National Taipei University of Technology

Computer Science and Information Engineering

Principles and Applications of Data Science

Spring 2022

Semester Group Project Report

The relationship between temperature, humidity and human stress

Name: 顏翊純、黃彥文

Sid: 110599002、109598113

Date:06/22/2022

**Content(目錄)**

**Abstract**

New York Times has written the article, << His College Knew of His Despair. His Parents Didn’t, Until It Was Too Late.>>, using “despair” to talking about the university students may go through in America. And as we realized that there are much people suffer from Depression or Bipolar Disorder in our daily life. We want to found that if there is something we can do before these happens. We start to search for human stress, and two papers written by L. Rachakonda, S. P. Mohanty, E. Kougianos, and P. Sundaravadivel were found. They create an IoT device which can detect the data from human body that can help to detect the stress immediately. We used the data provided by this project to create a model which can predict the human stress by temperature and humidity.

We found that most of them have some common personality, they suffer a big stress at the end, but they are not conscious of when these stresses started. We hoped that we can create a model which can predict the stress, by this way, people can do something relaxing at the beginning to make these stress don’t grow larger and larger.

1. **Introduction**

This part is a summary about the whole report and project. Please have the objectives of your project and the motivation in the introduction. In addition, the applications and contributions should be included.

**Motivation:**

New York Times has written the article, << His College Knew of His Despair. His Parents Didn’t, Until It Was Too Late.>>, using “despair” to talking about the university students may go through in America. And as we realized that there are much people suffer from Depression or Bipolar Disorder in our daily life. We want to found that if there is something we can do before these happens. We start to search for human stress, and two papers written by L. Rachakonda, S. P. Mohanty, E. Kougianos, and P. Sundaravadivel were found. They create an IoT device which can detect the data from human body that can help to detect the stress immediately. We used the data provided by this project to create a model which can predict the human stress by temperature and humidity.

We found that most of them have some common personality, they suffer a big stress at the end, but they are not conscious of when these stresses started. We hoped that we can create a model which can predict the stress, by this way, people can do something relaxing at the beginning to make these stress don’t grow larger and larger.

**Objectives:**

Create a model which can predict the human stress by temperature and humidity data.

Although the temperature and humidity are the body temperature and humidity in these papers, we’ve known that the weather will also influence the frequency of symptoms, which called “Seasonal affective disorder (SAD)”.

1. **Literature review and related works**

Please write your comments and comparisons to the related materials you have reviewed. If there are some related or similar works, please also state and have a discussion and comparison.

(之前找的2篇論文，或其他的)

1. **Problem statement**

Give a scenario of the problem you are working and state the problem. If possible, you can try to formulate your problem formally.

1. **Proposed models (approaches)**

Please present your proposed models (approaches) for the problem and give the reasons why your models or approaches are designed. Give one or two examples to illustrate how your models run.

Use the classifier algorithm instead of regression algorithm because...

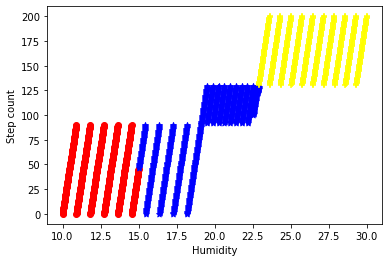
(regression是用連續數據去做，我們的比較像分類)

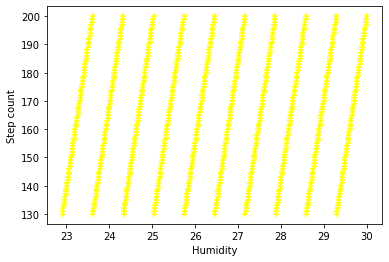
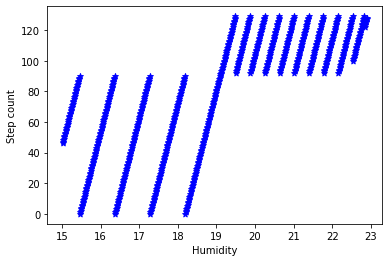
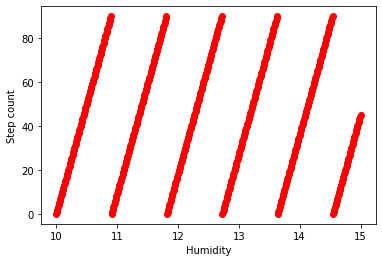
<https://ikala.cloud/supervised-learning-classification-regression/>

1. **Experiments**

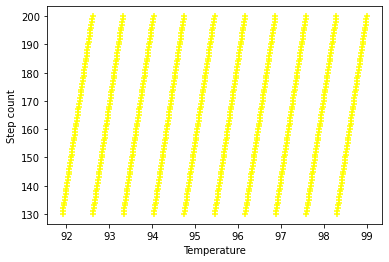
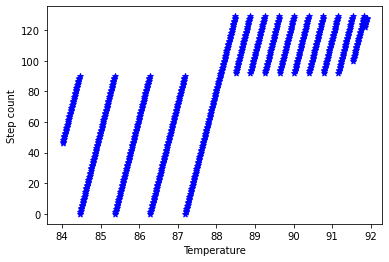
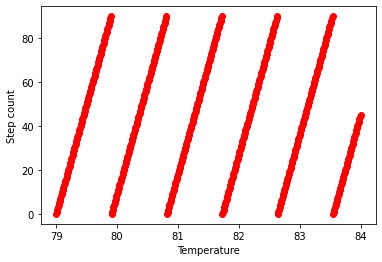
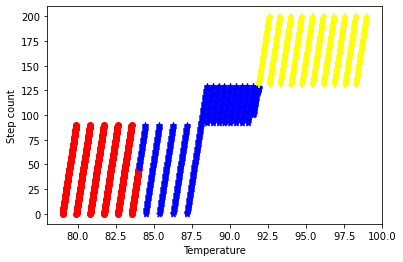
**This part is very important.** It first includes what the objectives of the experiments are, how you design them, ~~which tools are used~~, where the data are from, when you have the data, what the measurements are, and what you will compare with. Then, the results are shown with careful analysis and discussion. One should have reasons for all the trends presented in the results.

**Classify by Humidity and Step Count:**

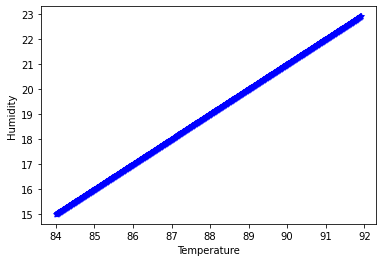
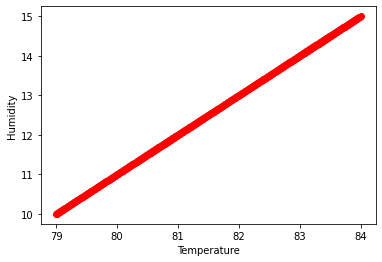
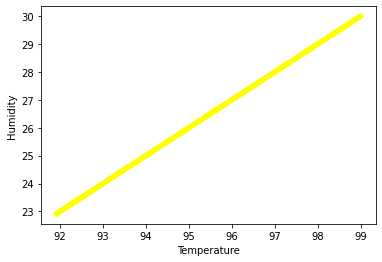
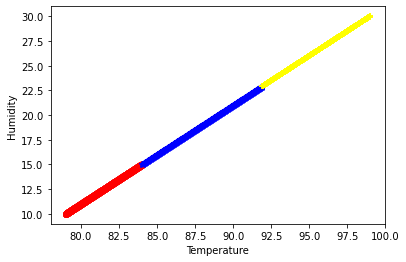




**Classify by Temperature and Step Count:**



**Classify by Temperature and Humidity:**



**Correlation coefficient between humidity and stress level is 0.93**



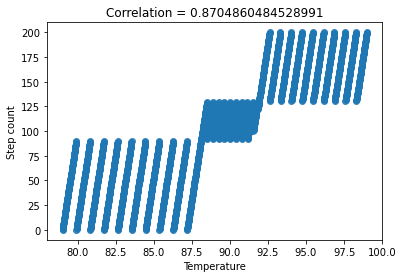
**correlation coefficient between temperature and stress level is 0.93**



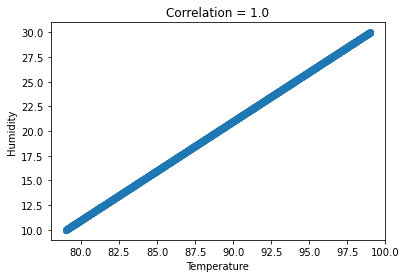
**correlation coefficient between step count and stress level is 0.83**



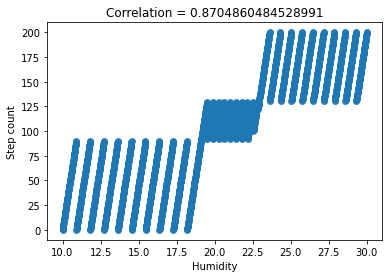
**correlation coefficient between temperature and step count is 0.87**



**correlation coefficient between temperature and humidity is 1.0**



**correlation coefficient between humidity and step count is 0.87**



Through the previous analysis, we found:

* the higher temperature, the higher pressure
* the higher humidity, the higher pressure
* the higher the pressure, the more steps you take.

**K-NN**

hum\_temp\_data = lysis\_data[['Humidity', 'Temperature']]

str\_lev\_data = lysis\_data['Stress Level']

train\_data , test\_data , train\_label , test\_label = train\_test\_split(hum\_temp\_data, str\_lev\_data, test\_size=0.2)

train\_data , ver\_data , train\_label , ver\_label = train\_test\_split(train\_data, train\_label, test\_size=0.25)

model = KNeighborsClassifier(n\_neighbors=3)

# Train the model using the training sets

model.fit(train\_data, train\_label)

# Predict Output

predicted = model.predict(ver\_data) # 0:Overcast, 2:Mild

print("accuracy: ", accuracy\_score(ver\_label, predicted))

    accuracy:  0.9975

predicted = model.predict(test\_data)

print("accuracy: ", accuracy\_score(test\_label, predicted))

    accuracy:  1.0

**Decision\_Tree\_with\_Ada\_Boost\_Classifier**

hum\_temp\_data = lysis\_data[['Humidity', 'Temperature']]

str\_lev\_data = lysis\_data['Stress Level']

train\_data , test\_data , train\_label , test\_label = train\_test\_split(hum\_temp\_data, str\_lev\_data, test\_size=0.2)

train\_data , ver\_data , train\_label , ver\_label = train\_test\_split(train\_data, train\_label, test\_size=0.25)

model =  AdaBoostClassifier(DecisionTreeClassifier(max\_depth=2), n\_estimators=50)

model.fit(train\_data, train\_label)

pred\_ver = model.predict(ver\_data)

acc\_train = accuracy\_score(ver\_label, pred\_ver)

pred\_test = model.predict(test\_data)

acc\_test = accuracy\_score(test\_label, pred\_test)

print("Verification Accuracy: {} \nTest Accuracy: {}".format(acc\_train, acc\_test))

Verification Accuracy: 1.0

Test Accuracy: 0.9975062344139651

model =  AdaBoostClassifier(DecisionTreeClassifier(max\_depth=1), n\_estimators=50)

model.fit(train\_data, train\_label)

pred\_ver = model.predict(ver\_data)

acc\_train = accuracy\_score(ver\_label, pred\_ver)

pred\_test = model.predict(test\_data)

acc\_test = accuracy\_score(test\_label, pred\_test)

print("Verification Accuracy: {} \nTest Accuracy: {}".format(acc\_train, acc\_test))

Verification Accuracy: 1.0

Test Accuracy: 0.9975062344139651

model = AdaBoostClassifier(n\_estimators=50, random\_state=0)

model.fit(train\_data, train\_label)

pred\_ver = model.predict(ver\_data)

acc\_train = accuracy\_score(ver\_label, pred\_ver)

pred\_test = model.predict(test\_data)

acc\_test = accuracy\_score(test\_label, pred\_test)

print("Verification Accuracy: {} \nTest Accuracy: {}".format(acc\_train, acc\_test))

Verification Accuracy: 1.0

Test Accuracy: 0.9975062344139651

**SVM**

hum\_temp\_data = lysis\_data[['Humidity', 'Temperature']]

str\_lev\_data = lysis\_data['Stress Level']

train\_data , test\_data , train\_label , test\_label = train\_test\_split(hum\_temp\_data, str\_lev\_data, test\_size=0.2)

train\_data , ver\_data , train\_label , ver\_label = train\_test\_split(train\_data, train\_label, test\_size=0.25)

C = 2 # SVM regularization parameter

svc = svm.SVC(kernel='linear', C=C).fit(train\_data, train\_label)

pred\_ver = svc.predict(ver\_data)

acc\_train = accuracy\_score(ver\_label, pred\_ver)

pred\_test = svc.predict(test\_data)

acc\_test = accuracy\_score(test\_label, pred\_test)

print("Verification Accuracy: {} \nTest Accuracy: {}".format(acc\_train, acc\_test))

Verification Accuracy: 0.9975

Test Accuracy: 0.9925187032418953

C = 2 # SVM regularization parameter

rbf\_svc = svm.SVC(kernel='rbf', gamma=0.7, C=C).fit(train\_data, train\_label)

pred\_ver = rbf\_svc.predict(ver\_data)

acc\_train = accuracy\_score(ver\_label, pred\_ver)

pred\_test = rbf\_svc.predict(test\_data)

acc\_test = accuracy\_score(test\_label, pred\_test)

print("Verification Accuracy: {} \nTest Accuracy: {}".format(acc\_train, acc\_test))

Verification Accuracy: 0.9975

Test Accuracy: 0.9950124688279302

C = 2 # SVM regularization parameter

poly\_svc = svm.SVC(kernel='poly', degree=3, C=C).fit(train\_data, train\_label)

pred\_ver = poly\_svc.predict(ver\_data)

acc\_train = accuracy\_score(ver\_label, pred\_ver)

pred\_test = poly\_svc.predict(test\_data)

acc\_test = accuracy\_score(test\_label, pred\_test)

print("Verification Accuracy: {} \nTest Accuracy: {}".format(acc\_train, acc\_test))

Verification Accuracy: 1.0

Test Accuracy: 0.9950124688279302

C = 2 # SVM regularization parameter

lin\_svc = svm.LinearSVC(C=C, dual=False).fit(train\_data, train\_label)

pred\_ver = lin\_svc.predict(ver\_data)

acc\_train = accuracy\_score(ver\_label, pred\_ver)

pred\_test = lin\_svc.predict(test\_data)

acc\_test = accuracy\_score(test\_label, pred\_test)

print("Verification Accuracy: {} \nTest Accuracy: {}".format(acc\_train, acc\_test))

Verification Accuracy: 0.9975

Test Accuracy: 0.9975062344139651

**Tools:**

1. Jupyter notebook

2. Python Library

**Data Sources:**

1. **Conclusion**

Please include what the project has done and whether the objectives are achieved. What do you get in the project and what can be done **in the future**.

(收獲+未來展望)

1. **Others**

Please state the workload and role of each member in your team for the project. Besides, show the timeline for the project and check whether all the proposed works have been done.

(分工+timeline)

**Reference**