Implement Regression With AWS

In [10]:
```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
```

In [ ]:
```
```

In [18]:
```
## Read the data
import s3fs
df = pd.read_csv("https://nehagoudbaddam.s3.amazonaws.com/Data/weatherAlbury.csv'
print('Size of weather data frame is :',df.shape)
df.head()
```

```
Size of weather data frame is : (3011, 13)
```

Out[18]:

	Date	Location	MinTemp	MaxTemp	Rainfall	Humidity9am	Humidity3pm	Pressure9am	P
0	01/12/2008	Albury	13.4	22.9	0.6	71.0	22.0	1007.7	
1	02/12/2008	Albury	7.4	25.1	0.0	44.0	25.0	1010.6	
2	03/12/2008	Albury	12.9	25.7	0.0	38.0	30.0	1007.6	
3	04/12/2008	Albury	9.2	28.0	0.0	45.0	16.0	1017.6	
4	05/12/2008	Albury	17.5	32.3	1.0	82.0	33.0	1010.8	

In [19]:
```
## Preprocess the data

df = df.dropna()
print("new shape:" ,df.shape)
df['RainToday'].replace({'No': 0, 'Yes': 1},inplace = True)
df['RainTomorrow'].replace({'No': 0, 'Yes': 1},inplace = True)
```

```
new shape: (2981, 13)
```

In [13]:
```
## Implement Logistic Regression Model

X = df[["MinTemp", "MaxTemp", "Rainfall", "Humidity9am", "Humidity3pm", "Pressure
        "Pressure3pm", "Temp9am", "Temp3pm", "RainToday"]]
y = df.RainTomorrow

X_train,X_test,y_train,y_test = train_test_split(X,y,test_size=0.20, random_state


logreg = LogisticRegression(max_iter=1000)
logreg.fit(X_train,y_train)

y_pred = logreg.predict(X_test)

print("done")
```

```
done
```

In [14]:
```
## Evaluate the model

from sklearn.metrics import accuracy_score

score = accuracy_score(y_test,y_pred)
print('Accuracy :',score)
```

```
Accuracy : 0.8676716917922948
```

In [15]:
```
## Making predictions

Obs = pd.read_csv("https://nehagoudbaddam.s3.amazonaws.com/Data/weatherObs.csv")

Obs['RainToday'].replace({'No': 0, 'Yes': 1},inplace = True)
newData = Obs.values

y_pred = logreg.predict(newData)
y_pred
```

Out[15]:
```
array([0, 1, 1], dtype=int64)
```

Task 1: Execute the code properly with given sample data

Task 2: Explain what you analyzed in the code. Make a detailed report.

Task 3: Use any other dataset to run the tasks above again.

Task 4: Perform any other machine learning code with putting data in AWS S3.

In [ ]: