

```
In [ ]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import folium
import imageio
from tqdm import tqdm_notebook
from folium.plugins import MarkerCluster
import imageio
import statsmodels.api as sm
import scipy
from itertools import product
import seaborn as sns
from numpy.random import seed
from numpy.random import randint
from numpy import mean
import matplotlib.pyplot as plt
from scipy.stats import shapiro
import collections
from sklearn import preprocessing

from sklearn.model_selection import train_test_split
import tensorflow as tf
from keras.models import Sequential
from keras.layers import Conv2D ,Dense, Flatten , MaxPool2D , Input
```

```
In [ ]: def draw_corr(data,figsize = (12,12)):
    corr =data.corr()
    fig = plt.figure(figsize = figsize)
    # plt.matshow(correlations)
    plt.matshow(corr, cmap= 'RdBu', fignum=fig.number)
    plt.xticks(range(len(corr.columns)), corr.columns, rotation = 'vertical');
    plt.yticks(range(len(corr.columns)), corr.columns);
    plt.grid(False)
    plt.colorbar()

def minmaxscalar_pd(df):
    x = df.values #returns a numpy array
    min_max_scaler = preprocessing.MinMaxScaler()
    x_scaled = min_max_scaler.fit_transform(x)
    df = pd.DataFrame(x_scaled,columns=df.columns)
    return df
```

```
In [ ]: data= pd.read_csv('weather1.csv')
```

```
In [ ]: data.head()
```

Out[ ]:	MinTemp	MaxTemp	Rainfall	Evaporation	Sunshine	WindGustDir	WindGustSpeed	WindDir9am
0	8.0	24.3	0.0	3.4	6.3	NW	30.0	SW
1	14.0	26.9	3.6	4.4	9.7	ENE	39.0	E
2	13.7	23.4	3.6	5.8	3.3	NW	85.0	N
3	13.3	15.5	39.8	7.2	9.1	NW	54.0	WNW
4	7.6	16.1	2.8	5.6	10.6	SSE	50.0	SSE

5 rows × 22 columns

In [ ]: `data.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 366 entries, 0 to 365
Data columns (total 22 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   MinTemp          366 non-null    float64
 1   MaxTemp          366 non-null    float64
 2   Rainfall         366 non-null    float64
 3   Evaporation     366 non-null    float64
 4   Sunshine         363 non-null    float64
 5   WindGustDir      363 non-null    object  
 6   WindGustSpeed   364 non-null    float64
 7   WindDir9am      335 non-null    object  
 8   WindDir3pm       365 non-null    object  
 9   WindSpeed9am    359 non-null    float64
 10  WindSpeed3pm    366 non-null    int64  
 11  Humidity9am     366 non-null    int64  
 12  Humidity3pm     366 non-null    int64  
 13  Pressure9am     366 non-null    float64
 14  Pressure3pm     366 non-null    float64
 15  Cloud9am        366 non-null    int64  
 16  Cloud3pm        366 non-null    int64  
 17  Temp9am         366 non-null    float64
 18  Temp3pm         366 non-null    float64
 19  RainToday        366 non-null    object  
 20  RISK_MM          366 non-null    float64
 21  RainTomorrow    366 non-null    object  
dtypes: float64(12), int64(5), object(5)
memory usage: 63.0+ KB
```

In [ ]: `data.describe()`

Out[ ]:	MinTemp	MaxTemp	Rainfall	Evaporation	Sunshine	WindGustSpeed	WindSpeed9am
<b>count</b>	366.000000	366.000000	366.000000	366.000000	363.000000	364.000000	359.000000
<b>mean</b>	7.265574	20.550273	1.428415	4.521858	7.909366	39.840659	9.651811
<b>std</b>	6.025800	6.690516	4.225800	2.669383	3.481517	13.059807	7.951929
<b>min</b>	-5.300000	7.600000	0.000000	0.200000	0.000000	13.000000	0.000000
<b>25%</b>	2.300000	15.025000	0.000000	2.200000	5.950000	31.000000	6.000000
<b>50%</b>	7.450000	19.650000	0.000000	4.200000	8.600000	39.000000	7.000000
<b>75%</b>	12.500000	25.500000	0.200000	6.400000	10.500000	46.000000	13.000000
<b>max</b>	20.900000	35.800000	39.800000	13.800000	13.600000	98.000000	41.000000

◀ ▶

In [ ]: `data2 = data.copy()`

In [ ]: `data2['RainToday'] = data2['RainToday'].replace(['No', 'Yes'], [0,1])`

In [ ]: `data2['RainToday']`

Out[ ]:

0	0
1	1
2	1
3	1
4	1
..	
361	0
362	0
363	0
364	0
365	0

Name: RainToday, Length: 366, dtype: int64

In [ ]: `data2.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 366 entries, 0 to 365
Data columns (total 22 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   MinTemp          366 non-null    float64
 1   MaxTemp          366 non-null    float64
 2   Rainfall         366 non-null    float64
 3   Evaporation     366 non-null    float64
 4   Sunshine         363 non-null    float64
 5   WindGustDir      363 non-null    object  
 6   WindGustSpeed    364 non-null    float64
 7   WindDir9am       335 non-null    object  
 8   WindDir3pm       365 non-null    object  
 9   WindSpeed9am     359 non-null    float64
 10  WindSpeed3pm     366 non-null    int64  
 11  Humidity9am      366 non-null    int64  
 12  Humidity3pm      366 non-null    int64  
 13  Pressure9am      366 non-null    float64
 14  Pressure3pm      366 non-null    float64
 15  Cloud9am         366 non-null    int64  
 16  Cloud3pm         366 non-null    int64  
 17  Temp9am          366 non-null    float64
 18  Temp3pm          366 non-null    float64
 19  RainToday         366 non-null    int64  
 20  RISK_MM          366 non-null    float64
 21  RainTomorrow     366 non-null    object  
dtypes: float64(12), int64(6), object(4)
memory usage: 63.0+ KB
```

```
In [ ]: i = 'RainTomorrow'
data2[i] = data2[i].replace(['No', 'Yes'], [0, 1])
```

```
In [ ]: data3 = data2.select_dtypes(include=np.number)
```

```
In [ ]: data3.head()
```

	MinTemp	MaxTemp	Rainfall	Evaporation	Sunshine	WindGustSpeed	WindSpeed9am	WindSpeed3pm
<b>0</b>	8.0	24.3	0.0	3.4	6.3	30.0	6.0	1.0
<b>1</b>	14.0	26.9	3.6	4.4	9.7	39.0	4.0	2.0
<b>2</b>	13.7	23.4	3.6	5.8	3.3	85.0	6.0	2.0
<b>3</b>	13.3	15.5	39.8	7.2	9.1	54.0	30.0	1.0
<b>4</b>	7.6	16.1	2.8	5.6	10.6	50.0	20.0	1.0

```
In [ ]: #discrete random variable
def get_frequencies(values):
    frequencies = {}
    for v in values:
        if v in frequencies:
            frequencies[v] += 1
        else:
            frequencies[v] = 1
    return frequencies
```

```
def get_probabilities(sampledata, freqs):
    probabilities = []
    for k, v in freqs.items():
        probabilities.append(round(v / len(sampledata), 1))
    return probabilities
```

In [ ]: `data["Cloud9am"]`

Out[ ]:

0	7
1	5
2	8
3	2
4	7
	..
361	1
362	0
363	3
364	6
365	1

Name: Cloud9am, Length: 366, dtype: int64

In [ ]:

```
plt.style.use('ggplot')
plt.rcParams['font.family'] = 'sans-serif'
plt.rcParams['font.serif'] = 'Ubuntu'
plt.rcParams['font.monospace'] = 'Ubuntu Mono'
plt.rcParams['font.size'] = 14
plt.rcParams['axes.labelsize'] = 12
plt.rcParams['axes.labelweight'] = 'bold'
plt.rcParams['axes.titlesize'] = 12
plt.rcParams['xtick.labelsize'] = 12
plt.rcParams['ytick.labelsize'] = 12
plt.rcParams['legend.fontsize'] = 12
plt.rcParams['figure.titlesize'] = 12
plt.rcParams['image.cmap'] = 'jet'
plt.rcParams['image.interpolation'] = 'none'
plt.rcParams['figure.figsize'] = (12, 10)
plt.rcParams['axes.grid']=True
plt.rcParams['lines.linewidth'] = 2
plt.rcParams['lines.markersize'] = 8
colors = ['xkcd:pale orange', 'xkcd:sea blue', 'xkcd:pale red', 'xkcd:sage green', 'xkcd:scarlet']
```

```
sample = data["Cloud9am"]
calculated_frequencies = get_frequencies(sample)
print(calculated_frequencies)
calculate_probabilities = get_probabilities(sample, calculated_frequencies)
print("prob", calculate_probabilities)
x_axis = list(set(sample))

plt.bar(x_axis, calculate_probabilities)
plt.legend(["Cloud thickness"])
plt.show()

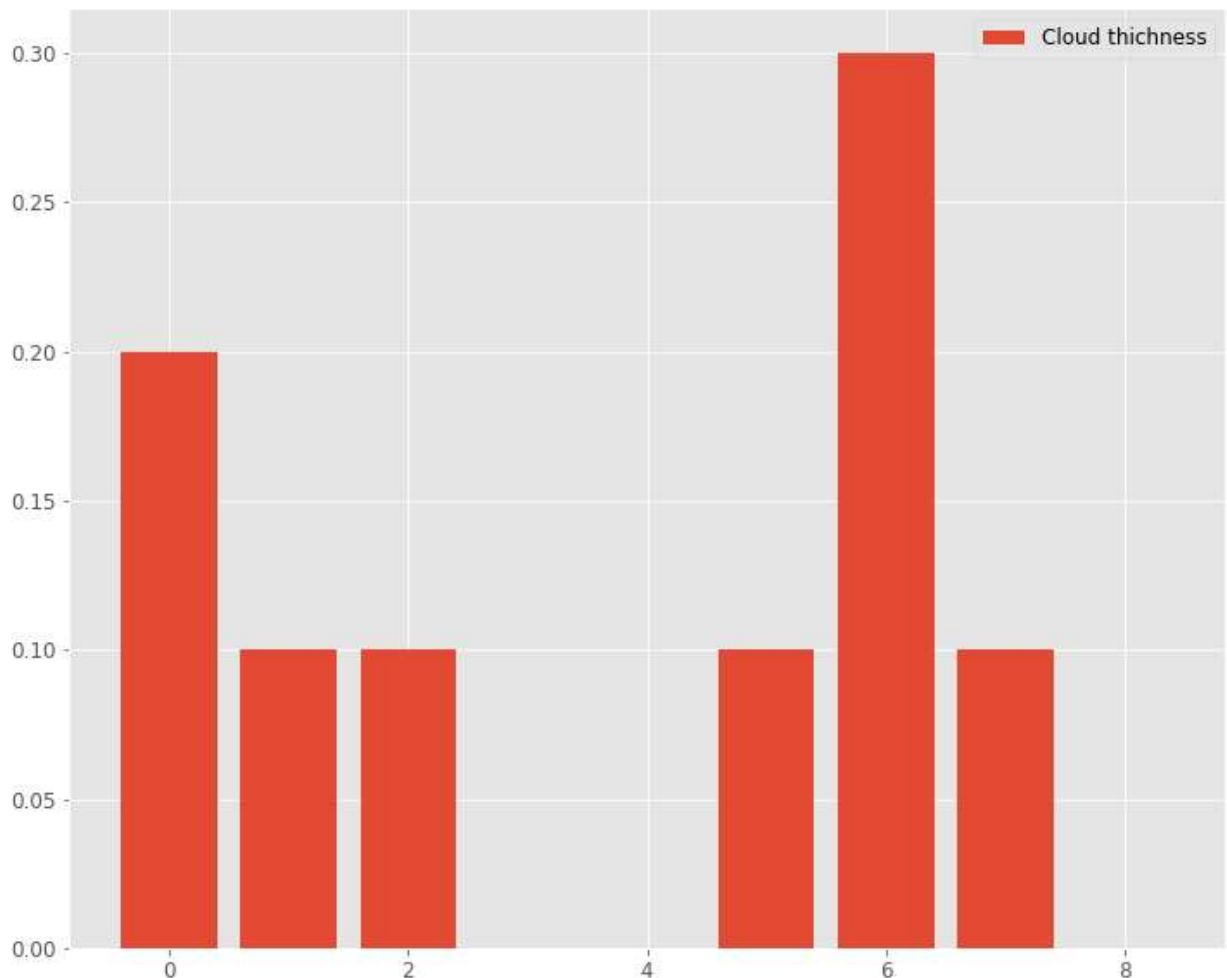
plt.scatter(data['WindSpeed3pm'] , data[ "WindSpeed9am" ] )
plt.title("Wind Speeds 'at 3pm' to 'at 9am' ")
```

```
plt.xlabel("Wind Speeds at 3pm")
plt.ylabel("Wind Speeds at 9am")
plt.show()

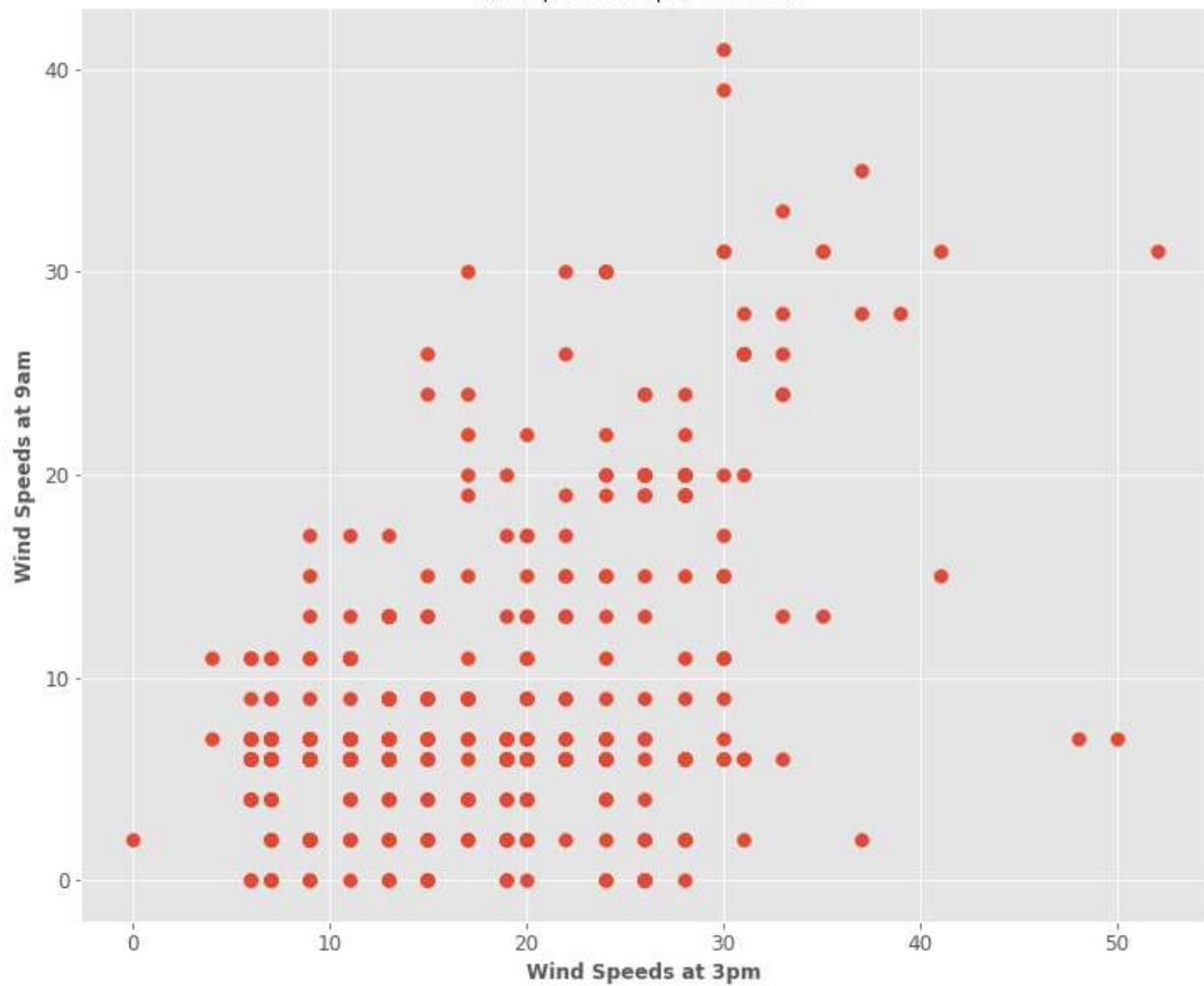
plt.scatter(data['WindSpeed3pm'] ,data[ "Sunshine"] )
plt.title("Wind Speeds at 3pm to Sunshine")
plt.xlabel("Wind Speeds at 3pm")
plt.ylabel("Sunshine")
plt.show()

mylabels= ["Sunny day","No cloud","Very little cloud","-","Cloudy","Very cloudy","Heavy clouds"]
myexplode = [0.15,0,0,0,0,0,0.2,0,0]
plt.pie(calculate_probabilities,labels =mylabels,explode = myexplode, shadow = True)
plt.title("Cloud thickness")
plt.show()
```

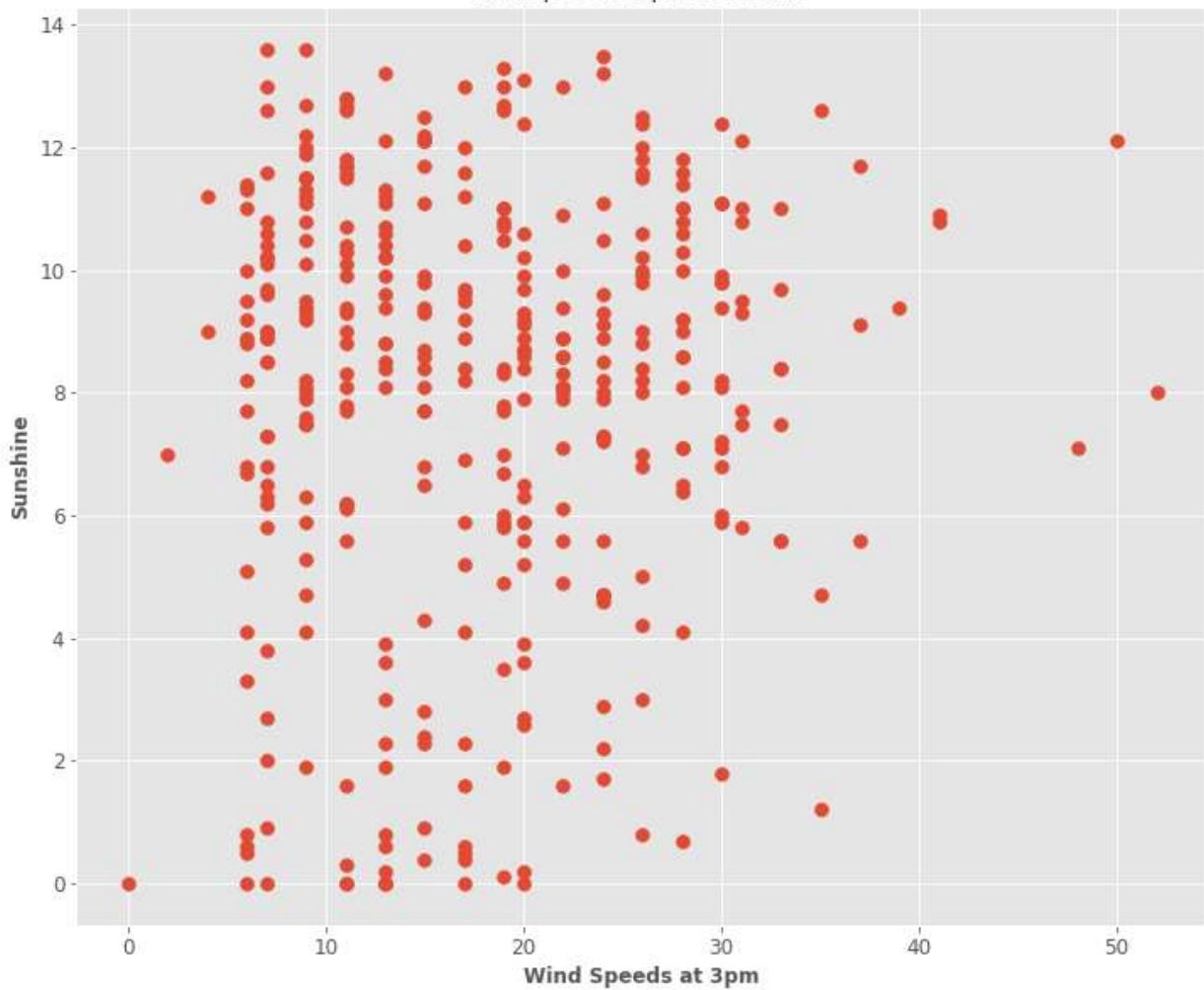
{7: 86, 5: 23, 8: 40, 2: 17, 4: 9, 6: 25, 1: 116, 0: 33, 3: 17}  
prob [0.2, 0.1, 0.1, 0.0, 0.0, 0.1, 0.3, 0.1, 0.0]

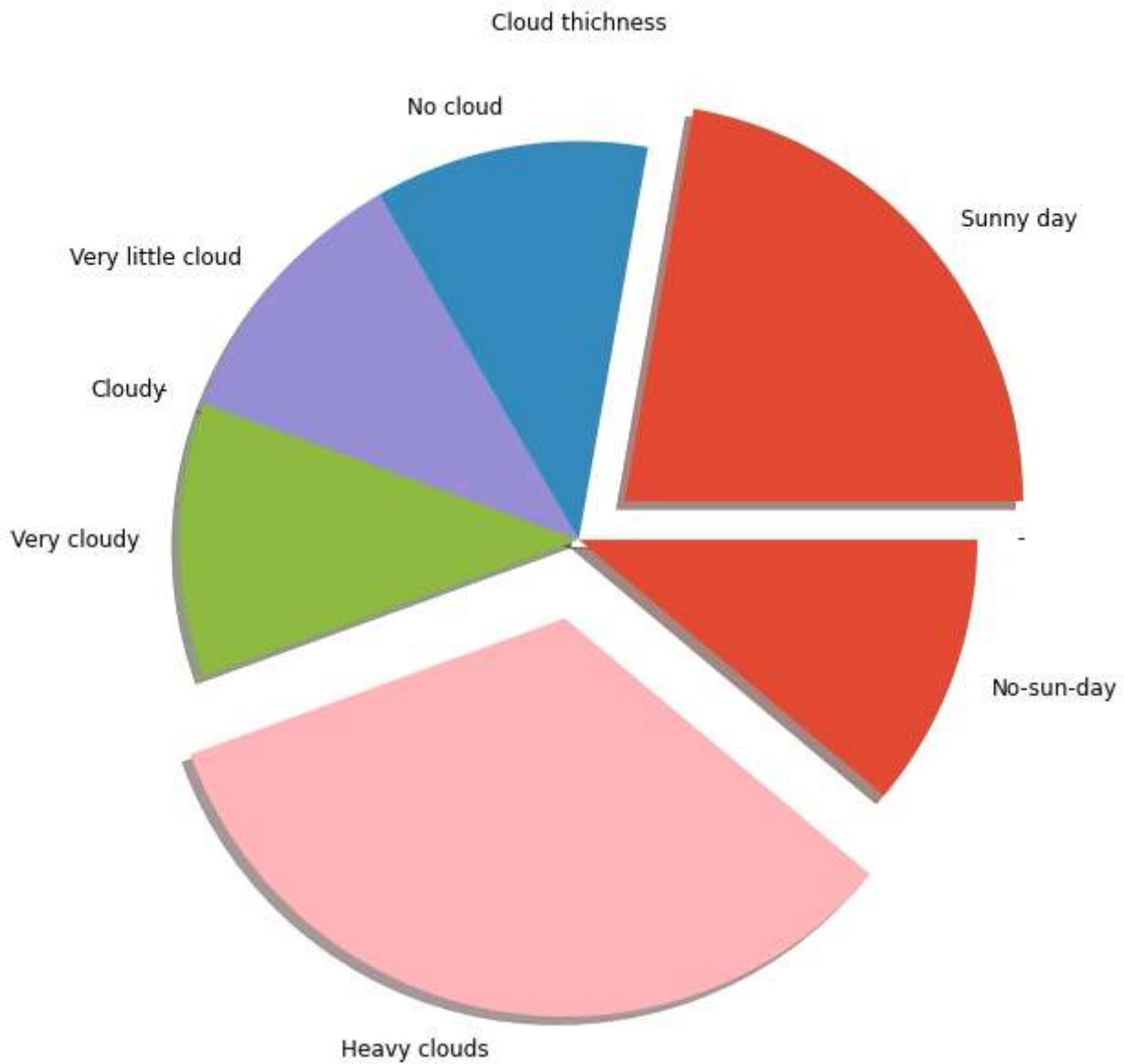


## Wind Speeds 'at 3pm' to 'at 9am'



## Wind Speeds at 3pm to Sunshine





```
In [ ]: #Central Limit theorem
```

```
# seed the random number generator, so that the experiment is #replicable
seed(1)
# generate a sample of men's weights
Rainfall = data['Rainfall']
# print(Rainfall)
print('The average probability of rain is {} %'.format(mean(Rainfall)))

plt.hist(Rainfall)
plt.title('Probability of Rain')
plt.show()
stat, p = shapiro(Rainfall)

print('Statistics={}, p={}'.format(stat, p))
alpha = 0.05
if p > alpha:
    print('Sample looks Normal (do not reject H0)')
else:
    print('Sample does not look Normal(rejectH0)')
```

```
# initializing
Temp3pm = data['Temp3pm']

# getting data of the histogram
count, bins_count = np.histogram(Temp3pm, bins=20)

# finding the PDF of the histogram using count values
pdf = count / sum(count)

# using numpy np.cumsum to calculate the CDF
# We can also find using the PDF values by looping and adding
cdf = np.cumsum(pdf)

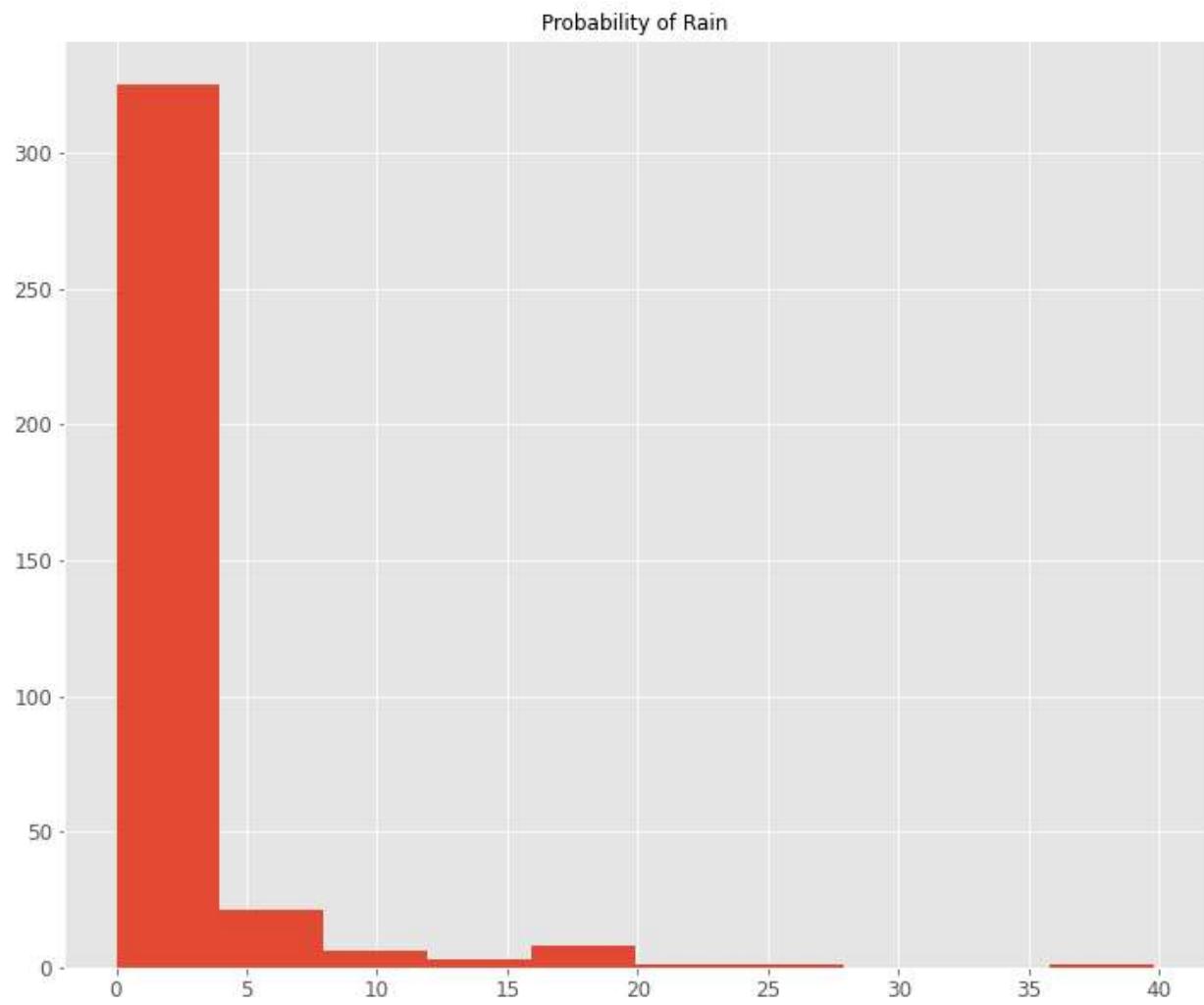
# plotting PDF and CDF
plt.plot(bins_count[1:], pdf, color="red", label="PDF")
plt.plot(bins_count[1:], cdf, label="CDF")
plt.legend()
plt.show()

#NAIVE BAYES
def bayesTheorem(pA, pB, pBA):
    return pA * pBA / pB
#define function for Bayes' theorem
def bayesTheorem(pA, pB, pBA):
    return pA * pBA / pB

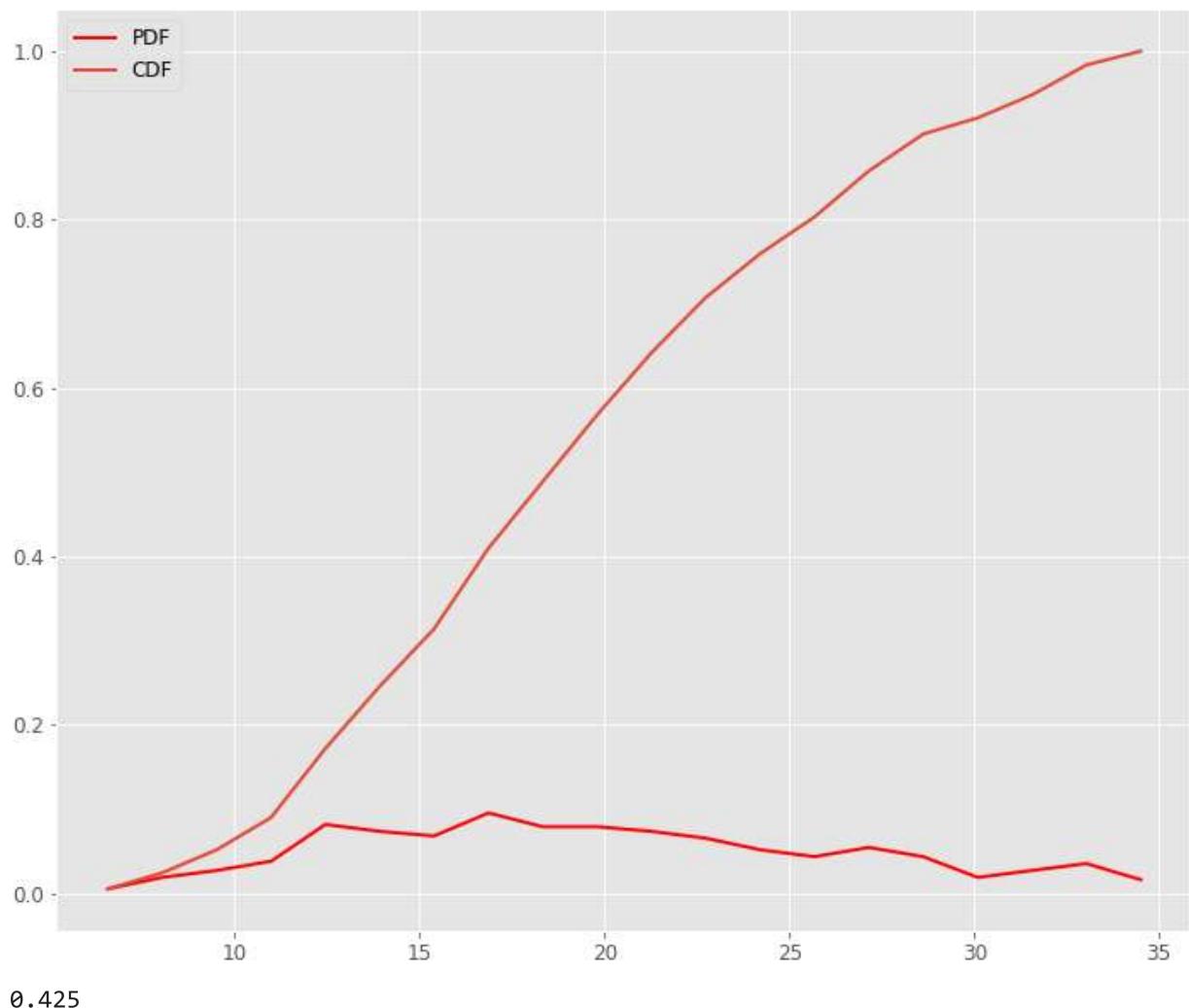
#define probabilities
pRain = 0.2
pCloudy = 0.4
pCloudyRain = 0.85

#use function to calculate conditional probability
print(bayesTheorem(pRain, pCloudy, pCloudyRain))
```

The average probability of rain is 1.4284153005464482 %



Statistics=0.38883382081985474, p=2.9762369019118345e-33  
Sample does not look Normal(rejectH0)



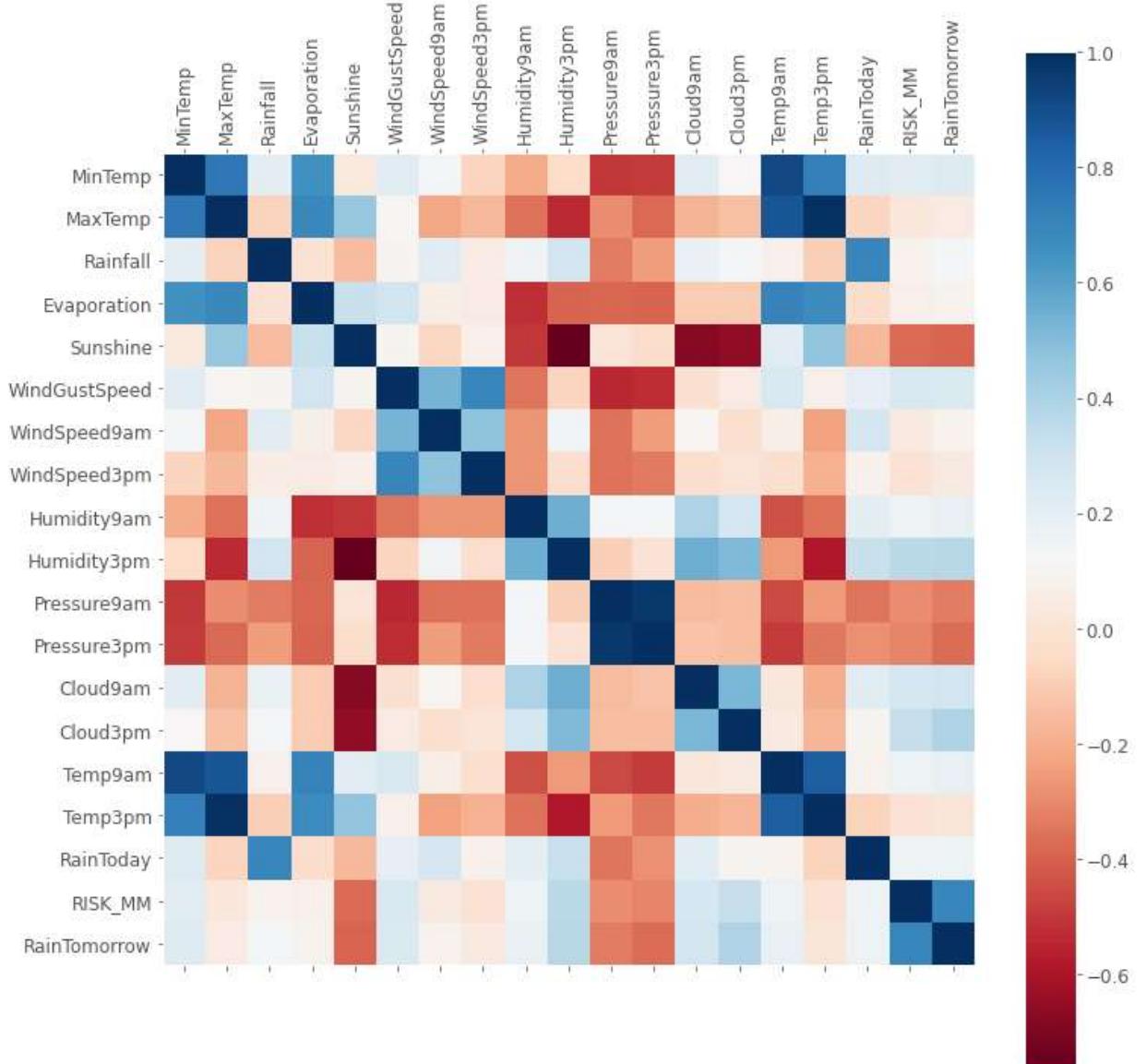
0.425

In [ ]: `data3.head()`

	MinTemp	MaxTemp	Rainfall	Evaporation	Sunshine	WindGustSpeed	WindSpeed9am	WindSpeed
0	8.0	24.3	0.0	3.4	6.3	30.0	6.0	
1	14.0	26.9	3.6	4.4	9.7	39.0	4.0	
2	13.7	23.4	3.6	5.8	3.3	85.0	6.0	
3	13.3	15.5	39.8	7.2	9.1	54.0	30.0	
4	7.6	16.1	2.8	5.6	10.6	50.0	20.0	

In [ ]: `draw_corr(data3)`

C:\Users\Mario\AppData\Local\Temp\ipykernel\_16792\3479777238.py:9: MatplotlibDeprecationWarning: Auto-removal of grids by pcolor() and pcolormesh() is deprecated since 3.5 and will be removed two minor releases later; please call grid(False) first.  
plt.colorbar()



```
In [ ]: data3 = minmaxscalar_pd(data3)
```

```
In [ ]: outputs = data3["RainTomorrow"]
inputs = data3.drop(["RainTomorrow"], axis=1)
```

```
In [ ]: print(inputs.shape,outputs.shape)
inputs.head()
```

```
(366, 18) (366, )
```

	MinTemp	MaxTemp	Rainfall	Evaporation	Sunshine	WindGustSpeed	WindSpeed9am	WindSpe
<b>0</b>	0.507634	0.592199	0.000000	0.235294	0.463235	0.200000	0.146341	0.3
<b>1</b>	0.736641	0.684397	0.090452	0.308824	0.713235	0.305882	0.097561	0.3
<b>2</b>	0.725191	0.560284	0.090452	0.411765	0.242647	0.847059	0.146341	0.1
<b>3</b>	0.709924	0.280142	1.000000	0.514706	0.669118	0.482353	0.731707	0.4
<b>4</b>	0.492366	0.301418	0.070352	0.397059	0.779412	0.435294	0.487805	0.5

```
In [ ]: y = outputs.to_numpy()  
X = inputs.to_numpy()
```

```
In [ ]: len(X[np.isnan(X)])
```

```
Out[ ]: 12
```

```
In [ ]: X[np.isnan(X)] = 0
```

```
In [ ]: X[np.isnan(X)]
```

```
Out[ ]: array([], dtype=float64)
```

```
In [ ]: X_train, X_test, y_train, y_test = train_test_split( X, y, test_size=0.23, random_stat
```

```
In [ ]: print([i.shape for i in [ X_train, X_test, y_train, y_test]])  
[(281, 18), (85, 18), (281,), (85,)]
```

```
In [ ]: model_predict_rain_1 = Sequential(  
    [  
        tf.keras.Input(shape=(18,)),      #specify input size  
        Dense(25,activation='relu', name = 'layer1'),  
        Dense(15,activation='relu', name = 'layer2'),  
        Dense(2,activation='softmax', name = 'layer3')  
    ], name = "rain_prediction" )
```

```
In [ ]: tf.keras.utils.enable_interactive_logging()  
model_predict_rain_1.summary()
```

Model: "rain\_prediction"

Layer (type)	Output Shape	Param #
<hr/>		
layer1 (Dense)	(None, 25)	475
layer2 (Dense)	(None, 15)	390
layer3 (Dense)	(None, 2)	32
<hr/>		
Total params: 897		
Trainable params: 897		

Layer (type)	Output Shape	Param #
<hr/>		
layer1 (Dense)	(None, 25)	475
layer2 (Dense)	(None, 15)	390
layer3 (Dense)	(None, 2)	32
<hr/>		
Total params: 897		
Trainable params: 897		
Non-trainable params: 0		

---

```
In [ ]: model_predict_rain_1.compile(
    metrics=['accuracy'],
    loss=tf.keras.losses.SparseCategoricalCrossentropy(from_logits=True),
    optimizer=tf.keras.optimizers.Adam(0.001),
)
```

```
In [ ]: epochs = 200
tf.get_logger().setLevel('WARNING')
tf.keras.utils.disable_interactive_logging()
model_predict_rain_1.fit(X_train, y_train, epochs=epochs, verbose=1, validation_split=0.2)
tf.keras.utils.enable_interactive_logging()
```

```
In [ ]: acc =model_predict_rain_1.evaluate(X_test,y_test)
print(f'accuracy = {acc[1]}')
```

3/3 [=====] - 0s 6ms/step - loss: 0.4361 - accuracy: 0.9176  
accuracy = 0.9176470637321472

```
In [ ]: model_predict_rain_2 = Sequential(
    [
        tf.keras.Input(shape=(18,)),      #specify input size
        Dense(128,activation='relu'),
        Dense(64,activation='relu'),
        Dense(15,activation='relu'),
        Dense(2,activation='softmax')
    ], name = "rain_prediction2" )
```

```
In [ ]: model_predict_rain_2.summary()
```

Model: "rain\_prediction2"

Layer (type)	Output Shape	Param #
<hr/>		
dense_12 (Dense)	(None, 128)	2432
dense_13 (Dense)	(None, 64)	8256
dense_14 (Dense)	(None, 15)	975
dense_15 (Dense)	(None, 2)	32
<hr/>		
Total params: 11,695		
Trainable params: 11,695		
Non-trainable params: 0		
<hr/>		
Layer (type)	Output Shape	Param #
<hr/>		
dense_12 (Dense)	(None, 128)	2432
dense_13 (Dense)	(None, 64)	8256
dense_14 (Dense)	(None, 15)	975
dense_15 (Dense)	(None, 2)	32
<hr/>		
Total params: 11,695		
Trainable params: 11,695		
Non-trainable params: 0		

---

In [ ]: model\_predict\_rain\_2.compile(

```
#     metrics=['accuracy'],
#     loss=tf.keras.losses.SparseCategoricalCrossentropy(),
#     metrics=[tf.keras.metrics.SparseCategoricalAccuracy()],
#     optimizer=tf.keras.optimizers.Adam(0.0001),
# )
```

In [ ]: epochs = 1000  
tf.keras.utils.disable\_interactive\_logging()  
model\_predict\_rain\_2.fit(X\_train, y\_train, epochs=epochs, verbose=1, validation\_split=0.2)  
tf.keras.utils.enable\_interactive\_logging()In [ ]: acc =model\_predict\_rain\_2.evaluate(X\_test,y\_test)  
print(f'accuracy = {acc[1]}')

```
3/3 [=====] - 0s 4ms/step - loss: 0.4690 - sparse_categorical_accuracy: 0.9176
3/3 [=====] - 0s 4ms/step - loss: 0.4690 - sparse_categorical_accuracy: 0.9176
accuracy = 0.9176470637321472
```

In [ ]: y\_preds = model\_predict\_rain\_2.predict(X)

12/12 [=====] - 0s 2ms/step

```
In [ ]: got_wrong= 0
for i in range(y.shape[0]):
    y_pred_value = np.where(y_preds[i]==np.max(y_preds[i]))[0][0]
    got_wrong += abs(y_pred_value-y[i])
    print(f'The Real Value is: {y[i]}\t The Predicted value is: {y_pred_value}')
print(f'got {got_wrong} wrong from {y.shape[0]} example')
```













```
The Real Value is: 0.0    The Predicted value is: 0
The Real Value is: 0.0    The Predicted value is: 0
The Real Value is: 0.0    The Predicted value is: 0
The Real Value is: 0.0    The Predicted value is: 0
The Real Value is: 0.0    The Predicted value is: 0
The Real Value is: 0.0    The Predicted value is: 0
The Real Value is: 0.0    The Predicted value is: 0
got 8.0 wrong from 366 example
```

```
In [ ]: y_preds[211]
```

```
Out[ ]: array([0.49364486, 0.5063551 ], dtype=float32)
```