

Explain Modeling

By learning data in 16+17+18(70%) and predicting 30% in 2018.

The label value of the data showed a value of 3:1, and it was divided into 0.2:1.0 through sampling.

The accuracy of the results was not high, but the accuracy of data 0 and prediction 0 and the accuracy of data 1 and prediction 1 were close to 90%

You can check it out.

So, when we looked at CART depth5, we found out which of these prediction variables played an important role

It was found that retained earnings, net profit, equity capital growth rate, and the number of employees acted as major variables.

16+17+18년도 70% 데이터를 학습시켜 18년도 30%를 예측해보았다.

데이터의 label 값이 3:1 의 수치를 보여 샘플링을 통해서 0.2 : 1.0 으로 나누어 보았다.

결과값의 정확도는 높지 않았지만 데이터0과 예측0 의 정확도와 데이터1과 예측1의 정확도는 90% 가까이 나온것을

확인할 수 있다.

그래서 CART depth5를 통해 이러한 예측의 변수중에 어떤 변수가 중요한 요인으로 작용하였는지 확인해보니

이익잉여금, 당기순이익, 자기자본증가율, 종업원수가 주요변수로 작용함을 알 수 있었다.