step 0 : Import Data
(5511 rows, 8 cols)

step 1 : Data Manipulation

Go to test set

remove outlier Q1, Q3

Split to Train
(2,332 rows)

Split to Test
(994 rows)

step 2 : stat plot

Box plot
for each feature

step 3 : Fitted Model
Baseline & Quadratic

Step 4 : Log the AIC score and choose top 3

Step 5 : Evaluation
The Accuracy
TN + TF / Total

Log model performance
Then
Sort

	Model	Cut_off	Accuracy
1	model_quad19	0.5	0.8306773
2	model_quad18	0.5	0.7988048
3	model_quad11	0.5	0.7958167
4	model_quad11	0.7	0.7818725
5	model_quad18	0.7	0.7808765
6	model_quad19	0.7	0.7798805
7	model_quad19	0.9	0.5219124
8	model_quad18	0.9	0.5079681
9	model_quad11	0.9	0.5029880

Step 6. Predict new data by our best model

Finish

	Model	AIC_Score
1	model_quad11	2102.120
2	model_quad19	2108.766
3	model_quad18	2113.795
4	model_quad17	2116.842
5	model_quad14	2125.234
6	model_quad15	2127.229
7	model_quad16	2132.018
8	model_quad10	2133.670
9	model_quad9	2134.905
10	model_quad13	2136.589
11	model_quad12	2151.254
12	model_quad23	2164.579
13	model_quad22	2164.633
14	model_quad21	2164.636
15	model_quad7	2175.036
16	model_quad6	2176.207
17	model_quad20	2176.462
18	model_quad 0	2179.125
19	model_quad4	2180.249
20	baseline_model 1	2180.276
21	model_quad2	2180.576
22	model_quad1	2181.084
23	model_quad8	2181.520
24	model_quad5	2182.200
25	model_quad3	2182.302

Quadratic Experiment

```
model_quad11 <- glm(formula = fund_return_2020_q3 ~ equity_style_score +
                      equity_size_score +
                      environmental_score + I(environmental_score^2) + I(environmental_score^3) +  I(environmental_score^4) +  I(environmental_score^5) +
                      social_score +
                      governance_score +
                      sustainability_score,
                    family = "binomial",
                    data = data)
```

```
model_quad19 <- glm(formula = fund_return_2020_q3 ~ equity_style_score +
                      equity_size_score +
                      environmental_score +
                      social_score +
                      governance_score + I(governance_score^2) + I(governance_score^3) + I(governance_score^4) + I(governance_score^5) +
                      sustainability_score,
                    family = "binomial",
                    data = data)
```

```
model_quad18 <- glm(formula = fund_return_2020_q3 ~ equity_style_score +
                      equity_size_score +
                      environmental_score +
                      social_score +
                      governance_score + I(governance_score^2) + I(governance_score^3) + I(governance_score^4) +
                      sustainability_score,
                    family = "binomial",
                    data = data)
```

Baseline

```
# model Base line
model1 <- glm(fund_return_2020_q3 ~ equity_style_score + equity_size_score + environmental_score + social_score + governance_score + sustainability_score ,
              data = data,
              family="binomial")
```