RESEARCH

Introduction of Profile Areas of Data Science :Project 8

Suresh Kumar Choudhary, Frenny Macwan and Aman Jain

Full list of author information is available at the end of the article

Abstract

Goal of the Project: To find the co-relation between the activity and the stress for the students studying in a university, along with the other stress correlated features and training a classifier model to classify the mental state of a student.

Main Result of the Project: We were able to find the co relation between the stress and activity for the students U10 and U16, and later derived a negative corelation between stress and conversation.

Personal Key Leanings: Prophet Library, Using SQLLite, SQLAlchemy

Estimate working hours: 48 Hours in Total

Project Evaluation: 2

Number of Words: 1026

Keywords: Time Series data trend; Seasonality; Corelation; Classification; Studentlife

Scientific Background

Student-Life continuous sensing app assess day to day activity of the students of a batch of 48 students who were using android devices and were voluntarily recruited from a CS batch . The app tracks the sleep, mood, social life, academic performance almost everything, things which are already informed to the students and before signing the consent they are given a walk through of the system, they get some incentives in return like Google Nexus Phones, T-Shirts etc. The general trend what is being observed is term starts with good sleep patterns, good confidence, but when we move towards the end of the term, the sleep pattern, changes, stress increases. The App integrates MobileEMA STATES stress mood across the term. There was a strong co-relation between automatic sensing data and a broad data of Mental Well being, depression etc. Using call records, Bluetooth proximity detect the social life and daily activity. and If there is no response to EMA, a mail was being sent to the student to connect over the wifi. The data were SSL encrypted so there is no privacy breach and also the data would be removed once the people leave the study. Physical activity detetor detects the movements, and classifies run, jog, walk cycling etc. Conversation Detection, which detect a conversation. Sleep Detection sees the light features, phone usage features including the phone lock state during the night. Mobile EMA is like a short survey system, which is being popped on the screen regularly on the student's phone and if the student misses filling the feedback questionare an email is being sent to him/her for reminder. We discuss a number of insights into behavioral trends, and importantly, correlations between objective

sensor data from smartphones and mental well-being and academic performance for a set of students at Dartmouth College.

Goal

To reanalyze parts of the Student Life Study by :

1 Analyzing if Stress is co-related to Activity.
2 What else is stress correlated to?
3 Predicting a Student's current state.

Data

The StudentLife dataset contains the EMA (Survey Responses), the Automatic Sensing data which collects a lot of information.The data set contain four types of data : Sensor Data, EMA Data, Pre and Port Survey Responses and Educational Data. Sensor Data is further divided into 10 different sensor data, physical activity, audio inferences, conversation inferences, Bluetooth scan, light sensor, GPS, phone charge, phone lock, WiFi, WiFi location which are all stored in CSV Files. Inside each sensor data, the information for each student like activity_u01 will be provided in the seperate csv files. EMA Responses are stored in EMA folder where the responses of stress are been given. All the Pre and Post Survey responses are being stored in the survey folder. Educational data, which include classes taken during 2013 Spring term, deadlines for each participants, grades and Piazza usage for CS65, is stored under education folder.

Result

Task 1 : Iss stress corelated to activity level?

To import the data, we have used the Google Big Query. We imported the CSV files for activity_u10, activity_u16 and perceivedstressscale. The schema was created and table with the names activity_u10, activity_u16 and stressScale was create respectively in a dataset we named as Students. The stressScale table was amputated with only two columns, the *student id* and the column *In the last month, how often have you felt nervous and "stressed"?*.

- Distribution of the stress level of the students is showed in Figure : 1, where we can get the exact count of the students and their stress level for the past one month by the responses they have sent via EMA Api.
- Activity Distribution of the student_u10 is showed in Figure : 2
- Activity Distribution of the student_u16 is showed in Figure : 3
- Through the data set we also concluded that Student U_16 was more active than StudentU_10, which can be seen in Figure : 4
- Corelation between stress and activity can be found in the Table 1

Table 1 Stress vs Activity Corelation for Student U10 and U16

Students	Stress	Activity %
U10	Fairyy Often	4.88
U16	Very Often	5.99

Figure 1 Stress Distribution of all the Students

```
print('List of frequency of the students who felt stressful in the last one month : ')
df_stressCounts = df_stressscale['In the last month, how often have you felt nervous and "stressed"?'].value_counts()
print(df_stressCounts)
```

```
List of frequency of the students who felt stressful in the last one month :
Sometime        31
Very often      23
Fairly often    19
Almost never    11
Name: In the last month, how often have you felt nervous and "stressed"?, dtype: int64
```

```
labels = df_stressCounts.index
counts = df_stressCounts.values
fig, ax = plt.subplots(figsize = (5,5))
ax.pie(counts, labels = labels)
plt.show()
```

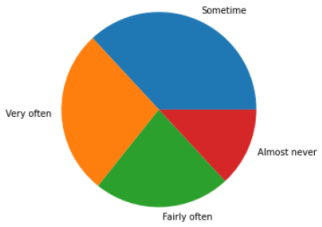
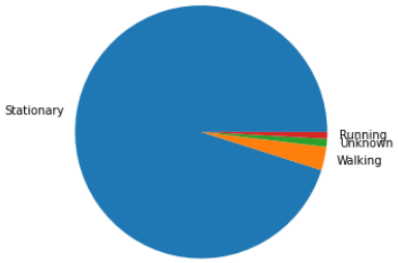


Figure 2 Activity Distribution of Student U10

```
df_activity_u10_counts = df_activity_u10['_activity_inference'].value_counts()
labels = ['Stationary', 'Walking','Unknown','Running']
counts = df_activity_u10_counts.values
fig, ax = plt.subplots(figsize = (5,5))
ax.pie(counts, labels = labels)
plt.show()
```



Task 2 : What else is stress correlated to?

For correlating the stress (Except for Activity), we have also used the feature of conversation. The stress scale was encoded in numbers, the conversation information for 10 students were imported in the GoogleBigQuery and tables/schemas were created for each of the student, and we performed the co-relation between the conversation and the stress where we found a negative co-relation between them. This would in general mean, a student indulged in more conversations have less stress and students who don't converse a lot, feel more depressed or alone. This has been shown in the Figure : 5. There could be more features which are directly

Figure 3 Activity Distribution of Student U16

```
df_activity_u16_counts = df_activity_u16['_activity_inference'].value_counts()
labels = ['Stationary', 'Walking','Unknown','Running']
counts = df_activity_u16_counts.values
fig, ax = plt.subplots(figsize = (5,5))
ax.pie(counts, labels = labels)
plt.show()
```

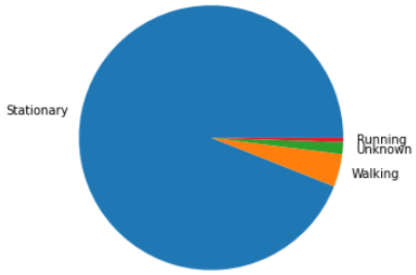


Figure 4 Activity Comparison of U10 and U16

```
U10_Stationary = (df_activity_u10['_activity_inference'].value_counts()[0]/df_activity_u10.shape[0])*100
U16_Stationary = (df_activity_u16['_activity_inference'].value_counts()[0]/df_activity_u16.shape[0])*100
U10_Activness = 100 - U10_Stationary
U16_Activness = 100 - U16_Stationary
from tabulate import tabulate
print(tabulate([['U10', U10_Stationary,U10_Activness], ['U16', U16_Stationary,U16_Activness]], headers=['Name', 'Stationary','Activity']))
```

```
Name      Stationary    Activity
------  ------------  ----------
U10          95.1169     4.8831
U16          94.0075     5.99247
```

CONCLUSION : STUDENT U16 IS MORE ACTIVE THAN U10

correlated to the stress, like the student's social life. We found the same negative corelation between the phone lock timings, phone charge as well.

Figure 5 Corelation Between Stress and Conversation

Assignment8.ipynb

File Edit View Insert Runtime Tools Help All changes saved

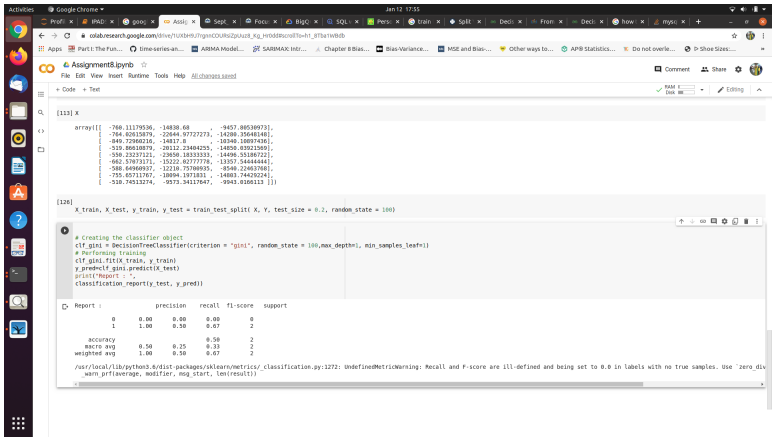
+ Code + Text

```
 'conversation_u08.csv',
 'conversation_u09.csv',
 'conversation_u10.csv']
avg_list=[]
for table in file_list:
  query='SELECT avg(start_timestamp-_end_timestamp) FROM Students.'+table.split('.')[0]
  temp=gbq.read_gbq(query, project_id='ipads2020assignment8').head()
  avg_list.append([x[1] for x in temp.itertuples()][0])
avg_list
stress_dict={}
for i in df_stressscale.itertuples():
  if i[1] in st_list:
    stress_dict[i[1]]=i[2]
stress_class=[]
for key,val in stress_dict.items():
  stress_dict[key]=stress_val_con[val]
  stress_class.append(stress_val_con[val])
#correlation between stress and conversation
np.corrcoef(avg_list, stress_class)
```

```
array([[ 1.        , -0.42761519],
       [-0.42761519,  1.        ]])
```

Task 3 : Predicting Student's Current State We were planing the use Decision Tree Classifier Algorithm, with various features like conversation, phonelock timings, phonecharge timings, due to lack of time we were not able to add other features. Please find the output in the Figure **??**

Figure 6 Corelation Between Stress and Conversation



Discussion

Mental well being of a student is of the most priority. University can surely take this study and find out how the mental state of the student vary during the course of the semester. Through the data the university can get ample amount of data or about the lifestyle of the students. As we move towards the end of semester the habits break, the sleep patterns become very irregular, the diets becomes bad, and due to which the students become more stressful. Studies might show, a good amount of activity, talking to personals over phone calls, having a good social life, less social media can reduce the stress of the student. University this way, can surely circulate some guidelines to the students during the course of semester.

Appendix

Code -

- Aman - Task1
- Suresh - Task3
- Frenny - Task2

Report -

- Aman - Scientific background, Data, Task1
- Suresh - Discussion, Task 3
- Frenny - Abstract, Goal and Task 2.

Author details

References