Prediction of Stress Score

Shelly Chang

Overview

- Stress is the physical or mental response to an external cause. A stressor may be a one-time or short-term occurrence, or it can happen repeatedly over a long time. Anxiety is the body's reaction to stress and can occur even if there is no current threat (i.e., chronic stress). A person may be at risk for an anxiety disorder if it feels like he/she can’t manage the stress and if the symptoms of stress start to feel persistent and pervasive, interfere with daily life and cause the person to avoid doing things.

- This project is built on biofeedback metrics collected through Fitbit devices as part of the European H2020 RAIS (Real-time Analytics for Internet of Sports) project.

- One of the metrics in this dataset is the Stress Score, which reflects the respondent’s level of stress. Understanding which factors can help predict this stress score will be beneficial to users and potentially guide them to make lifestyle changes to manage stress more effectively.

Problem Statement

- Between a score of 1-100, a high score indicates that body is showing fewer signers of stress; a low score indicates more signs of stress. This score serves as a reminder for us to take on more challenges or take a break to relax.

- Fitbit calculates the Stress Score from factors that cover 3 areas:
  - Responsiveness: heart-rate data and electrodermal activity (EDA)
  - Exertion balance: impact of physical activity on your stress level.
  - Sleep patterns: effect of sleep duration and quality on your stress level.

- This research will explore whether a selection of metrics in these 3 areas tracked by Fitbit are statistically significant in their ability to explain variance in the Stress Score and to predict the Stress Score.

- Audience of interest may include medical professionals, health/fitness coaches and anyone in the general public that is interested in understanding how cardiovascular health, physical activity and sleep quality impact level of stress on the body.

Data Analyzed

Dataset

- Fitbit data tracked under the European H2020 RAIS (Real-time Analytics for Internet of Sports) project.
- Involved 71 participants in the European region, and collected from mid-2021 to early-2022.
- Data accessible via: https://www.kaggle.com/datasets/skywescar/lifesnaps-fitbit-dataset

Feature Engineering

The model includes approximately 30 features that encompass physical exercise, sleep and heart related metrics. These 3 categories reflect how the stress score is calculated by fitbit. Few exclusions are:

- Select features that have close to as many observations as the target, therefore we will skip min_goal and max_goal as they only have half of the observations.
- We will also exclude exertion_points_percentage sleep_points_percentage, and responsiveness_points_percentage as they feed directly into the calculation of the stress score and will cause the model to be overfit.
- We will also exclude spo2 and scl_avg which contain significant amount of NaN values as well as subjective measures that are not directly tracked by fitbit (i.e., participants' evaluation of their mood).

Features Overview

- Features that carry more weight in explaining variability in Stress Score and have p-value < 0.05 are:
  - Heart health: nremhr, rmssd, resting_hr
  - Exercise: distance, lightly_active_min, lightly_active_minutes, very_active_minutes
  - Sleep: minutesAwake, sleep_efficiency, sleep_rem_ratio
- It's counterintuitive that more sedentary_minutes contribute to lower stress level, unless there are underlying factors such as it tracks users resting or practicing mindfulness.

	coef	std err	t	P>|t|
const	0.0334	0.045	0.750	0.453
nightly_temperature	-0.0611	0.021	-2.853	0.004
nremhr	-0.2289	0.038	-6.099	0.000
rmssd	0.2557	0.030	8.451	0.000
full_sleep_breathing_rate	-0.1079	0.027	-4.047	0.000
calories	-0.0888	0.034	-2.594	0.010
filteredDemographicVO2Max	0.0264	0.025	1.047	0.295
distance	-0.3498	0.166	-2.102	0.036
bpm	-0.1187	0.045	-2.656	0.008
lightly_active_minutes	0.1387	0.042	3.338	0.001
moderately_active_minutes	0.1122	0.035	3.203	0.001
very_active_minutes	0.0820	0.037	2.221	0.027
sedentary_minutes	0.1356	0.039	3.520	0.000
resting_hr	0.1959	0.045	4.352	0.000
minutesToFallAsleep	1.475e-16	5.08e-17	2.906	0.004
minutesAsleep	0.2862	0.258	1.111	0.267
minutesAwake	-0.2751	0.073	-3.772	0.000
minutesAfterWakeup	-0.0413	0.019	-2.202	0.028
sleep_efficiency	0.1700	0.023	7.306	0.000
sleep_deep_ratio	0.0440	0.026	1.712	0.087
sleep_wake_ratio	0.0957	0.043	2.245	0.025
sleep_light_ratio	-0.0301	0.034	-0.886	0.376
sleep_rem_ratio	0.2517	0.026	9.569	0.000
steps	0.4681	0.173	2.711	0.007
minutes_in_default_zone_1	0.1007	0.039	2.587	0.010
minutes_below_default_zone_1	-0.0026	0.029	-0.090	0.928
minutes_in_default_zone_2	0.1202	0.024	4.952	0.000
minutes_in_default_zone_3	0	0	nan	nan
sleep_duration_min	-0.0943	0.286	-0.330	0.742

Model Selection

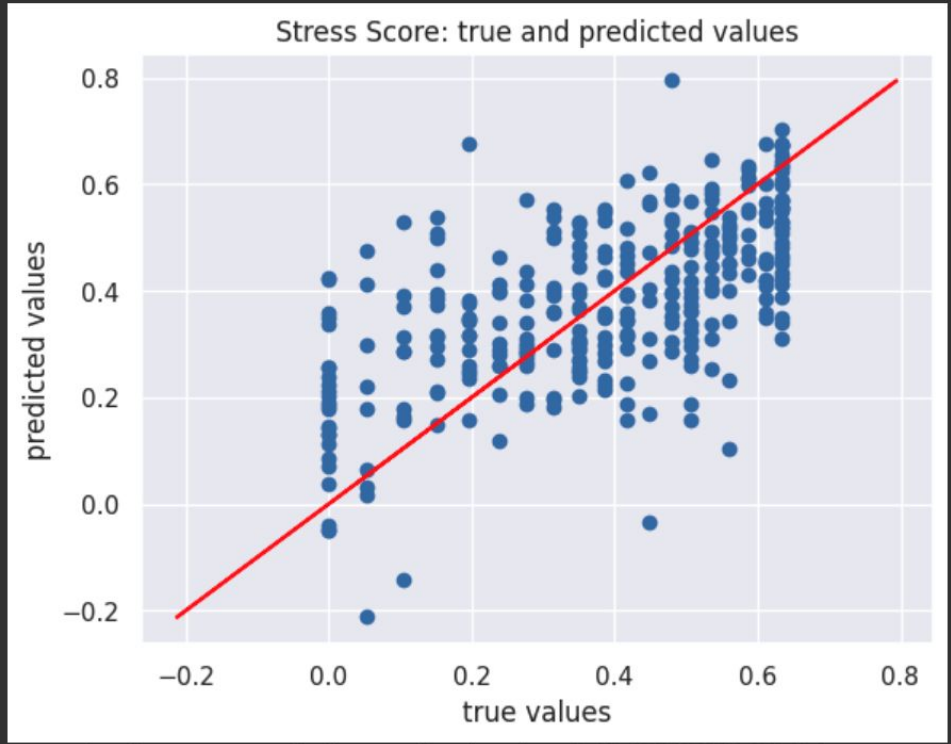
- OLS Regression model and Support Vector Machine Regression model are selected to evaluate whether the feature set is statistically significant in explaining the variance in the Target and predicting the Target.
- **OLS Regression Model**
  - While the p-value associated with the F-test is less than 0.05, Adjusted R-squared value is 0.35. This means that the model only explains 37% of the variance in stress score. We will run another model Support Vector Machine to determine if it is better at explaining the variance in our Target and predicting the Target.

OLS Regression Results

Dep. Variable:	stress_score	R-squared:	0.376
Model:	OLS	Adj. R-squared:	0.366
Method:	Least Squares	F-statistic:	36.06
Date:	Mon, 18 Sep 2023	Prob (F-statistic):	4.69e-139
Time:	19:09:21	Log-Likelihood:	645.23
No. Observations:	1580	AIC:	-1236.
Df Residuals:	1553	BIC:	-1092.
Df Model:	26		
Covariance Type:	nonrobust		

Support Vector Machine Model

- As our dataset is non-linear, the RBF (radial basis function) kernel functions was use to predict the target variable.
- The SVM model score measures how well a statistical model predicts an outcome. In our RBF model, the feature set can explain approximately 70% of the variability in the Target which is statistically significant and better than the OLS Regression model.
- All 3 evaluation metrics are low,and lower values indicate better performance of the model.




```
Mean absolute error of the prediction is: 0.12437451791392186
Mean squared error of the prediction is: 0.025005362138808388
Root mean squared error of the prediction is: 0.15813083867104605
```

Challenges and Next Steps

- Data and Model may not be as applicable to the general public because people that use fitbit may already be more health-oriented to begin with.
- More analysis is required to determine how heart rate, exercise and sleep interact together to impact users' overall stress level.
    - For example, someone who does not have optimal heart health but exercises strenuously may introduce more stress to his/her body.
- Additional features for analysis can include users' subjective assessment of how they feel emotionally and mentally.