Cognitive State Predicting of Pilots with Machine Learning of Physiological Data

IE7275 – Data Mining in Engineering

Jeffrey Johnston

Introduction

Aviation is considered one of the most dangerous career fields and minor errors can result in major accidents and in commercial aviation with hundreds of passengers on board, the stakes can be even greater. In most aviation accidents, pilot error is almost always a main cause or contributing factor. In 2019, Booz Allen Hamilton hosted a competition for predicting the cognitive state of pilots in an effort to model and detect when pilots may be experiencing troubling events.

The dataset consists of physiological data from pilots during controlled experiments outside of the aircraft. During the experiments the pilots were subjected to intentional distractions intended to induce one of the three following cognitive states: Channelized Attention, Diverted Attention and Startle/Surprise. Channelized Attention is the state of being focused on one task to the exclusion of all others. Diverted Attention is the state of having one's attention diverted by actions or thought processes associated with a decision. Startle/Surprise is the state of shock from an unexpected event.

The data was collected from a variety of sensors operating at a sample rate of 256 Hz recording data over time while subjecting a pair of pilots as a crew to one of three experiments. Each experiment was intended to put the pilots in one of the three cognitive states and collect the physiological data from that state as well as their baseline state.

Exploratory Data Analysis

The raw data file was downloaded from Kaggle as a .csv file and read in python for data inspection and analysis. It consists of 4,867,421 rows by 28 columns. Each row represents all sensor readings taken for a specific experiment, crew and pilot for a specific point in time as well as the corresponding cognitive state. The columns are as follows:

> crew – One of the nine unique crews throughout the experiment indicated by an integer. (1, 2, 3, 4, 5, 6, 7, 8, 13)
>
> experiment – One of the three experiments conducted. (CA, DA, SS)
>
> time – The specific timestamp from the experiment that all sensor readings in that row are from in terms of seconds into the experiment.
>
> seat – One of the two pilots in the crew, left seat (0) or right seat (1).
>
> eeg_fp1 – Readings in microvolts from one of the 20 eeg sensors placed in specific locations around the head.

eeg_f7 – Readings in microvolts from one of the 20 eeg sensors placed in specific locations around the head.

eeg_f8 – Readings in microvolts from one of the 20 eeg sensors placed in specific locations around the head.

eeg_t4 – Readings in microvolts from one of the 20 eeg sensors placed in specific locations around the head.

eeg_t6 – Readings in microvolts from one of the 20 eeg sensors placed in specific locations around the head.

eeg_t5 – Readings in microvolts from one of the 20 eeg sensors placed in specific locations around the head.

eeg_t3 – Readings in microvolts from one of the 20 eeg sensors placed in specific locations around the head.

eeg_fp2 – Readings in microvolts from one of the 20 eeg sensors placed in specific locations around the head.

eeg_o1 – Readings in microvolts from one of the 20 eeg sensors placed in specific locations around the head.

eeg_p3 – Readings in microvolts from one of the 20 eeg sensors placed in specific locations around the head.

eeg_pz – Readings in microvolts from one of the 20 eeg sensors placed in specific locations around the head.

eeg_f3 – Readings in microvolts from one of the 20 eeg sensors placed in specific locations around the head.

eeg_fz – Readings in microvolts from one of the 20 eeg sensors placed in specific locations around the head.

eeg_f4 – Readings in microvolts from one of the 20 eeg sensors placed in specific locations around the head.

eeg_c4 – Readings in microvolts from one of the 20 eeg sensors placed in specific locations around the head.

eeg_p4 – Readings in microvolts from one of the 20 eeg sensors placed in specific locations around the head.

eeg_poz – Readings in microvolts from one of the 20 eeg sensors placed in specific locations around the head.

eeg_c3 – Readings in microvolts from one of the 20 eeg sensors placed in specific locations around the head.

eeg_cz – Readings in microvolts from one of the 20 eeg sensors placed in specific locations around the head.

eeg_o2 – Readings in microvolts from one of the 20 eeg sensors placed in specific locations around the head.

ecg – 3-point electrocardiogram signal reading in microvolts.

r – Respiration, measure of the rise and fall of the chest, reading in microvolts.

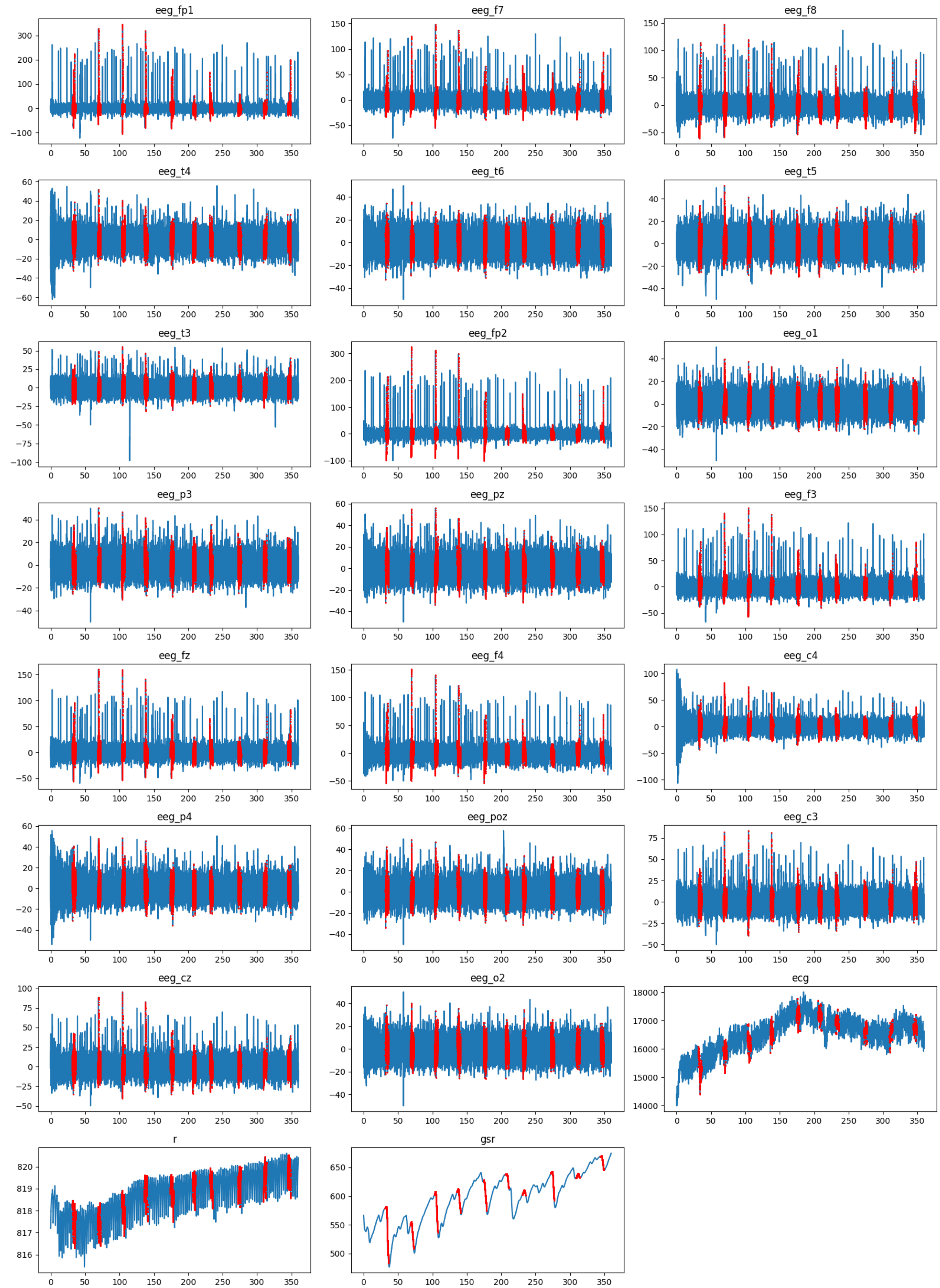
gsr – Galvanic Skin Response, measure of electrodermal activity in microvolts.

event – Cognitive state of the pilot, baseline (A), SS (B), CA (C), DA (D).

The data was further analyzed to determine any null values as well as descriptive statistics for each column such as the overall count of values, mean, standard deviation, min value, 25^{th} percentile, 5oth percentile, 75^{th} percentile and max value. Although none of the values were reported as null values there is still noise that may need further analysis to process out.

Raw data was plotted for various crews and experiment conditions to get a general visualization of the data for an experiment. The following plot is for Crew 3, Pilot 1 throughout the Diverted Attention experiment, with the baseline cognitive state in blue and the DA cognitive state in red.

Figure 1: Crew 3, Pilot 1, Diverted Attention plot



The data was grouped by crew, seat and event to then calculate entropy and complexity per feature for each individual pilot and event. For example, Crew 1, Seat 0, Event A would have an entropy and complexity value for each of the 23 physiological features. This was done for each group at a set window length in order to allow for more samples to be broken down and calculate within each experiment. For example, of the 117,442 rows of data from crew 1, seat 0 with event A was split up into multiple smaller windows. With this a new dataframe was created with 1149 rows and 49 columns. This data was again checked for null values which were then replaced with the column's mean value.

Modeling

The overall goal was to use machine learning models to classify the cognitive state of a pilot given the physiological data. The data was modeled using multiple supervised machine learning algorithms and the train test split method with 20% of the data being set aside and used for the model testing. For each model all 46 features, entropy and complexity for each of the 23 physiological factors (20 channels of eeg, ecg, r and gsr), were used as the independent predictor variables and the cognitive state event as the dependent response variable. Baseline models were completed for Logistic Regression, Neural Networks Multilayer Perceptron (MLP) and Decision Tree methods where each model was evaluated using precision, recall, f1-score and accuracy. The same three methods were used again using the GridsearchCV technique for hyperparameter tuning.

For the hyperparameter tuning of the Logistic Regression model the parameters varied were C from logspace(-4, 4, 20), penalty between l1 and l2, and the solver between sag, saga, liblinear and newton-cg.

For the hyperparameter tuning of the Neural Network MLP model the parameters varied were hidden layer size from one hidden layer with 50 neurons, one hidden layer with 100 neurons, two hidden layers with 50 neurons each and two hidden layers one with 100 neurons and the other with 50 neurons. The activation function was varied from relu, to tanh and logistic, the solver parameter varied between sgd and adam, alpha parameter varied between 0.0001, 0.001 and 0.01, and lastly the learning rate from constant to adaptive.

For the hyperparameter tuning of the Decision Tree model the max depth parameter varied from none to 10, 15, 20 and 30, minimum sample split from 2, 5, and 10, minimum samples per leaf varied from 1 to 2 and 4, and the criterion varied between gini and entropy.

Results

The baseline Logistic Regression model was 0.72 accurate and took 0.0514 seconds to run. The classification report for the test set can be seen below in Figure 2.

Figure 2: Baseline Logistic Regression Classification Report

	precision	recall	f1-score	support
Baseline	0.72	0.84	0.78	131
Channelized-Attention	0.73	0.67	0.70	84
Diverted-Attention	0.00	0.00	0.00	12
Startle/Surprise	0.00	0.00	0.00	3
accuracy			0.72	230
macro avg	0.36	0.38	0.37	230
weighted avg	0.68	0.72	0.70	230

The baseline Neural Network MLP model was 0.84 accurate and took 0.8474 seconds to run. The classification report for the test set can be seen below in Figure 3.

Figure 3: Baseline Neural Network MLP Classification Report

	precision	recall	f1-score	support
Baseline	0.83	0.92	0.87	131
Channelized-Attention	0.86	0.85	0.85	84
Diverted-Attention	0.67	0.17	0.27	12
Startle/Surprise	0.00	0.00	0.00	3
accuracy			0.84	230
macro avg	0.59	0.48	0.50	230
weighted avg	0.82	0.84	0.82	230

The baseline Decision Tree model was 0.70 accurate and took 0.1183 seconds to run. The classification report for the test set can be seen below in Figure 4.

Figure 4: Baseline Decision Tree Classification Report

	precision	recall	f1-score	support
Baseline	0.77	0.76	0.76	131
Channelized-Attention	0.70	0.74	0.72	84
Diverted-Attention	0.12	0.08	0.10	12
Startle/Surprise	0.00	0.00	0.00	3
accuracy			0.70	230
macro avg	0.40	0.39	0.39	230
weighted avg	0.70	0.70	0.70	230

The hyperparameter tuned Logistic Regression model with the best accuracy was with the parameters of C = 29.764, penalty = l2, and solver = liblinear and had an accuracy of 0.726 with 459.77 seconds of run time.

The hyperparameter tuned MLP model with the best accuracy had the parameters of activation = relu, alpha = 0.0001, hidden layer size = (100, 50), learning rate = adaptive and solver = sgd achieving an accuracy of 0.848 with 2817.18 seconds of run time.

The hyperparameter tuned Decision Tree model with the best accuracy had the parameters of criterion = gini, max depth = 15, min samples per leaf = 4, and min samples per split = 2 achieving an accuracy of 0.729 with a run time of 37.60 seconds.

The best overall model as the baseline Multilayer Perceptron model with 84% accuracy. The models did not perform as well predicting the Diverted Attention and Startle/Surprise states most likely due to the lower sample sizes of those conditions.

Future Work

Due to the environment in which pilots operate it would be ideal to determine a physiological factor that can not only accurately help predict the cognitive state but one that can also be measured and recorded in a simpler manner. For example, it is more convenient and likely to have a pilot wear a wrist or chest device with an ecg sensor similar to an apple watch, whoop strap or polar chest strap than have their head covered in eeg electrodes. Therefore it will be most beneficial to focus on the ecg, r and gsr features as the predictors, enhance the feature extraction methods and apply additional modeling methods. Additionally being able to record, model, predict and interpret the result in real time could help pilots identify potential unknown issues they are experiencing and allow them to make a change before an accident were to occur.

Project Code

https://github.com/JeffJ22/IE7275-project/tree/main

One file includes code for models predicting all four of the original cognitive states (baseline, channelized attention, diverted attention and startle/surprise) and one file includes code modeling predictions for binary states (baseline, non-baseline).

References

Johnson, A., Sarle, J., Maggie, Terwilliger, P., Walker, S., Cukierski, W., & Espenschied, W. (2018). Reducing Commercial Aviation Fatalities. https://www.kaggle.com/competitions/reducing-commercial-aviation-fatalities/overview

Johnston, J., Karnik, S., Randad, H., Raut, D. A., & Xavier, J. A. (2023). *IE6400 Foundations of Data Analytics, Project 3*.

Radhakrishnan, S. (2018). Permutation-entropy/p_entropy.py . https://github.com/srk-srinivasan/Permutation-Entropy/blob/master/p_entropy.py