

[공모전 안내문]

SAMSUNG CARD DATA COMPETITION

당신의 생각이 세상을 바꿉니다.



SAMSUNG CARD DATA COMPETITION 안내문

SCDC 2020 운영사무국

SAMSUNG CARD

삼성카드는
데이터분석에 관심있는 대학생 및 대학원생에게
카드사 데이터 분석 기회를 제공하고,
문제 해결을 위한 참신한 아이디어 및 인사이트를 얻고자
SAMSUNG CARD DATA COMPETITION
(이하 SCDC)을 개최합니다.

1. 참가자격

국내외 대학생 및 대학원생

- (!) 개인 또는 팀(3인 이내)으로 지원 가능합니다.
- (!) 2020.8.24을 기준으로 참가 자격을 갖추어야 지원 가능합니다.(참가 접수 시 재학증명서 제출)
- (!) 휴학생 및 졸업예정자도 가능하며, 졸업생은 제외됩니다.

2. 주요일정

Track1	Track2
① 참가접수 : 2020.8.24(월) ~ 9.25(금) ② 1차 심사 합격자 발표 및 데이터 공개 : 2020.10.13(화) ③ 최종 결과를 제출 : 2020.11.10(화) ④ 2차 심사 합격자 발표 : 2020.11.19(목) ⑤ 최종 PT : 2020.11.25(수) 예정 ⑥ 수상자 발표 및 시상 : 2020.11.30(월) 예정	① 참가접수 : 2020.8.24(월) ~ 9.25(금) ② 최종 결과를 제출 : 2020.11.10(화) ③ PT대상자 발표 : 2020.11.19(목) ④ 최종 PT : 2020.11.25(수) 예정 ⑤ 수상자 발표 및 시상 : 2020.11.30(월) 예정

- (!) 합격자 발표는 각 신청자 문자메시지로 개별 안내해 드립니다.
- (!) 최종 프레젠테이션 상세 일정 및 장소는 합격자에 한해 개별 안내해 드립니다.
- (!) 참가자 심사 및 결과물 심사 단계에서 추가 자료를 요구할 수 있습니다.
- (!) 자세한 심사기준 및 심사결과는 비공개로 원칙으로 합니다.
- (!) 제출된 결과물이 심사기준에 부합하지 않을 경우, 수상작을 선정하지 않을 수 있습니다.
- (!) 공모전을 위해 사용하는 기법, 데이터, 제출되는 결과물 등은 타인의 소유권, 지식재산권(특허권·실용신안권·디자인권·저작권 등) 등 권리를 침해하지 않아야 합니다. 만일 권리 침해 문제가 발견되는 경우 그 책임은 참가자에게 있으며, 평가에 불이익을 받거나 수상 취소 등의 불이익을 받을 수 있습니다.

- (!) 본 공모전의 수상자가 삼성카드에 제출한 결과물에 대한 권리는 삼성카드에 귀속되며, 삼성카드에서 수정하여 분석 또는 마케팅 용도로 활용할 수 있습니다.
- (!) 상금의 제세공과금(22%)은 수상자 본인 부담이며, 상금은 수상자의 수령일 기탁 소득으로 산정됩니다.
- (!) 제세공과금 처리를 위해 개인정보 확인을 요청할 수 있습니다.
- (!) 삼성카드에서 제공하는 데이터는 대회 출품을 위한 목적 외에 이용할 수 없으며, 대회 종료 후 파기해야 합니다.
- (!) 삼성카드에서 제공하는 데이터와 응모작 관련 내용은 다른 사람에게 누설할 수 없으며, 비밀 유지를 위한 서약서 내용을 준수해야 합니다.
- (!) 대회 일정 및 세부 내용은 변경될 수 있으며, 그 내용을 사전에 알려 드립니다.

3. 결과물 제출 일정

① 참가 접수	2020.9.25(금)까지
· SCDC 참가 신청서 -----	[별첨1]
· SCDC 참가 서약서 -----	[별첨2]
· 개인정보 필수 수집·이용 동의서 -----	[별첨3]
· SCDC의 아이디어, 분석 기법 등 권리 인계 동의서 -----	[별첨4]
· 재학증명서 -----	[스캔하여 제출]
· SCDC 과제 계획서(Track1만 해당) -----	[자유 양식, 파워포인트 10장(표지 포함) 이내] (!) 데이터 모델링 전반 프로세스 및 마케팅 아이디어 개요 등 기재
② 최종 결과물 제출	2020.11.10(화)까지
· 결과 보고서 -----	기재 권장 사항 - 데이터 전처리, 모델링 과정, 모델 선정기준 및 분석 결과, 주제와 관련된 세부 질문을 고려한 데이터 해석 결과 및 관련 시각화 자료, 마케팅 아이디어 제안 (!) 파워포인트 파일로 제출 (필요 시 결과 보고서 외 설명자료는 워드 파일로 제출 가능) (!) 자유 양식, 분량 제한 없음
· 프로모션 결과 예측 파일 (quiz_s.csv/Track1만 해당)	(!) 제출 포맷은 참가자 이메일주소로, 분석 시 필요한 데이터와 함께 제공
· 데이터 분석 프로그램 코드 파일 (Track1만 해당)	
· 분석에 활용한 외부 데이터 및 데이터 출처	
③ 2차 심사 합격자 자료 제출	2020.11.23(월)까지
· 최종 프레젠테이션 발표용 파워포인트 파일 -----	(!) 결과 보고서를 15분 발표 분량으로 수정하여 제출

4. 최종 결과물 제출 시 유의사항

- ① Track1의 경우, 분석 프로그램은 Python입니다. 다른 프로그램 이용 시 심사에서 제외됩니다.
- ② Track2의 경우, 분석 프로그램의 제한은 없습니다.
- ③ 파워포인트 파일 작성을 위해 별도의 합법적인 소프트웨어를 이용하는 것은 무방합니다.
- ④ 분석을 위해 제공된 데이터 외에도 모델링 및 마케팅 아이디어 기획을 위해 외부 데이터 및 자료(날씨, 인구통계, 소비자 트렌드 등)를 자유롭게 활용할 수 있습니다. 다만, 이 경우에는 분석결과 산출을 위한 추가 데이터 및 해당 데이터 활용방법과 출처에 대해 설명 자료에 명시해야 합니다.
- ⑤ 분석을 위해 제공된 데이터와 다른 데이터를 결합하여 재식별화에 사용하는 것 또는 개인을 알아보기 위한 목적으로 가명 정보를 처리하는 행위는 금지됨을 안내드립니다.

5. 평가기준

Track1. 온라인 가맹점 방문 고객 예측 알고리즘 개발

영역	상세 내용
모델링	모델 정확성 · 예측 결과 파일을 기준으로 LIFT@20, AUROC 산출 · 예측 정확도 일정 수준(LIFT ≥ 2.5 , AUROC ≥ 0.80)이상 달성 시 최종 프레젠테이션 대상에 포함
	데이터 전처리 · 데이터 클렌징, 데이터 변환, Outlier 처리 등
	모델 개발 완성도 · 유효변수 선택 과정, 모델 후보군 및 해당 모델 고려 이유, 재현성 등
종합	데이터 해석력 및 구현 모델 설명력 · 주어진 데이터 해석/설명
	외부 데이터 활용 · 모델링 제안에 활용된 외부 데이터 적합도
마케팅 아이디어 제안	실현 가능성 및 구체성 · 삼성카드 업무 활용도 및 부합도

Track2. 팬데믹 시대를 극복할 마케팅 전략 수립

영역	상세 내용
데이터 분석력	데이터 해석력 · 주어진 데이터 해석/설명
	데이터 시각화 전달력 · 데이터 및 분석 방법 설명을 위한 자료 시각화
	외부 데이터 활용 · 마케팅 아이디어 제안에 활용된 외부 데이터 적합도
마케팅 아이디어 제안 (마케팅 기획)	사업 적정성 · 고객의 요구(needs)반영, 사업 추진 및 적용 용이성, 수익 창출 가능성
	논리성/설득력 · 자료 준비성, 내용 및 구성의 설득력/타당성/논리성
	창의성/독창성 · 참신한 아이디어, 서비스 기획안의 독창성 및 차별성

6. 알고리즘 평가 기준 (Track1만 해당)

6.1 알고리즘 성능평가 Metric

quiz.csv에 있는 3개 가맹점별 LIFT@20, AUROC의 평균 값 계산 후
아래 식으로 Final Score 계산

```
# FinalScore 계산
LIFT = Average(LIFT가맹점1, LIFT가맹점2, LIFT가맹점3)
AUROC = Average(AUROC가맹점1, AUROC가맹점2, AUROC가맹점3)

Final Score = 0.7(LIFT/5) + 0.3(AUROC)
```

· LIFT@20

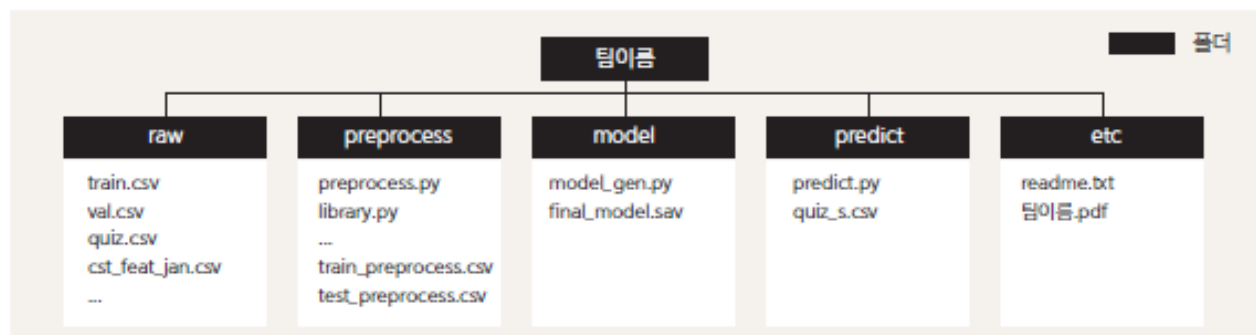
$$\text{LIFT} = \frac{\frac{\text{\# of 1 in top 20\%}}{\text{\# of 20\% sampled dataset}}}{\frac{\text{\# of 1 in total dataset}}{\text{\# of total dataset}}} = \frac{\text{예측한 데이터 기준 상위 20\% 데이터 중 1의 비율}}{\text{전체 데이터에서 1의 비율}}$$

예시 : LIFT_가맹점1

- 예측 score 기준 내림차순 상위 20% 데이터(2만) 중 1의 개수 : 550개
- 전체 데이터(10만) 중 1의 개수 : 750개
- LIFT@20 : $(550/20000)/(750/100000) = 3.67$

6.2 코드제출규칙

- 재현성 및 코드 검증 표준화를 위해 <그림1>과 같은 형식으로 파일 제출
- 제출 코드는 최종 LIFT와 AUROC를 복원할 수 있어야 함
- 코드와 주석의 인코딩은 UTF-8 사용
- 전체모델(preprocess.py + library.py + model_gen.py + final_model.py + predict.py)의 크기가 1GB를 넘지 않음
- 파일 제출 형식



<그림1>

6.3 Track1 데이터 설명 (Track2 데이터 설명은 "SCDC_공모전 테이블 설명.xlsx" 파일 참조)

본 공모전에서 제공하는 정보는 삼성카드 회원의 온라인 가맹점 이용 이력을 가공하여 만든 샘플 데이터입니다.

회원의 re-identification 및 민감 정보 보호를 위해 변수명을 비공개하는 점 양해 부탁드립니다.

1차 심사 때는 샘플 데이터가 주어지며, 1차 심사를 통과한 팀에게는 전체 데이터셋이 제공됩니다.

2차 심사 중 팀별로 총 5번까지 결과 제출(quiz_s.csv)이 가능합니다. 별도의 리더보드가 운영되지 않으며, 제출해 주신 결과에 따른 순위는 2차 진출 팀들에게 이메일로 개별 안내 드릴 예정입니다.

1차 심사용 샘플 데이터 목록(Method Proposal용)

samp_cst_feat.csv(19MB)	- 온라인 가맹점 이용 회원의 특징 정보 샘플
samp_train.csv(143.4KB)	- 온라인 가맹점 이용 이력
mrc_info.csv(5KB)	- 온라인 가맹점 업종 정보

2차 진출팀 제공 데이터 목록

trian.csv(7MB), val.csv(1MB)	- 1월 동안 회원 28만명이 이용한 10개의 온라인 가맹점 이력
cst_feat_jan.csv(776MB)	- 1월 동안 온라인 가맹점을 이용한 회원의 특징 데이터 - 회원 28만명에 대한 226개의 feature값 제공 - 회원의 익명성 및 민감 정보 보호를 위해 변수명 및 변수정보 비공개
cst_feat_feb_quiz.csv(200MB), cst_feat_feb_test.csv(209MB)	- 2월 온라인 가맹점 이용 여부 예측을 위한 회원 특징 데이터 - 회원 13만명에 대한 226개의 feature값 제공 - 회원의 익명성 및 민감정보 보호를 위해 변수명 및 변수정보 비공개
mrc_info.csv(5KB)	- 10개의 온라인 가맹점별 업종 정보
quiz.csv(5.4MB), test.csv(5.4MB)	- 회원 10만명이 해당 온라인 가맹점을 2월에 방문할 score 산출(1로 label될 확률) - quiz.csv : 3개 가맹점마다 10만명씩 매칭(1차 심사 통과 후 제공) - test.csv : 3개 가맹점마다 10만명씩 매칭(2차 심사 후 상위 6개 팀에게 제공)

사용가능 라이브러리(평가 환경의 라이브러리 버전)

- General DL/ML library/framework: tensorflow(2.1), pytorch(1.3.1), keras(2.3.1), scikit-learn(0.22.1)
- Scientific Computation: numpy(1.18.1), pandas(0.25.3), scipy(1.4.1)
- Visualization/Interactive: seaborn(0.10.0), matplotlib (3.1.3)
- Fast/Parallel computation: dask(2.11.0), cython(0.29.15)
- GBT: xgboost(1.0), lightgbm(2.3.1), catboost(0.22)
- Python Version : 3.6

7. 시상

구분	순위(팀)	혜택
Track 1	1등(1개팀)	500만원
	2등(2개팀)	300만원
	3등(3개팀)	100만원
Track 2	1등(1개팀)	300만원
	2등(2개팀)	150만원

8. 접수방법

삼성카드 홈페이지 (www.samsungcard.com)

→ 전체메뉴 → 고객센터 → 뉴스/공지사항

→ 공지사항 내 'SAMSUNG CARD DATA COMPETITION'게시글 클릭

→ 첨부 문서 작성 후 이메일(SCDC2020@samsung.com)로 참가접수

9. 문의

SCDC2020 운영사무국(SCDC2020@samsung.com)



[주제]

Track1. 온라인 가맹점 방문 고객 예측 알고리즘 개발

Track2. Pandemic 시대를 극복할 마케팅 전략 수립

(!) 중복 선택 및 중복 시상이 가능합니다.

Track1. 온라인 가맹점 방문 고객 예측 알고리즘 개발

온라인 업종 이용고객이 점차 늘어나고 있습니다. 1, 2차로 나누어서 주어지는 데이터에 기반하여, 이전 달 온라인 가맹점을 이용한 고객정보로 다음 달 온라인 가맹점을 이용할 고객들을 예측하는 모델을 만들어 보세요. 추가적으로 온라인 가맹점의 이용률을 높일 수 있는 마케팅 전략도 수립해 보세요

(!) 1차 데이터 (Track1_데이터 1~4)는 참가자 전원에게 주어지며, 2차 데이터는 1차 심사 통과자에 한하여 주어집니다.

Track2. Pandemic 시대를 극복할 마케팅 전략 수립

코로나19로 촉발된 팬데믹 시대의 도래는 삶의 많은 부분에 영향을 미치고 있습니다.
코로나19 이전과 이후의 소비시장에도 많은 변화가 있었습니다. [Track2_데이터 1~2]는 2019년 4~5월과 2020년 4~5월 업종별 카드 이용건수입니다. 주어진 데이터를 분석 및 해석하고, 이를 바탕으로 팬데믹 시대를 극복할 마케팅 전략을 수립해 보세요.

[별첨 1]

삼성카드 데이터분석 공모전(SCDC2020)

참가신청서

1. 참가자 정보 (개인일 경우 참가자 1 에만 기재)

참가자1	성명		학교명		전공	
	생년월일			휴대전화		
	이메일					
참가자2	성명		학교명		전공	
	생년월일			휴대전화		
	이메일					
참가자3	성명		학교명		전공	
	생년월일			휴대전화		
	이메일					

※ 원활한 공모전 운영을 위하여 휴대전화 번호와 이메일까지 모두 기입해주시기 바랍니다.

2. 참가 주제 선택(중복 선택 가능)

구 분	주 제	선택여부(Y/N)
Track 1	온라인 가맹점 방문 고객 예측 알고리즘 개발	
Track 2	팬데믹 시대를 극복할 마케팅 전략 수립	

※ 공모전 관련된 자세한 사항은 공모전 안내문을 참조 부탁드립니다.

2020 년 월 일

참가자 1 (서명)

참가자 2 (서명)

참가자 3 (서명)

삼성카드 데이터분석 공모전(SCDC2020)

참 가 서 약 서

삼성카드 데이터분석 공모전(SCDC2020)에 참여함에 있어,

아래의 사항을 서약합니다.

1. 공모전을 위해 사용하는 기법, 데이터, 제출되는 결과물 등은 타인의 소유권, 지식재산권(특허권·실용신안권·디자인권·저작권 등) 등 권리를 침해하지 않아야 합니다.
2. 삼성카드에서 제공하는 데이터는 대회 참가를 위한 목적 외에는 이용할 수 없으며, 공모전 종료 후에 반드시 파기해야 합니다.
3. 삼성카드에서 제공하는 데이터와 응모작 관련 내용은 다른 사람에게 누설할 수 없으며, 비밀 유지 의무를 성실히 이행해야 합니다.
4. 삼성카드에서 제공하는 데이터를 삼성카드가 제공하지 않은 다른 데이터와 결합 및 특정 개인을 알아볼 수 있도록 재식별화 하지 않습니다.
5. 상기 사항 위반 시 그에 대한 책임은 본인에게 있으며, 평가에 불이익을 받거나 수상 취소 등의 불이익을 받을 수 있습니다.

2020년 월 일

생년월일

서약인

(서명)

[별첨 3]

[개인정보 필수 수집 · 이용 동의서]

삼성카드주식회사 귀중

본인은 삼성카드주식회사(이하 “회사”라 함)가 아래의 내용과 같이 본인의 개인(신용)정보를 수집 · 이용하는데 동의합니다.

1. 수집 · 이용하는 개인정보의 항목 및 수집 · 이용 목적

항목	수집 · 이용 목적	보유기간
*성명, 학교명, 전공, 생 년월일, 휴대전화, 이메일, 기타 재학증명서 내 기재 사항	<input type="checkbox"/> 데이터분석 공모전 운영	동의서 제출일부터 본인이 “회사” 의 공모전에 참가하는 기간까지 (※ 단, 관계 법령상 별도의 정함이 있는 경우 그에 따라 보유 · 이용, 보안사고 및 저작권 침해 등 기타 문제가 발생시에는 사고 조사 · 처 리가 완료될 때까지 보유 · 이용함)

2. 동의를 거부할 권리

귀하는 본 개인정보 수집 · 이용에 대한 동의를 거부하실 권리가 있으나, 동의를 거부하실 경우
공모전 참가가 불가할 수 있음을 알려 드립니다.

◆ 본인은 귀사가 상기 수집 · 이용을 위하여 다음과 같은 본인의 정보를 처리하는 것에 동의합니다.

동의함 ☐ / 동의하지 않음 ☐

※ 개인정보 처리에 대한 상세한 사항은 삼성카드 홈페이지 (<http://www.samsungcard.com/>)에 공개한
'개인정보 처리방침'을 참조하십시오.
다만, 본 동의서 내용과 상충되는 부분은 본 동의서의 내용이 우선합니다.

성 명*		휴대전화번호*	- -	생년월일*	년 월 일
------	--	---------	-----	-------	-------

본인은 상기 동의서 내용을 충분히 설명 듣고 이해하여 이에 동의합니다.

_____년 _____월 _____일 성명 : _____ (서명 또는 날인)

삼성카드 데이터분석 공모전(SCDC2020)

아이디어, 분석기법 등 권리 인계 동의서

삼성카드주식회사 귀중

본 삼성카드 데이터분석 공모전(이하 “공모전” 이라 함)은 삼성카드 마케팅 및 홍보 등을 위하여 삼성카드의 주관 아래 진행되는 바,

“공모전”에 참여하는 본인은 “공모전”에서 수상자로 선정될 경우, “공모전” 진행 과정에서 삼성카드에 제출하는 본인의 모든 제작물 및 그에 포함된 분석기법 등 일체의 아이디어(이하 “아이디어 등”이라 함)를 삼성카드가 복제, 배포, 공연, 공중송신, 전시, 대여, 2차적 제작 또는 편집(색상, 형태, 규격 변경 등)할 수 있도록 아래에 해당하는 저작권, 소유권 기타 “아이디어 등”과 관련된 모든 법적 권리를 “공모전”과 관련하여 제공하는 상금 및 기타 고지된 혜택 외 별도 금원의 지급없이 삼성카드에게 양도합니다.

◆ 아이디어 등에 대한 소유권 및 저작권 등

- 저작권, 디자인권, 퍼블리시티권 등 일체의 지식재산권, 소유권 등
제반 권리 및 상업적 사용권한
- 아이디어 등을 앱, 출판, 광고, 영상물 등 모든 종류의 마케팅Tool로 자유롭게 활용

상기 정보를 모두 확인 하였으며 이에 동의 합니다.

2020년 월 일

성 명 : (서명)

생년월일 :