**2. Objective**

The project aims to perform Binary Classification AND variable selection Analysis on EEG Data.

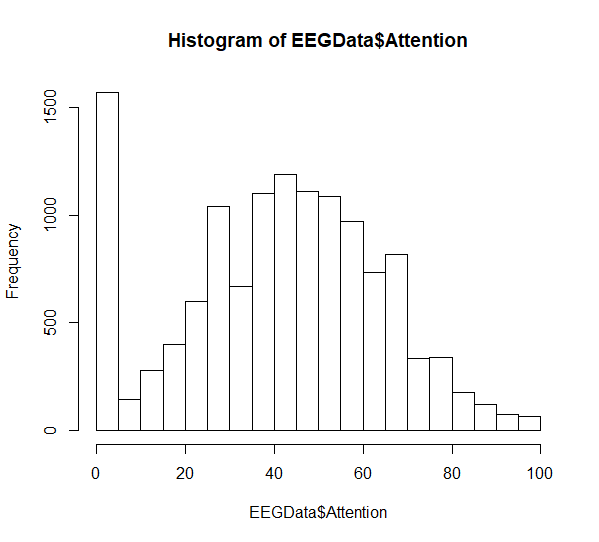
**4. Associated Data Files**

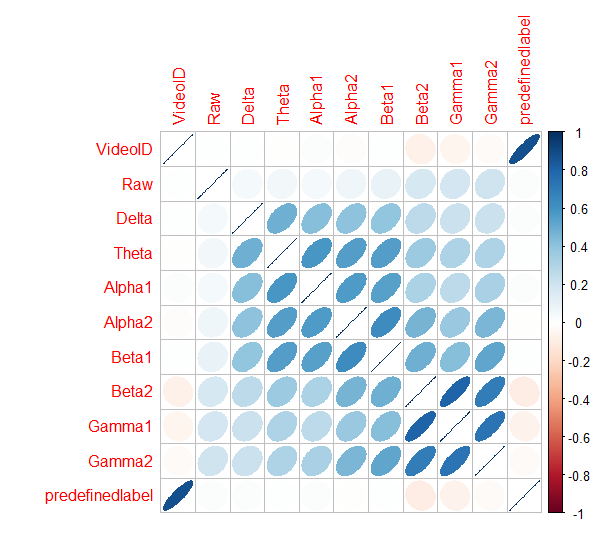
https://drive.google.com/drive/folders/1uePUTXQRuWOz9gWVRRH7P2Dr3E0HrPiP?usp=sharing

**5. Problem Statement**

Use R Functions

1. Perform Binary Classification in this data set . 65% of prediction accuracy is quite decent.





> logistic<- glm(Attention~VideoID+Mediation+Raw+Delta+Theta+Alpha1+Alpha2+Beta1+Beta2+Gamma1+Gamma2+predefinedlabel+user.definedlabeln, data = train\_proj)

> summary(logistic)

Call:

glm(formula = Attention ~ VideoID + Mediation + Raw + Delta +

Theta + Alpha1 + Alpha2 + Beta1 + Beta2 + Gamma1 + Gamma2 +

predefinedlabel + user.definedlabeln, data = train\_proj)

Deviance Residuals:

Min 1Q Median 3Q Max

-54.732 -13.007 -1.616 11.799 75.756

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 2.326e+01 6.143e-01 37.856 < 2e-16 \*\*\*

VideoID -1.453e+00 1.246e-01 -11.663 < 2e-16 \*\*\*

Mediation 5.721e-01 8.501e-03 67.302 < 2e-16 \*\*\*

Raw -4.534e-04 2.762e-04 -1.642 0.100678

Delta -3.578e-06 3.483e-07 -10.272 < 2e-16 \*\*\*

Theta -3.768e-06 1.070e-06 -3.523 0.000429 \*\*\*

Alpha1 -2.620e-05 3.577e-06 -7.325 2.56e-13 \*\*\*

Alpha2 -2.921e-05 4.727e-06 -6.180 6.67e-10 \*\*\*

Beta1 2.986e-05 6.775e-06 4.407 1.06e-05 \*\*\*

Beta2 5.344e-06 4.367e-06 1.224 0.221065

Gamma1 -1.875e-05 4.086e-06 -4.590 4.48e-06 \*\*\*

Gamma2 -3.857e-06 7.707e-06 -0.501 0.616705

predefinedlabel 6.694e+00 7.292e-01 9.181 < 2e-16 \*\*\*

user.definedlabeln -2.953e+00 3.653e-01 -8.084 6.97e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 324.7618)

Null deviance: 5925067 on 10248 degrees of freedom

Residual deviance: 3323937 on 10235 degrees of freedom

AIC: 88372

Number of Fisher Scoring iterations: 2

> logistic1<- glm(Attention~VideoID+Mediation+Raw+Delta+Theta+Alpha1+Alpha2+Beta1+Beta2+Gamma1+Gamma2+predefinedlabel+user.definedlabeln, data = test\_proj)

> summary(logistic1)

Call:

glm(formula = Attention ~ VideoID + Mediation + Raw + Delta +

Theta + Alpha1 + Alpha2 + Beta1 + Beta2 + Gamma1 + Gamma2 +

predefinedlabel + user.definedlabeln, data = test\_proj)

Deviance Residuals:

Min 1Q Median 3Q Max

-70.54 -12.31 -0.98 10.59 60.82

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 4.824e+01 1.420e+00 33.977 < 2e-16 \*\*\*

VideoID -2.606e-01 2.497e-01 -1.044 0.296780

Mediation 4.002e-02 2.169e-02 1.845 0.065092 .

Raw -1.618e-03 1.840e-03 -0.880 0.379131

Delta -1.388e-06 5.890e-07 -2.356 0.018539 \*

Theta -1.785e-05 1.818e-06 -9.817 < 2e-16 \*\*\*

Alpha1 -2.231e-05 5.764e-06 -3.871 0.000111 \*\*\*

Alpha2 -4.488e-05 7.887e-06 -5.691 1.41e-08 \*\*\*

Beta1 3.178e-06 1.475e-05 0.215 0.829452

Beta2 8.248e-05 7.257e-06 11.366 < 2e-16 \*\*\*

Gamma1 3.764e-05 1.560e-05 2.413 0.015898 \*

Gamma2 2.562e-04 3.441e-05 7.446 1.31e-13 \*\*\*

predefinedlabel -2.087e+00 1.464e+00 -1.426 0.154073

user.definedlabeln -5.919e+00 7.335e-01 -8.069 1.08e-15 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 305.4749)

Null deviance: 928323 on 2561 degrees of freedom

Residual deviance: 778350 on 2548 degrees of freedom

(1 observation deleted due to missingness)

AIC: 21946

Number of Fisher Scoring iterations: 2

2. Carry out the variable selection (causal inference) task that may help further research

Note: You need to submit associate R project along with the screenshot.