

Data Mining

Do You Have A Girl-Friend?

201203146 임승한



인천대학교 컴퓨터공학부
INCHEON NATIONAL UNIVERSITY
Dept. of Computer Science & Engineering

CONTENTS

1

Data Mining

2

통계 분석

3

J48 알고리즘 분석

1. Data Mining

1. Data Mining

1.1 개요

- 인천대학교 대학생을 대상으로 현대 대학생들의 연애 현황을 알아본다.
- 연인이 되기 위해 갖춰야 할 조건이 무엇인지 Data Mining을 통하여 알아본다.

1.2 Data Mining Tool



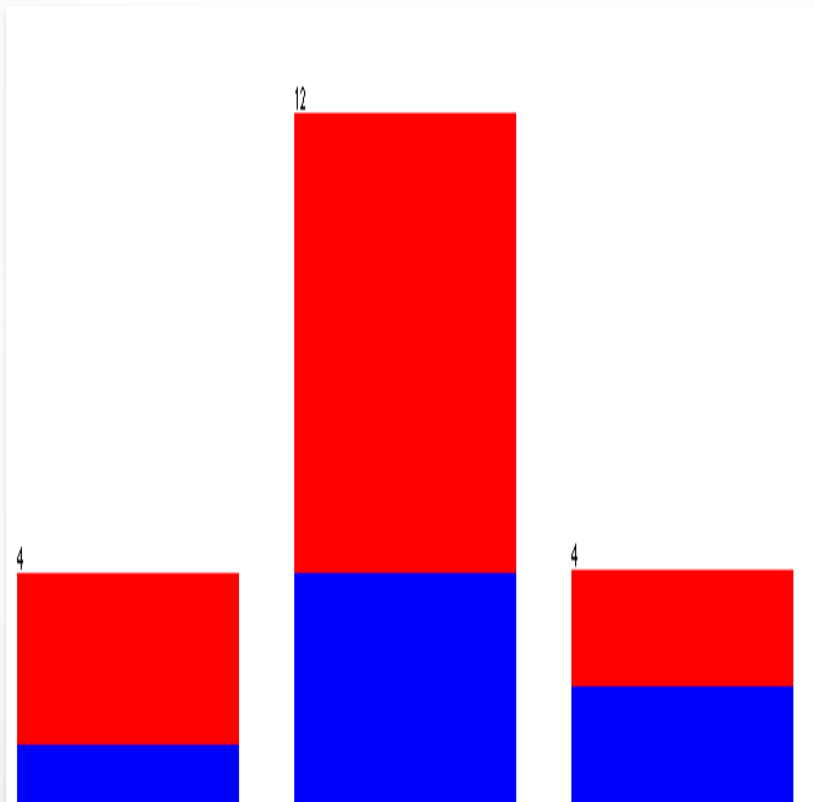
- WEKA (Version 3.6.12)

2. 통계 분석

2. 통계 분석

2015년 인천대학교 대학생 20명의 설문조사 결과

2.1 돈과 여자친구의 상관 관계



- graph 01

형편의 여유가 있는 학생 4명 중
여자친구가 있는 학생의 수는 1명

- graph 02

중산층인 학생 12명 중
여자친구가 있는 학생의 수는 4명

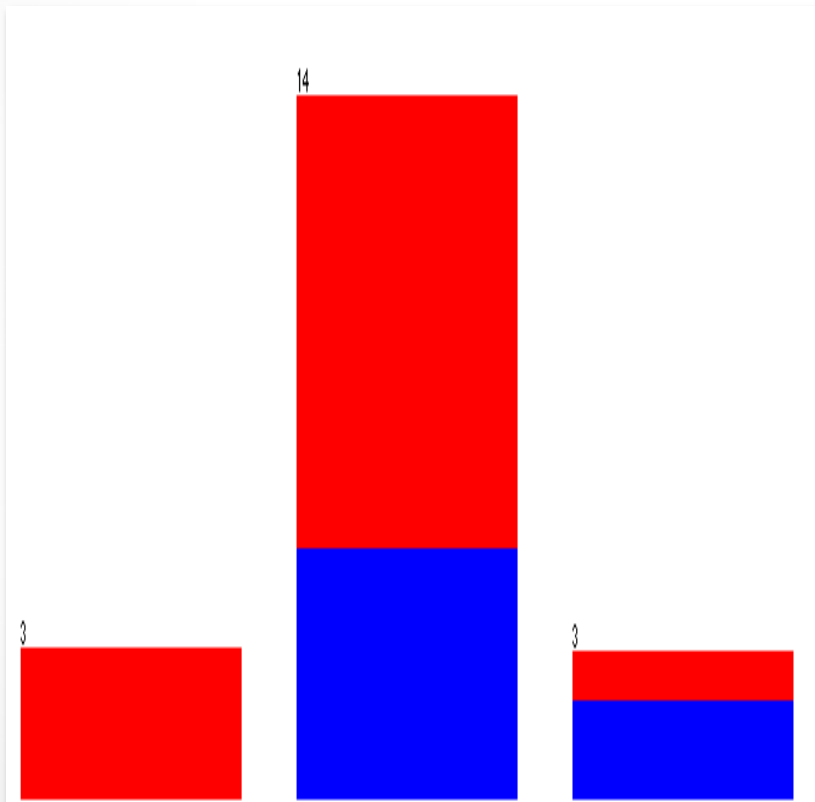
- graph 03

형편의 여유가 없는 학생 4명 중
여자친구가 있는 학생의 수는 2명

2. 통계 분석

2015년 인천대학교 대학생 20명의 설문조사 결과

2.2 신장과 여자친구의 상관 관계



- graph 01

신장이 180cm이상인 학생 3명 중
여자친구가 있는 학생의 수는 0명

- graph 02 (170cm ~ 179cm)

신장이 170cm이상인 학생 14명 중
여자친구가 있는 학생의 수는 5명

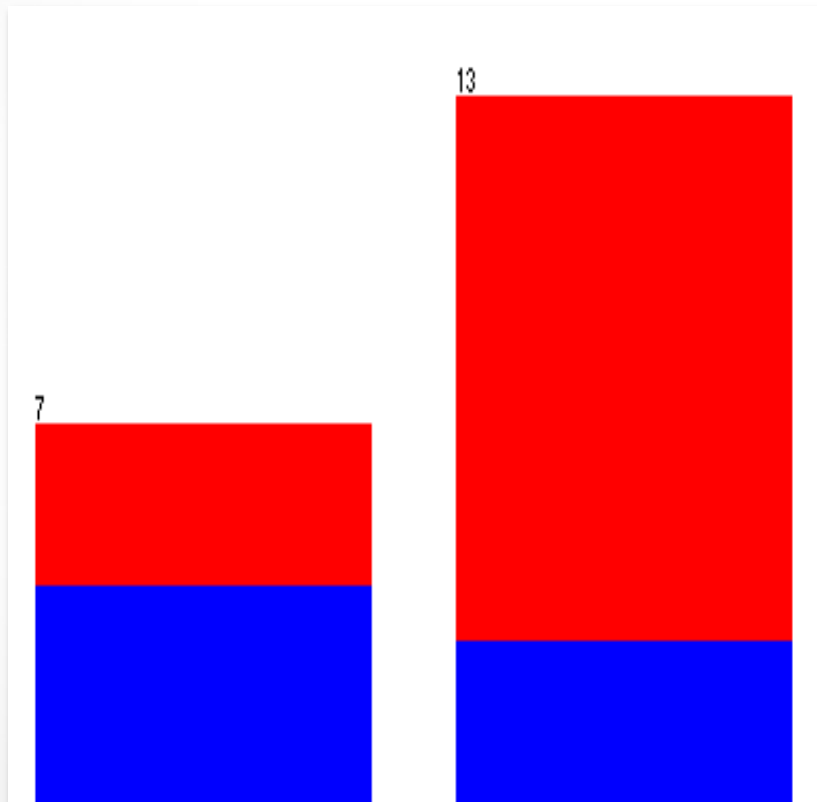
- graph 03 (160cm ~ 169cm)

신장이 160cm이상인 학생 3명 중
여자친구가 있는 학생의 수는 2명

2. 통계 분석

2015년 인천대학교 대학생 20명의 설문조사 결과

2.3 자동차와 여자친구의 상관 관계



- graph 01 (아버지 자동차 사용 가능한 경우 포함)

자동차를 소유한 학생 7명 중

여자친구가 있는 학생의 수는 4명

- graph 02

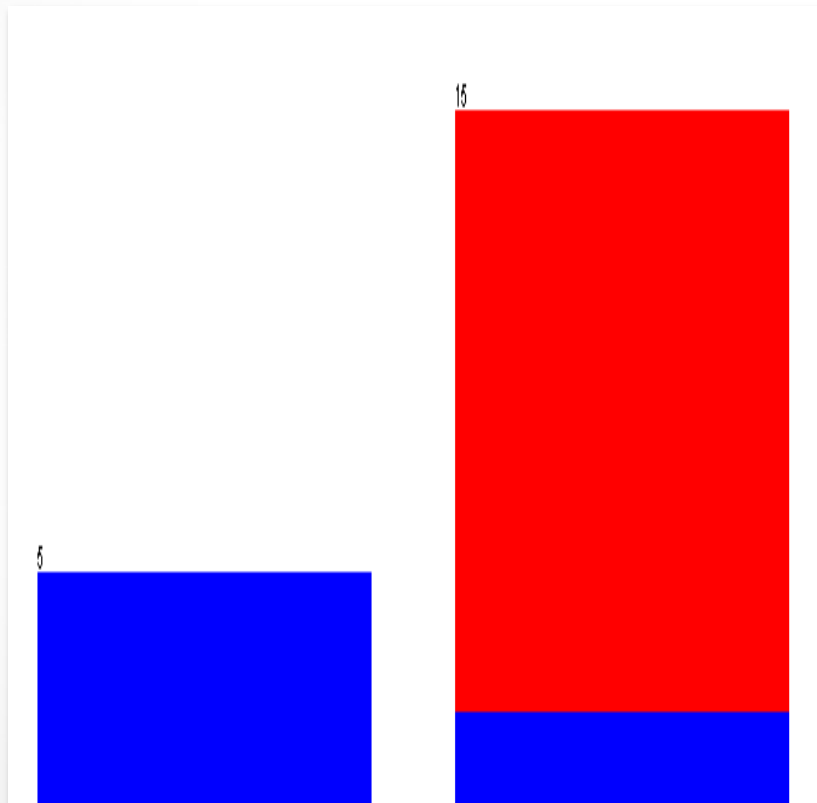
자동차를 소유하지 않은 학생 13명 중

여자친구가 있는 학생의 수는 3명

2. 통계 분석

2015년 인천대학교 대학생 20명의 설문조사 결과

2.4 얼굴과 여자친구의 상관 관계



- graph 01 (미남)

이목구비가 뚜렷한 학생 5명 중
여자친구가 있는 학생의 수는 5명

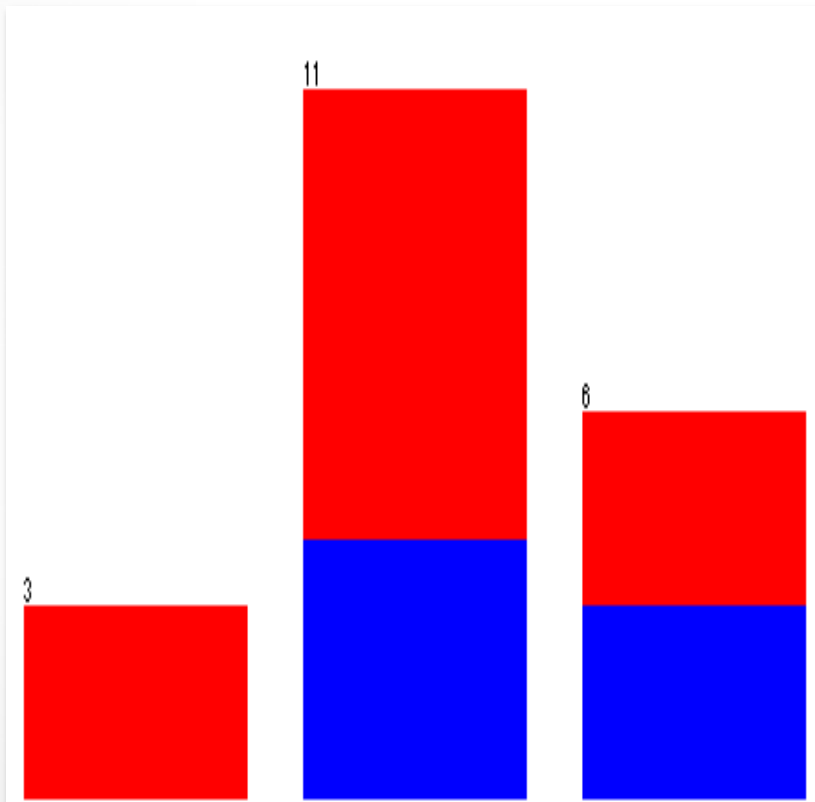
- graph 02

이목구비가 뚜렷하지 않은 학생 15명 중
여자친구가 있는 학생의 수는 2명

2. 통계 분석

2015년 인천대학교 대학생 20명의 설문조사 결과

2.5 학점과 여자친구의 상관 관계



- graph 01 (4.0 ~ 4.5)

평균학점이 4.0이상인 학생 3명 중
여자친구가 있는 학생의 수는 0명

- graph 02 (3.0 ~ 3.9)

평균학점이 3.0이상인 학생 11명 중
여자친구가 있는 학생의 수는 4명

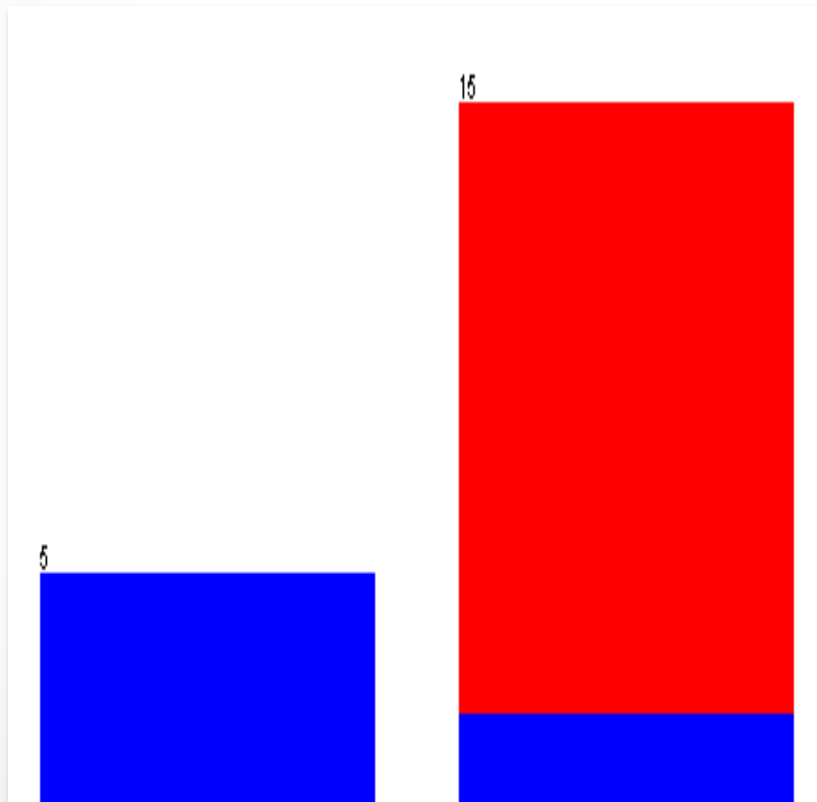
- graph 03 (2.0 ~ 2.9)

평균학점이 2.0이상인 학생 6명 중
여자친구가 있는 학생의 수는 3명

2. 통계 분석

2015년 인천대학교 대학생 20명의 설문조사 결과

2.6 유머감각과 여자친구의 상관 관계



- graph 01

유머 감각이 있는 학생 5명 중
여자친구가 있는 학생의 수는 5명

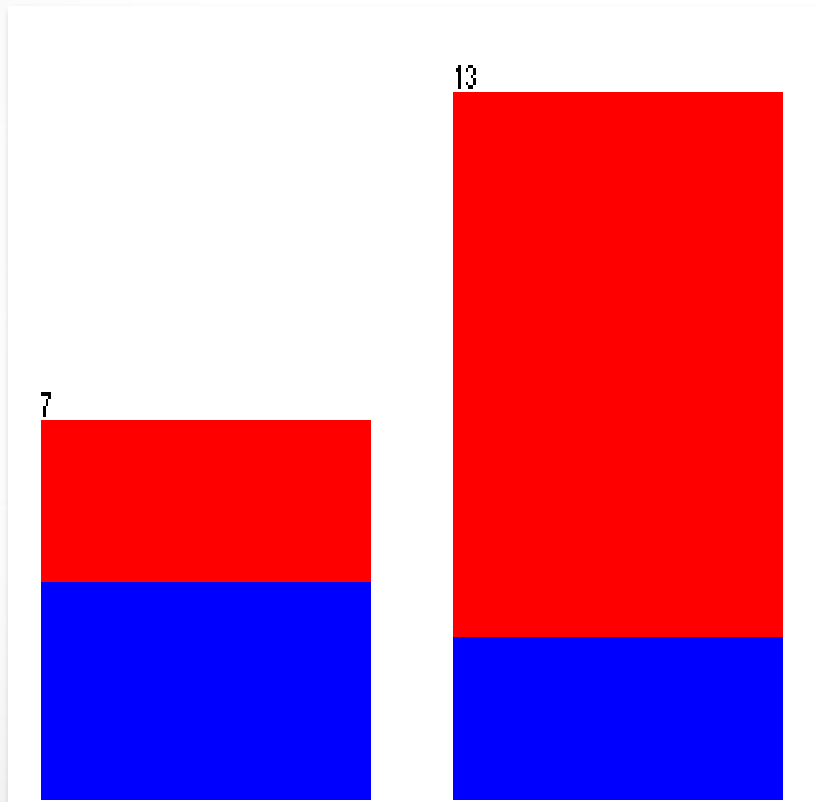
- graph 02

유머 감각이 부족한 학생 15명 중
여자친구가 있는 학생의 수는 2명

2. 통계 분석

2015년 인천대학교 대학생 20명의 설문조사 결과

2.7 군대와 여자친구의 상관 관계



- graph 01

군 복무를 마친 학생 7명 중

여자친구가 있는 학생의 수는 4명

- graph 02

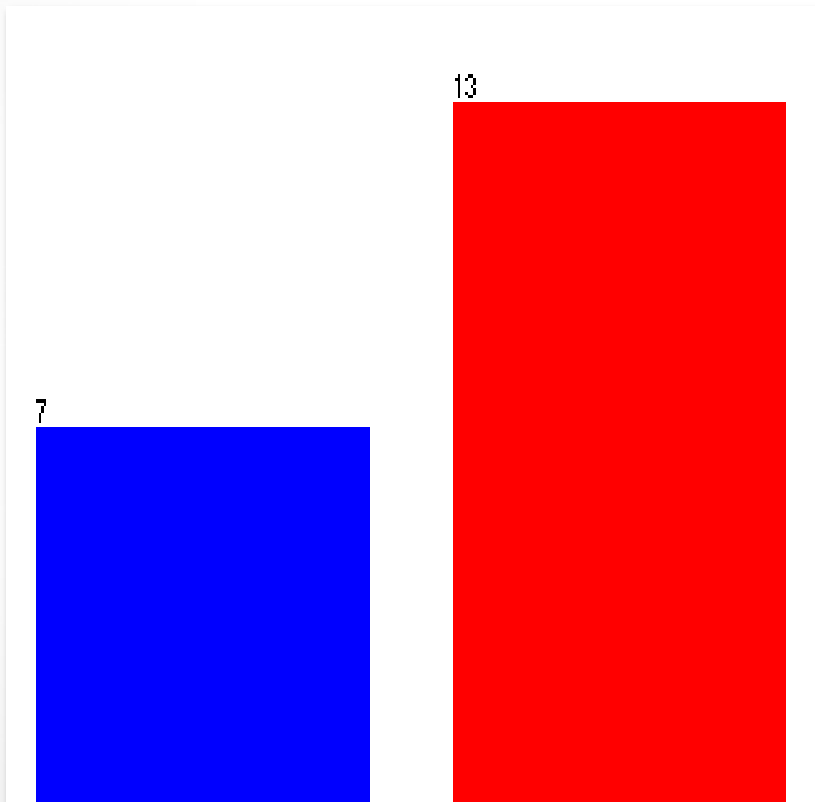
군 복무를 마치지 않은 학생 13명 중

여자친구가 있는 학생의 수는 3명

2. 통계 분석

2015년 인천대학교 대학생 20명의 설문조사 결과

2.8 여자친구 유무의 가장 중요한 속성



- graph 01

이목구비가 뚜렷한 학생 7명 중
여자친구가 있는 학생의 수는 7명

- graph 02

이목구비가 뚜렷하지 않은 학생 13명 중
여자친구가 있는 학생의 수는 0명

※ 통계 분석을 통하여 여자친구의 유무 상관 관계를 통하여
가장 영향 많이 주는 속성은 '얼굴' 이라는 사실을 알 수 있다.

3. J48 알고리즘 분석

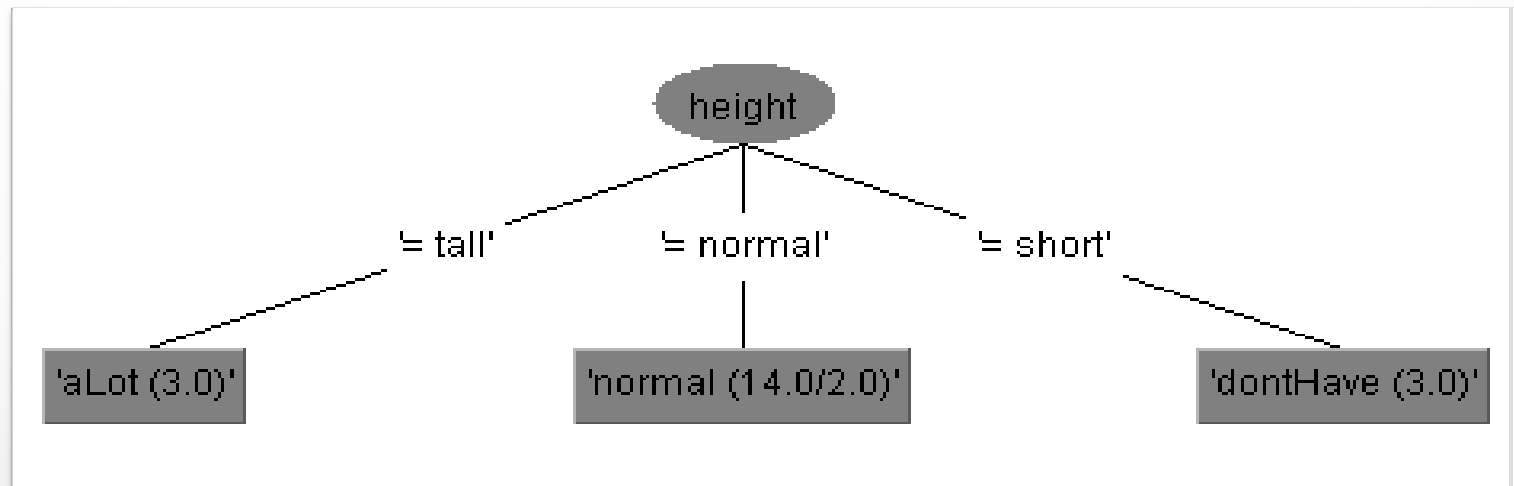
3. J48 알고리즘 분석

2015년 인천대학교 대학생 20명의 설문조사 결과

3.1 돈

수행 결과 분석

- 형편의 여유가 있고 신장이 180cm 이상인 학생은 3명이다.
- 중산층이고 신장이 170cm 이상 180미만인 학생은 2명이다.
- 형편의 여유가 없고 신장이 160cm 이상 170미만인 학생은 3명이다.



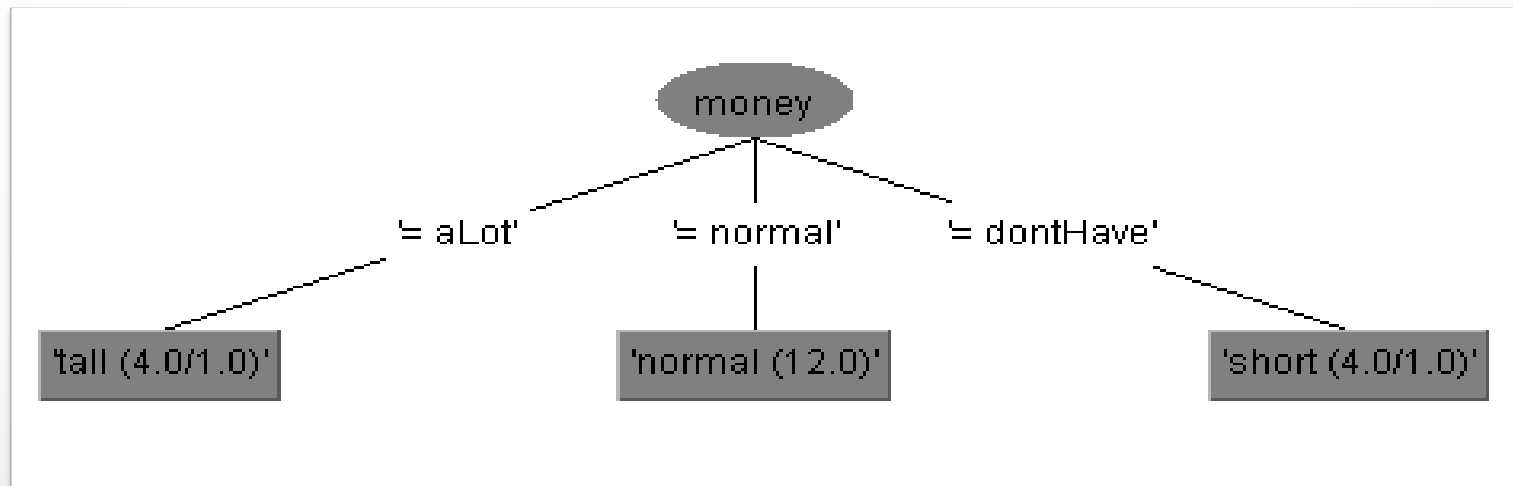
3. J48 알고리즘 분석

2015년 인천대학교 대학생 20명의 설문조사 결과

3.2 신장

수행 결과 분석

- 신장이 180cm 이상이고 형편의 여유가 있는 학생은 3명이다.
- 신장이 170cm 이상 180미만이고 중산층인 학생은 12명이다.
- 신장이 160cm 이상 170미만이고 형편의 여유가 없는 학생은 3명이다.



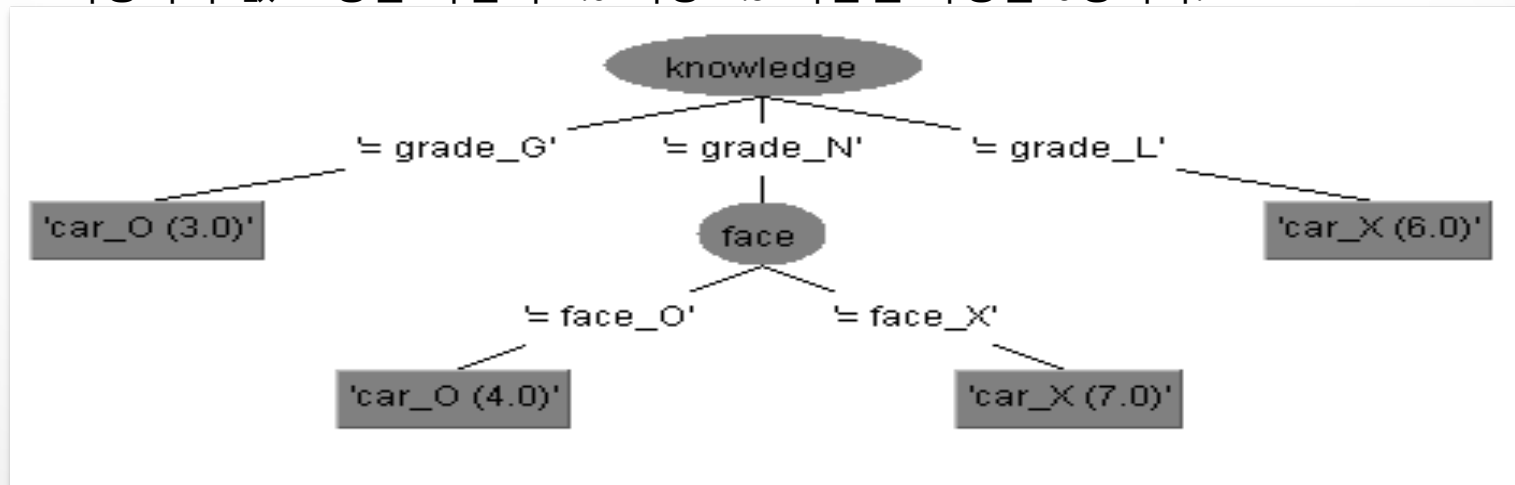
3. J48 알고리즘 분석

2015년 인천대학교 대학생 20명의 설문조사 결과

3.3 자동차

수행 결과 분석

- 자동차를 소유하고 평균 학점이 4.0 이상인 학생은 3명이다.
- 자동차를 소유하고 평균 학점이 3.0 이상 3.9 미만이며 미남인 학생은 4명이다.
- 자동차가 없고 평균 학점이 3.0 이상 3.9 미만이며 미남이지 못한 학생은 7명이다.
- 자동차가 없고 평균 학점이 2.0 이상 2.9 미만인 학생은 6명이다.



※ grade_G = 평균학점 4.0 ~ 4.5, grade_N = 3.0 ~ 3.9, grade_L = 2.0 ~ 2.9
(grade_Good) (grade_Normal) (grade_Low)

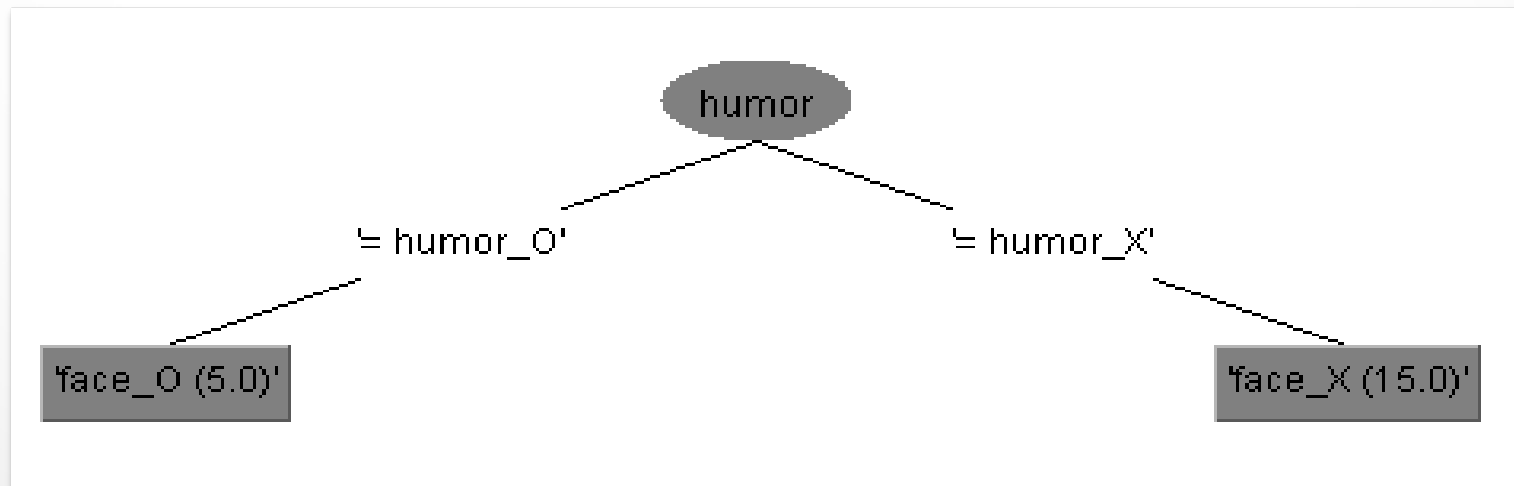
3. J48 알고리즘 분석

2015년 인천대학교 대학생 20명의 설문조사 결과

3.4 얼굴

수행 결과 분석

- 미남이고 유머 감각이 있는 학생은 5명이다.
- 미남이 아니고 유머 감각이 없는 학생은 15명이다.



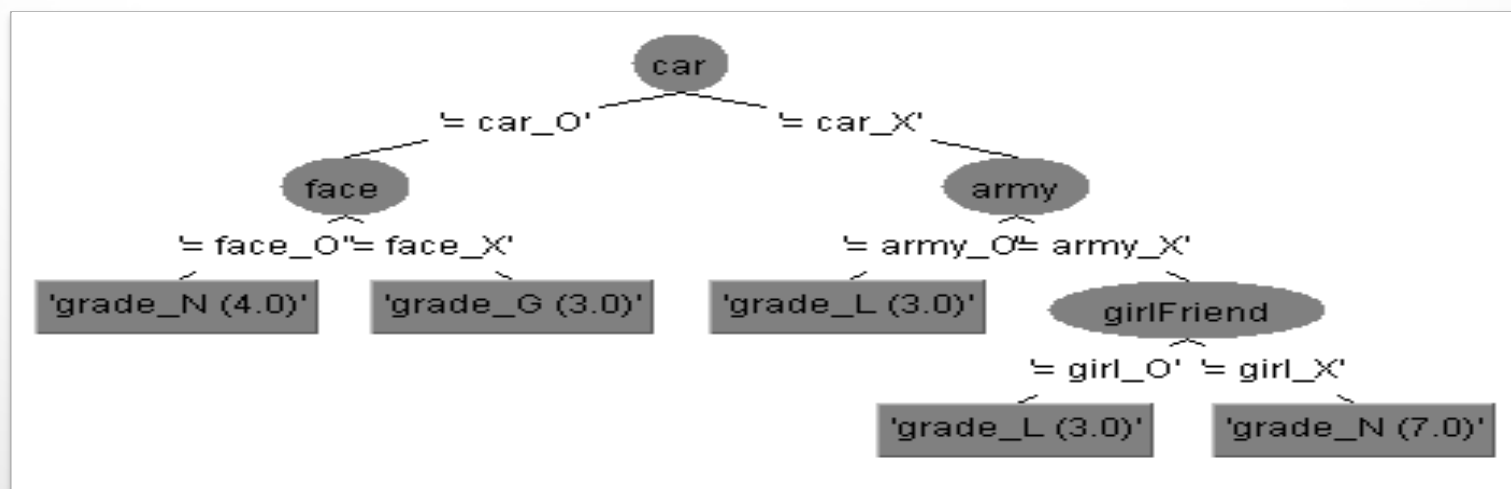
3. J48 알고리즘 분석

2015년 인천대학교 대학생 20명의 설문조사 결과

3.5 학점

수행 결과 분석

- 평균 학점이 3.0 이상 3.9 미만이고 미남이며 자동차가 있는 학생은 4명이다.
- 평균 학점이 4.0 이상이고 미남이지 않으며 자동차가 있는 학생은 3명이다.
- 평균 학점이 2.0 이상 2.9 미만이고 여자친구가 있으며 군대를 가지 않고 자동차가 없는 학생은 3명이다.
- 평균 학점이 3.0 이상 3.9 미만이고 여자친구가 없으며 군대를 가지 않고 자동차가 없는 학생은 7명이다.
- 평균 학점이 2.0 이상 2.9 미만이고 전역을 했으며 자동차가 없는 학생은 3명이다.



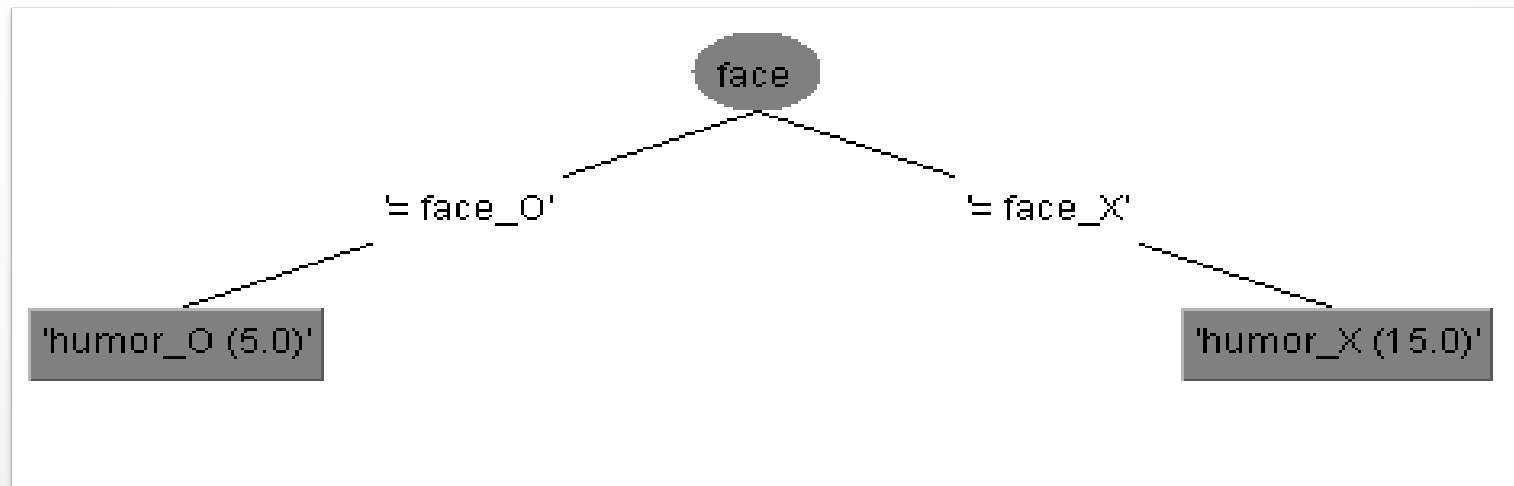
3. J48 알고리즘 분석

2015년 인천대학교 대학생 20명의 설문조사 결과

3.6 유머 감각

수행 결과 분석

- 유머 감각이 있고 잘생긴 학생은 5명이다.
- 유머 감각이 없고 못생긴 학생은 15명이다.



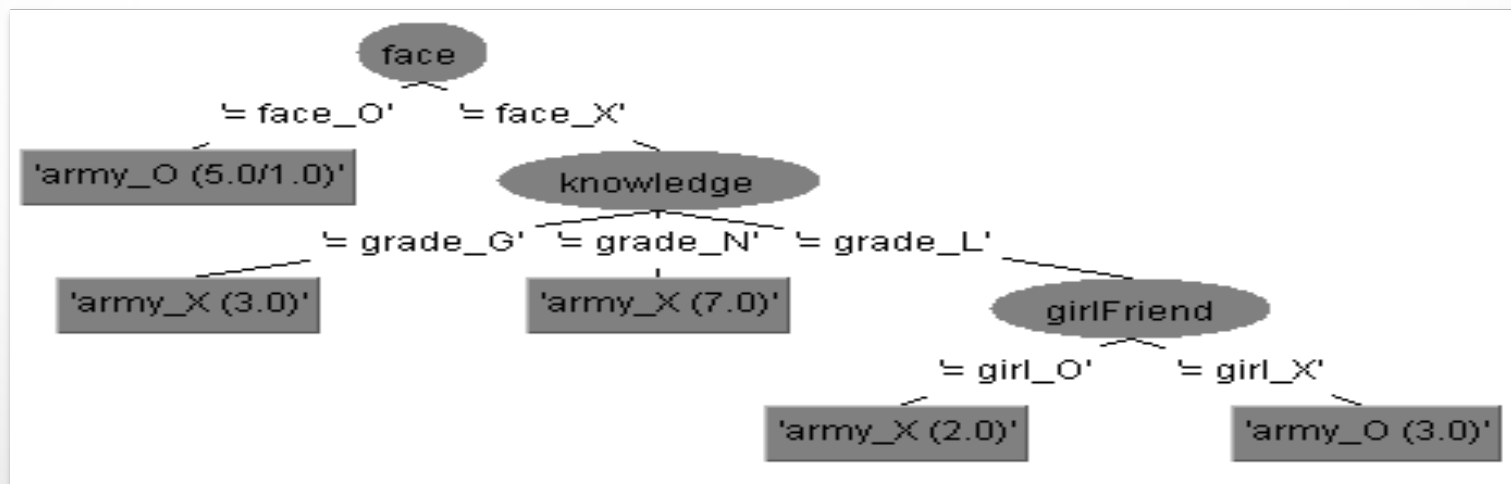
3. J48 알고리즘 분석

2015년 인천대학교 대학생 20명의 설문조사 결과

3.7 군대

수행 결과 분석

- 전역했고 미남인 학생은 5명이다.
- 군대를 가지 않았고 평균 학점이 4.0 이상이며 못생긴 학생은 3명이다.
- 군대를 가지 않았고 평균 학점이 3.0 이상 3.9 미만이고 못생긴 학생은 7명이다.
- 군대를 가지 않았고 여자친구가 있으며 평균 학점이 2.0 이상 2.9 미만이고 못생긴 학생은 2명이다.
- 전역했고 여자친구가 없으며 평균 학점이 2.0 이상 2.9 미만이고 못생긴 학생은 3명이다.



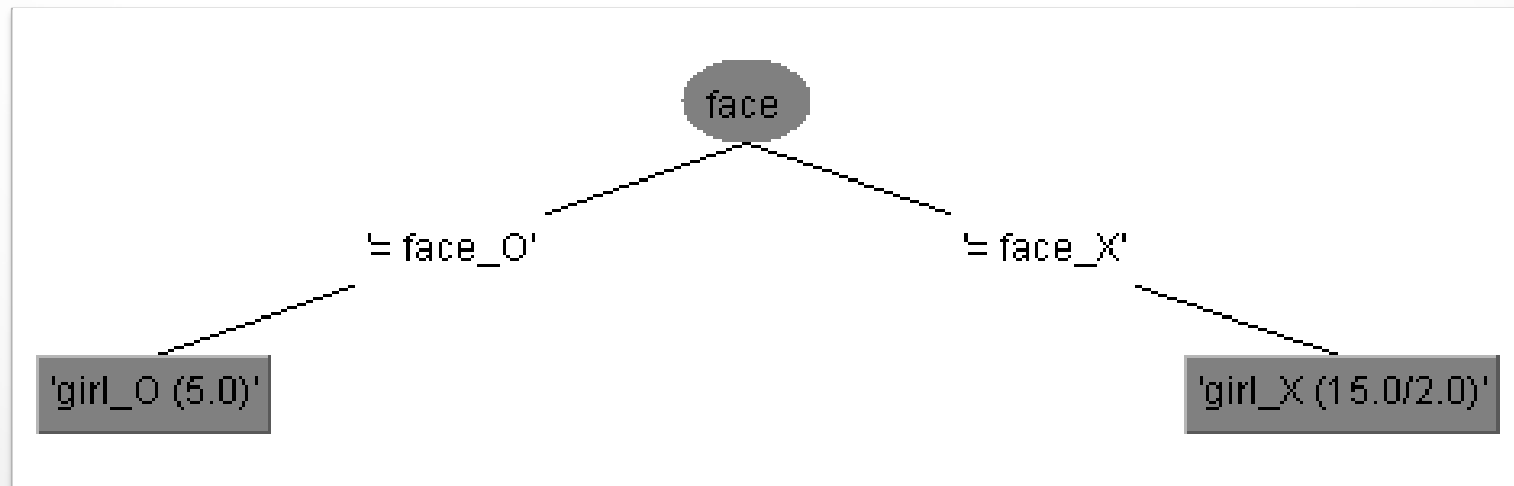
3. J48 알고리즘 분석

2015년 인천대학교 대학생 20명의 설문조사 결과

3.8 여자친구

수행 결과 분석

- 여자친구가 있고 잘생긴 학생은 5명이다.
- 여자친구가 없고 잘생기지 못한 학생은 13명이다.



3. J48 알고리즘 분석

2015년 인천대학교 대학생 20명의 설문조사 결과

3.9 수행 결과 분석

=== Summary ===

Correctly Classified Instances
Incorrectly Classified Instances
Kappa statistic
Mean absolute error
Root mean squared error
Relative absolute error
Root relative squared error
Total Number of Instances

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Preci
	0.667	0	1
	0.909	0.222	0.
	0.833	0.071	0.
Weighted Avg.	0.85	0.144	0.

=== Confusion Matrix ===

```
a b c <-- classified as
2 1 0 | a = grade_G
0 10 1 | b = grade_N
0 1 5 | c = grade_L
```

• 학점

=== Summary ===

Correctly Classified Instances	14	70
Incorrectly Classified Instances	6	30
Kappa statistic	0.3407	
Mean absolute error	0.3125	
Root mean squared error	0.476	
Relative absolute error	66.4894 %	
Root relative squared error	97.4162 %	
Total Number of Instances	20	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measu
	0.571	0.231	0.571	0.571	0.57
	0.769	0.429	0.769	0.769	0.76
Weighted Avg.	0.7	0.359	0.7	0.7	0.7

=== Confusion Matrix ===

```
a b <-- classified as
4 3 | a = army_0
3 10 | b = army_X
```

• 군대

=== Summary ===

Correctly Classified Instances	20	100	%
Incorrectly Classified Instances	0	0	%
Kappa statistic	1		
Mean absolute error	0		
Root mean squared error	0		
Relative absolute error	0	%	
Root relative squared error	0	%	
Total Number of Instances	20		

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	1	0	1	1	1	1	face_0
	1	0	1	1	1	1	face_X
Weighted Avg.	1	0	1	1	1	1	

=== Confusion Matrix ===

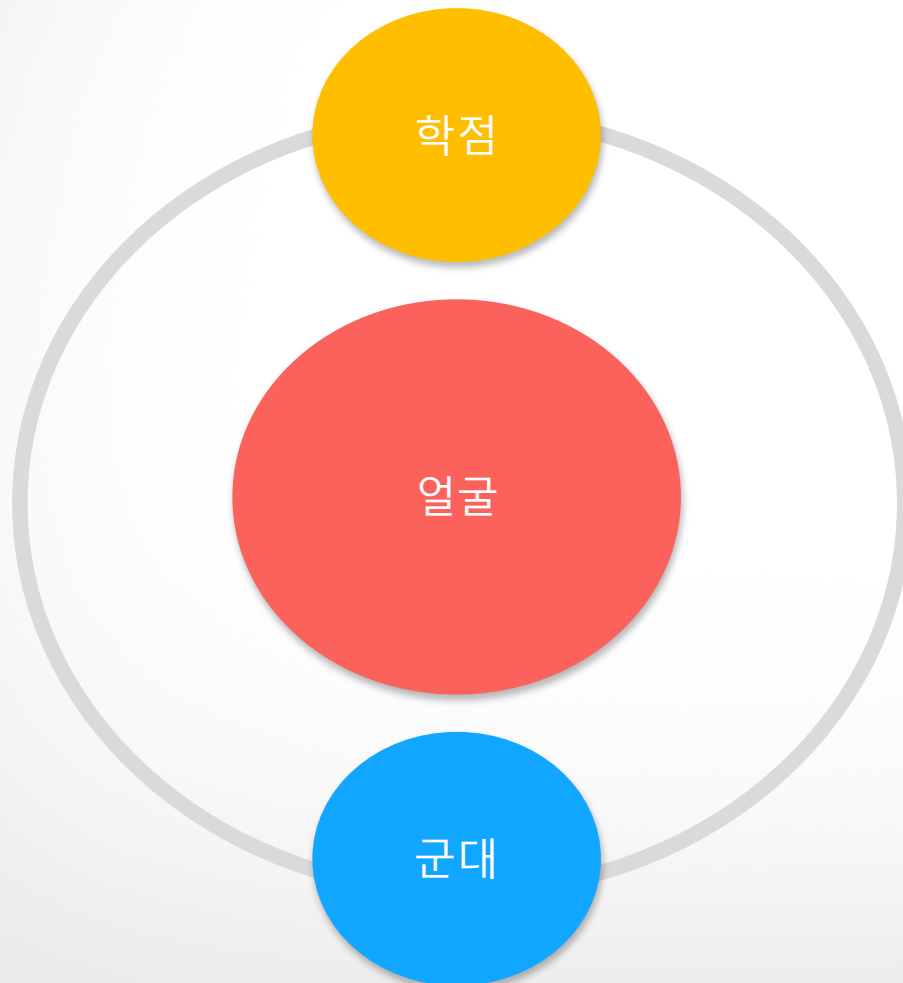
```
a b <-- classified as
5 0 | a = face_0
0 15 | b = face_X
```

• 얼굴

3. J48 알고리즘 분석

2015년 인천대학교 대학생 20명의 설문조사 결과

3.9 수행 결과 분석



- 여자친구의 유무의 영향을 주는 속성은 학점, 군대, 얼굴로 분석되었다.
- 가장 중요한 속성은 **얼굴**이다.

감사합니다



인천대학교 컴퓨터공학부

INCHEON NATIONAL UNIVERSITY

Dept. of Computer Science & Engineering