Emotional State Classification with Biometric Sensor Data

David Havrilla   Andrew Rendler   Justin Teufel

PROBLEM STATEMENT/MOTIVATION

How do we identify an emotional state like stress preemptively, as to alert a user that they may be approaching a stressed state, and corrective action should be taken? This type of technology could be applied in the area of workplace ergonomics, for example in high stress environments like police work.

How do we describe emotional states like amusement, with a quantitative approach? With an understanding of how stress manifests itself in our biological systems, we can use this information to make better design decisions. This could be applied to A/B testing in environmental or UX design; Think, designing a dentist or doctor's office to optimize neutral emotional states.

Is temperature a good predictor for stress? What is a better predictor of stress?

Since laugher/amusement are generally associated with better health, what sensor modalities are predictors of amusement?

How effective of an indicator can respiration, and/or body temperature be in order in roughly classifying levels of anxiety, nervousness, and other similar feelings, and is there a strong enough correlation to develop a predictive classification model? If a model can be developed, what level of precision can be achieved, and what is an ideal classification scale (ie. rating anxiety levels on a scale of 1-5, or 1-10).

LITERATURE SURVEY

Alberdi, A., Aztiria, A., & Basarab, A. (2015, November 28). Towards an automatic early stress recognition system for office environments based on multimodal measurements: A review. Retrieved July 8, 2020, from https://www.sciencedirect.com/science/article/pii/S1532046415002750

In this study a framework for office stress monitoring was developed using contextual, physiological, and psychological signals – as well as providing evidence to support that, out of all signals captured, which were the most indicative of stress states. Additionally, with the incorporation of contextual and psychological data, distinctions are made between cognitive stimulation and stress.

Dawans, B., Kirschbaum, C., & Heinrichs, M. (2010, September 16). The Trier Social Stress Test for Groups (TSST-G): A new research tool for controlled simultaneous social stress exposure in a group format. Retrieved July 9, 2020, from https://www.sciencedirect.com/science/article/abs/pii/S0306453010002088?via=ihub

This study focused on group behavior, taking groups of 6 out of a total pool of 25 participants to apply the Trier Social Stress Test – involving public speaking and mental math recitation scenarios. The test collected physiological data including ECG, respiration, heartrate, and cortisol level measurements. This study provides a valuable foundation for future experiments involving group based physiological response stress research.

Ruensuk M. (2018, October 8-12). Detecting Emotions using Smartphone Sensors: Technique to Raise Self-Awareness for Social Media Users. UbiComp/ISWC '18. ACM 978-981. https://doi.org/10.1145/3267305.3277825

The study proposal suggests research is needed in investigating emotional self-awareness. The author shares her feelings on the importance of transferring of emotions over social media platforms. Her proposal is to use smartphone commodity sensors to detect physiological responses of social media users. The goal of completing such research would provide social media users with greater emotional self-awareness from the physiological data collected. This emotional self-awareness would improve human interaction between social media users.

H. Vögel et al., "Emotion-Awareness for Intelligent Vehicle Assistants: A Research Agenda," 2018 IEEE/ACM 1st International Workshop on Software Engineering for AI in Autonomous Systems (SEFAIAS), Gothenburg, 2018, pp. 11-15.

Emotion-aware Vehicle Assistant (EVA) is a design to be used as personal assistance within autonomous vehicles. The design idea is to interact on a contextual basis by utilizing environmental information including: "sights and objects, sound, car sensor input, intonation voice and sentiment of language, direction of gaze, and gestures" and other forms of sensory input. The purpose is to allow a user to make more emotionally aware decision utilizing EVA's interpretation of contextual information.

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6468793/

The following paper titled *Assessing Anxiety Disorders Using Wearable Devices: Challenges and Future Directions* explores the reliability of using wearable devices (specifically those that measure electrocardiogram signals (ECG) to predict anxiety levels. The paper analyzes a collection of past studies and uses these to establish a set of criteria for future studies to adhere do in order to be effective.

https://www.sciencedirect.com/science/article/abs/pii/S1532046419300693

The following study titled *Predicting anxiety state using smartphone-based passive sensing* aims to use "behavioral features" (usage statistics, brightness, holding angle) from cellphone logs to predict stress/anxiety levels.

PROPOSED WORK

Our project will require no additional data collection, we are using the WESAD dataset, described later in our summary. The dataset fully describes the system that we are working to analyze, so there is no need to enrich our information or integrate any additional features.

First, we have read the source files (.pkl 'pickle') into the defined data structure or class. We will include the subject ID along with the ECG, EDA, EMG, respiration, and temperature measurements. Our team has decided to eliminate the use of the data collected from the accelerometers, as the analysis would require more time than available for this activity.

The data is collected at disparate rates – some at a rate of 70 per second, down to 1 every 2 seconds. We will want to summarize some of the readings to several per second, as the collection rates likely are high enough to support this reduction. We will make this determination during the initial EDA process.

During the EDA process we will try to identify correlation between the sensor outputs and the emotional response, **this is our primary objective for this project**. An interesting python library was highlighted during research; Neurokit. This package can be used to read or analyze variables such as ECG and EDA, that may return more valuable features that can be applied to our predictive model.

Another approach that could be valuable would be time series. If a generally correlated feature is identified, some window of historical data could be transposed onto each incoming 'new' record, where the sequence of

states for the given variable all become useful features.

Once we have identified the highest quality features generated during the process, we will attempt to build a predictive model. Though our overarching objective is to identify features that are correlated to emotional state, our stretch goal is to build a model that could ingest incoming data and provide a user a response of their predicted emotional state (amusement, neutral, or stress).

DATA SET

The data being used is the Wearable Stress and Affect Detection (WESAD) Data Set by Schmidt et al (https://ubicomp.eti.uni-siegen.de/home/datasets/icmi18/). The data set is publicly available and features sensor data measured from both a chest and wrist-worn device.

The dataset contains information for 15 different subjects.

The devices used measured: blood volume pulse, electrocardiogram, electrodermal activity, electromyogram, respiration, body temperature, and three-axis acceleration.

From this dataset, we plan to focus on respirations and body temperature. Once some data analysis is established, we plan to broaden our predictive search to other measured biometric data in ability to predict various emotions.

The emotions being studied within this data set were neutral, stress, and amusement.

EVALUATION METHODS

We will apply the following evaluation methods:

Chi Square or 2 Way ANOVA during our exploratory data analysis. This can help us generalize the relationship between some selected explanatory variable and our response to evaluate which features should be allowed further consideration.

During the model development phase of the project we will apply a cross validation cross validation approach, there is a scikit learn workflow that was identified during our research that can be easily integrated into our process. We will use the cross-validation score to evaluate our success.

We will use a confusion matrix to understand the performance of our models as we test various configurations. With this analysis we can generate the standard metrics of precision, recall, and accuracy. And as suggested by one of our classmates during the feedback session, we will also incorporate the F1 metric.

TOOLS

As our primary programming language, we will use Python 3.7.x. We chose Python as our programming language for this project based on each group member's familiarity with the language, as well as comfort using Python. We plan to use additional packages in Python that are commonly used in data science, including: pandas, numpy, matplotlib, and sklearn.

Pandas (source: https://pandas.pydata.org/) is an open source package for statistical analysis of data sets. The package allows the user a much more efficient way of manipulating data structures. Another benefit of panadas is the ability to work with multiple data formats. We plan to use pickle files, which can be imported with pandas.

Numpy (https://numpy.org/) is a package that makes Python computation much more user-friendly. What makes it so valuable is the ability to perform more complex mathematic computations and its use in building multi-dimensional arrays. Numpy is often used in fields of study such as: bioinformatics, mathematical analysis, multi-variate analysis, and more.

Matplotlib (https://matplotlib.org/) is a library used for plotting graphs within Python. Again, it can utilize numpy as an extension. We plan to use it as one of a few different ways to plot data points from our dataset.

Sklearn (https://scikit-learn.org/stable/), or more formally known as scikit-learn, is a machine learning library used within Python. To best optimize sklearn, it is used congruently with numpy for complex calculations and operations with arrays. One of the benefits of using sklearn is the fact that it is also an open source package.

Tableau (https://www.tableau.com/) is an additional tool we may use based on its ease of use and ability to generate complete and customizable visualizations. Granted, it is not nearly as customizable as other options. The primary benefit is the ease of importing datasets, while having general control over the visualizations created.

MILESTONES

As a group, we plan to meet once a week, currently every Sunday, to discuss work to be completed and make any necessary changes to milestones or other portions of the group project. We plan to meet via video conference and will continue to collaborate between meetings via e-mail or text message.

We plan to have the remaining portions of our project, parts 4-6, completed by Sunday, August 3rd. Setting an earlier due date will allow us to complete the project with enough time to make last minute changes. Additionally, a Sunday due date will allow us to meet via video conference and review last minute edits as a group.

Since Part 7 is an individual piece for peer evaluation, it will be a separately determined milestone for each group member.

PROGRESS - CODE

Up to this point our focus has been on data cleaning/analysis prep and preliminary exploratory analysis. In preparation for our exploratory analysis, we've developed a python class that interacts with the raw data - the .pkl files associated with each subject.

Our class has a series of methods that we developed to visualize and 'smooth' the data where necessary. Our methods include several graphing functions - some of which focus on a single sensor attribute, and some for correlation studies. Some functions have optional arguments which allow the user to smooth the data where necessary.

PROGRESS - EDA

Our exploratory data analysis at this point includes time series visualizations of the raw data (with some smoothing), just to get an understanding of how the sensor data is

changing over the duration of the study, and how it relates to the subjects 'protocol condition'.

For this submission we have included exploratory graphs from 3 subjects (1,4,8), with 4 time series graphs for each subject. The first graph shows the subjects entries for the protocol condition (0 = not defined / transient, 1 = baseline, 2 = stress, 3 = amusement, 4 = meditation, 5/6/7 = should be ignored). The next three graphs show sensor readings for respiration, temperature, and electrodermal activity respectively - these graphs 'smoothed' the raw data, which was sampled at 700Hz, by averaging every 100 entries. As you'll see below, the electrodermal and temperature time series graphs are more legible in the sense that you can easily identify a sensor reading at a particular time during the study, where the variation in the raw respiration data doesn't allow for that legibility. The visualization is telling though, because the increased variability indicates greater displacement from the subject's inhaling and exhaling.
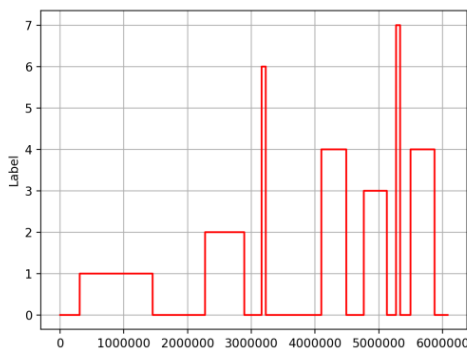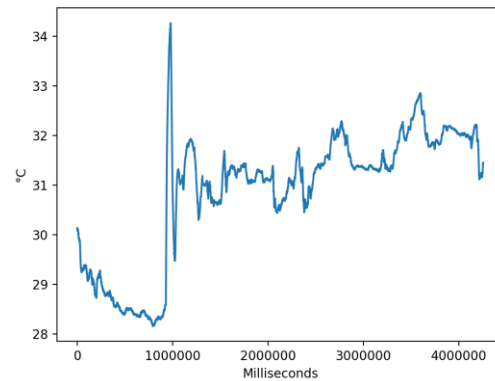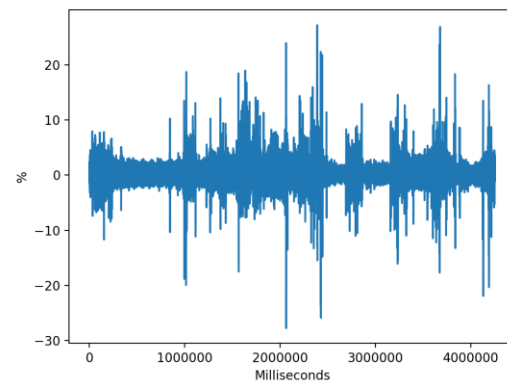
Subject 1:
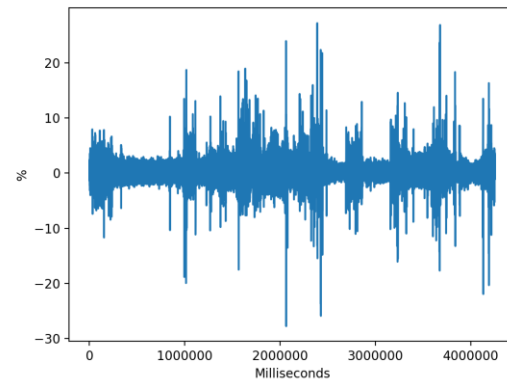
Respiration - Subject 1

Temperature - Subject 1

Protocols - Subject 1

EDA - Subject 1

Subject 4:
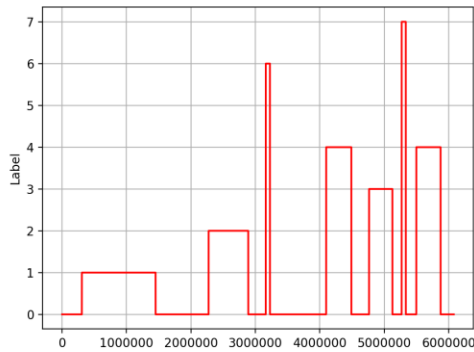
Protocols - Subject 4

EDA - Subject 4
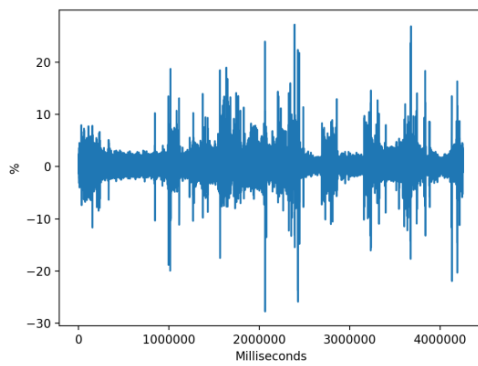
Subject 8:

Respiration - Subject 1

Protocols - Subject 8
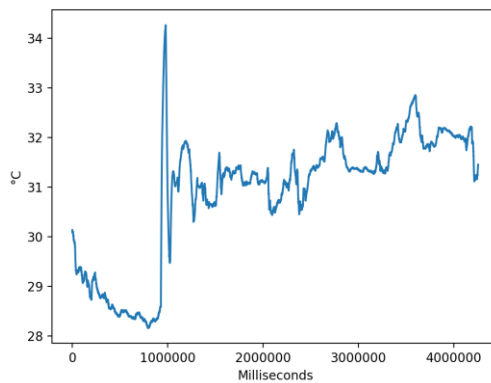
Temperature - Subject 4
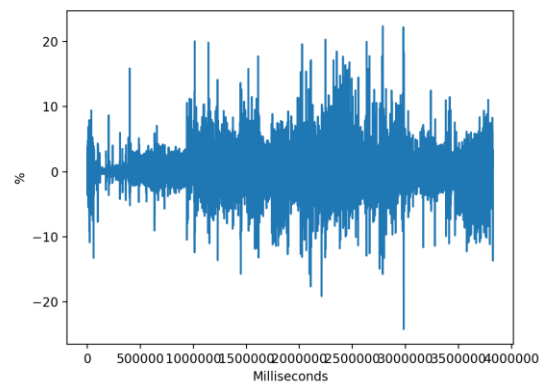
Respiration - Subject 8
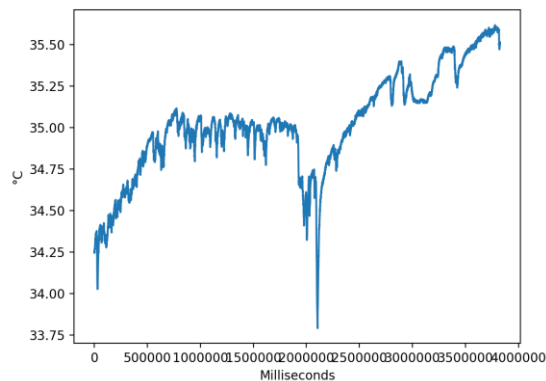
Temperature - Subject 8
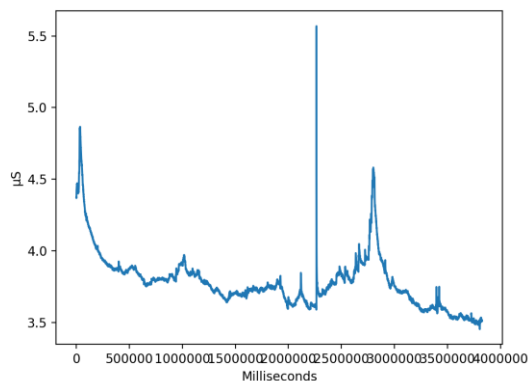
EDA - Subject 8

NEXT STEPS - EDA

To further refine our analysis, we will take additional steps to clean the data - which will involve removal of outliers, possibly deriving additional statistics from the raw data, and excluding irrelevant data (such as protocols 5,6 and 7). Additionally, we will work on developing correlation graphs to identify any relationships between sensor readings and the protocols.
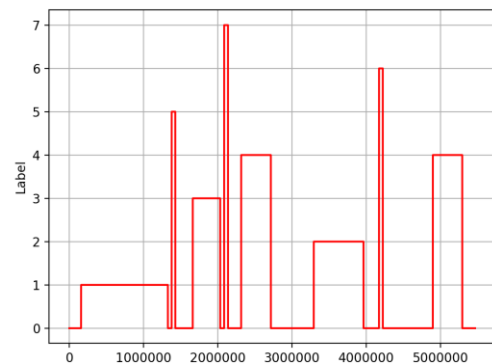
ABSTRACT

We had several questions related mostly to the sensor modalities and for any predictive values they made provide. The focus of our questions revolved around improving predictive knowledge of one's health and interacting preemptively.

Our first question focuses exactly on that modality prediction for preemptive intervention. How do we identify an emotional state like stress preemptively, as to alert a user that they may be approaching a stressed state, and corrective action should be taken? From this question, our goal was to be able to provide an answer as far as which sensor modalities were able to provide a reliable preemptive prediction for stress. This would allow a sensor to notify a user prior to a stressful event time for intervention.

The next question falls in line quite similarly to the prior question. Is temperature a good predictor for stress? What is a better predictor of stress? Again, trying to identify which modalities correlate well with stress and if not temperature, identify which other modalities are we able to utilize.

To contrast those questions, we also found it important to focus on the amusement rather than stress to identify some of the healthier emotions and behaviors. Since laugher/amusement are generally associated with better health, what sensor modalities are predictors of amusement? Rather than prevention of stressful events or events that may lead to poor health, we want to be able to help users identify which behaviors might be promote good health or elicit positive feelings. We felt this would encourage users to focus on those positive behaviors if the user knows how to identify them.

We discovered a few different important results, some about utilizing the dataset itself, and other results directly related to our questions.

Although not directly related to the questions, some additional results are important to note. We found that in the raw state, the 70hz readings had too much variance to generate strong predictions. A 700:1 ratio for reduction

allowed us to identify stronger relationships. It should also be pointed out that data from the chest sensors was much more accurate and reliable than the data from wrist sensors.

Elaborating on the results of the questions, it was noted with our model and preliminary data that the strongest relationships between a stress emotional response and modalities included ECG, temperature, EDA, and EMG. We were able to create a k-nearest neighbor model using those modalities with an approximately 99.9% level of accuracy.

INTRODUCTION

Many of our questions were developed with an intent to identify specific meaningful relationships or correlations between sensor modalities and responses, if there even were any significant relationships. The additional intent behind our questions is to see if the emotional response can be predicted with some degree of accuracy.

Our questions are important, as they relate back to improving health and workplace safety. We believe sensor modalities in either a chest sensor or wrist sensor can be used in predictive modeling to preemptively notify a user if that user is experiencing a stressful event or an amusing event.

Sensors can be used in either the professional work setting, or by a consumer. Given the expense of a chest sensor, it is much more likely to be used in either a workplace or as a research tool. Meanwhile, the wrist sensor, as is already a commonplace, can be easily accessible to the consumer.

One of the sensors in the workplace may help provide information to the employee or other pre-emptive knowledge of a stressful event. Take policework for example, prior to a stressful event, a sensor may be able to notify that individual to modify either their behavior or modify their environment to prevent a stressful event from occurring.

The other scenario such sensor modalities can be used is with consumers. As already noted, wrist sensors are already a commonplace but may only track a couple modalities such as heartrate or steps taken. Utilizing a sensor that can track additional modalities and use the data with predictive modeling can provide the consumer and their respective physician with more information about one's health.

RELATED WORK

As previously discussed, there is significant work contributions surrounding the ability to identify or predict stress. Much of the data and research is to focus on improving health through diminishing the impact of stress.

Many of the work being done also uses different sensor modalities.

One of the especially interesting aspects to the work being completed are the myriad scenarios. Some of the work being done is to look at stress in the workplace, while other sensors focus on identifying stress in the driver of a car.

Other research and data focus on larger spectrums of human emotion in uses such as social media and its impact on the user.

DATA SET

The data being used is the Wearable Stress and Affect Detection (WESAD) Data Set by Schmidt et al (https://ubicomp.eti.uni-siegen.de/home/datasets/icmi18/). The data set is publicly available and features sensor data measured from both a chest and wrist-worn device.

The dataset contains information for 15 different subjects. It is important to note that a couple of the subjects were removed from use due to inaccurate sensor readings.

The researchers in the study collected sensor modalities using two different types of sensors, a wrist sensor and a chest sensor.

The devices used measured: blood volume pulse, electrocardiogram, electrodermal activity, electromyogram, respiration, body temperature, and three-axis acceleration.
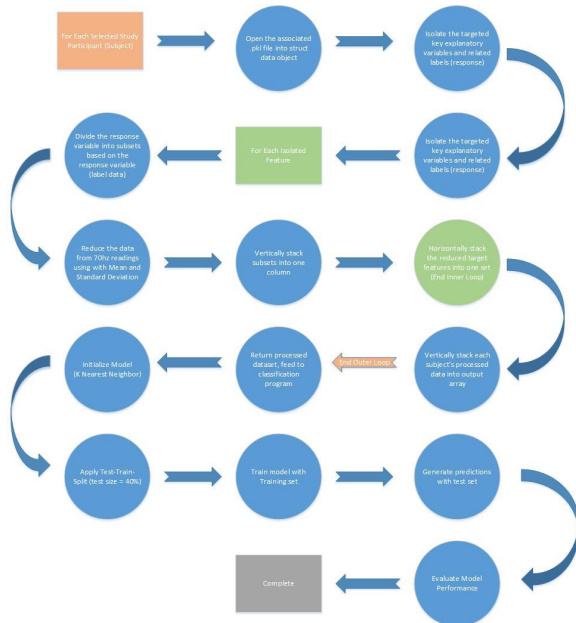
MAIN TECHNIQUES APPLIED

Figure 1: Data Ingestion – Processing – Classification – Model Evaluation [Larger image on page 11]

We applied several course relevant techniques during each phase of our data mining project, to briefly summarize the key techniques applied.

During the data acquisition phase of our project we developed a program that first reads the data in from the source pickle (.pkl) files, and stores the data contained within the file in a user defined python class. For this, and each subsequent step until the classification or modeling step, will be performed in a loop - with each iteration working with the data from an individual subject from the study (there were 12 active participants in the study)

During the processing phase, we extract an individual raw feature and the label data (response) and process it by first subdividing the array based on the label data classes (1, 2, 3) using numpy methods. The result is 3

subarrays that we summarize by stepping through the array grabbing an n sized chunk of records at each step, calculating the standard deviation and mean, and storing this value in a new array. Once the subarray has been reduced, we vertically stack the subarrays to produce again a single feature array. We found that 700:1 or 10 second averages produced a high-quality feature. Next, we performed this step for all the features in the dataset, for both chest and wrist sensors, and evaluated the relationship or correlation to the response variable. From here we implemented a forward stepwise development process, we were able to identify 4 explanatory variables that consistently produced models scoring ~99.9% accuracy. We employed another loop here to assemble the individually processed feature arrays into one main table for analysis. It is also at this point that we vertically stacked the processed data from all queried individuals before passing it along to the mining or classification step.

During the classification step we apply a concept of breaking the dataset into testing and training subsets, through the random selection. This helps produce better performing models when it relates to 'unseen data' or testing the models performance with what could potentially be disparate characteristics in the test and training sets, as would be seen in a real life application of the model. For our exercise we use a 60/40 split, with 40% of the records allocated to testing. Once the data has been split, we initialize a Scikit Learn model to classify our data, and then generate a set of predictions with our test set. Finally, we evaluate the results of the test predictions using a confusion matrix and simple pass/fail count rates (accuracy).

KEY RESULTS

Through this process we were able to develop the following conclusions:

- During our initial EDA we identified that analyzing the data in its raw state, 70hz readings, the data was too variable to generate strong predictions.

- We tested various reduction strategies to eliminate some of the noise or variability in the sensor data. We found that we were able to see a much stronger relationship as we approached a 700:1 ratio for reduction.

- Not all of the data available to us was necessary. We found the strongest relationship between ecg, temp, etc. sensor data and the response variable, emotional state.

- The Chest sensor readings, though supplying some of the same measurements, was more reliable.

- We tested several potential model approaches; linear regression, logistic regression, linear discriminant analysis, decision tree, and K Nearest Neighbors. We found the best results with K Nearest Neighbor (Tied with decision tree for performance)

- We were able to generate a model with an accuracy of ~100%, consistently, with just 4 explanatory features

APPLICATIONS:

This type of model has applications in several domain and could be used almost anywhere. The team spent a fair amount of time discussing how a basic sensor array and a model similar to this could be applied. A few of the 'applications' were more geared towards the product design perspective, "If we were going to build a product like this, how could we use our findings to build the *best* product". An example of this would be: The EDA portion of this study could be applied to hardware design and sensor specifications. We were

working with data collected at 70hz. The definition was very high and could likely be decreased considerably. This would impact the energy and computational resource requirements for any biofeedback hardware (some wearable device).

A few examples of where we saw that this figurative 'product' could be utilized are more intuitive, such as; This type of classification could be applied to wearable devices that would alert a user whether or not they are stressed, or inform them once they have reached a state of relaxation. From another perspective, collecting emotional state feedback about some environment or situation. A product similar to what we have put together could be applied to design settings to test the response of participants. Such as, designing workspaces that maintain stress to a minimum, or entertainment spaces to optimize for amusement.

For Each Selected Study Participant (Subject)

Open the associated .pkl file into struct data object

Isolate the targeted key explanatory variables and related labels (response)

Divide the response variable into subsets based on the response variable (label data)

For Each Isolated Feature

Isolate the targeted key explanatory variables and related labels (response)

Reduce the data from 70hz readings using with Mean and Standard Deviation

Vertically stack subsets into one column

Horizontally stack the reduced target features into one set (End Inner Loop)

Initialize Model (K Nearest Neighbor)

Return processed dataset, feed to classification program

End Outer Loop

Vertically stack each subject's processed data into output array

Apply Test-Train-Split (test size = 40%)

Train model with Training set

Generate predictions with test set

Complete

Evaluate Model Performance