ISSN No:-2456-2165

Rainfall Prediction Study on Prediction Is Tomorrow Rain Yes or Not

Abdullah Mohammad Alaraj
Khalil Mohammad Alharbi
Hamad Suliman Almani

TABLE OF CONTENTS

ISSN No:-2456-2165

LIST OF FIGURES

ISSN No:-2456-2165

ABSTRACT

The user can enter values and prediction rain. Weather predicting is an application of science and technology to predict the state of the atmosphere for a particular location. Ancient weather prediction methods usually relied on the observed patterns of events, also called pattern recognition. For example, ifthe humidity increases with the formation of clouds at sunset, then tomorrow it will be rainy. However, notall of these predictions are reliable. Here this system will predict the weather based on parameters liketemperature, humidity, wind etc. (htt1)

The user will enter some data such as temperature, humidity and wind into the system and then *RainTomorrw* will predict the weather based on training and testing the dataset on the previous data in the dataset. Our role in this system is to add the previous weather data in the database and clean it of missingvalues and errors either by deleting ineffective values or filling in the average or most common values andthen processing outliers using z-Score and IQr. Then search for the relationships between the variable through correlation coefficients. Accordingly, a decision tree was built based on the value of entropy, andthen training and testing were performed using scientific methods used in machine learning, including Random Forest and Gradient Boosting algorithm.

After we finished training and testing the model and reviewing the accuracy, a simple website was designedso that the user could know the *RainTomorrw*.

Thus, this prediction will prove to be reliable. The system can be used in transportation, shipping, agriculture, education, sports, open events, etc. through the form via the link.

ISSN No:-2456-2165

CHAPTER ONE
INTRODUCTION

Several times the challenging weather condition makes life miserable and also affects the businessesdirectly or indirectly. Nowadays, everybody wants to start their day with a good plan so that most ofthe people often used to check the weather report before preparing their daily plan. Although the weather forecasts are not always perfect but using Machine Learning based data mining method it ispossible to forecast future weather conditions based on the historical data. (htt1)

The project represents the machine learning-based predictive analysis to predict the *RainTomorrw* for upcoming day based on the given data. The proposed system will help to predict the future trend of the weather considering the historical data. Through this model, the probability of the *RainTomorrw* canbe analyzed so that all the people can make their plans without any confusion.

We have proposed an experimental approach to develop a *RainTomorrw* prediction classifier to predict future weather condition based on the other features.

The data has been collected based on the weather records of different cities in Australia along with atmospheric parameters. This dataset has been available at Kaggle. (htt2) (htt4)

1.1 Define the problem

It is important to accurately determine RainTomorrw in order to preserve water resources and benefit from rainfall in the agricultural field, as well as in the areas of transportation and shipping so as not to sufferlosses, as well as in education so that we maintain the safety of students and other areas. (htt1)

1.2 Scope

It tells us whether there will be RainTomorrw or not.

1.3 Purpose

There are several reasons why weather forecasts are important. They would certainly be missed if theywere not there. It is a product of science that impacts the lives of many people. The following is a listof various reasons why weather predictions are important:

- Helps people prepare for how to dress (i.e. warm weather, cold weather, windyweather, rainy weather)
- Helps businesses and people plan for power production and how much power to use(i.e. power companies, where to set thermostat)
- Helps people prepare if they need to take extra gear to prepare for the weather(i.e.umbrella, rain coat, sun screen)
- Helps people plan outdoor activities (i.e. to see if rain/storms/cold weather willimpact outdoor event)
- Helps curious people to know what sort of weather can be expected (i.e. a snow onthe way, severe storms)
- Helps businesses plan for transportation hazards that can result from the weather(i.e. fog, snow, ice, storms, clouds as it relates to driving and flying for example)
- Helps people with health related issues to plan the day (i.e. allergies, asthma, and heatstress).
- Helps businesses and people plan for severe weather and other weatherhazards (lightning, hail, tornadoes, hurricanes, ice storms).
- Helps farmers and gardeners plan for crop irrigation and protection (irrigationscheduling, freeze protection).

1.4 Selection and Testing:

After the topic of the research was identified and viewed from interested sources, the database was selected from kaggle.com under the title **Australia Weather Data** and the dataset as **WeatherTraining Data.csv.** (htt4) (htt5)

The dataset has been checked and ensured of its quality and suitable for the study, analysis and productionwork through that it addresses the basic problem of the project subject, which is production *RainTomorrow.* It is also complete and large in size and describes the data and is relevant to the projectand the data is historically appropriate because there was no change in the climate after taking the data andit is sufficient to carry out the study, the strong and neglected data are very few, and the influential climatic information in the dataset has been identified and sufficient experience has been created to know the causes of rain.

1.5 Source

Observations were drawn from numerous weather stations. The daily observations are available from http://www.bom.gov.au/climate/data.

ISSN No:-2456-2165

Definitions adapted from http://www.bom.gov.au/climate/dwo/IDCJDW0000.shtml.
Data source - http://www.bom.gov.au/climate/dwo/ and http://www.bom.gov.au/climate/data. Copyright Commonwealth of Australia 2010, Bureau of Meteorology

1.6 Work environment

Programming language : python
Application environment: Jupyter , Excel

1.7 Context

Predict next-day rain by training classification models on the target variable *RainTomorrow.*

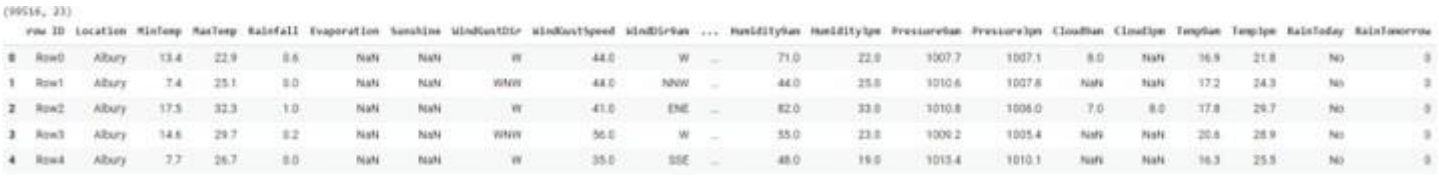
1.8 Content

This dataset contains about 10 years of daily weather observations from many locations across Australia.

ISSN No:-2456-2165

CHAPTER TWO
DATASET DESCRIPTION

The dataset consists of 23 columns and 99516 rows. (htt2) (htt1)



(99516, 23)

	row ID	Location	MinTemp	MaxTemp	Rainfall	Evaporation	Sunshine	WindGustDir	WindGustSpeed	WindDir9am	...	Humidity9am	Humidity3pm	Pressure9am	Pressure3pm	Cloud9am	Cloud3pm	Temp9am	Temp3pm	RainToday	RainTomorrow
0	Row0	Albury	13.4	22.9	0.6	NaN	NaN	W	44.0	W	...	71.0	22.0	1007.7	1007.1	8.0	NaN	16.9	21.8	No	0
1	Row1	Albury	7.4	25.1	0.0	NaN	NaN	WNW	44.0	NNW	...	44.0	25.0	1010.6	1007.8	NaN	NaN	17.2	24.3	No	0
2	Row2	Albury	17.5	32.3	1.0	NaN	NaN	W	41.0	ENE	...	82.0	33.0	1010.8	1006.0	7.0	8.0	17.8	29.7	No	0
3	Row3	Albury	14.6	29.7	0.2	NaN	NaN	WNW	56.0	W	...	55.0	23.0	1009.2	1005.4	NaN	NaN	20.6	28.9	No	0
4	Row4	Albury	7.7	26.7	0.0	NaN	NaN	W	35.0	SSE	...	48.0	19.0	1013.4	1010.1	NaN	NaN	16.3	25.5	No	0

Fig 1: All the features and the number of data

We find that the data in the database is of quality through its accuracy and sincerity of its source, as it isfrom a government source, and that the data is clearly explained, has value and can be accessed, and thereis no conflict and therefore we can start the processing stage to process the data .

We will get acquainted with all the features and their data type, as well as the range of numeric values andtheir unit of measurement as follows:

Feature Name	Description	Type	Range
row ID	The number of rows of data represents a sequential order	object	
Location	Name of the city from Australia	object	
MinTemp	The Minimum temperature during a particular day	float	-8.2 : 31.8
MaxTemp	The maximum temperature during a particular day	float	-4.8 : 47
Rainfall	Rainfall during a particular day. (millimeters).	float	0 : 278.4
Evaporation	Evaporation during a particular day	object	0 :145
Sunshine	Evaporation during a particular day. (millimeters)	float	0 :14.3
WindGusDir	Bright sunshine during a particular day", (compass points")	object	
WindGuSpeed	Speed of strongest gust during a particular day. (kilometers per hour)	float	7 :122
WindDir9am	The direction of the wind for 10 min prior to 9 am (compass points)	object	
WindDir3pm	The direction of the wind for 10 min prior to 3 pm (compass points)	float	

ISSN No:-2456-2165

WindSpeed9am	Speed of the wind for 10 min prior to 3 pm. (kilometers per hour)	object	0 :74
WindSpeed3pm	The direction of the wind for 10 min prior to 3 pm	object	0 :74
Humidity9am	Rainfall during a particular day	float	0 : 83
Humidity9am	The humidity of the wind at 9 am. (percent)	float	1 :100
Humidity3pm	The humidity of the wind at 3 pm	float	1 :100
Pressure9am	Atmospheric pressure at 9 am. (hectopascals)	float	982.2 :1040.4
Pressure3pm	Atmospheric pressure at 3 pm. (hectopascals)	float	977.1 :1038.9
Cloud9am	Cloud obscured portions of the sky at 9 am (eighths) [2]	object	
Cloud3pm	Cloud obscured portions of the sky at 3 pm ,(eighths) [2]	object	
Temp9am	The temperature at 9 am. (degree Celsius)	object	-7.2 :39.4
Temp3pm	The temperature at 3 pm. (degree Celsius)	float	-5.4 :45.4
RainToday	If today is rainy then 'Yes'. If today is not rainy then 'No'.	object	
RainTomorrow	If tomorrow is rainy then '1' . If tomorrow is not rainy then '0	in	

Table1: Descriptive table of the dataset.

20 No cloude ,1/8th sky cover (few) , 2/8th sky cover (Scattered) , 3/8th sky cover , 4/8th sky cover , 5/8th sky cover , 6/8th sky cover (Broken) , 7/8th sky cover , 8/8th sky cover (Overcast) (htt)

ISSN No:-2456-2165

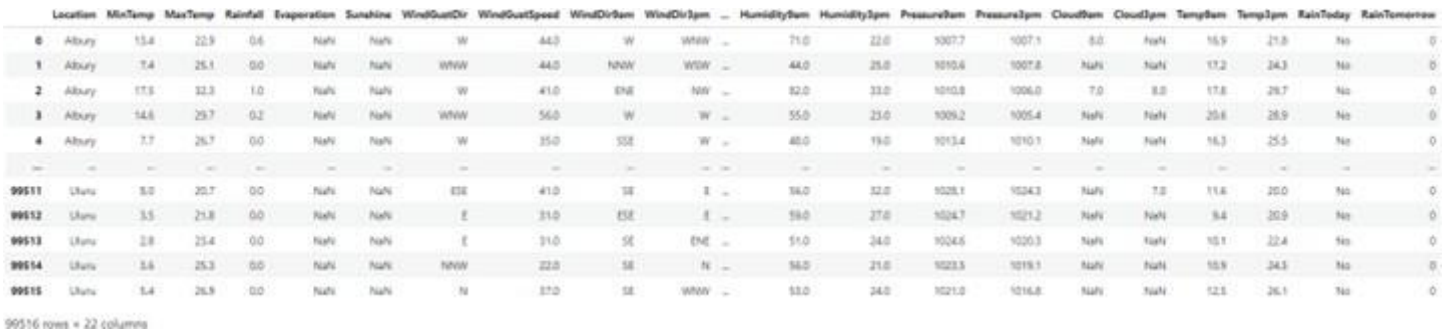
#	Column	Non-Null Count	Dtype
0	row ID	99516 non-null	object
1	Location	99516 non-null	object
2	MinTemp	99073 non-null	float64
3	MaxTemp	99286 non-null	float64
4	Rainfall	98537 non-null	float64
5	Evaporation	56985 non-null	float64
6	Sunshine	52199 non-null	float64
7	WindGustDir	92995 non-null	object
8	WindGustSpeed	93036 non-null	float64
9	WindDir9am	92510 non-null	object
10	WindDir3pm	96868 non-null	object
11	WindSpeed9am	98581 non-null	float64
12	WindSpeed3pm	97681 non-null	float64
13	Humidity9am	98283 non-null	float64
14	Humidity3pm	97010 non-null	float64
15	Pressure9am	89768 non-null	float64
16	Pressure3pm	89780 non-null	float64
17	Cloud9am	61944 non-null	float64
18	Cloud3pm	59514 non-null	float64
19	Temp9am	98902 non-null	float64
20	Temp3pm	97612 non-null	float64
21	RainToday	98537 non-null	object
22	RainTomorrow	99516 non-null	int64

Figure 2: Attributes, their type, and non-null values (htt1) (htt2)

ISSN No:-2456-2165

CHAPTER THREE
DATA PREPARATION

After defining the database and its contents, we start the preprocessing of the data set to start the analysis operations, after understanding the data set, we found that the row ID of the attribute has no correlation or significance in the analysis, so it was removed in order to reduce the size of the data set because thesevalues have no analytical meaning.



	Location	MinTemp	MaxTemp	Rainfall	Evaporation	Sunshine	WindGustDir	WindGustSpeed	WindDir9am	WindDir3pm	...	Humidity9am	Humidity3pm	Pressure9am	Pressure3pm	Cloud9am	Cloud3pm	Temp9am	Temp3pm	RainToday	RainTomorrow
0	Albury	13.4	22.9	0.6	NaN	NaN	W	44.0	W	WNW	...	71.0	22.0	1007.7	1007.1	8.0	NaN	16.9	21.8	No	0
1	Albury	7.4	25.1	0.0	NaN	NaN	WNW	44.0	NNW	WSW	...	44.0	25.0	1010.6	1007.8	NaN	NaN	17.2	24.3	No	0
2	Albury	17.5	32.3	1.0	NaN	NaN	W	41.0	ENE	NW	...	82.0	33.0	1010.8	1006.0	7.0	8.0	17.8	29.7	No	0
3	Albury	14.6	29.7	0.2	NaN	NaN	WNW	56.0	W	W	...	55.0	23.0	1009.2	1005.4	NaN	NaN	20.6	28.9	No	0
4	Albury	7.7	26.7	0.0	NaN	NaN	W	35.0	SSE	W	...	48.0	19.0	1013.4	1010.1	NaN	NaN	16.3	25.5	No	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
99511	Uluru	5.0	20.7	0.0	NaN	NaN	ESE	41.0	SE	E	...	56.0	32.0	1028.1	1024.3	NaN	7.0	11.4	20.0	No	0
99512	Uluru	3.5	21.8	0.0	NaN	NaN	E	31.0	ESE	E	...	59.0	27.0	1024.7	1021.2	NaN	NaN	9.4	20.9	No	0
99513	Uluru	2.8	23.4	0.0	NaN	NaN	E	31.0	SE	ENE	...	51.0	24.0	1024.6	1020.3	NaN	NaN	10.1	22.4	No	0
99514	Uluru	3.6	25.3	0.0	NaN	NaN	NNW	22.0	SE	N	...	56.0	21.0	1023.5	1019.1	NaN	NaN	10.9	24.5	No	0
99515	Uluru	5.4	26.9	0.0	NaN	NaN	N	37.0	SE	WNW	...	53.0	24.0	1021.0	1016.8	NaN	NaN	12.5	26.1	No	0

96516 rows × 22 columns

Figure 3: shows the data set after deleting the column (htt1) (htt2)

According to the initial observation, the data set contains many missing values which will affect theanalysis and they are as in the figure and so we have to deal with them

	Total	Percent
Sunshine	47317	0.475471
Evaporation	42531	0.427379
Cloud3pm	40002	0.401966
Cloud9am	37572	0.377547
Pressure9am	9748	0.097954
Pressure3pm	9736	0.097834
WindDir9am	7006	0.070401
WindGustDir	6521	0.065527
WindGustSpeed	6480	0.065115
WindDir3pm	2648	0.026609
Humidity3pm	2506	0.025182
Temp3pm	1904	0.019133
WindSpeed3pm	1835	0.018439
Humidity9am	1233	0.012390
RainToday	979	0.009838
Rainfall	979	0.009838
WindSpeed9am	935	0.009395
Temp9am	614	0.006170
MinTemp	443	0.004452
MaxTemp	230	0.002311
Location	0	0.000000
RainTomorrow	0	0.000000

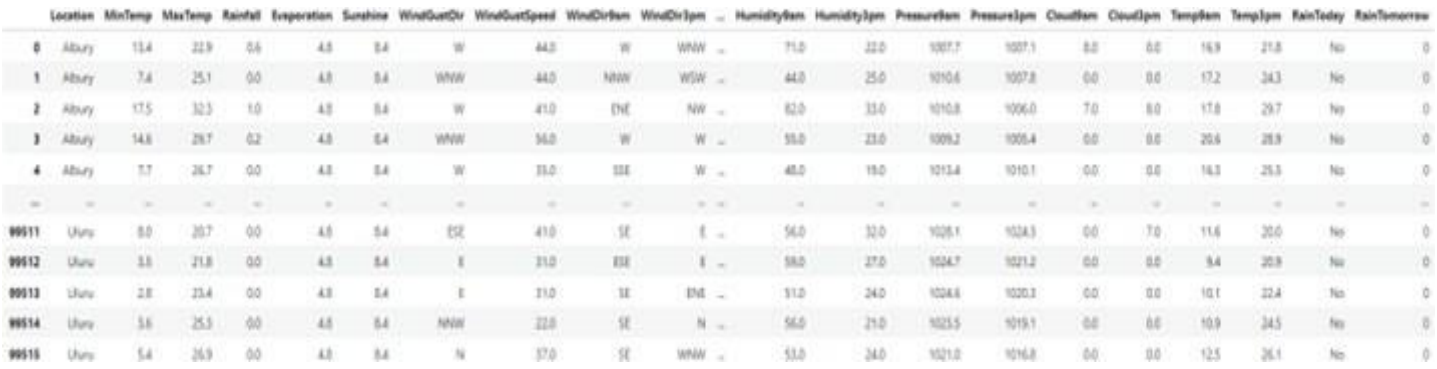
Figure 4: Shows the traits and the number of missing values along with their ratio (htt1) (htt2)

The empty character fields are as follows (*WindGustDir, WindDir9am, WindDir3pm*).These fields havebeen filled with the most frequent values of the respective attributes except RainToday .

Numeric values (*MinTemp, MinTemp, MaxTemp, Rainfall, Evaporation, Sunshine, WindGustSpeed, Wind- Speed9am, WindSpeed3pm, Humidity9am, Humidity3pm, Pressure9am, Pressure3pm, Temp9am, Temp3pm*) are processed by taking the median of the class values that which is one of the ways to processnull values .

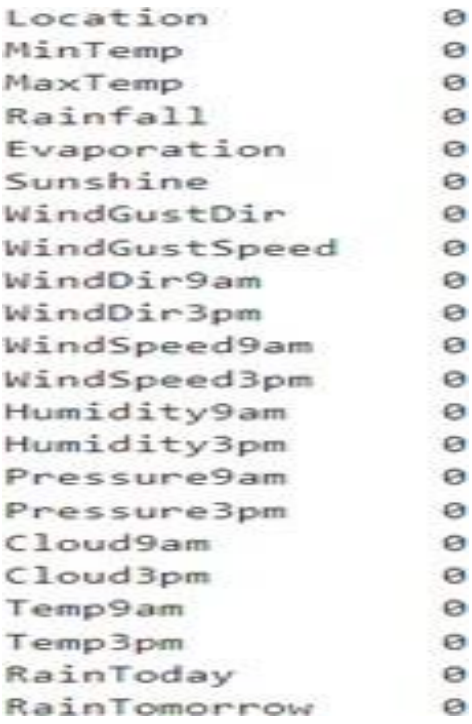
ISSN No:-2456-2165

As for the two fields (*Cloud9am, Cloud3pm*) so it was filled with the value 0 because, according to our understanding of the dataset, it shows that when there are no cloud, the value is placed empty and it represents 0 in the cloud criteria eighths .



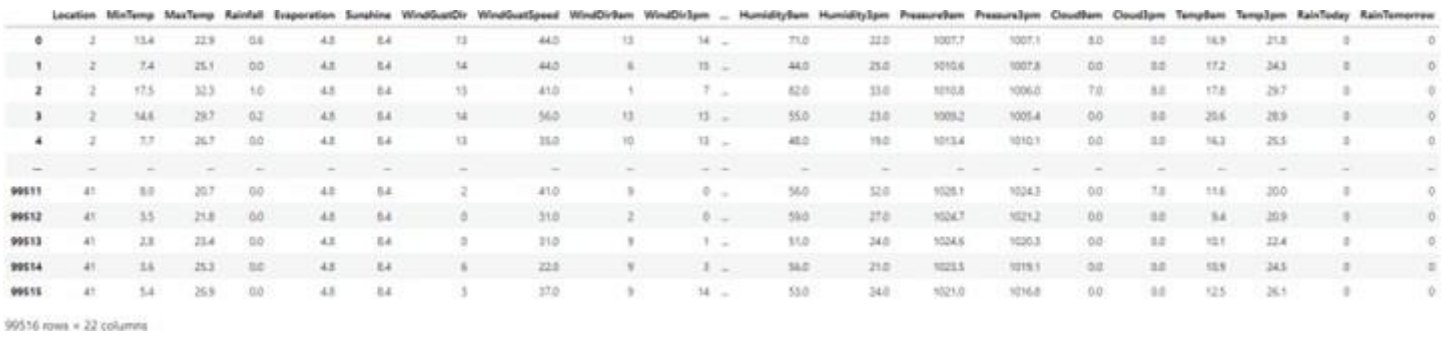
	Location	MinTemp	MaxTemp	Rainfall	Evaporation	Sunshine	WindGustDir	WindGustSpeed	WindDir9am	WindDir3pm	...	Humidity9am	Humidity3pm	Pressure9am	Pressure3pm	Cloud9am	Cloud3pm	Temp9am	Temp3pm	RainToday	RainTomorrow
0	Albury	13.4	22.9	0.6	4.8	8.4	W	44.0	W	WNW	...	71.0	22.0	1007.7	1007.1	8.0	0.0	16.9	21.8	No	0
1	Albury	7.4	25.1	0.0	4.8	8.4	WNW	44.0	NNW	WSW	...	44.0	25.0	1010.6	1007.8	0.0	0.0	17.2	24.3	No	0
2	Albury	17.5	32.3	1.0	4.8	8.4	W	41.0	ENE	NW	...	82.0	33.0	1010.8	1006.0	7.0	8.0	17.8	29.7	No	0
3	Albury	14.6	29.7	0.2	4.8	8.4	WNW	56.0	W	W	...	55.0	23.0	1009.2	1005.4	0.0	0.0	20.6	28.9	No	0
4	Albury	7.7	26.7	0.0	4.8	8.4	W	35.0	ESE	W	...	48.0	19.0	1013.4	1010.1	0.0	0.0	16.3	25.3	No	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
99511	Uluru	8.0	20.7	0.0	4.8	8.4	ESE	41.0	SE	E	...	56.0	32.0	1028.1	1024.3	0.0	7.0	11.6	20.0	No	0
99512	Uluru	3.8	21.8	0.0	4.8	8.4	E	31.0	ESE	E	...	59.0	27.0	1024.7	1021.2	0.0	0.0	9.4	20.9	No	0
99513	Uluru	2.8	23.4	0.0	4.8	8.4	E	31.0	SE	ENE	...	51.0	24.0	1024.6	1020.3	0.0	0.0	10.1	22.4	No	0
99514	Uluru	5.6	25.3	0.0	4.8	8.4	NNW	22.0	SE	N	...	56.0	21.0	1023.5	1019.1	0.0	0.0	10.9	24.5	No	0
99515	Uluru	5.4	26.9	0.0	4.8	8.4	N	37.0	SE	WNW	...	53.0	24.0	1021.0	1016.8	0.0	0.0	12.5	26.1	No	0

Figure 5: Data after missing value processing (htt1) (htt2)



Location	0
MinTemp	0
MaxTemp	0
Rainfall	0
Evaporation	0
Sunshine	0
WindGustDir	0
WindGustSpeed	0
WindDir9am	0
WindDir3pm	0
WindSpeed9am	0
WindSpeed3pm	0
Humidity9am	0
Humidity3pm	0
Pressure9am	0
Pressure3pm	0
Cloud9am	0
Cloud3pm	0
Temp9am	0
Temp3pm	0
RainToday	0
RainTomorrow	0

In the dataset, the *RainToday* have binary values (Yes and No) so this variable should be converted into 0(No) and 1 (Yes). In the Unknown Weather dataset, the records having the empty values for the RainToday attribute have been dropped from the dataset since it is quite difficult to decide whether the day was rainyor not and their number is small and does not affect the analysis process. The rest of the attributes of theobject type were also converted to numeric so that we can deal with it in the analysis .



	Location	MinTemp	MaxTemp	Rainfall	Evaporation	Sunshine	WindGustDir	WindGustSpeed	WindDir9am	WindDir3pm	...	Humidity9am	Humidity3pm	Pressure9am	Pressure3pm	Cloud9am	Cloud3pm	Temp9am	Temp3pm	RainToday	RainTomorrow
0	2	13.4	22.9	0.6	4.8	8.4	13	44.0	13	14	...	71.0	22.0	1007.7	1007.1	8.0	0.0	16.9	21.8	0	0
1	2	7.4	25.1	0.0	4.8	8.4	14	44.0	6	15	...	44.0	25.0	1010.6	1007.8	0.0	0.0	17.2	24.3	0	0
2	2	17.5	32.3	1.0	4.8	8.4	15	41.0	1	7	...	82.0	33.0	1010.8	1006.0	7.0	8.0	17.8	29.7	0	0
3	2	14.6	29.7	0.2	4.8	8.4	14	56.0	13	13	...	55.0	23.0	1009.2	1005.4	0.0	0.0	20.6	28.9	0	0
4	2	7.7	26.7	0.0	4.8	8.4	13	35.0	10	13	...	48.0	19.0	1013.4	1010.1	0.0	0.0	16.3	25.5	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
99511	41	8.0	20.7	0.0	4.8	8.4	2	41.0	9	0	...	56.0	32.0	1025.1	1024.3	0.0	7.0	11.6	20.0	0	0
99512	41	3.5	21.8	0.0	4.8	8.4	0	31.0	2	0	...	59.0	27.0	1024.7	1021.2	0.0	0.0	9.4	20.9	0	0
99513	41	2.8	23.4	0.0	4.8	8.4	3	31.0	9	1	...	51.0	24.0	1024.6	1020.3	0.0	0.0	10.1	22.4	0	0
99514	41	3.6	25.3	0.0	4.8	8.4	6	22.0	9	3	...	56.0	21.0	1023.5	1019.1	0.0	0.0	10.9	24.5	0	0
99515	41	5.4	26.9	0.0	4.8	8.4	3	37.0	9	14	...	53.0	24.0	1021.0	1016.8	0.0	0.0	12.5	26.1	0	0

99516 rows × 22 columns

Figure 6: Convert all attributes to numeric (htt1) (htt2)

Thus, all data is null and numeric

```
Location        0
MinTemp         0
MaxTemp         0
Rainfall        0
Evaporation     0
Sunshine        0
WindGustDir     0
WindGustSpeed   0
WindDir9am      0
WindDir3pm      0
WindSpeed9am    0
WindSpeed3pm    0
Humidity9am     0
Humidity3pm     0
Pressure9am     0
Pressure3pm     0
Cloud9am        0
Cloud3pm        0
Temp9am         0
Temp3pm         0
RainToday       0
RainTomorrow    0
```

Figure 7: No missing values (htt1) (htt2)

With regard to outlier and errors, the data set was reviewed in two ways, the first is manual, which is tosearch all the fields of the data set, which resulted in a human error when entering, which is to write thevalue NA instead of NO and it was manually corrected as in the figure

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	row ID	Location	MinTemp	MaxTemp	Rainfall	Evaporatic	Sunshine	WindGustI	WindGust	WindDir9a	WindDir3p	WindSpee	WindSpee	Humidity9	Humidity3	Pressure9	Pressure3	Cloud9am	Cloud3pm	Temp9am	Temp3pm	RainToday	RainTomorro
2	Row0	Albury	13.4	22.9	0.6			W	44	W	WNW	20	24	71	22	1007.7	1007.1	8		16.9	21.8	No	0
3	Row1	Albury	7.4	25.1	0			WNW	44	NNW	WSW	4	22	44	25	1010.6	1007.8			17.2	24.3	No	0
4	Row2	Albury	17.5	32.3	1			W	41	ENE	NW	7	20	82	33	1010.8	1006	7	8	17.8	29.7	No	0
5	Row3	Albury	14.6	29.7	0.2			WNW	56	W	W	19	24	55	23	1009.2	1005.4			20.6	28.9	No	0
6	Row4	Albury	7.7	26.7	0			W	35	SSE	W	6	17	48	19	1013.4	1010.1			16.3	25.5	No	0
7	Row5	Albury	13.1	30.1	1.4			W	28	S	SSE	15	11	58	27	1007	1005.7			20.1	28.2	Yes	0
8	Row6	Albury	13.4	30.4	0			N	30	SSE	ESE	17	6	48	22	1011.8	1008.7			20.4	28.8	No	1
9	Row7	Albury	15.9	21.7	2.2			NNE	31	NE	ENE	15	13	89	91	1010.5	1004.2	8	8	15.9	17	Yes	1
10	Row8	Albury	12.6	21	3.6			SW	44	W	SSW	24	20	65	43	1001.2	1001.8		7	15.8	19.8	Yes	0
11	Row9	Albury	9.8	27.7				WNW	50	NA	WNW		22	50	28	1013.4	1010.3	0		17.3	26.2	NA	0
12	Row10	Albury	14.1	20.9	0			ENE	22	SSW	E	11	9	69	82	1012.2	1010.4	8	1	17.2	18.1	No	1
13	Row11	Albury	13.5	22.9	16.8			W	63	N	WNW	6	20	80	65	1005.8	1002.2	8	1	18	21.5	Yes	1
14	Row12	Albury	11.2	22.5	10.6			SSE	43	WSW	SW	24	17	47	32	1009.4	1009.7		2	15.5	21	Yes	0

Figure 8: Detect input error (htt1) (htt2)

One way of searching for outliers values is that we detect them through graphics via the IQR account andsome features that have been processed by z score The attributes that have been detected are the MaxTempand that have occurred unusually the rest of the days so that they have been processed to not affect theresults of the analysis

Before *After*

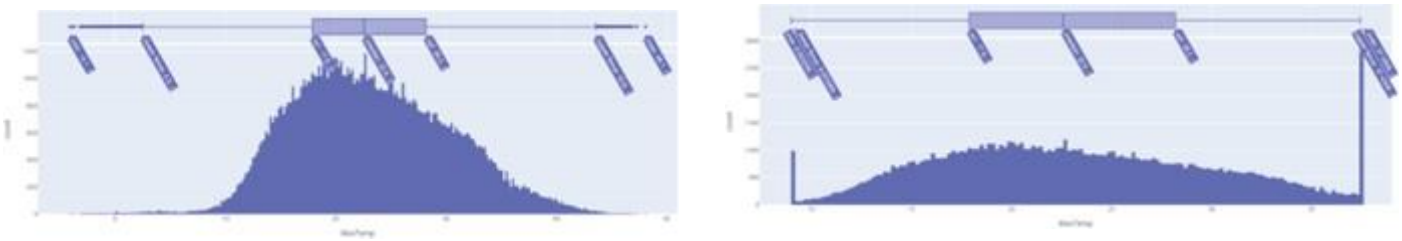


Figure 9: Comparison of *MaxTemp* before and after processing (htt1) (htt2)

Also with MinTemp :

Before *After*

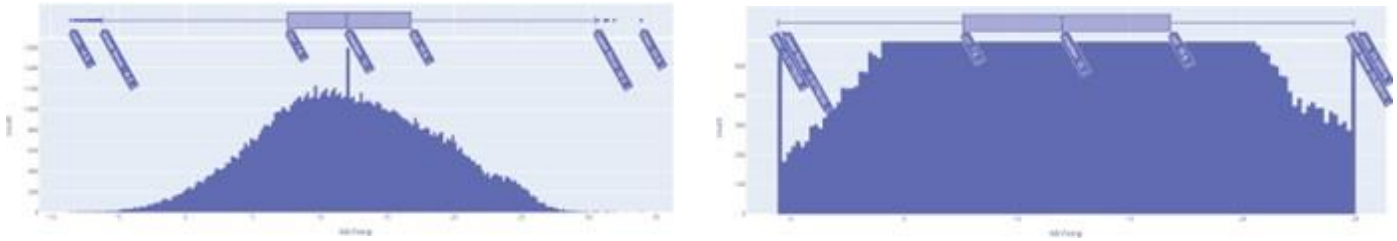


Figure 10: Comparison of *MinTemp* before and after processing (htt1) (htt2)

We also found in wind speed outlier values as in figure :

Before *After*

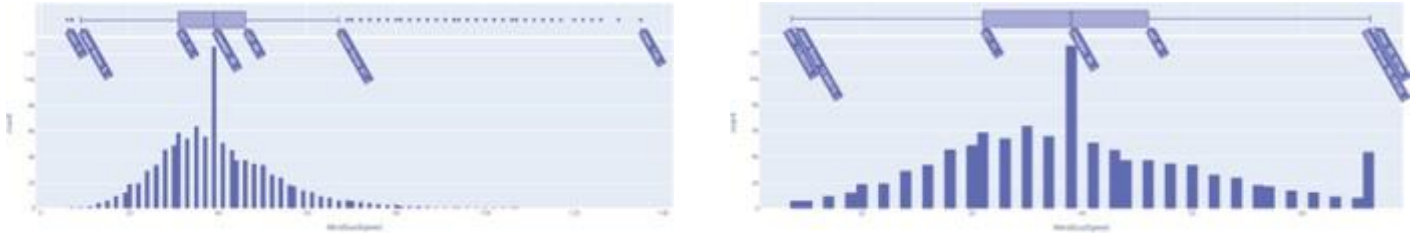


Figure 11: Comparison of *WindGustSpeed* before and after processing (htt1) (htt2)

We also found in Temp3pm outlier values as in figure :

Before *After*

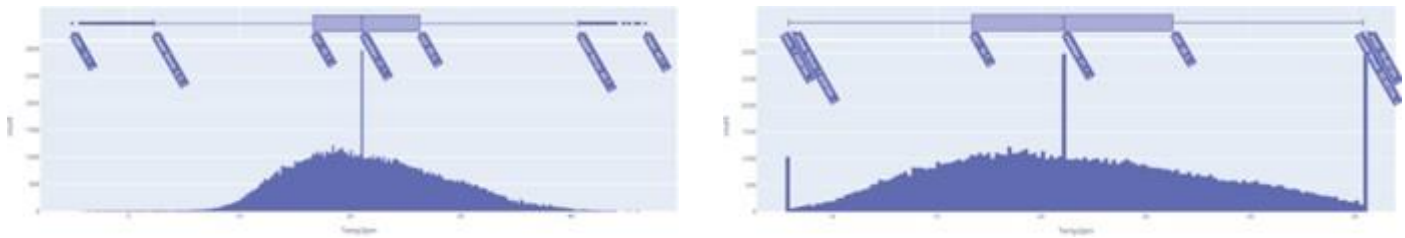


Figure 12: Comparison of *Temp3pm* before and after processing (htt1) (htt2)

We also found in Temp9am outlier values as in figure :

Before *After*

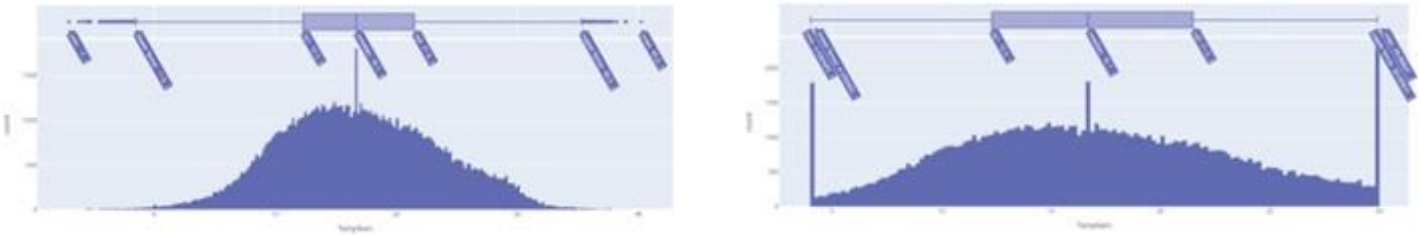


Figure 13: Comparison of *Temp9am* before and after processing (htt1) (htt2)

We also found in Pressure3pm outlier values as in figure :

Before *After*

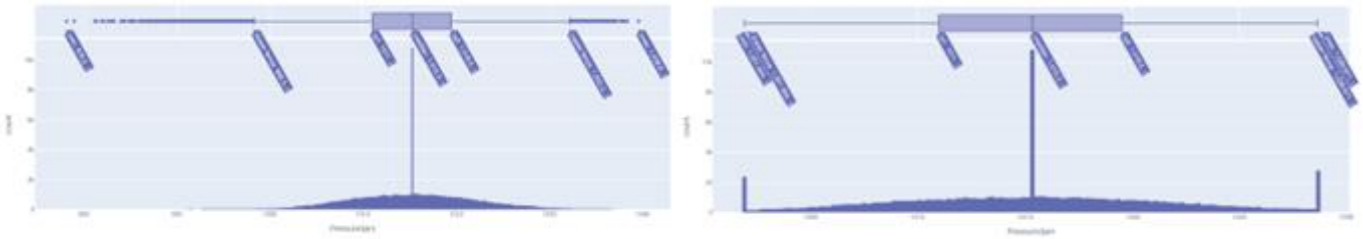


Figure 14: Comparison of *Pressure3pm* before and after processing (htt1) (htt2)

We also found in Pressure9am outlier values as in figure

:Before *After*

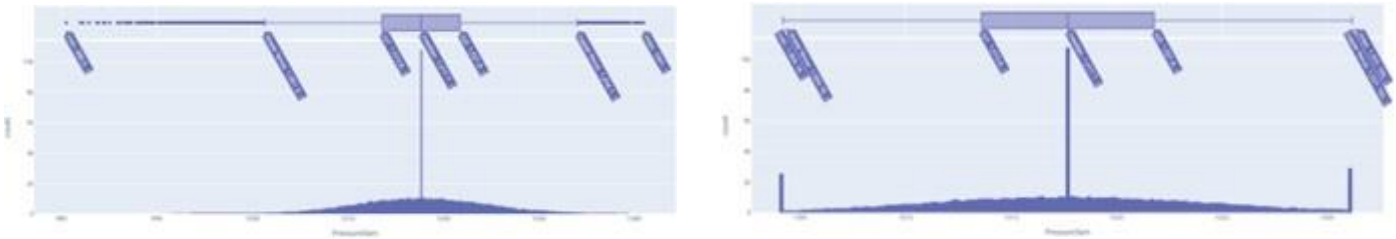


Figure 15: Comparison of *Pressure9apm* before and after processing (htt1) (htt2)

We also found in Pressure3pm outlier values as in figure :

Before *After*

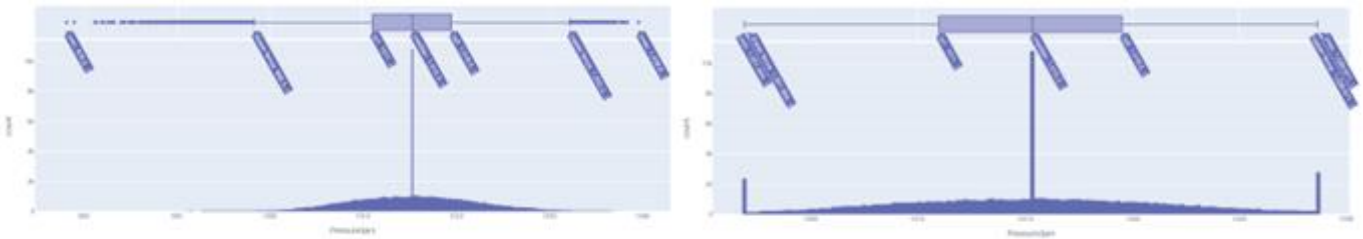


Figure 17: Comparison of *Pressure3pm* before and after processing (htt1) (htt2)

We also found in Humidity9am outlier values as in figure

Before *After*

Figure 16: Comparison of *Humidity9am* before and after processing (htt1) (htt1)

We also found in WindSpeed3pm outlier values as in figure :

Before *After*

Figure 18: Comparison of *WindSpeed3pm* before and after processing (htt1) (htt2)

We also found in WindSpeed9am outlier values as in figure :

Before *After*

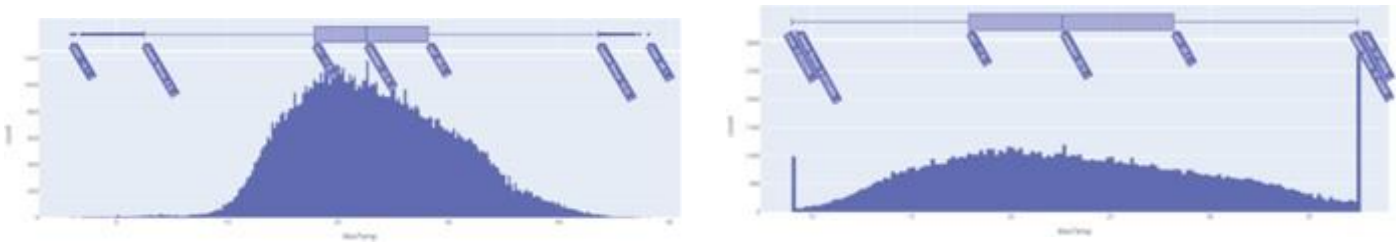


Figure 19: Comparison of *WindSpeed9am* before and after processing (htt1) (htt2)

After examining the data set again, we find that the data set is ready to start the analysis, as the increase in attributes, null values, error values, and outlier values and some data have been converted into numeri- cal values to help us in the analysis stage.

ISSN No:-2456-2165

CHAPTER FOUR
EXPLORATORY DATA ANALYSIS

After making sure that the data and its quality, we will begin to delve into the understanding of the dataand the relationships between them through several methods, including the correlation matrix that is simply a table which displays the correlation coefficients for different variables. The matrix depicts thecorrelation between all the possible pairs of values in a table. It is a powerful tool to summarize a largedataset and to identify and visualize patterns in the given data, as shown in the figure.

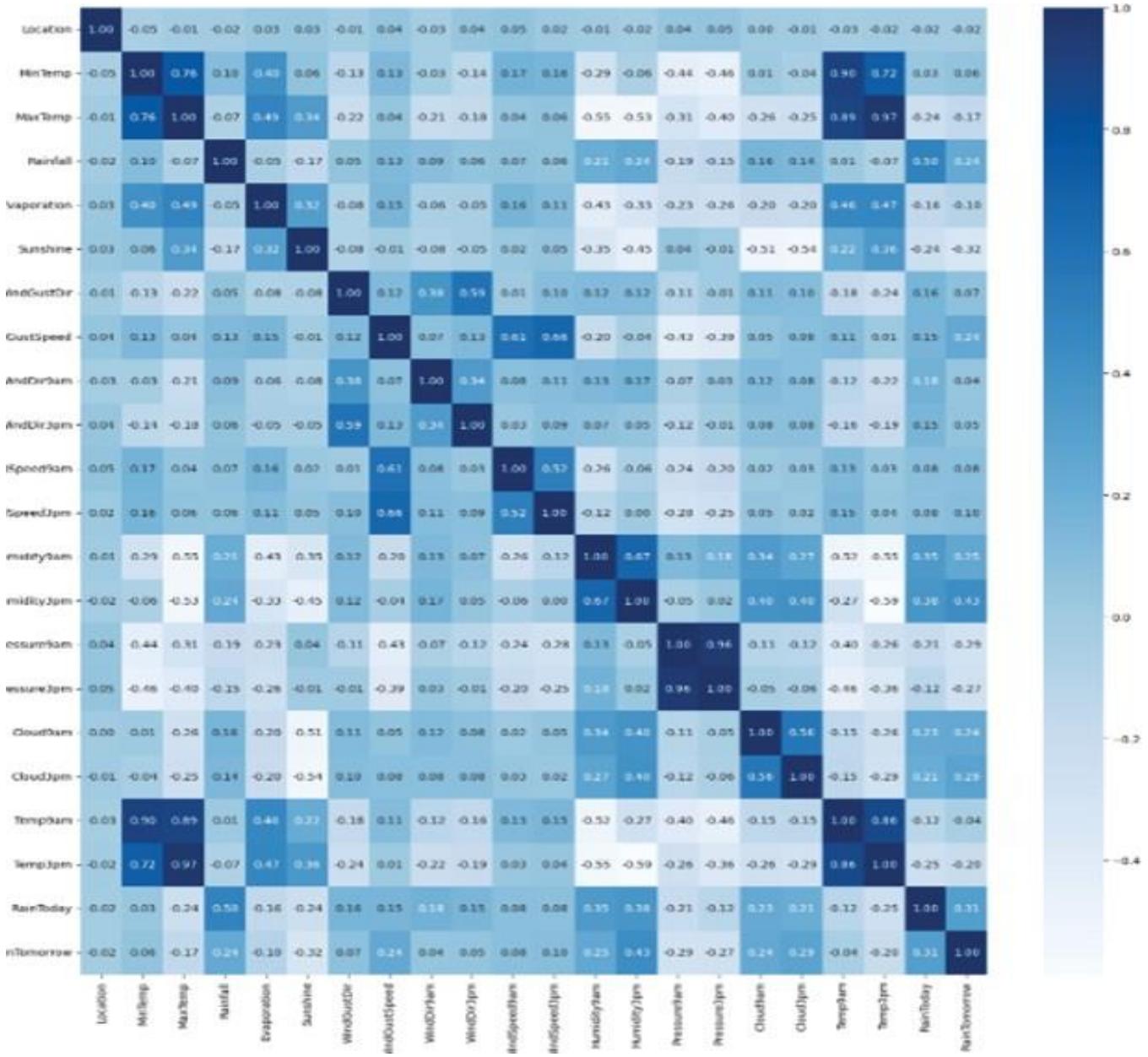


Figure 20: Correlation coefficients (htt1) (htt2)

ISSN No:-2456-2165

We find that there is a relationship between *RainTomorrow* and between the variables as follows: :

- Humidity3pm as its effect is considered the strongest in predicting *RainTomorrow* .
- *RainToday* & *Cloud3pm* come second in forecasting .
- *Humidity9am* & *Cloud9am* & *Rainfall* & WindGustSpeed are among the least influential factorscompared to the previous ones .
- We also find that *SunShine* has an inverse relationship with RainTomorrow as well as *Pressure9am* & *Pressure3pm.*

All key variables were also taken and their effect studied with *RainTomorrow.* For example:

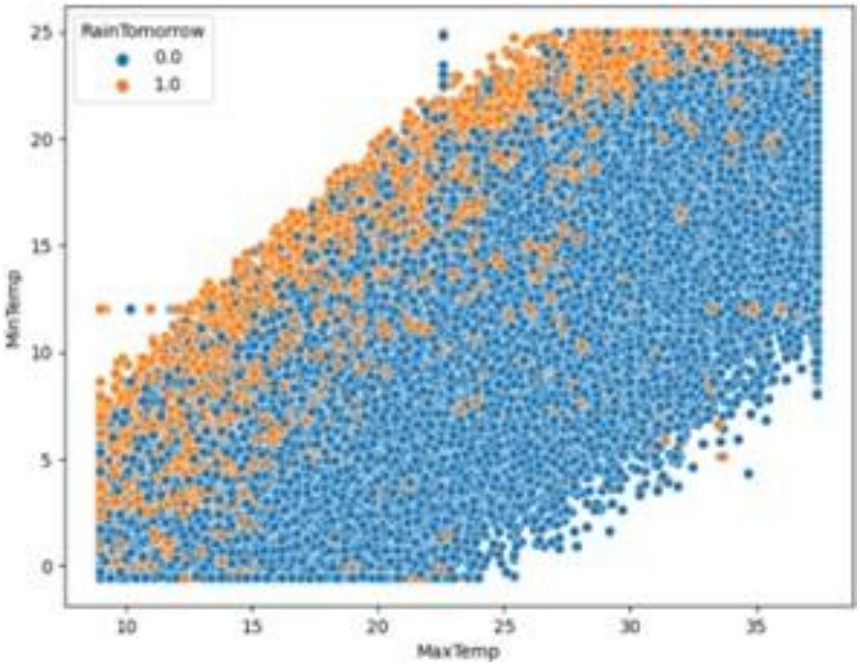


Figure 21: The relationship between *MaxTemp* and *MinTemp* and *RainTomorrow* (htt1) (htt2)

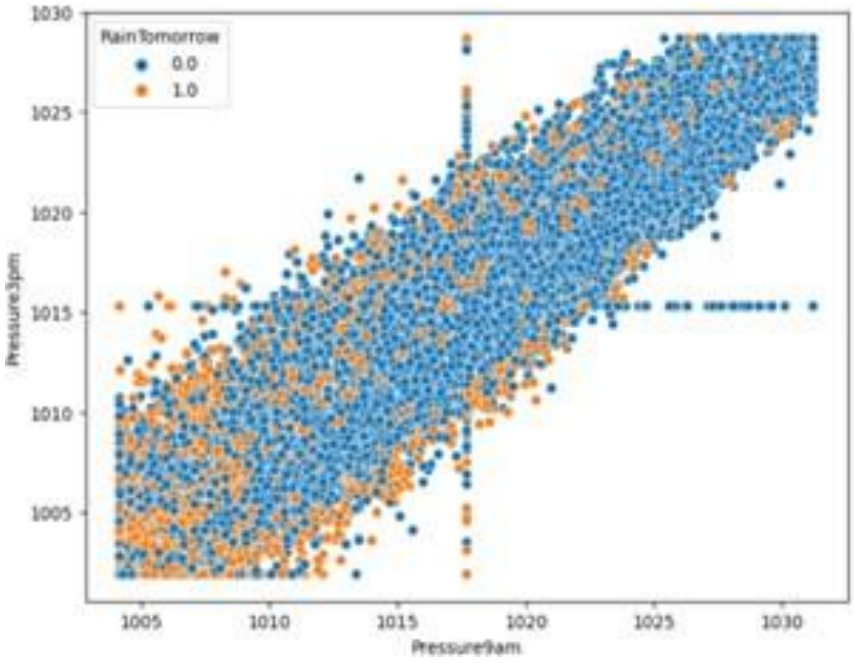


Figure 22: The relationship between *Pressure9am* and pressure3pm and *RainTomorrow* (htt1) (htt2)

ISSN No:-2456-2165

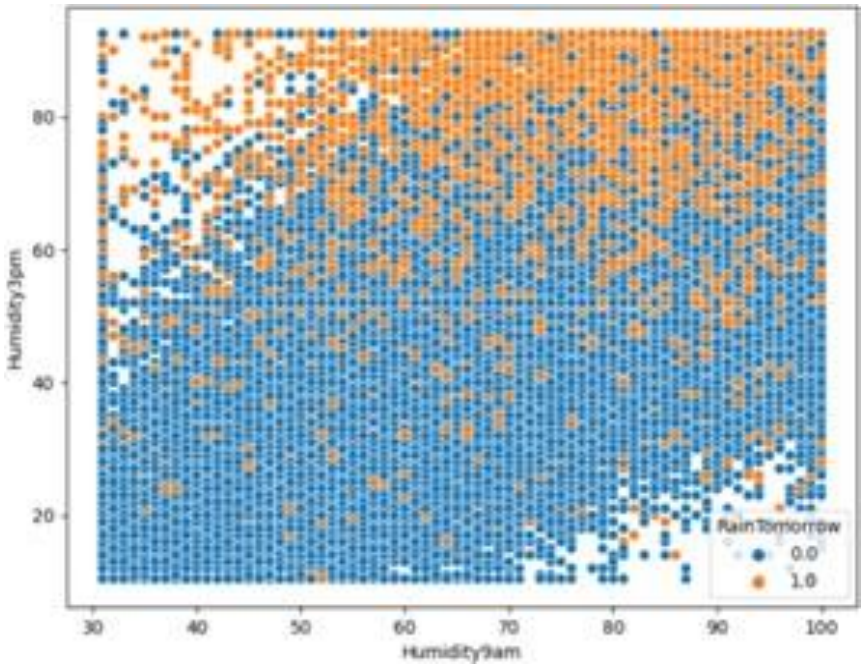


Figure 23: The relationship between *Humidity9am* and Humidity3pm and *RainTomorrow* (htt1) (htt2)

The figure shows the relationship of Humidity with RainTomorrow

ISSN No:-2456-2165

CHAPTER FIVE
PROCESSING

After we get to the stage that we understand the dataset, cleaned it up, checked its quality, and understandthe features and relationships between them and for *RainTomorrow* in particular, we'll start the process of processing, creating patterns, testing and training the dataset on it using the decision tree that we'll use tomodel all possibilities and demonstrate Results.

In order to make the decision tree scientifically to give accurate results, this was done by knowing Entropy
.

```
entropy of Humidity3pm= 6.351025995628162
entropy of Humidity9am= 6.180343213805931
entropy of Cloud3pm= 2.589448017428498
entropy of Cloud9am= 2.574991593923537
entropy of RainToday= 0.7622139573615083
entropy of Pressure3pm= 4.778579609965022
entropy of Pressure9am= 4.78981286224143
entropy of WindSpeed3pm= 4.180834434584423
entropy of WindDir3pm= 3.9773945102064108
entropy of WindDir9am= 3.927349249045117
entropy of WindDir3pm= 3.9773945102064108
entropy of WindGustSpeed= 4.758014375278549
entropy of WindGustDir= 3.9542742795556864
entropy of Sunshine= 2.8959812854963474
entropy of Evaporation= 2.9158532867044857
entropy of Rainfall= 1.7842693536416923
entropy of Location= 5.106169531106806
```

Figure 24: entropy value for attributes (htt1) (htt2)

This shows that the root of the tree will be Humidity3pm

In light of this, a decision tree was built, as shown in the following figure:

Figure 25: Decision Tree (htt1) (htt2) (htt3)

It is considered very very large, so it has been optimization to obtain the most powerful elements affectingthe prediction of tomorrow's rain

ISSN No:-2456-2165

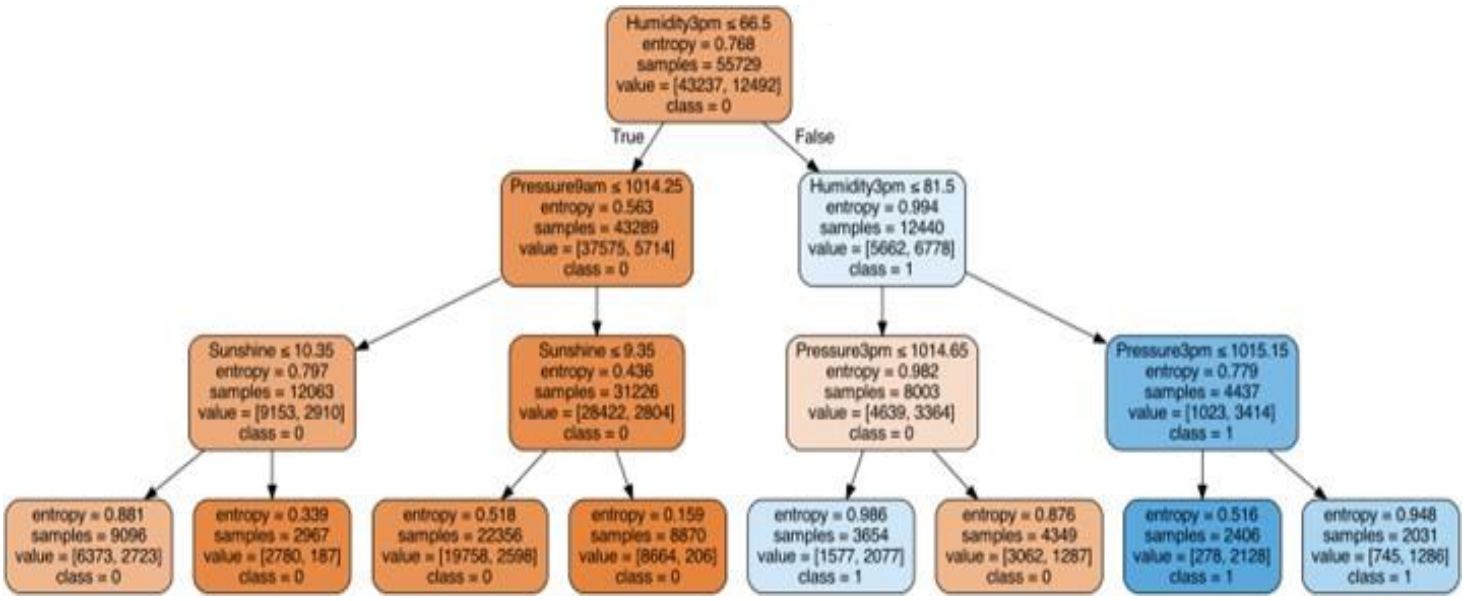


Figure 26: Decision Tree optimization (htt3)

ISSN No:-2456-2165

CHAPTER SIX
TRAINING AND TESTING:

A random forest algorithm was chosen to train the data set on the model, which gave results withan accuracy of :

```
RandomForestClassifierModel Train Score is :  0.9107825369197365
RandomForestClassifierModel Test Score is :  0.853943874255365
```

Figure 27: Accuracy of testing and training the model (htt1) (htt2) (htt3)

KNN algorithm was chosen to train the data set on the model, which gave results with an accuracy of :

```
KNNClassifierModel Train Score is :  0.8738179403901021
KNNClassifierModel Test Score is :  0.8439675590325127
```

Figure 28: Accuracy of KNN (htt1) (htt2) (htt3)

After we prepared the model, trained it, and reached satisfactory accuracy results in the test, we relied onit. We entered values to show the efficiency of the model. We took two samples:
The first is with the same values as one of the rows and shows the result of the prediction similar to theoriginal data

```
Enter the value of Location
Location value=47.0
Enter the value of MinTemp
MinTemp value=9.7
Enter the value of MaxTemp
MaxTemp value=14.1
Enter the value of Rainfall
Rainfall value=0.2
Enter the value of Evaporation
Evaporation value=4.8
Enter the value of Sunshine
Sunshine value=8.4
Enter the value of WindGustDir
WindGustDir value=11.0
Enter the value of WindGustSpeed
WindGustSpeed value=52.0
Enter the value of WindDir9am
WindDir9am value=12.0
Enter the value of WindDir3pm
WindDir3pm value=11.0
Enter the value of WindSpeed9am
WindSpeed9am value=24.0
Enter the value of WindSpeed3pm
WindSpeed3pm value=19.0
Enter the value of Humidity9am
Humidity9am value=71.0
Enter the value of Humidity3pm
Humidity3pm value=84.0
Enter the value of Pressure9am
Pressure9am value=1031.1920554503115
Enter the value of Pressure3pm
Pressure3pm value=1028.6709342266977
Enter the value of Cloud9am
Cloud9am value=8.0
Enter the value of Cloud3pm
Cloud3pm value=8.0
Enter the value of Temp9am
Temp9am value=10.9
Enter the value of Temp3pm
Temp3pm value=11.7
Enter the value of RainToday
RainToday value=0.0
```

Figure 29: Introduction of model values (htt3)

And the result came out.

```
In [290]: y_pred_test = GBCModel.predict(inputtestdf)
          y_pred_test

Out[290]: array([1.])

In [291]: if y_pred_test[0]==1:
              print("There will be a high chance of rain ")
          else:
              print("There will be a slight chance of rain")

          There will be a high chance of rain
```

Figure 30: the result to appear (htt3)

The second, we entered random values within the range of each column and the result of the prediction wascorrect

```
Enter the value of Location
Location value=47.0
Enter the value of MinTemp
MinTemp value=8.7
Enter the value of MaxTemp
MaxTemp value=15.7
Enter the value of Rainfall
Rainfall value=0.0
Enter the value of Evaporation
Evaporation value=4.8
Enter the value of Sunshine
Sunshine value=8.4
Enter the value of WindGustDir
WindGustDir value=14.0
Enter the value of WindGustSpeed
WindGustSpeed value=52.0
Enter the value of WindDir9am
WindDir9am value=14.0
Enter the value of WindDir3pm
WindDir3pm value=13.0
Enter the value of WindSpeed9am
WindSpeed9am value=19.0
Enter the value of WindSpeed3pm
WindSpeed3pm value=19.0
Enter the value of Humidity9am
Humidity9am value=69.0
Enter the value of Humidity3pm
Humidity3pm value=42.0
Enter the value of Pressure9am
Pressure9am value=1015.4
Enter the value of Pressure3pm
Pressure3pm value=1012.7
Enter the value of Cloud9am
Cloud9am value=0.0
Enter the value of Cloud3pm
Cloud3pm value=0.0
Enter the value of Temp9am
Temp9am value=9.9
Enter the value of Temp3pm
Temp3pm value=14.2
Enter the value of RainToday
RainToday value=0.0
```

Figure 31: Introduction of model values (htt3)

And the result came out .

```
In [294]: y_pred_test = GBCModel.predict(inputtestdf)
          y_pred_test

Out[294]: array([0.])

In [295]: if y_pred_test[0]==1:
              print("There will be a high chance of rain ")
          else:
              print("There will be a slight chance of rain")

          There will be a slight chance of rain
```

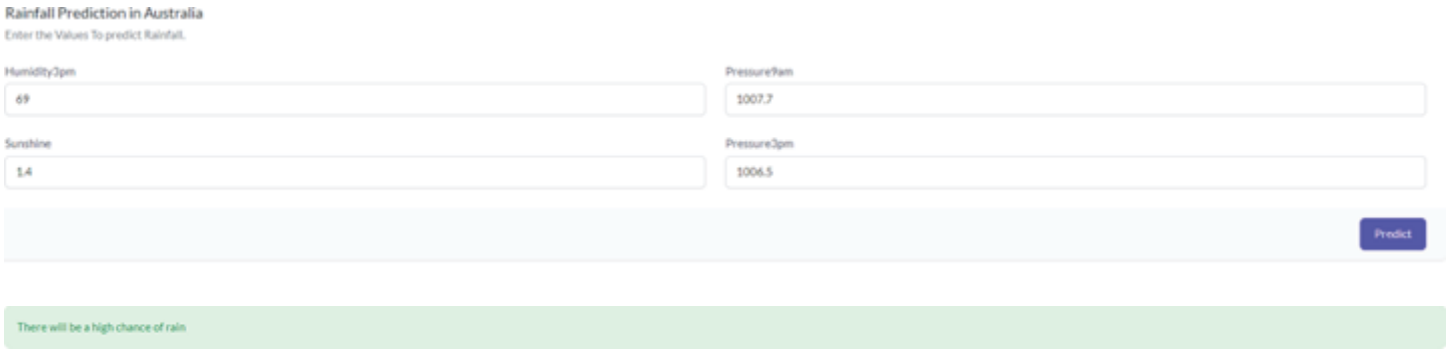
Figure 32: the result to appear (htt3)

ISSN No:-2456-2165

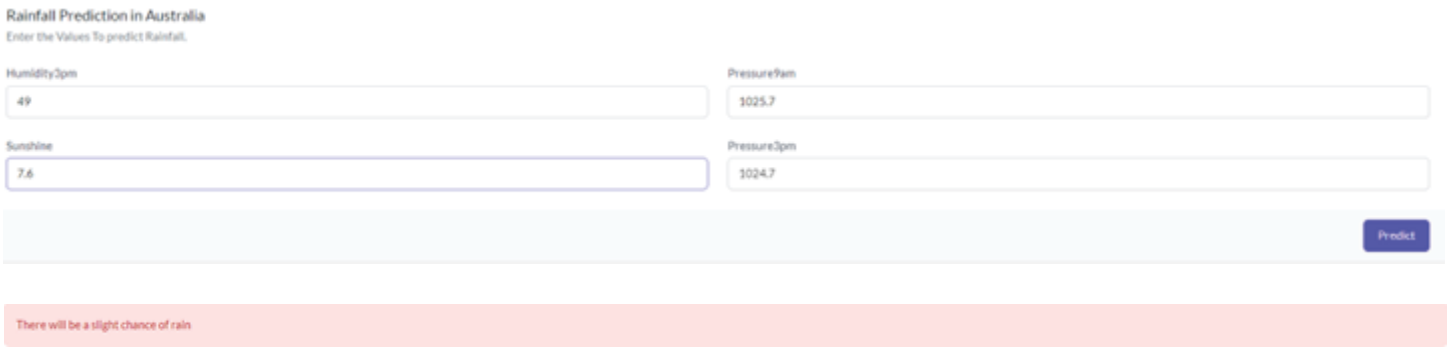
Here are the original values of the dataset and show that there is *RainTomorrow.* (htt4)

	MaxTemp	Rainfall	Evaporati	Sunshine	WindGust	WindGust	WindDir9am	WindDir3	WindSpee	WindSpee	Humidity9am	Humidity3pm	Pressure9am	Pressure3pm	Cloud9am	Cloud3pm	Temp9am	Temp3pm	RainToda	RainTomorrow
4541	16.7	0.4	3.8	0.1	NNE	37	NE	ESE	13	11	58	96	1014.6	1010.9	7	8	14.7	12.6	No	1
4542	20.4	9.2	0.4	2.6	WSW	50	N	NNE	7	11	99	84	1008.1	1002.4	8	8	14.4	17.3	Yes	1
4543	18.6	0	2.4	1.5	NNE	31	NE	NE	11	17	84	57	1017.5	1012.2	7	7	14.3	17.6	No	1
4544	15.3	3.6	2.6	1.4	NNE	48	NW	W	15	20	85	69	1007.7	1006.5	8	7	12.2	14.5	Yes	1

We introduced the same values to test the model on which the data was studied, trained and tested andshow the same result. (htt4)



This is another predict test



ISSN No:-2456-2165

CHAPTER SEVEN
FINDINGS AND RECOMMENDATIONS

Based on the foregoing, and after processing, testing and training the data set on the possible patterns, wereach the following results based on humidity, which according to the entory are considered to be the strongest influencer, as follows:

- The *Humidity3pm* variable is the strongest influence on the forecasting process for tomorrow's rain, which makes up 43% of the rest of the variables combined. Therefore, if the *Humidity3pm* exceeds 66.5%,then this is the first indicator of the possibility of *RainTomorrow*, and if the *Humidity3pm* in the evening exceeds 81.5%, compared to the fact that the *Pressure3pm* decreases or exceeds 1015.15, the prediction of RainTomorrow is 100%. (htt4) (htt5)
- *Humidity3pm* exceeded 81.5% and *Pressure3pm* dropped below 1014.65, so the possibility of *RainTomorrow* is strong.
- But if the *Humidity3pm* is less than 61.5%, the *Pressure9am* is greater than 1014.25, and the *Sunshine* exceeds 9.35, then it is impossible for there to be *RainTomorrow*.

Based on the foregoing and according to what has been reached, we offerthe following recommendations:

The agricultural sector, through the Sunshine and the lack of Humidity3pm, we recommend reaping thefruits at any time and we recommend irrigating the crops because there is no possibility of *RainTomorrow*.But if the *Humidity3pm* is more than 81.5%, we recommend picking the fruits before nightfall and cover- ing the young plants that are affected by the amount of rain and keeping the grains in warehouses for the possibility of *RainTomorrow* high

Shipping and transportation sector, They continue their routine work normally unless the Humidity3pm ismore than 81.5%, so they must take precautions when stopping at night and cover the goods because *RainTomorrow* will be

Education sector: Whenever the *Sunshine* during the day at a degree of 9.35, then tomorrow you can go to schools as in the rest of the days, either by walking, cycling or other things.

But if the Humidity3pm in the evening exceeds 81.5% and the Pressure3pm is less than 1015.15, werecommend going to school early and using public or private transportation.

ISSN No:-2456-2165

CHAPTER EIGHT
CONCLUSION

The project illustrates the use of machine learning-based predictive analysis to forecast the likelihood of rain for the next day using the provided data. By taking into account historical data, the suggested system will assist in forecasting the weather's future tendency. This model makes it possible to analyse the likelihood of RainTomorrow so that everyone can make arrangements without being confused. In order to create a RainTomorrow prediction classifier to forecast future weather conditions based on the other parameters, we have suggested an experimental methodology. The information was gathered using atmospheric parameters and weather records from various Australian cities. Kaggle has had access to this dataset. All the findings and recommendations have been clearly illustrated.

REFERENCES

[1]. http://www.bom.gov.au/climate/data/
[2]. http://www.bom.gov.au/climate/dwo/IDCJDW0000.shtml
[3]. http://www.bom.gov.au/climate/dwo/
[4]. https://www.kaggle.com/datasets/arunavakrchakraborty/australia-weather-data
[5]. https://www.kaggle.com/code/rajashreerd/australia-rainfall
[6]. https://towardsdatascience.com/introduction-to-machine-learning-algorithms-linear-regression-14c4e325882a
[7]. https://www.researchgate.net/publication/336797575_Weather_Forecasting
[8]. http://assiilah.com/predict
[9]. https://www.researchgate.net/publication/336797575_Weather_Forecasting