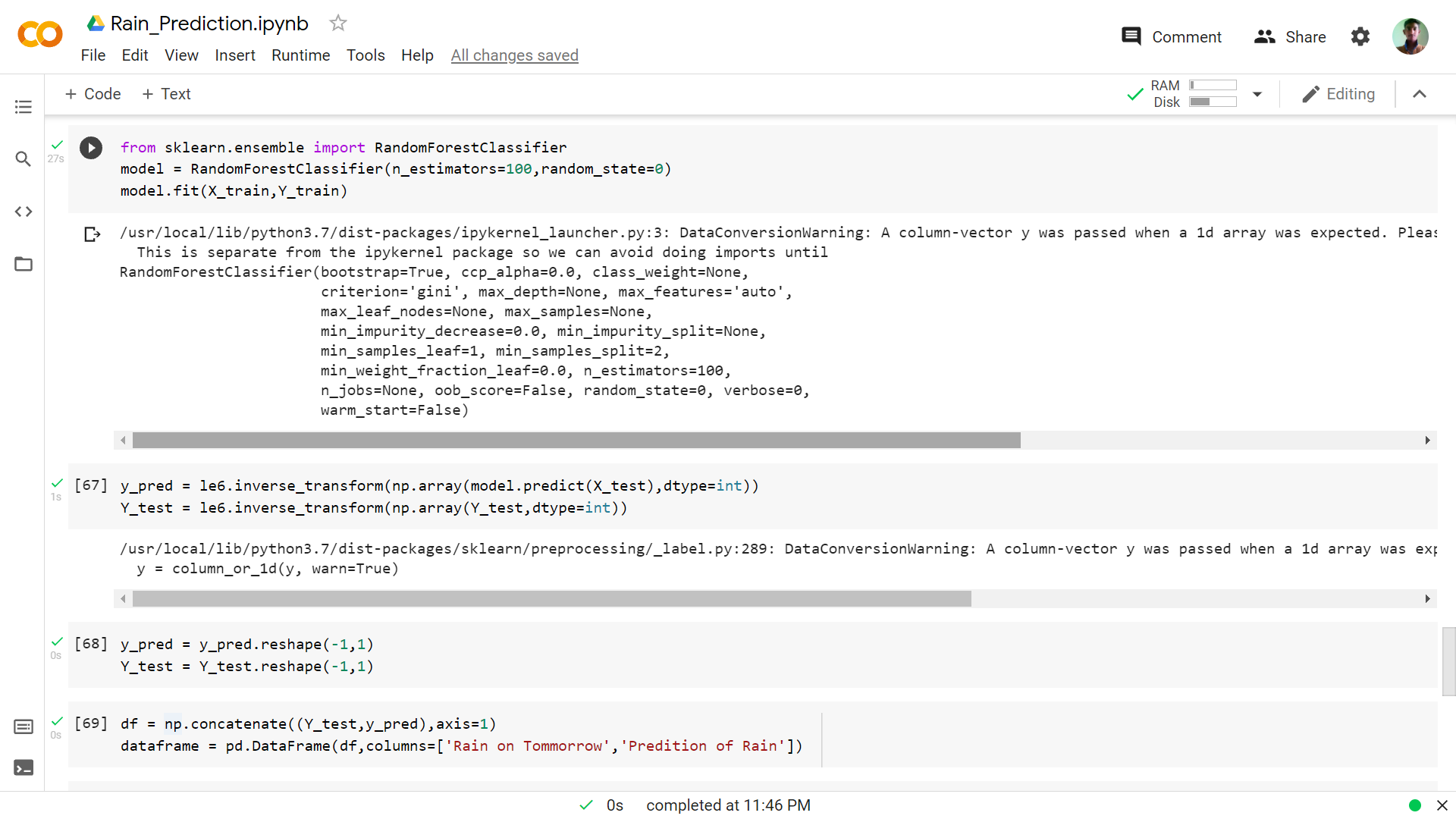
****

1. **Training Model**

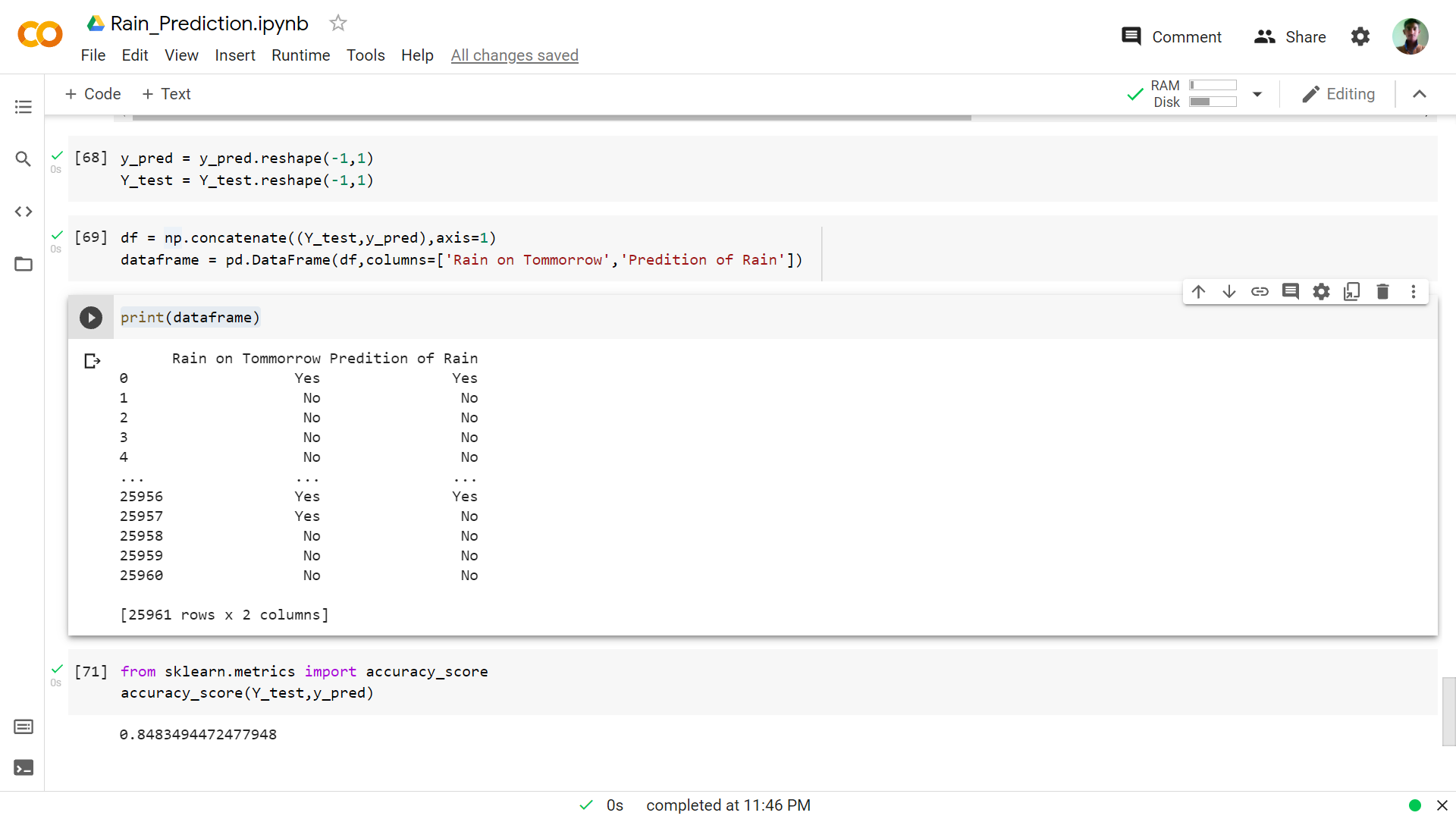
**7.1)** Here I train my Machine Learning model which can predict whether it will rain tomorrow or not by using Random Forest Classifier. I took n\_estimators=100 in model fitting.

1. **Result**

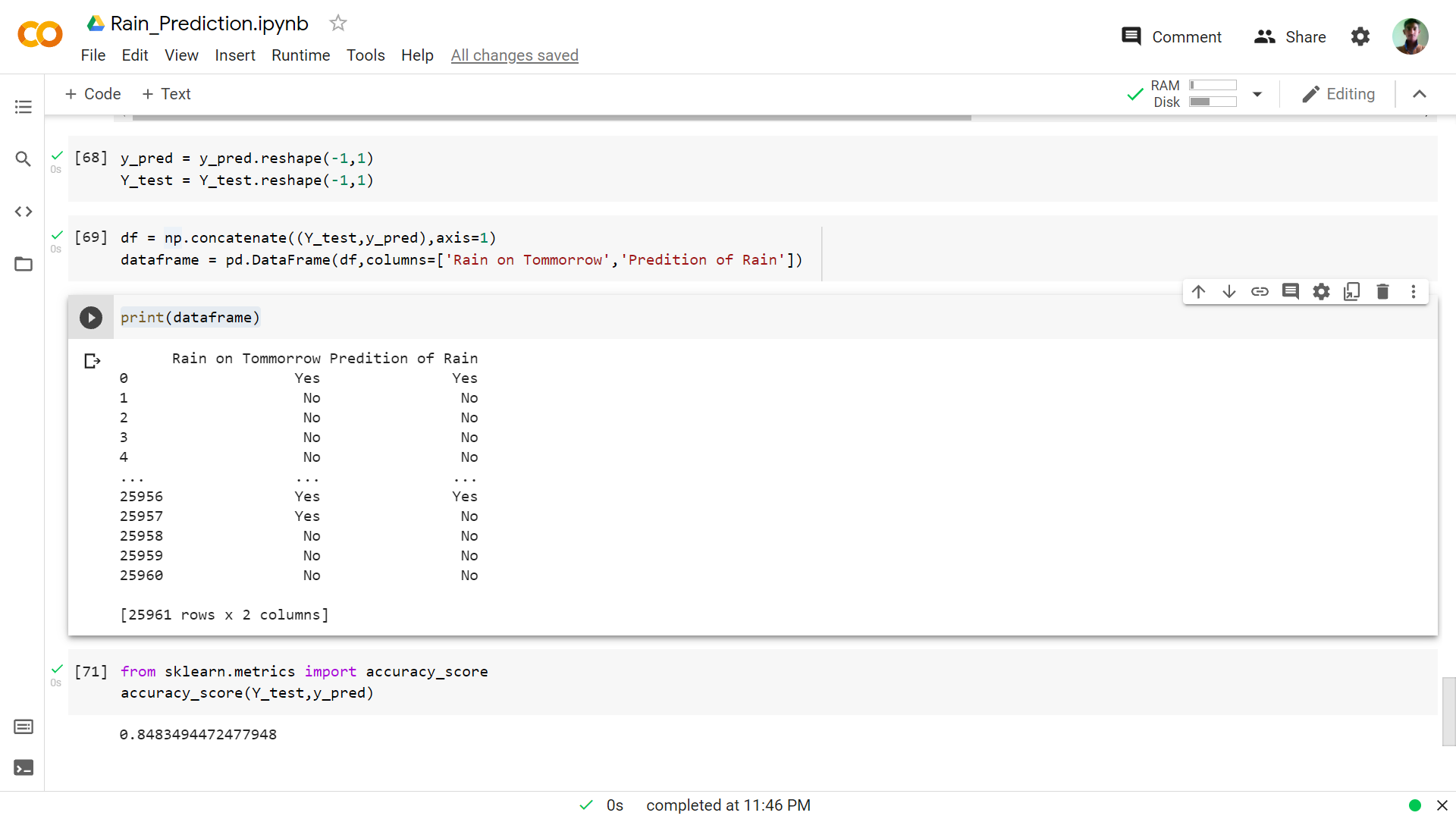
**8.1**) Here I convert my numerical number of YES and NO back to Yes and No using Inverse Transform.

****

**8.2)** Observing result with available data

****

1. **Calculating Accuracy**

The accuracy of this model is 84.83%.

**Conclusions**

This model predicts whether it will rain tomorrow or not with an accuracy score of 84.83%. This model uses Random Forest Classifier to predict this. It gives answer Yes if it rains tomorrow or gives No if it doesn’t rain tomorrow.

**References**

https://www.kaggle.com