



빅데이터 예측분석 12주

서강대학교 정보통신 대학원 정 화민 교수 (MIS Ph.D.)



Decision Tree

* 의사결정 흐름

- 의사결정자는 불확실한 상황에서 끈임없이 신속하고 정확한 의사결정을 해야함.
- 의사결정자는 고객관련 수많은 행동 결과의 자료를 이용하여 자료 간의 관련성, 유사성 등을 고려하여 고객을 분류하고 예측할 필요가 있다.
- 의사결정자는 고객이 우량 고객인지 불량 고객인지 분류할 수 있다.
- 이들 고객군마다 상이한 전략을 구사할 수 있다.
- 고객관련 자료를 분류하고 예측하는 것을 넘어 고객과의 관계를 강화하는 것이 추세이다
- 고객과의 관계를 강화하여 고객에게는 만족을 제고하고 이결과 고객 충성도를 유도하고 수익을 창출하려고 하는 고객관계 경영이다.

* 의사결정나무분석(Decision Tree Analysis)

- 자료를 탐색하여 분류/예측하여 이를 모형화 하여 고객과의 관계를 강화하는데 사용되는 의사결정 방법이다.

Decision Tree

- 의사결정나무는 말 그대로 나무를 거꾸로 세워 놓은 구조로 뿌리가 상단에 위치하고 하단에는 나뭇가지와 잎이 연결되어 있는 형태를 이룬다.
- 의사결정나무 상단에 놓여있는 뿌리를 뿌리마디(Root node) 또는 부모마디(Parent node)라고 한다.
- 나무에서 뿌리 마디와 끝 마디 사이를 중간마디(Internal node)라고 한다.
- 의사결정 나무에서는 이를 자식마디(Child node)라고 한다.
- 의사결정 나무에서 마디와 마디는 가지(Branch)로 연결되어 있다.
- 의사결정나무에서 고객관련 정보가 어느 수준까지 이루어졌는지를 알아보기 위해서는 깊이(Depth)라는 단어를 쓴다.

* 의사결정나무분석(Decision Tree Analysis)의 장점/단점

- 질적변수나 양적변수의 자료 분석이 가능하며 다변량 분석의 기본 가정인 선형성, 정규분포성, 등분산성을 따르지 않아도 된다.
- 분석결과가 나무구조로 되어 있어 해석이 용이하다.
- 단점은 분석결과가 표본의 크기에 영향을 받기 쉽고 연속변수 사용이 많으면 많을 수록 모형의 예측력이 떨어질 수 있다.
- 분석에 무리하게 많은 변수를 투입하면 과대 적합이 발생할 수 있다.

* 의사결정 나무 분석의 종류

- CHAID(Chi-square automatic interaction detection):
목표변수가 질적 변수이거나 양적 변수이며 예측변수는 질적 변수인 경우에 사용된다.
분리기준으로 카이제곱 통계량이나 F검증을 사용한다.
목표변수가 질적 변수인 경우는 카이제곱 통계량을 이용하는데 카이제곱 통계량이 크려 이에 대한 확률이 0.05 미만일 경우 부모 마디는 자식마디를 형성하게 된다.
- 그 외 Exhaustive CHAID, CRT(Classification and Regression tree), QUEST(Quick, Unbiased, Efficient Statistical Tree)등이 있다

Decision Tree

* 사례

석사학위논문

전신마취 수술 후 수술실에서의 마취각성 지연 영향요인

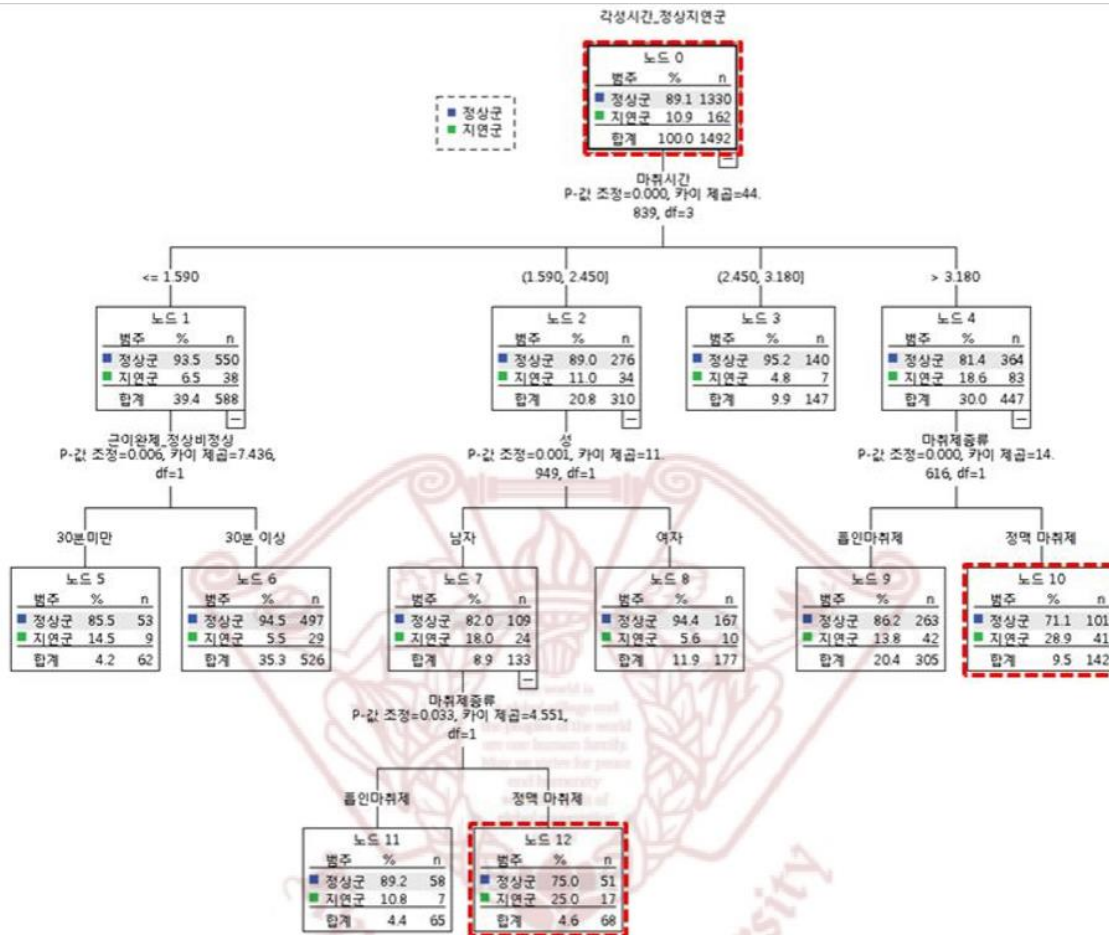
지도교수 윤 은 경

경희대학교 대학원

간호학과

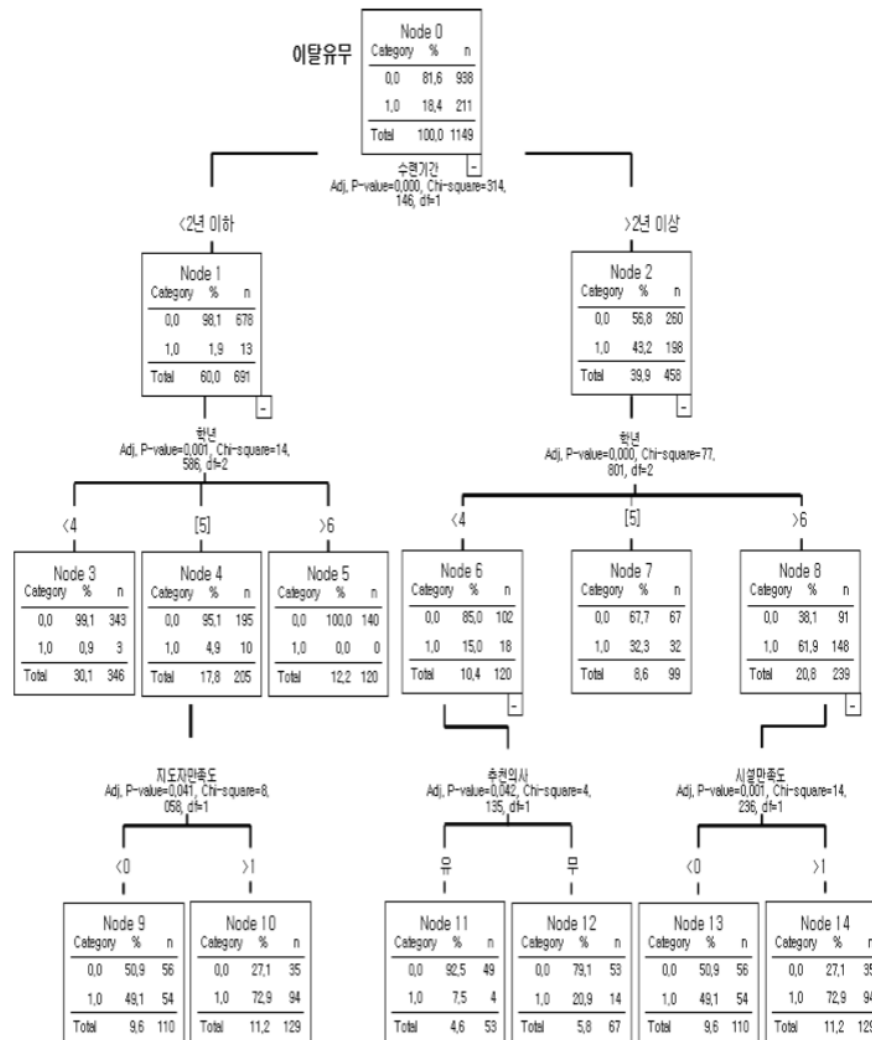
김 심 선

2015년 8월



Decision Tree

* 사례



석사학위논문

의사결정나무와 로지스틱 회귀분석을
이용한 태권도 수련생 이탈 예측을
위한 비교 연구

Comparative Analysis of Prediction Taekwondo Trainee's
Defection using Decision Tree and Logistic Regression.

지도교수 권태원

이 논문을 체육학 석사학위논문으로 제출합니다.

2007년 8월

한양대학교 일반대학원
생활스포츠학과
구유희

Decision Tree

* 의사결정나무 기초 동영상 (유튜브)사례/ 머신러닝 개념 적용

<https://www.youtube.com/watch?v=jFAXaFbPApI>

<https://rpubs.com/jmhome/DecisionTree> (R을 사용한 의사결정나무분석)

1. 의사결정나무 (Decision Tree)

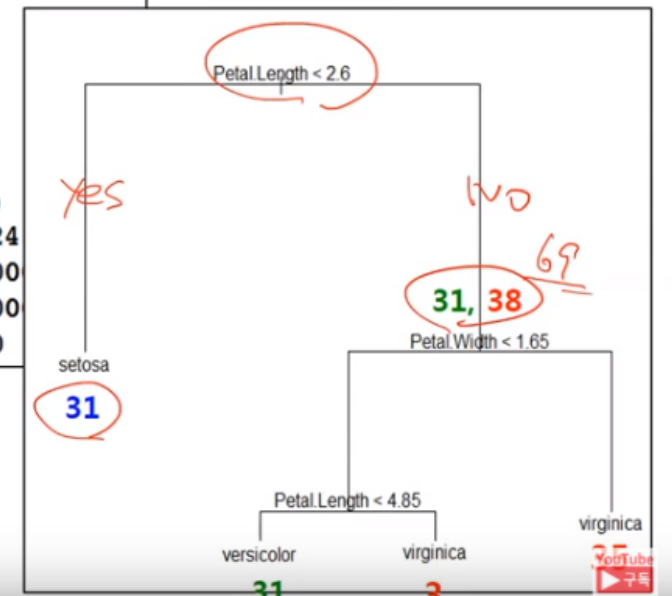
12.1 Decision Tree

• 학습데이터의 tree결과

• tree의 결과 (*는 터미널노드) : 마디 6에서는 더이상 분지할 필요 없음

```
> treemod<-tree(Species~., data=train)
> treemod
node), split, n, deviance, yval, (yprob)
* denotes terminal node
```

```
1) root 100 218.80 virginica ( 0.31000 0.31000 0.38000 )
2) Petal.Length < 2.6 31 0.00 setosa ( 1.00000 0.00000 0.00000 ) *
3) Petal.Length > 2.6 69 94.94 virginica ( 0.00000 0.44928 0.55072 )
6) Petal.Width < 1.65 34 20.29 versicolor ( 0.00000 0.91176 0.08824 )
12) Petal.Length < 4.85 29 0.00 versicolor ( 0.00000 1.00000 0.00 )
13) Petal.Length > 4.85 5 6.73 virginica ( 0.00000 0.40000 0.6000 )
7) Petal.Width > 1.65 35 0.00 virginica ( 0.00000 0.00000 1.00000 )
```



Decision Tree

* 의사결정나무 기초 (머신러닝) 동영상 (유튜브)사례
<https://www.youtube.com/watch?v=n0p0120Gxqk>

We know family photo has more than 1 person

