

Analysis of Students Leaving Their Majors Using Decision Tree¹⁾

Cheolyong Park²⁾ and Gyu Moon Song³⁾

Abstract

Since 1997, when a new educational system that encourages faculties instead of departments in universities is first introduced, students have much more chance to choose and leave their majors than before. As a result, colleges of basic arts and sciences confront with a serious problem since lots of students have left their majors at the colleges. In this paper, we analyze and provide a predictive model for those students in a university using decision trees.

Keywords : , ,

1.

2006 가

가 .

가 ,

. 1997

가

/

가

-
1. The present research has been conducted by the attached research institute Research Grant of Keimyung University in 2001.
 2. Associate Professor, Department of Statistics, Keimyung University, Taegu 704-701, Korea
E-Mail: cypark1@kmu.ac.kr
 3. Professor, Department of Statistics, Keimyung University, Taegu 704-701, Korea
E-mail: kms252@kmu.ac.kr

2 , /

가 가 가 가 가

가 가 가 가 가

가 가 가 가 가

(generalization)가 가 (interaction)

2

1 가

3

2.

2.1

1995 2002 9

4 , , , () , (censored 2088)

data) 56.37% 1177 43.63% 911

1

, , , ,

, 1 1 4 2
1.

	: 4 [1, 2, 3, 4, 5]
	: 9 [6, 7, 8, 9, 10, 11, 12, 13, 14, 15]
	= 1, = 2
	= 1, = 2
	= 1, = 2
	= 1, = 0
	: 1, 2, 3
	1 1 4 2
	1 1 4 2 / *100%
	= 1, = 0

1996

1997

가

, 0.3717, 1 1
0.5667
0.001
1 1

가

(factor analysis)

8

. 8

가

가

78.5% 65.4%

2

Varimax

(factor loadings matrix)

	1	2
11	0.36240	0.86240
12	0.47185	0.79360
21	0.61797	0.62531
22	0.74998	0.48643
31	0.82091	0.43798
32	0.85319	0.39320
41	0.82525	0.44351
42	0.81690	0.38966

	1	2
11	0.42796	0.42865
12	0.23154	0.90354
21	0.37322	0.75905
22	0.62889	0.52583
31	0.73983	0.41402
32	0.83248	0.34972
41	0.86560	0.26689
42	0.85889	0.27980

2 2 4 2 ,
 1 2 1 .
 2 2 4 2
 , 1 2 2 1 1 1
 . / 가
 2 1
 / 가
 가 가

2.2

2.1

CART (entropy) CHAID (Hatigan, 1975), (Gini index) (Breiman, Friedman, Olshen, and Stone, 1984)
 C4.5 (Quinlan, 1993) 가
 (correct classification rate) C4.5
 (training data) 82.1%, 가 (validation data) 80.4%
 C4.5 가
 2 .

가
 가
 C4.5 75.6%
 . C4.5
 3
 2.
 CHAID, CART C4.5
 78.7%, 가

				가		
	455	155	610	210	91	301
	143	646	789	77	311	388
	598	801	1399	287	402	689

C4.5
 가 가 12
 , 1, 2
 가
 1) ;
 2) 가 ;
 3) 1, 2 .

가 75.6%
 , 10% 85.7% ,
 20% 82.5% /
 10% 20%

3.

: 1) 4 1, 2 ; 2)

가 ; 3) ; 4) ; 5)
 ; 6) 2 2 가
 , 가
 1
 1) ; 2) 가 ; 3) 1 1, 2
 가
 가
 가 / 가 가
 가 가 가
 가 가 가 가
 가 가 가 가
 가 가 가 가

1. Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J. (1984). *Classification and Regression Trees*. Chapman & Hall/CRC, New York.
2. Hatigan, J.A. (1975). *Clustering Algorithms*. John Wiley & Sons, Inc, New York.
3. Quinlan, J.R. (1993). *C4.5 Programs for Machine Learning*. Morgan Kaufmann, San Mateo.