Predicting the stress level:

Youssef ARRASSEN

Abstract

The aim of this project is to predict the stress of the user, we get the data from Kaggle, and we did some cleaning and optimization of data. And trained a Model to predict the stress level.

Keywords: DASS, Stress, Predicting Stress.

1. Introduction

When we look at stress it plays a factor in performance, and we need to minimize it, for this a set of question in place to determine the level of a person stress level, there is a test called DASS (Depression Anxiety Stress Scales), we created a simple ML that can predict if the user is stressed or not.

The dataset is from Kaggle, and updated last 2 months (version 13), and I took a copy and hosted it on my GitHub account just updating the model and improving it in the future, more details in “Dataset”

1. Dataset Description

The Dataset is from Kaggle ([https://www.kaggle.com/)](https://www.kaggle.com/), and it include 47 features we kept only what is related to stress:

Q1 I found myself getting upset by quite trivial things.

Q6 I tended to over-react to situations.

Q8 I found it difficult to relax.

Q11 I found myself getting upset rather easily.

Q12 I felt that I was using a lot of nervous energy.

Q14 I found myself getting impatient when I was delayed in any way (eg, elevators, traffic lights, being kept waiting).

Q18 I felt that I was rather touchy.

Q22 I found it hard to wind down.

Q27 I found that I was very irritable.

Q29 I found it hard to calm down after something upset me.

Q32 I found it difficult to tolerate interruptions to what I was doing.

Q33 I was in a state of nervous tension.

Q35 I was intolerant of anything that kept me from getting on with what I was doing.

Q39 I found myself getting agitated.

Each item was presented one at a time in a random order for each new participant along with a 4 point rating scale asking the user to indicate how often that had been true of them in the past week where

1 = Normal

2 = Mild

3 = Moderate

4 = Severe

1+2 = Not stressed

3+4 = Stressed

On the next page was a generic demographics survey with many different questions.

TIPI1 Extraverted, enthusiastic.

TIPI2 Critical, quarrelsome.

TIPI3 Dependable, self-disciplined.

TIPI4 Anxious, easily upset.

TIPI5 Open to new experiences, complex.

TIPI6 Reserved, quiet.

TIPI7 Sympathetic, warm.

TIPI8 Disorganized, careless.

TIPI9 Calm, emotionally stable.

TIPI10 Conventional, uncreative.

The TIPI items were rated "I see myself as:" \_\_\_\_\_ such that

1 = Disagree strongly

2 = Disagree moderately

3 = Disagree a little

4 = Neither agree nor disagree

5 = Agree a little

6 = Agree moderately

7 = Agree strongly

|  |  |
| --- | --- |
| Education:  "How much education have you completed?" | 1=Less than high school  2=High school  3=University degree  4=Graduate degree |
| Urban:  "What type of area did you live when you were a child?" | 1=Rural (country side), 2=Suburban, 3=Urban (town, city) |
| English native:  "Is English your native language?" | 1=Yes,  2=No |
| Age:  "How many years old are you?" | (Num) |
| Religion:  "What is your religion?" | 1=Agnostic,  2=Atheist,  3=Buddhist,  4=Christian (Catholic),  5=Christian (Mormon),  6=Christian (Protestant),  7=Christian (Other),  8=Hindu,  9=Jewish,  10=Muslim,  11=Sikh,  12=Other |
| Married:  "What is your marital status?" | 1=Never married,  2=Currently married,  3=Previously married |
| Family size:  "Including you, how many children did your mother have?" | (Num) |
| Screen size | 1=device with small screen (phone, etc)  2=device with big screen (laptop, desktop, etc) |

\*We achieved higher accuracy when working with QA alone

1. Methods

The Methods I used are Decision Tree, K-NN, Random Forest, and then I compared the 3 models with different parameters,

**With** **Decision Tree**, we achieved an accuracy of 0.8, then I got the feature importance plotted

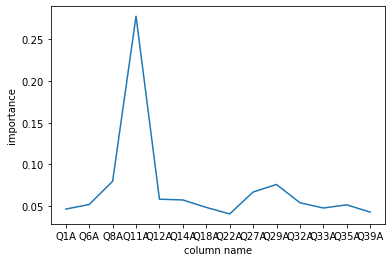


Figure 1: Feature Importance for the Stress Questions

These 3 are the top and very important:

* Q8 I found it difficult to relax.
* Q11 I found myself getting upset rather easily.
* Q29 I found it hard to calm down after something upset me.

**With K-NN**, we tested the dataset with multiple to check for the best one, accuracy is lower from 0.89 to 0.94.

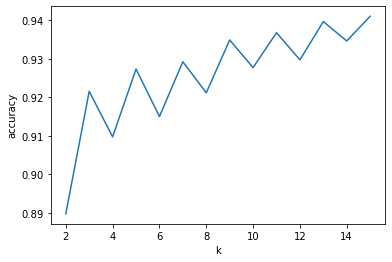


Figure 2: K-NN with different K

With Random Forest

1. Experimental Results

Equations and formulae should be typed and numbered consecutively with Arabic numerals in parentheses on the right hand side of the page (if referred to explicitly in the text),

 (1)

1. Conclusions

They should also be separated from the surrounding text by one space.

Acknowledgements

These and the Reference headings are in bold but have no numbers. Text below continues as normal.

References

Clark, T., Woodley, R., De Halas, D., 1962. Gas-Graphite Systems, in *“Nuclear Graphite”* R. Nightingale, Editor. Academic Press, New York, p. 387.

Deal, B., Grove, A., 1965. General Relationship for the Thermal Oxidation of Silicon, Journal of Applied Physics 36, p. 3770.

Deep-Burn Project: Annual Report for 2009, Idaho National Laboratory, Sept. 2009.

Fachinger, J., den Exter, M., Grambow, B., Holgerson, S., Landesmann, C., Titov, M., Podruhzina, T., 2004.“Behavior of spent HTR fuel elements in aquatic phases of repository host rock formations,” 2nd International Topical Meeting on High Temperature Reactor Technology. Beijing, China, paper #B08.

Fachinger, J., 2006. Behavior of HTR Fuel Elements in Aquatic Phases of Repository Host Rock Formations. Nuclear Engineering & Design 236, p. 54.

The source code will be given here.