Mental Health at Work (PS6)

By DATA PIRATES
(PSA CODESPRINT 2021)

Problem Statement

To develop a system or process to detect personnel going through mental health issues or potentially suffering from a burn-out, and intervention program to address this group.

OUR SOLUTION

Survey

The specified survey should be conducted at regular intervals.

Analysis of Survey Data

The data collected by the survey will now be analysed and uses Data Science to identify potential burn-outs.

Intervention Programs

Depending upon the results of the survey data analysis, recommended groups shall be adviced intervention programs.

PERIODIC MENTAL HEALTH SURVEY

Periodic Mental Health Survey

We incorporate the perceived stress scale(developed by Cohen, Kamarck, & Mermelstein 1983), the perceived loneliness scale for our analysis and identification of potential burn-outs.

Created by:
DATA PIRATES
Nipun Bhatia
Kanupriya Arora
Siddhant Pathak

siddhant.dpsn@gmail.com (not shared) Switch account

* Required

PSA CODESPRINT 2021 PRELIMINARY ROUND

Please state your age *

Your answer

Please select your gender *

- Female
- Male
- Other/would rather not say

Perceived Loneliness Scale

This scale was developed by Cohen, Kamarck, & Mermelstein in 1983.

The questions in this scale ask you about your feelings and thoughts during the last week. In each case, you will be asked to indicate your response by placing an "X" over the circle representing HOW OFTEN you felt or thought a certain way.
Although some of the questions are similar, there are differences between them and you should treat each one as a separate question. The best approach is to answer fairly quickly. That is, don't try to count up the number of times you felt a particular way, but rather indicate the alternative that seems like a reasonable estimate.

Only some specific questions are shown in this form, the entire inclusion of the question set is a scope of improvement for further exploration.

1. In the last week, how often have you been upset because of something that happened unexpectedly? *

- Never (0)
- Almost Never (1)
- Sometimes (2)
- Fairly Often (3)
- Very Often (4)

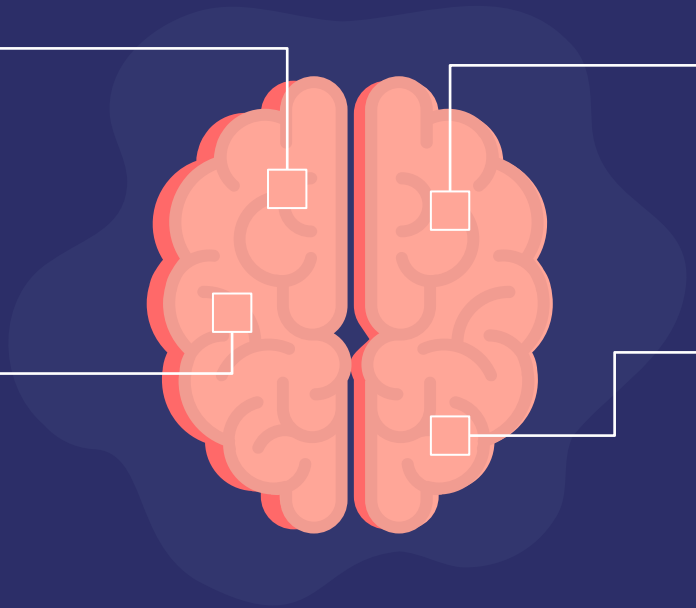
In the last week, how often have you felt that you were unable to control the important things in your life? *

- Never (0)
- Almost Never (1)
- Sometimes (2)
- Fairly Often (3)
- Very Often (4)

ANALYSIS OF SURVEY DATA

Collection and Curation of Dataset

Introduction of essential variables



Random Forest Classifier – Supervised Learning

Testing the model

COVIDiSTRESS Global Survey dataset on psychological and behavioural consequences of the COVID-19 outbreak

Yuki Yamada[1], Dominik-Borna Ćepulić[2], Tao Coll-Martín[3], Stéphane Debove[4], Guillaume Gautreau[5], Hyemin Han[6], Jesper Rasmussen[7], Thao P. Tran[8], Giovanni A. Travaglino[9], COVIDiSTRESS Global Survey Consortium* & Andreas Lieberoth[7]

	Unnamed: 0	Duration..in.seconds.	RecordedDate	UserLanguage	Dem_age	Dem_gender	Dem_edu	Dem_edu_mom	Dem_employment	Country	...	
0	1	777	2020-06-17 01:01:39	NL	48	Female	College degree, bachelor, master	NaN	Full time employed	Netherlands	...	zie is
1	2	952	2020-06-16 23:19:18	PL	36	Female	College degree, bachelor, master	Up to 12 years of school	Full time employed	Poland	...	
2	3	2396	2020-06-16 22:47:36	EN	45	Female	College degree, bachelor, master	Up to 12 years of school	Full time employed	Finland	...	
3	4	1365	2020-06-16 22:02:45	SME	35	Female	PhD/Doctorate	Up to 12 years of school	Self-employed	Mexico	...	
4	5	902	2020-06-16 21:11:53	SME	46	Female	College degree, bachelor, master	None	Full time employed	Mexico	...	

Raw Data

	Dem_age	Dem_gender	Dem_edu	Dem_maritalstatus	PSS10_avg	SLON3_avg	SPS_avg
0	48	Female	College degree, bachelor, master	Married/cohabiting	2.4	2.666667	5.1
1	36	Female	College degree, bachelor, master	Married/cohabiting	1.7	1.000000	5.3
2	45	Female	College degree, bachelor, master	Married/cohabiting	1.7	2.000000	5.2
4	46	Female	College degree, bachelor, master	Married/cohabiting	2.8	2.333333	4.6
6	60	Female	Some College, short continuing education or eq...	Divorced/widowed	2.1	1.666667	6.0

Clean Data

Collection and Curation of Dataset

The data is obtained from the COVIDiSTRESS Global Survey dataset on psychological and behavioural consequences of the COVID-19 outbreak.

The link to the article and the dataset is as follows:
https://rdcu.be/czet7

[SPS]	Available social provisions in critical/distressing situations	Social Provisions Scale short form PSP-10. 10 items. 6-Point Likert Scale (1 =strongly disagree, 6 = strongly agree). Validated language versions where available. Back-translated where necessary. Mean scores will be computed.	Steigen & Bergh (2019)

[Scale_PSS10_UCLA] items 1-10	Perceived stress for the past week	Perceived Stress Scale (PSS-10). 10 items. 5-Point Likert Scale (1 =never, 5 = very often). Validated language versions where available. Back-translated where necessary. Mean score will be computed.	Cohen, Kamarck, & Mermelstein (1983)

Introduction of essential variables: Burnout?

We have analysed research papers and come up to conclusion of using the three scales of mental health to derive a categorical variable "**Burnout?"** that determines whether the individual is at risk or not.

```
In [15]: df['Burnout?'] = (((df["PSS10_avg"]) <3) & (df["SPS_avg"].astype(float) < 3) & (df["SLON3_avg"].astype(float)<=3))
```

```
In [16]: df.head()
```

Out[16]:

	Dem_age	Dem_gender	Dem_edu	Dem_maritalstatus	PSS10_avg	SLON3_avg	SPS_avg	Burnout?
0	48	Female	College degree, bachelor, master	Married/cohabiting	2.4	2.666667	5.1	False
1	36	Female	College degree, bachelor, master	Married/cohabiting	1.7	1.000000	5.3	False
2	45	Female	College degree, bachelor, master	Married/cohabiting	1.7	2.000000	5.2	False
4	46	Female	College degree, bachelor, master	Married/cohabiting	2.8	2.333333	4.6	False
6	60	Female	Some College, short continuing education or eq...	Divorced/widowed	2.1	1.666667	6.0	False

Random Forest Classifier: Supervised Learning

Splitting Dataset

Splitting the dataset into test and train, we made the test size 30%

```
# Import essential models and functions from sklearn
from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import train_test_split
from sklearn.metrics import confusion_matrix
from sklearn.tree import plot_tree

# Extract Response and Predictors
y = pd.DataFrame(df_ohe['Burnout?'])
X = pd.DataFrame(df_ohe.drop('Burnout?', axis = 1))


# Split the Dataset into Train and Test
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.3)
```

```
# Import GridSearch for hyperparameter tuning using Cross-Validation (CV)
from sklearn.model_selection import GridSearchCV

# Define the Hyper-parameter Grid to search on, in case of Random Forest
param_grid = {'n_estimators': np.arange(100,901,150),
              'max_depth': np.arange(5, 11)}

# Create the Hyper-parameter Grid
hpGrid = GridSearchCV(RandomForestClassifier(),
                      param_grid,
                      cv = 5,
                      scoring = 'accuracy')

# Train the models using Cross-Validation
hpGrid.fit(X_train, y_train["Burnout?"].ravel())
```

Fiting the Model

We fit the model on test and train data

```
# Decision Tree using Train Data
dectree = DecisionTreeClassifier(max_depth = 4)
dectree.fit(X_train, y_train)

# Plot the trained Decision Tree
f = plt.figure(figsize=(24,24))
plot_tree(dectree, filled=True, rounded=True,
          feature_names=X_train.columns,
          class_names=["No","Yes"])
```

```
# Fetch the best Model or the best set of Hyper-parameters
print(hpGrid.best_estimator_)

# Print the score (accuracy) of the best Model after CV
print(np.abs(hpGrid.best_score_))
```

```
RandomForestClassifier(max_depth=9, n_estimators=550)
0.9988195615514334
```

Prediction

We predict the response variable **Burnout?**

```
# Predict the Response corresponding to Predictors
y_test_pred = dectree.predict(X_test)
```

```
# Predict the Response corresponding to Predictors
y_train_pred = dectree.predict(X_train)
```

```
Train Data
Accuracy  :      1.0
```

```
Test Data
Accuracy  :      0.9986511549485753
```

Checking Accuracy

We check the accuracy and goodness of fit of model on the test and train predictions.

AREAS FOR IMPROVEMENT

There are still a lot of areas where this project can be improved such as with the inclusion of more and more data points, Binary Logistic Regression based Deep Learning techniques can be used in place of Random Forest Classifier for the sake of accuracy. Other areas include creation of intervention programs and updating survey questions, catering more to the needs of employees. A web portal can be developed for the similar purpose too for aesthetic purposes along with proper creation and maintenance of databases.

REFERENCES

ĆEpulić, D. B., Coll-Martín, T., Debove, S., Gautreau, and others. (2021). COVIDiSTRESS global survey dataset on psychological and behavioural consequences of the COVID-19 outbreak. Scientific Data, 8(1). https://doi.org/10.1038/s41597-020-00784-9

10 Surprising Mental Health Statistics From 2020. (2020, November 5). Mental Health First Aid. https://www.mentalhealthfirstaid.org/external/2020/11/10-surprising-mental-health-statistics-from-2020/

Perceived Stress Scale (PSS). (n.d.). Measurement Instrument Database for the Social Sciences. Retrieved October 10, 2021, from https://www.midss.org/content/perceived-stress-scale-pss

Workplace Mental Health Program Get In Touch. (n.d.). MindBeacon. Retrieved October 10, 2021, from https://info.mindbeacon.com/workplace-mental-health-program-get-in-touch-00