In [ ]:
```
import pandas as pd

# Try colab notebook first, handle exception when working locally using VS Code and support Git
try:
    dataset = pd.read_csv('/content/weather_prediction_dataset/seattle-weather.csv')
except FileNotFoundError:
    dataset = pd.read_csv("weather_prediction_dataset/seattle-weather.csv")
```

In [8]:
```
# Your code here. Add more cells if needed
dataset = dataset.drop_duplicates()

print(dataset.shape)
```

```
(1456, 6)
```

In [12]:
```
# Your code here. Add more cells if needed
def getDayInfo(row):
    day = row.split("/")[2]
    return day

dataset["day"] = dataset["date"].apply(lambda row : getDayInfo(row))

print(dataset.shape)
print(len(dataset['day'].drop_duplicates()))
```

```
(1456, 8)
31
```

In [20]:
```
# Your code here. Add more cells if needed
svm = SVC()
svm.fit(X_train, y_train)

svm_y_pred = svm.predict(X_test)

my_accuracy = accuracy_score(y_test, svm_y_pred)
print("My accuracy:", my_accuracy)
```

```
My accuracy: 0.7671232876712328
```

In [21]:
```
my_f1 = f1_score(y_test, svm_y_pred, average=None)
print("F1 score:", my_f1)
```

```
F1 score: [0.         0.         0.864      0.         0.78644068]
```

In [25]:
```
# Your code here. Add more cells if needed

# This section takes about 1 minute to run due to Grid Search

svm_param_grid = {
    "C": [0.001, 0.01, 0.1, 1, 10],
    "kernel": ["linear", "poly", "rbf", "sigmoid"]
}

svm_grid_search = GridSearchCV(SVC(), param_grid=svm_param_grid, scoring="accuracy")

svm_grid_search.fit(X_train, y_train)

svm_y_pred_optimal = svm_grid_search.predict(X_test)

my_accuracy_optimal = accuracy_score(y_test, svm_y_pred_optimal)

print("My accuracy:", my_accuracy_optimal)
print("Whether our model improves: ", (my_accuracy_optimal - my_accuracy) > 0)
```

```
My accuracy: 0.8424657534246576
Whether our model improves:  True
```