1. 랜덤포레스트

1-1 결측치 점수=0

1) train데이터 학습모델 -> train 데이터 예측

```
> confusionMatrix(predict(model),train$단계)
Confusion Matrix and Statistics

          Reference
Prediction 경계 위험 주의 초기
      경계   33    1    0    0
      위험    0    0    0    0
      주의    6    0  585    5
      초기    0    0   19 2021

Overall Statistics

               Accuracy : 0.9884
                 95% CI : (0.9836, 0.9921)
    No Information Rate : 0.7588
    P-Value [Acc > NIR] : < 0.00000000000000022

                  Kappa : 0.9686

 Mcnemar's Test P-Value : NA

Statistics by Class:

                     Class: 경계 Class: 위험 Class: 주의 Class: 초기
Sensitivity             0.84615  0.0000000      0.9685      0.9975
Specificity             0.99962  1.0000000      0.9947      0.9705
Pos Pred Value          0.97059        NaN      0.9815      0.9907
Neg Pred Value          0.99772  0.9996255      0.9908      0.9921
Prevalence              0.01461  0.0003745      0.2262      0.7588
Detection Rate          0.01236  0.0000000      0.2191      0.7569
Detection Prevalence    0.01273  0.0000000      0.2232      0.7640
Balanced Accuracy       0.92289  0.5000000      0.9816      0.9840
```

2) train데이터 학습모델 -> test 데이터 예측

```
> confusionMatrix(pred,test$단계)
Confusion Matrix and Statistics

          Reference
Prediction 경계 위험 주의 초기
      경계   14    1    1    0
      위험    0    0    0    0
      주의    3    0  253    6
      초기    0    0    5  863

Overall Statistics

               Accuracy : 0.986
                 95% CI : (0.9774, 0.992)
    No Information Rate : 0.7583
    P-Value [Acc > NIR] : < 0.00000000000000022

                  Kappa : 0.9626

 Mcnemar's Test P-Value : NA

Statistics by Class:

                     Class: 경계 Class: 위험 Class: 주의 Class: 초기
Sensitivity             0.82353  0.0000000      0.9768      0.9931
Specificity             0.99823  1.0000000      0.9899      0.9819
Pos Pred Value          0.87500        NaN      0.9656      0.9942
Neg Pred Value          0.99735  0.9991274      0.9932      0.9784
Prevalence              0.01483  0.0008726      0.2260      0.7583
Detection Rate          0.01222  0.0000000      0.2208      0.7531
Detection Prevalence    0.01396  0.0000000      0.2286      0.7574
Balanced Accuracy       0.91088  0.5000000      0.9833      0.9875
```
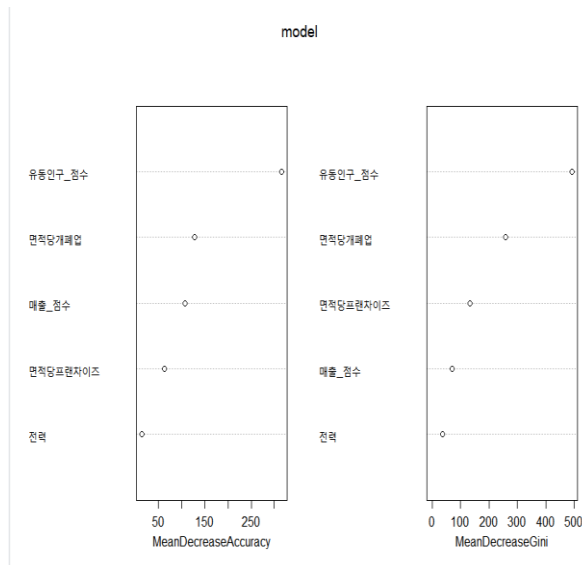
3) Varimpplot

1-2 결측치 = -1

1) model -> train 데이터 예측

```
> confusionMatrix(predict(model),train$단계)
Confusion Matrix and Statistics

           Reference
Prediction 경계 위험 주의 초기
     경계   29    1    1    0
     위험    0    0    0    0
     주의   10    0  574    8
     초기    0    0   17 2030

Overall Statistics

               Accuracy : 0.9861
                 95% CI : (0.9809, 0.9902)
    No Information Rate : 0.7633
    P-Value [Acc > NIR] : < 0.00000000000000022

                  Kappa : 0.9621

 Mcnemar's Test P-Value : NA

Statistics by Class:

                     Class: 경계 Class: 위험 Class: 주의 Class: 초기
Sensitivity             0.74359   0.0000000      0.9696      0.9961
Specificity             0.99924   1.0000000      0.9913      0.9731
Pos Pred Value          0.93548         NaN      0.9696      0.9917
Neg Pred Value          0.99621   0.9996255      0.9913      0.9872
Prevalence              0.01461   0.0003745      0.2217      0.7633
Detection Rate          0.01086   0.0000000      0.2150      0.7603
Detection Prevalence    0.01161   0.0000000      0.2217      0.7667
Balanced Accuracy       0.87141   0.5000000      0.9805      0.9846
```

2) model -> test 데이터 예측

```
> confusionMatrix(pred,test$단계)
Confusion Matrix and Statistics

          Reference
Prediction 경계 위험 주의 초기
     경계  16    1    0    0
     위험   0    0    0    0
     주의   1    0  251    4
     초기   0    0    3  870

Overall Statistics

              Accuracy : 0.9921
                95% CI : (0.9851, 0.9964)
    No Information Rate : 0.7627
    P-Value [Acc > NIR] : < 0.00000000000000022

                 Kappa : 0.9787

 Mcnemar's Test P-Value : NA

Statistics by Class:

                     Class: 경계 Class: 위험 Class: 주의 Class: 초기
Sensitivity            0.94118   0.0000000     0.9882     0.9954
Specificity            0.99911   1.0000000     0.9944     0.9890
Pos Pred Value         0.94118         NaN     0.9805     0.9966
Neg Pred Value         0.99911   0.9991274     0.9966     0.9853
Prevalence             0.01483   0.0008726     0.2216     0.7627
Detection Rate         0.01396   0.0000000     0.2190     0.7592
Detection Prevalence   0.01483   0.0000000     0.2234     0.7618
Balanced Accuracy      0.97015   0.5000000     0.9913     0.9922
```
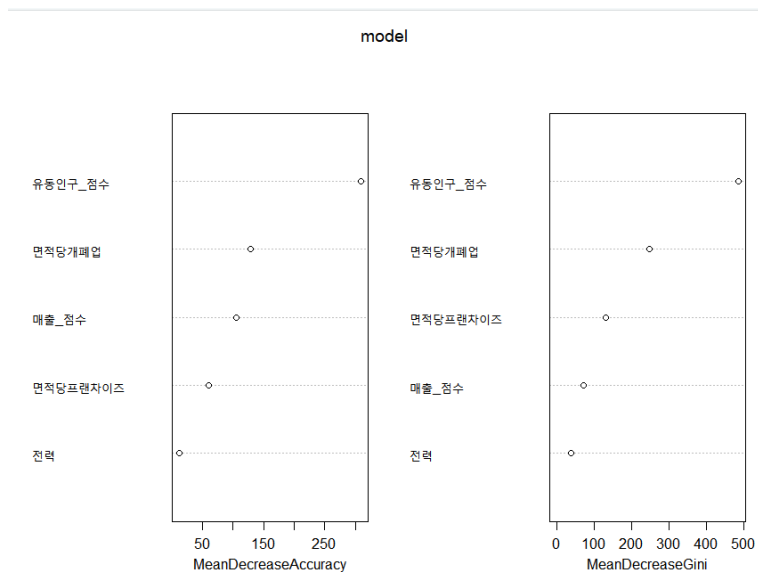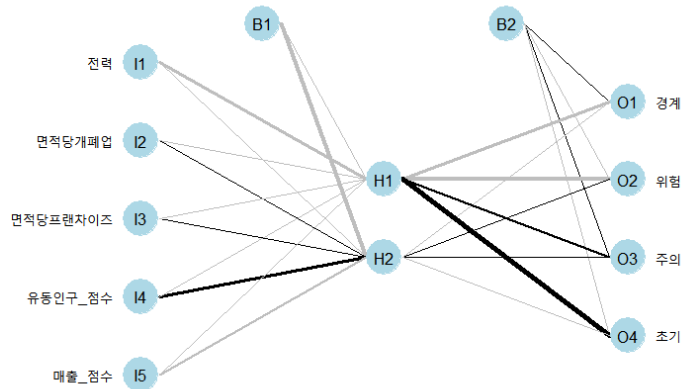
3) varimpplot

결론 : model 정확도가 굉장히 높지만 위험지역 분류에 대한 정확성이 낮음

근데 전력의 중요도가 굉장히 낮음 그냥 상관이 없는 수준

2. 인공신경망

2-1 결측치 = 0

Train 예측

```
> confusionMatrix(as.factor(pred),train$단계)
Confusion Matrix and Statistics

          Reference
Prediction 경계 위험 주의 초기
     경계   21    1   11    0
     위험    0    0    0    0
     주의   18    0  546   54
     초기    0    0   47 1972

Overall Statistics

               Accuracy : 0.9509
                 95% CI : (0.942, 0.9588)
    No Information Rate : 0.7588
    P-Value [Acc > NIR] : < 0.00000000000000022

                  Kappa : 0.8687

 Mcnemar's Test P-Value : NA

Statistics by Class:

                     Class: 경계 Class: 위험 Class: 주의 Class: 초기
Sensitivity            0.538462   0.0000000      0.9040      0.9733
Specificity            0.995439   1.0000000      0.9652      0.9270
Pos Pred Value         0.636364         NaN      0.8835      0.9767
Neg Pred Value         0.993174   0.9996255      0.9717      0.9171
Prevalence             0.014607   0.0003745      0.2262      0.7588
Detection Rate         0.007865   0.0000000      0.2045      0.7386
Detection Prevalence   0.012360   0.0000000      0.2315      0.7562
Balanced Accuracy      0.766950   0.5000000      0.9346      0.9502
```

Test 예측

```
> confusionMatrix(as.factor(pred),test$단계)
Confusion Matrix and Statistics

          Reference
Prediction 경계 위험 주의 초기
      경계   10    1    6    0
      위험    0    0    0    0
      주의    7    0  225   20
      초기    0    0   28  849

Overall Statistics

               Accuracy : 0.9459
                 95% CI : (0.9312, 0.9583)
    No Information Rate : 0.7583
    P-Value [Acc > NIR] : < 0.00000000000000022

                  Kappa : 0.8537

 Mcnemar's Test P-Value : NA

Statistics by Class:

                     Class: 경계 Class: 위험 Class: 주의 Class: 초기
Sensitivity             0.588235   0.0000000      0.8687      0.9770
Specificity             0.993800   1.0000000      0.9696      0.8989
Pos Pred Value          0.588235         NaN      0.8929      0.9681
Neg Pred Value          0.993800   0.9991274      0.9620      0.9257
Prevalence              0.014834   0.0008726      0.2260      0.7583
Detection Rate          0.008726   0.0000000      0.1963      0.7408
Detection Prevalence    0.014834   0.0000000      0.2199      0.7653
Balanced Accuracy       0.791018   0.5000000      0.9191      0.9380
```
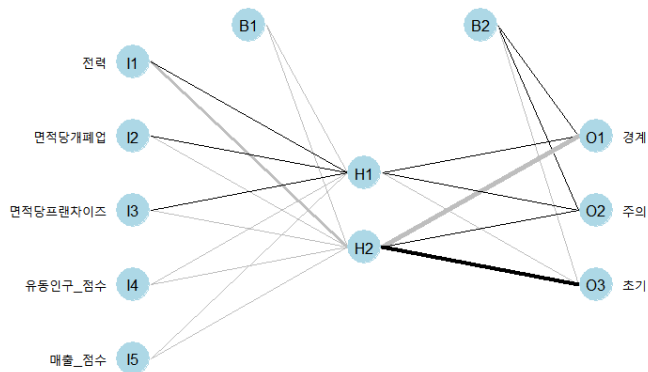
2-2 결측치 =-1

Train 예측

```
> confusionMatrix(as.factor(pred),train$단계)
Confusion Matrix and Statistics

          Reference
Prediction 경계 위험 주의 초기
      경계   16    1    5    0
      위험    0    0    0    0
      주의   23    0  533   53
      초기    0    0   54 1985

Overall Statistics

               Accuracy : 0.9491
                 95% CI : (0.94, 0.9571)
    No Information Rate : 0.7633
    P-Value [Acc > NIR] : < 0.00000000000000022

                  Kappa : 0.861

 Mcnemar's Test P-Value : NA

Statistics by Class:

                     Class: 경계 Class: 위험 Class: 주의 Class: 초기
Sensitivity             0.410256   0.0000000      0.9003      0.9740
Specificity             0.997719   1.0000000      0.9634      0.9146
Pos Pred Value          0.727273         NaN      0.8752      0.9735
Neg Pred Value          0.991314   0.9996255      0.9714      0.9160
Prevalence              0.014607   0.0003745      0.2217      0.7633
Detection Rate          0.005993   0.0000000      0.1996      0.7434
Detection Prevalence    0.008240   0.0000000      0.2281      0.7637
Balanced Accuracy       0.703988   0.5000000      0.9319      0.9443
```

Test 예측

```
> confusionMatrix(as.factor(pred),test$단계)
Confusion Matrix and Statistics

          Reference
Prediction 경계 위험 주의 초기
      경계   9    1    3    0
      위험   0    0    0    0
      주의   8    0  230   23
      초기   0    0   21  851

Overall Statistics

               Accuracy : 0.9511
                 95% CI : (0.937, 0.9629)
    No Information Rate : 0.7627
    P-Value [Acc > NIR] : < 0.00000000000000022

                  Kappa : 0.8676

 Mcnemar's Test P-Value : NA

Statistics by Class:

                     Class: 경계 Class: 위험 Class: 주의 Class: 초기
Sensitivity             0.529412   0.0000000     0.9055      0.9737
Specificity             0.996457   1.0000000     0.9652      0.9228
Pos Pred Value          0.692308         NaN     0.8812      0.9759
Neg Pred Value          0.992939   0.9991274     0.9729      0.9161
Prevalence              0.014834   0.0008726     0.2216      0.7627
Detection Rate          0.007853   0.0000000     0.2007      0.7426
Detection Prevalence    0.011344   0.0000000     0.2277      0.7609
Balanced Accuracy       0.762934   0.5000000     0.9354      0.9482
```

정확도 높음 그러나 여전히 위험 단계를 잘 분류하지 못함

3.