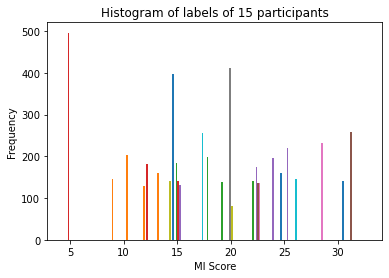
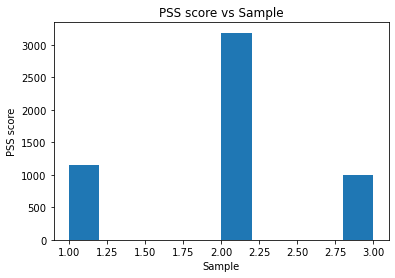
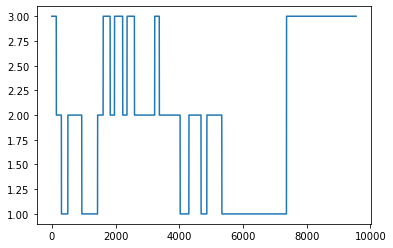
# PSS Analysis summary

* The study was collected

**Pre-brief** -> VR1 -> Educational Video -> VR2 -> **D ebrief**

* The PSS questionnaire is only collected during pre-brief and debrief
* We use the same analysis as the MIOS analysis
  + Pre-processing
    - Normalization
    - Data sanitation
    - Segmentation and windowing
      * NOTE: Conducted for
        + Beginning to End of pre-brief
        + Beginning to End of debrief
        + This does NOT include VR 1 and VR 2 since no data was collected there
  + Feature extraction
    - Same as before
  + Clustering
    - Same as before
    - Prior to clustering
    - 
    - After clustering
    - 
  + Classification
    - When mixing combining all of the data together, achieved 72% and 92% for SVM and DT respectively
    - LOSOCV achieved 34.55 +- 26.69 and 25.01 +- 25.00 for DT and SVM respectively
    - Weighted classification achieved very low (5-50%)
    - Downsample achieved very low (5-50%)
    - SMOTE achieved 95%
      * The label distribution can be seen as follows
      * 

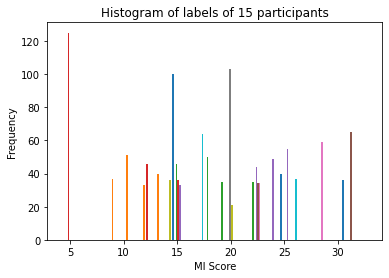
# PSS 1 oversampled

* Proposed idea is to oversample Sample 1 of PSS to Prebrief, VR1, and VR2
* Sample 2 of PSS is in Debrief
* Results were similar to the above

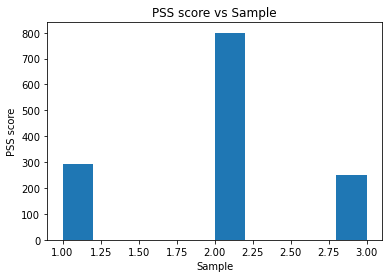
# Window sampling to 20 seconds

* Varying the window size to see the difference
* Reasoning:
  + PSS can evaluate overall stress
  + Changes in emotion takes longer to reflect in physiology
  + 5 seconds was an estimator for MIOS
* Length of Cluster 1: 428 Cluster 2: 800 Cluster 3: 292

Cluster of PSS scores



PSS scores after clustering [0-13, 13-27, 27<]



* Classification of all subjects together
  + DT: 89%
  + SVM: 70.79%
* LOSO single subject
  + DT: 54.81 +- 29.84
  + SVM: 29.92 +- 23.20
* Weighted classification
  + DT: 82.92%
  + SVM: 81.18%
* Downsample
  + DT: 93.59%
  + SVM: 84.75%

# 

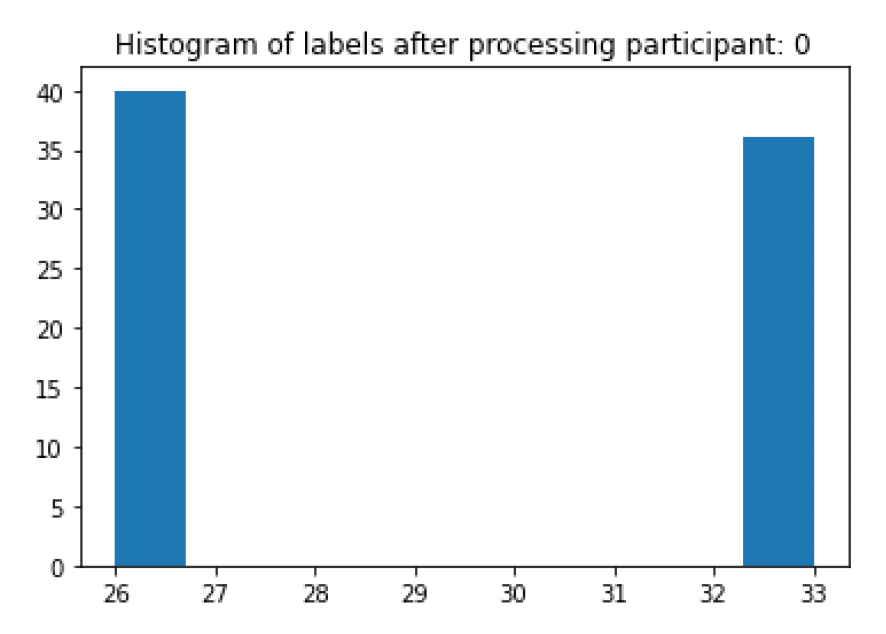
# Clustering

* We cannot classify everyone on one model
* We would cluster users that have similar data (e.g similar heart rates during experiment)
* Then we would do the analysis afterwards
  + With the previous approach we realized that building a general stress detection model for all the students is not feasible with our data. Therefore we wanted to try if clusters of students can be created for which accurate cluster-specific classification model can be built. First, the data of one student was removed to be used as test data. On the remaining data, a clustering algorithm (Weka’s expectation maximization) was used to create clusters of students. For each student, all instances were assigned to the cluster which contained most of his/hers instances. Next, a classification algorithm (SVM, j48, Bagging, Random Forest, or Ordinal classifier) was used to train a cluster-specific model using only the data of the specific cluster
  + [<https://ieeexplore-ieee-org.ezproxy.lib.ryerson.ca/stamp/stamp.jsp?tp=&arnumber=7194282>]

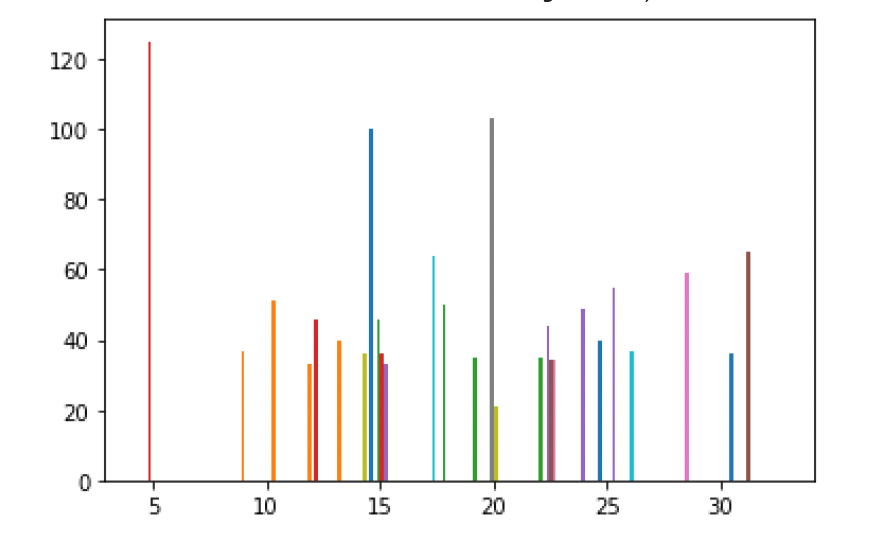
# Pseudo steps

1. Separate data to PSS1 (prebrief) and PSS2 (debrief)
2. Normalize and typical processing of data
3. Cluster the feature matrix data [unsupervised]
4. Apply analytic techniques to classify the clustered label

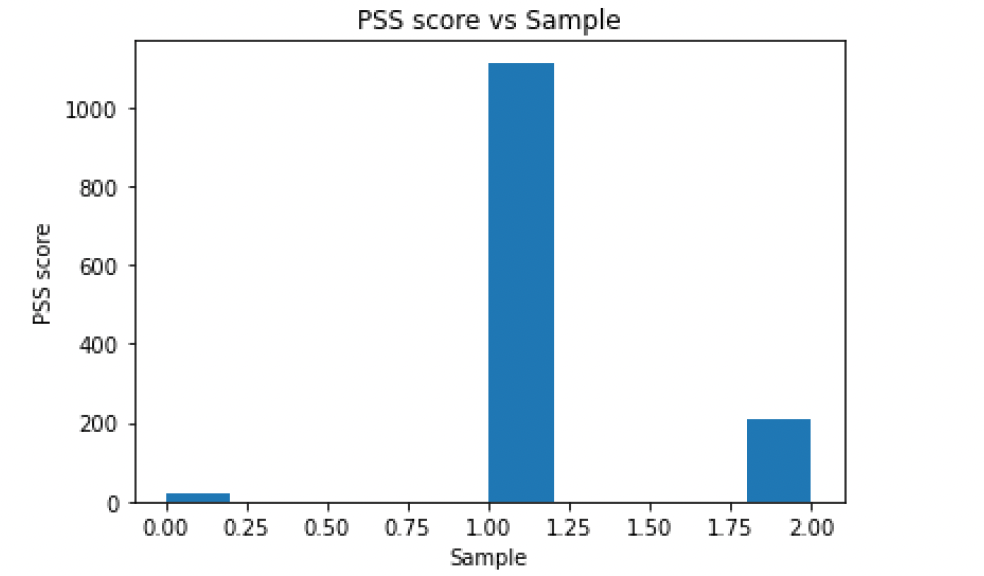
PSS value of one of the participants



PSS histogram of all participant data

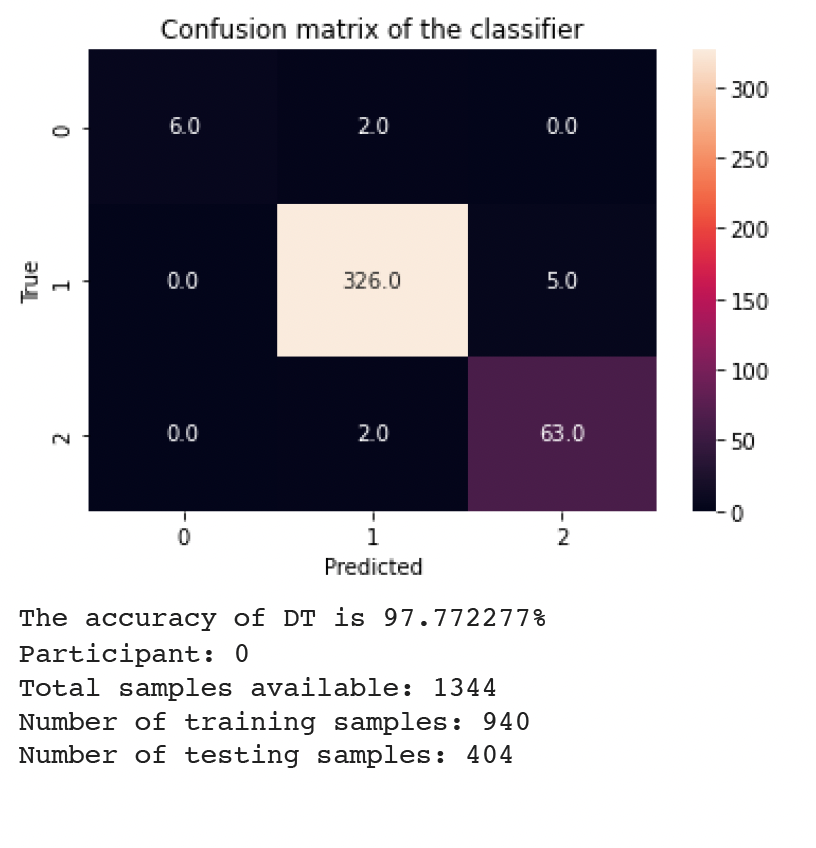


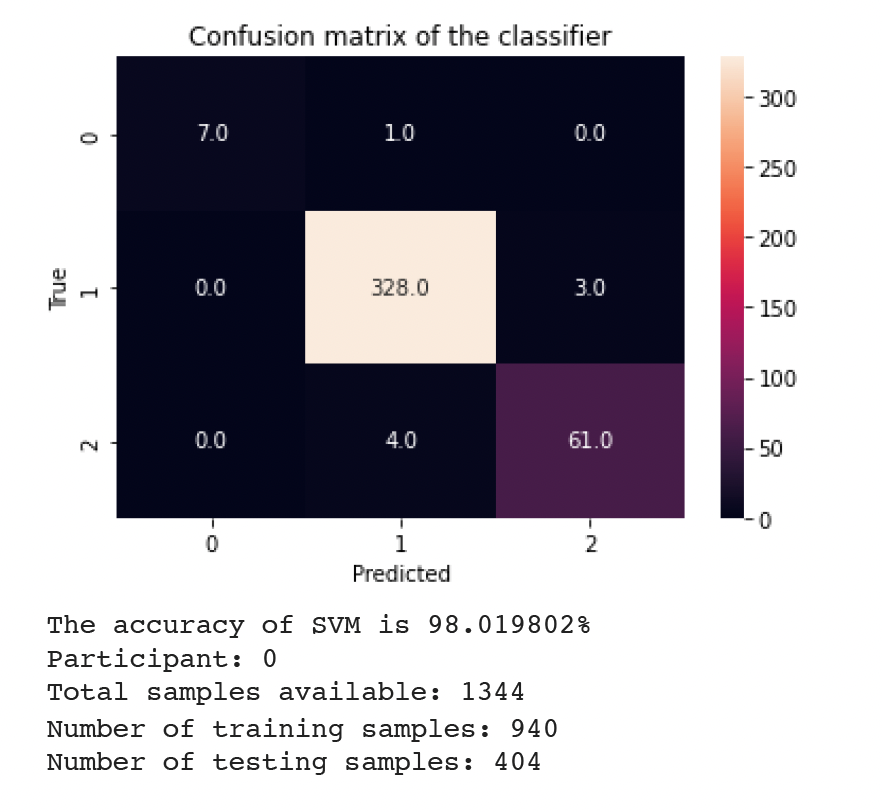
Histogram of participant data after clustering



DT and SVM achieving >98%

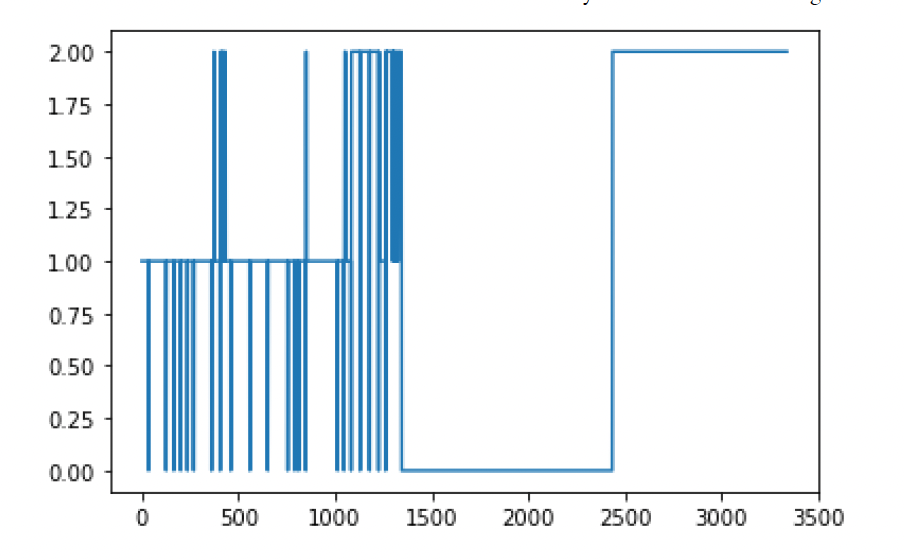
* Definitely overfitted





SMOTE analysis

* 99%



# 

# MVBC

* Proposed idea
  + View 1 would include AVG stats
  + View 2 would be trends