

분류분석 - 계약자행동가정(유지율)

성균관대 글로벌 보험·연금대학원
언더라이팅(UW)
윤진한&양동현
2018.12.15

CONTENTS

I. IFRS17 가정

II. 분류분석

II -1. 데이터 산출

-2. 결과

-3. 결과분석 및 개선과제

IFRS17 부채 평가시 사용하는 가정

- 해지율 가정, 사업비 배분 가정, 할인률 가정, 경제적 가정 등
- 대부분 회사 과거 경험통계 및 외부 시장지표를 토대로 산출
- 해지율 가정 산출 (LAT)
 - 최근 5년간의 통계를 사용함
 - 경과유지보험료 대비 실효해약보험료

상품군	납방	채널	1	2	3	4	5	6	7	8	9	10	11	12+
무배당 금리연동형(연금)	월납	FP												
		방카												
		GFP												
		direct												
	일시납	FP												
		방카												
		GFP												
		direct												
무배당 어린이보장	월납	FP												
		방카												
		GFP												
		direct												
	일시납	FP												
		방카												
		GFP												
		direct												

■ 분류분석 방법론 선정 배경

- 현재 가정산출은 과거 경험치를 관찰할 뿐 예측 모델 수립에는 취약함
- 검증된 모델이 수립되어 있으면 언더라이팅 측면에서 활용 가능

■ 데이터 산출

상품	대상기간	대상계약
CI보험	2013년 1월 1일~2018년 11월 30일	신계약 중 랜덤 추출

- 대상기간의 신계약 중 2018년 12월 10일 기준의 계약상태를 관찰(총 27,931건)

- 독립변수 : 계약월, 계약자나이, 성별, 시도, 직업군, 납입주기, 납입기간 등 17개

보험가입증서번호	일반계정계리코드	보험종류코드	계약월	계약자나이	계약자성별구분코드	계약자시도구분코드	주피보험자나이	주피보험자성별구분코드	주피보험자직업코드	표준구분코드	납입주기코드	납입기간	보험기간	보유계약금액	모집관영태코드	위험등급코드	직업군	보험료	보험상태코드	보험상태코드
213010000285	3084	1000495	201301	55	2	9	49	2	S323	1	1	20	99	20000	C	3	곡식작물재배	102650	1	4
213010000585	3084	1000621	201301	45	2	3	45	2	S969	1	1	35	99	100000	C	0	슈퍼마켓, 편	238000	1	6
213010002850	3084	1000621	201301	52	1	1	25	1	T220	1	1	20	99	30000	C	0	대학생	86750	1	6
213010002856	3084	1000621	201301	33	2	3	33	2	T218	1	1	32	99	70000	C	0	전업주부	144220	1	4
213010002885	3084	1000621	201301	52	2	1	52	2	S080	1	1	20	99	16000	C	0	고시학원강사	165810	1	6
213010004385	3084	1000621	201301	53	2	5	19	1	T221	1	1	20	99	40000	C	0	고등학생	120290	1	4
213010005855	3084	1000621	201301	23	2	3	20	2	T228	1	1	30	99	70000	C	0	학원생, 재수	91870	1	4
213010005859	3084	1000621	201301	23	1	1	23	1	T220	1	1	30	99	70000	C	0	대학생	143030	1	4
213010006851	3084	1000621	201301	52	2	1	52	2	S480	1	1	20	99	10000	C	0	패턴사 및 재	119455	1	4
213010007185	3084	1000621	201301	25	2	3	25	2	S066	1	1	35	99	100000	C	0	중고등학교 교	161260	1	4
213010007985	3084	1000621	201301	35	2	1	40	1	S948	1	1	30	99	100000	C	3	실내장식 각	525000	1	5
213010008554	3084	1000621	201301	48	2	1	48	2	T024	1	1	20	99	20000	C	4	음식 서비스	130400	1	6
213010008559	3084	1000621	201301	51	2	8	26	1	S067	1	1	29	99	30000	C	0	초등학교 교	84500	1	4

■ 데이터 클리닝

- 증권번호, 상품코드 등 계약상태와 무관한 데이터 삭제
- 직업코드, 납입주기코드, 보험기간 등 가지수가 너무 많거나 너무 적은 변수 제외
- 계약월, 계약자나이, 보험료, 가입금액 등 숫자형 변수에 대한 구간 설정
- 목적변수 허용값 조정 : 유지, 실효, 소멸 -> 유지(yes), non유지(no)

Year	C_Age	C_Gender	Region	B_Age	B_Gender	S_Code	P_Duratin	C_Amount	Branch_Code	Risk	Premium	P_Status	Status
2013	E.50대	2	전라남도	D.40대	2	1	C.20년~30년	B.2,000만원대	FP	3	B.10만원~20만원	1	yes
2013	D.40대	2	경기도	D.40대	2	1	D.30년~40년	J.1억이상	FP	0	C.20만원~30만원	1	no
2013	E.50대	1	서울시	B.20대	1	1	C.20년~30년	C.3,000만원대	FP	0	A.10만원미만	1	no
2013	C.30대	2	경기도	C.30대	2	1	D.30년~40년	G.7,000만원대	FP	0	B.10만원~20만원	1	yes
2013	E.50대	2	서울시	E.50대	2	1	C.20년~30년	A.2,000만원 미만	FP	0	B.10만원~20만원	1	no
2013	E.50대	2	강원도	A.20세미만	1	1	C.20년~30년	D.4,000만원대	FP	0	B.10만원~20만원	1	yes
2013	B.20대	2	경기도	B.20대	2	1	D.30년~40년	G.7,000만원대	FP	0	A.10만원미만	1	yes
2013	B.20대	1	서울시	B.20대	1	1	D.30년~40년	G.7,000만원대	FP	0	B.10만원~20만원	1	yes
2013	E.50대	2	서울시	E.50대	2	1	C.20년~30년	A.2,000만원 미만	FP	0	B.10만원~20만원	1	yes
2013	B.20대	2	경기도	B.20대	2	1	D.30년~40년	J.1억이상	FP	0	B.10만원~20만원	1	yes
2013	C.30대	2	서울시	D.40대	1	1	D.30년~40년	J.1억이상	FP	3	E.40만원이상	1	no
2013	D.40대	2	서울시	D.40대	2	1	C.20년~30년	B.2,000만원대	FP	4	B.10만원~20만원	1	no
2013	E.50대	2	전라북도	B.20대	1	1	C.20년~30년	C.3,000만원대	FP	0	A.10만원미만	1	yes
2013	D.40대	2	경기도	D.40대	2	1	C.20년~30년	B.2,000만원대	FP	0	B.10만원~20만원	1	yes
2013	C.30대	1	경기도	C.30대	1	1	C.20년~30년	E.5,000만원대	FP	0	B.10만원~20만원	1	yes
2013	D.40대	1	경기도	D.40대	1	1	C.20년~30년	D.4,000만원대	FP	0	B.10만원~20만원	1	yes
2013	D.40대	2	경기도	A.20세미만	2	1	C.20년~30년	E.5,000만원대	FP	0	A.10만원미만	1	no

■ 모델 훈련

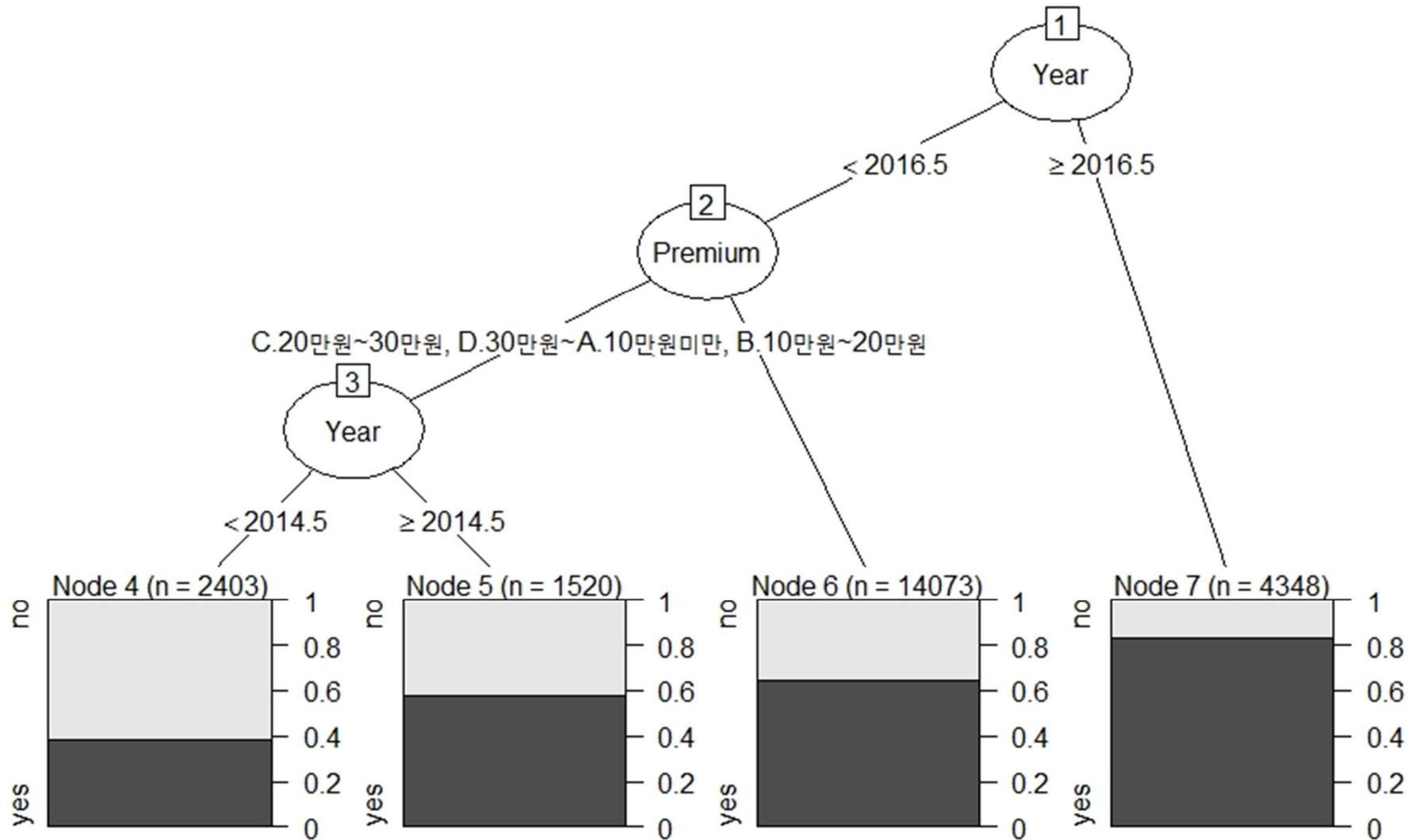
- RAW데이터의 목적변수 분포

Yes(유지)	No(非유지)	합계
18,088(64.8%)	9,843(35.2%)	27,931

- 훈련데이터와 테스트 데이터의 목적변수 분포

훈련데이터(80%)		테스트데이터(20%)	
Yes(유지)	No(非유지)	Yes(유지)	No(非유지)
64.6%	35.4%	65.3%	34.7%

모델 훈련



■ 모델 정확성

- Confusion Matrix

		Actual	
		No	Yes
Pred	No	375	249
	Yes	1,565	3,398

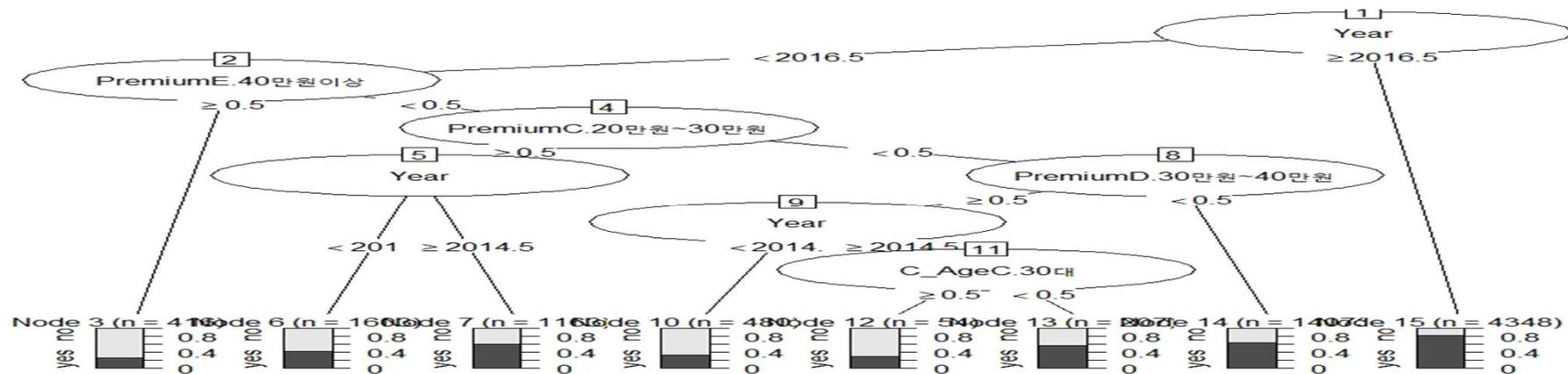
- Accuracy : $67.5\% = (375 + 3,398) / (375 + 249 + 1,565 + 3,398)$
- Sensitivity : $93.2\% = (3,398) / (249 + 3,398)$
- Specificity : $19.3\% = (375) / (375 + 1,565)$

■ 데이터 산출 문제

- 상품 선택의 문제 - 단일상품
 - 납입주기, 보험기간 등 핵심적인 독립변수 분석 불가능

■ 모델 훈련 결과의 의미

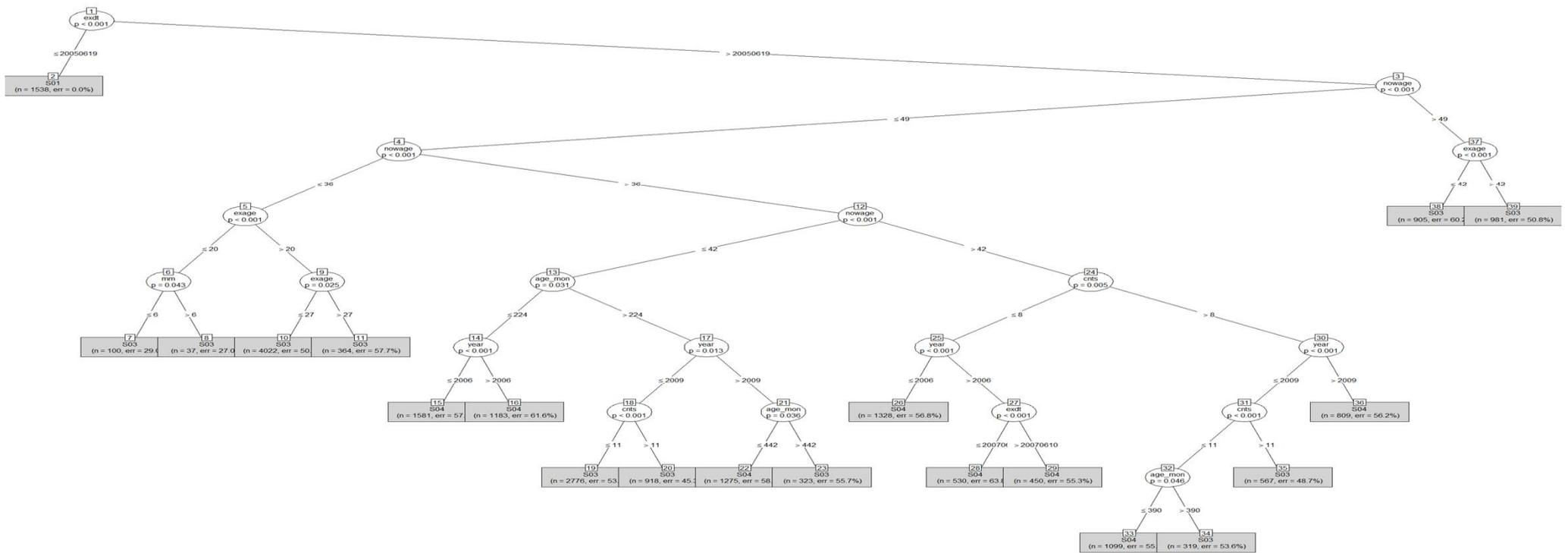
- 가지치기가 불필요



- 다른 독립변수의 영향 분석 불가능
 - 판매채널, 위험등급, 성별 등은 무의미?

[추가 분류분석]

year	cnts	exdt	exage	nowage	yy	mm	gubun
2002	1	20021124	36	52	67	5 S01	
2002	1	20021124	30	46	73	12 S01	
2002	1	20021124	31	47	72	3 S01	
2002	1	20021124	34	50	69	5 S01	
2002	1	20021124	33	49	70	2 S01	



Q & A