

# Project Description

## Project overview

How much the company can pay out the dividend to the investors. That's the one of the keys to decide the invest. So, we need to know what kind of financial statement effect to dividend is for wise invest decision making. Using the KOSPI, KOSDAQ listing company's financial statement to find out what kind of statements are sufficient to predict the dividend. And make the model to test the statements what we chose is important factor in investment decisions.

Program I use: Rapidminer, KISVALUE

KISVALUE – The program provides the company's financial data as CSV in the Korea stock market.

Model I use: Logistic Regression, Decision Tree(C5, CART), Neural Network

### 1. Logistic Regression

- Reason why I choose: Dependent variable is binary. So, it's easy to categorize the plus and minus value of the data

### 2. Decision Tree – C5

- The root node decision of C5 determines that the more pure the value of the node of the variable, the clearer the true and false values of the value, and that more accurate analysis is possible with less information.

### 3. Decision Tree – CART

- In CART, if the nodes are pure, they are grouped into the same class, and this probability is considered 1. Therefore, the larger the CART value, the greater the homogeneity and the greatest homogeneity is the root node.

### 4. Neural Network

- Through the number of hidden layers and the number of hidden nodes, prediction through learning between variables is possible. The most effective modeling is modeling using the least nodes while showing the highest predictive accuracy

Data: 2016, 2017, 2018 KOSPI, KOSDAQ-listed non-financial company

- Independent variable(binary): Increase or decrease of dividend (increase: 1, decrease: 0)
- Dependent variable: retained earnings, dividends per share, dividend, cash dividend, current ratio, quick ratio, net income, debt ratio, gross sales, gross sales per capita, accounts receivables turnover, inventory turnover, revenue growth, ROE, internal possession rate of stocks, beta(5Y Monthly), market cap, capita, credit score, company size

## 분석기법 설명

분석방법은 로지스틱 회귀분석, 의사결정나무를 통한 C5, CART, 신경망 분석을 사용하였다.

회귀분석은 독립변수와 종속 변수간의 관계를 모형으로 설정하여 두 변수간의 상관관계를 예측, 분류하는 방법이다. 그중 로지스틱 회귀분석을 선택한 이유는 종속변수가 이산변수(binary)로 증감의 여부(0, 1)인 경우이기에 이를 분류하는데 적합한 오퍼레이터이기 때문이다.

의사결정나무의 C5, CART를 사용하였다. 의사결정나무는 전체의 데이터를 변수에 따른 작은 집단으로 재귀 분할하여 예측을 하는 모델이다. 이러한 분석을 통해서 생기는 루트 노드가 종속변수와 관계 그리고 예측모델에서의 상관관계를 분석하기에 용이했다. 이러한 재귀적 분할에서 루트노드의 결정 이유에 대해 C5와 CART로 나누어 설명이 가능하다.

C5의 루트노드 결정은 변수의 노드의 값이 순수할수록 그 값에 대한 True와 False가 분명해진다고 판단하고 더 적은 정보를 가지고 더 정확한 분석이 가능하다고 판단한다.

CART에서는 노드가 순수하다면 같은 클래스로 묶어진다고 보고, 이 확률을 1로 본다. 그래서 CART값이 클수록 균질성이 크기에 가장 큰 균질성을 보여주는 것이 루트노드가 된다.

신경망 분석 같은 경우는 은닉층의 개수와 은닉노드의 개수를 통해서 변수간의 학습을 통한 예측을 가능케한다.

가장 효과적인 모델링은 가장 높은 예측정확도를 보이면서도 가장 적은 노드를 사용한 모델링이다.

