

제 10회 SAS 마이닝 챔피언십 DATA 분석

금융권(보험사) 고객이탈방지를 위한 Data Mining

SAS가 궁금하다면?


SAS
Software for Analyzing and Modeling

대한민국 데이터 마이닝 전문가가 되자!

제 9회 SAS 마이닝 챔피언십



등록기간 : 2011년 5월 30일 ~ 7월 6일
등록사이트 : www.sasmining.co.kr

공모지역
 전국 (지역 제한없이 전국에서 참가 가능)

대상인원
 전국 대학교 1학년 이상 대학생 (대학원생 포함)
 (전 학과, 전공, 학년, 학제에 상관없이)

참가인원

구분	팀	개인	총합
1회차부터 4회차까지	팀 100명	팀 400명	-
5회차부터 6회차까지	팀 100명	-	-
7회차부터 8회차까지	팀 100명	팀 200명	-
9회차부터 10회차까지	팀 100명	팀 200명	-

* 팀 100명씩 100개 팀을 구성하여 총 1,000명 참가 예정

이벤트기간
 5월 30일 ~ 7월 6일 (매주 수요일)
 5월 30일 ~ 6월 15일 (매주 수요일)
 6월 22일 ~ 7월 6일 (매주 수요일)
 * 2011년 7월 SAS Korea 컨퍼런스 (7월 6일, 7일)와 병행

참가방법
 1. 홈페이지(www.sasmining.co.kr)에서 참가신청서
 2. 신청서 제출
 3. 참가비 납부
 4. 참가자 등록
 5. 대회 진행

문의 : 02-5561-0000 / 02-5561-0100 / 02-5561-0101

 **SAS가 궁금하다면?**

www.sas.com/ko/Products/Software/Analytics/Miner/index.jsp

 **SAS가 궁금하다면?**

www.sas.com/ko/Products/Software/Analytics/Miner/index.jsp



이벤트 신청서 다운로드



제10회 SAS 마이닝 챔피언십




주최 | 한국데이터과학진흥원
후원 | (사)한국데이터마이닝학회



등록 기간
2022년 7월 25일~8월 30일

등록 사이트
www.sasmining.co.kr

등록 과정
가입 및 참가신청을 통해 게시 예정

등록 대상
국내 기업에서 근무 하거나 대학원생 등 대학원생
A. 기업 데이터 마이닝을 경험한 실무자
B. 학업 중인 대학원생
C. 2022. 7. 25일~8월 30일 기간에 20살 이상

참가 방법

7월 25일~7월 29일	신청서
7월 30일	자료 및 신청서 검토
7월 31일	선정된 참가자 발표
7월 31일~8월 30일	참가
7월 31일	결과 발표
8월 1일	결과 발표

* 신청서 제출 후 2022. 8월 10일 ~ 10월 10일

심사 내역

1차 심사 | 신청서-개인정보 동의 및 참가 신청서
2차 심사 | 1차 심사 신청서-개인정보 동의 및 참가 신청서
3차 심사 | 1차 심사 신청서-개인정보 동의 및 참가 신청서
4차 심사 | 1차 심사 신청서-개인정보 동의 및 참가 신청서

참가비
참가비용 | 참가비용 없음
www.sasmining.co.kr

[illegible]

Data 소개

● Purchase_info(구매이력)

CUST_ID	ITEM_CODE	PURCHASE_DATE
1	65444	201108
1	141208	201108
2	106532	201109
4	204575	201106
4	181287	201106
5	100913	201108
5	125459	201108
5	140916	201108
6	154165	201108
6	1514	201108
7	106532	201105

데이터셋 속성

서버:	로컬
SAS 참조:	
칼럼 개수:	3
행 개수:	222408

- 변수 소개

CUST_ID: 고객 ID

ITEM_CODE: ITEM 고유 코드

PURCHASE_DATE: 고객별 구매 월

- Data 수: 약 22만 개(222,408)

Data 소개

● Product_info(아이템 정보)

ITEM_CATE	ITEM_CODE	PRICE
C	100123	20400
C	100124	26700
A	100127	13300
C	100137	15600
C	100202	17000
A	100211	17500
C	100228	17200
C	100283	11600
C	100298	33000
C	100362	23000
C	100380	27600

데이터셋 속성

서버: 로컬

SAS 참조:

칼럼 개수: 3

행 개수: 3281

- 변수 소개

ITEM_CATE: ITEM 대 분류(예: 전자제품 & 식품 & 의류 등등)

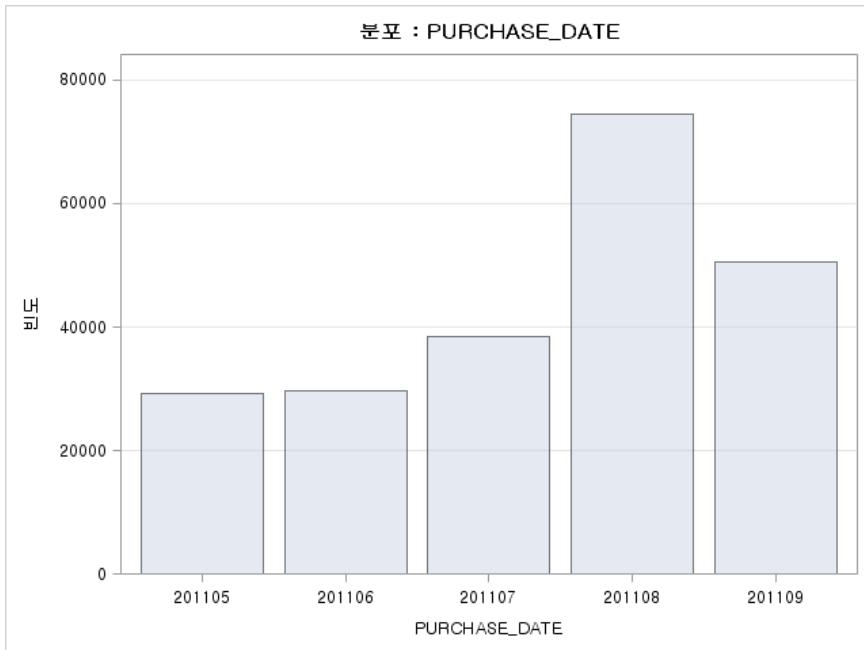
ITEM_CODE: ITEM 고유 코드

PRICE: 가격

- Data 수: 3,281 개

분석의 목적

- 10월 재 구매할 고객 예측



- 5월 ~ 9월 데이터를 가지고 10월에 재 구매 할 고객 10,000명을 예측
(기존의 데이터는 5월 ~ 10월 데이터로 11월에 재 구매 할 고객 예측)

각 Data에 대한 탐색(연속형 및 범주형)

- **Purchase_info Dataset**

- PURCHASE_DATE의 특징?
- 고객의 구매 횟수에 대한 분포를 구해보기!!
- 5, 6, 7, 8, 9월 전부 구매한 고객의 수 파악해보기!!






각 Data에 대한 탐색(연속형 및 범주형)

- **Product_info Dataset**

- PRICE에서 이상한 부분이 없는 지 확인해보기!
IF 이상한 부분이 있다면 어떤 ITEM_CATE에 속하는 지 확인!!
- 각각의 ITEM_CATE에 대해서 의미 있는 정보 파악!

Join

- 구매이력을 기준으로 조인을 하면 다음과 같이 생성됨.

	 CUST_ID	 ITEM_CODE	 PURCHASE_DATE	 ITEM_CATE	 PRICE
1	1	65444	201108	C	27100
2	1	141208	201108	B	12000
3	2	106532	201109	B	7000
4	4	204575	201106	C	28000
5	4	181287	201106	C	28000

112226	62072	220401	201109	A	14500
112227	62074	188598	201105	A	11100
112228	62074	105787	201106	A	8000
112229	62074	212635	201107	A	14900
112230	62074	211224	201107	A	8400
112231	62074	29275	201107	A	18700
112232	62074	182695	201109	A	7600

- 이를 고객별로 하나의 관측치로 표현(대부분의 의미가 들어갈 수 있도록)

Model(가장 많이 사용되는 모형)

- 연속형 예측
 - 회귀 분석
 - 의사결정 트리도 가능!
- 범주형 예측(주로: Yes / No or 1 / 0 ... 즉 맞다 아니다 의미!)
 - 로지스틱 회귀분석
 - 의사결정 트리

SAS Enterprise Miner

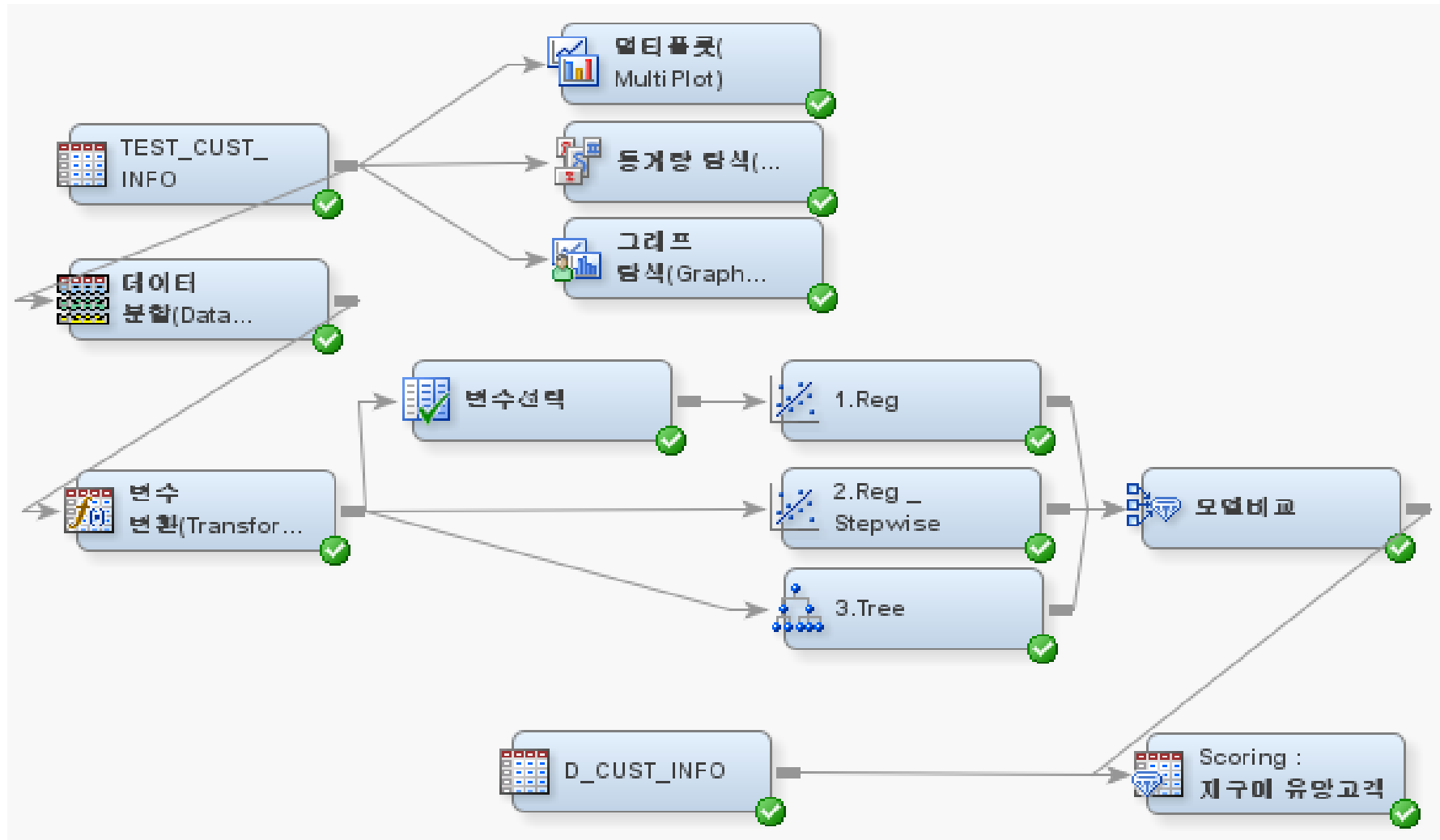
- 연속형(회귀분석)

The diagram illustrates a SAS Enterprise Miner workflow for continuous regression analysis. It starts with two input datasets: 'FL_RED_BLUE_GREEN' and 'SCROING_DATASET'. These are fed into a '회귀(Regression)' node. The output of the regression is then used in a '스코어(Score)' node. A '내보낸 데이터 - 스코어(Score)' window shows the output data, with the 'SCORE' port selected. A detailed view of the '스코어(Score)' output is shown in a table below.

관측값 번호	Predicted: fl	depth	density	P_red	mx	P_green
1	10.31121	0.2	0	0.543092052	69.5	36.67353189
2	4.377627	0.4	0	0.331837617	6.45	15.44922112
3	1.863398	0.6	0	0.202757901	0.755	6.508193268
4	0.795968	0.8	0	0.123888204	0.1225	2.741664403
5	0.3416	1	0	0.075697603	0.0415	1.154963196
6	0.147515	1.2	0	0.046252402	0.021	0.486543861
7	0.005716	2	0	0.006446782	0.003	0.015322774
8	17.91553	0.2	1	3.106226139	74.75	61.25781487
9	9.347423	0.4	1	2.024462403	15.8	31.59378469
10	4.887772	0.6	1	1.319430021	3.75	16.29452884
11	2.562521	0.8	1	0.859929815	1.155	8.403920981
12	1.347635	1	1	0.560453586	0.305	4.334331392
13	0.711326	1.2	1	0.365271929	0.115	2.235436132
14	28.06604	0.2	2	11.36502867	113.8	89.72027482
15	17.01515	0.4	2	7.77127507	40.75	53.79007272
16	10.32667	0.6	2	5.313907952	15.05	32.24880807
17	6.274821	0.8	2	3.633588758	4.7	19.33415535
18	3.817764	1	2	2.484605939	1.65	11.59142261
19	2.32615	1.2	2	1.698944786	0.6	6.94941546
20	1.419533	1.4	2	1.161718782	0.215	4.166388964
21	0.867748	1.6	2	0.794369858	0.11	2.497878721






SAS Enterprise Miner

- 범주형(로지스틱 회귀분석 & 의사결정 트리)



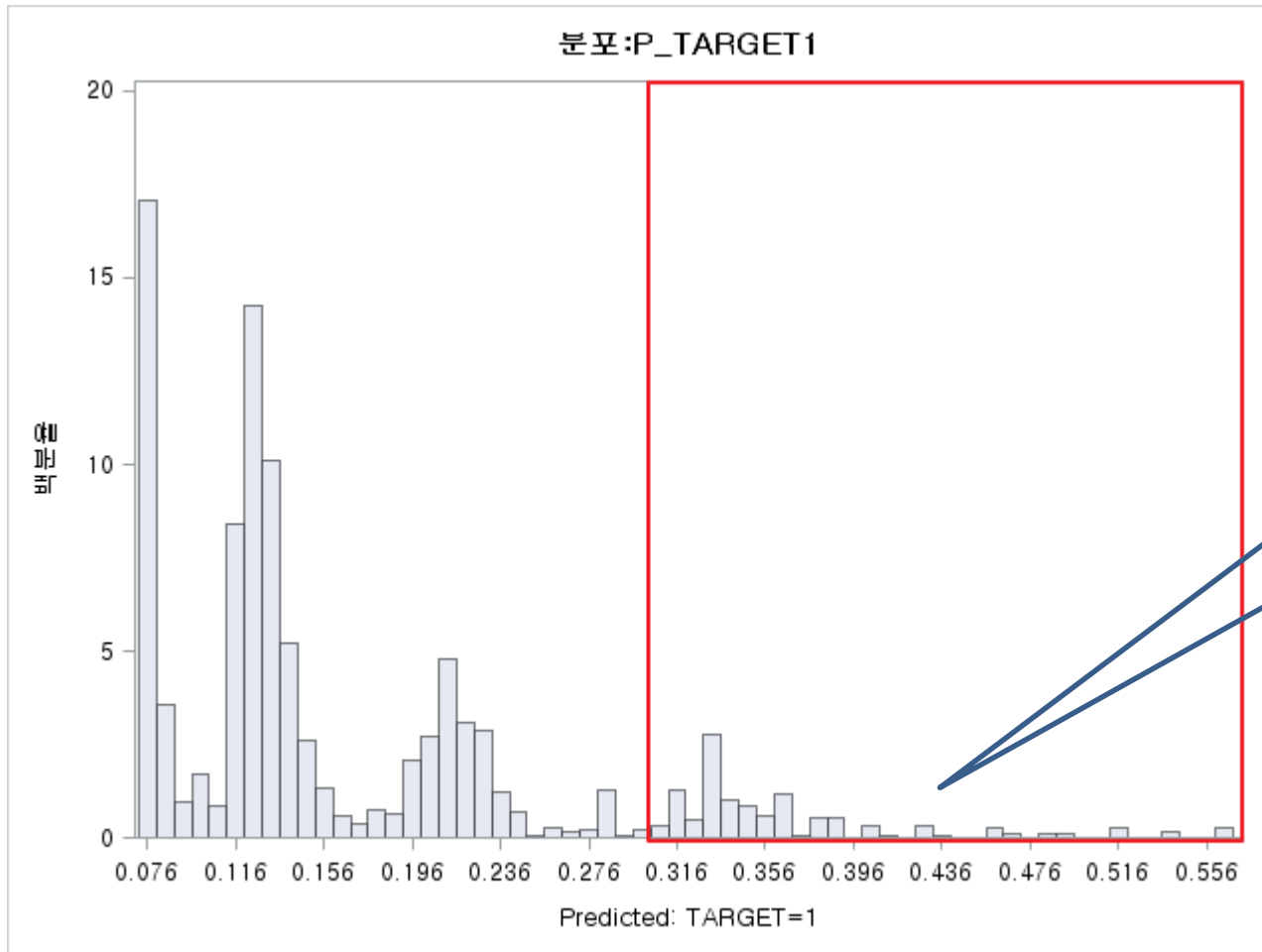
SAS Enterprise Miner

- 범주형(로지스틱 회귀분석 & 의사결정 트리) 결과

	 Warnings	 Into: TARGET	 Unnormalized Into: TARGET	 Predicted: TARGET=1	 Predicted: TARGET=0
1		0	0	0,0811490273	0,9188509727
2		0	0	0,0752780565	0,9247219435
3		0	0	0,1312239183	0,8687760817
4		0	0	0,0968337479	0,9031662521
5		0	0	0,1314389103	0,8685610897
6		0	0	0,1367863486	0,8632136514
7		0	0	0,1312239183	0,8687760817
8		0	0	0,0744718969	0,9255281031
9		0	0	0,1195291261	0,8804708739
10		0	0	0,2228098012	0,7771901988
11		0	0	0,2816619933	0,7183380067
12		0	0	0,2273625468	0,7726374532
13		0	0	0,1283721597	0,8716278403
14		0	0	0,0777448469	0,9222551531
15		0	0	0,3453470557	0,6546529443
16		0	0	0,1312239183	0,8687760817
17		0	0	0,2030832278	0,7969167722

SAS Enterprise Miner

- 범주형(로지스틱 회귀분석 & 의사결정 트리) 결과



상위 **10,000명** 추출!