

*Computer Science, CNU*

# Mini Project

**Dongil Kim**

**Dept. of Computer Science and Engineering**

**Chungnam National University**

# Mini Project

## ❖ 문제

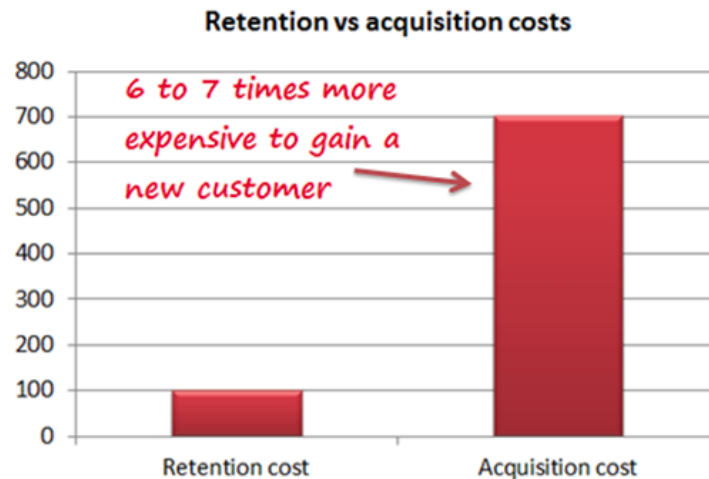
### ■ Response Modeling

- 누가 우리 캠페인에 반응할 것인가?
- E.g. 카탈로그, 전화, 쿠폰, donation 등등
- 우리나라에서 하루에 50만개의 쿠폰이 발행됨
- 누가 반응할 것인가?
- 누가 더 많이 return을 줄 것인가?

# Mini Project

## ❖ 배경이론

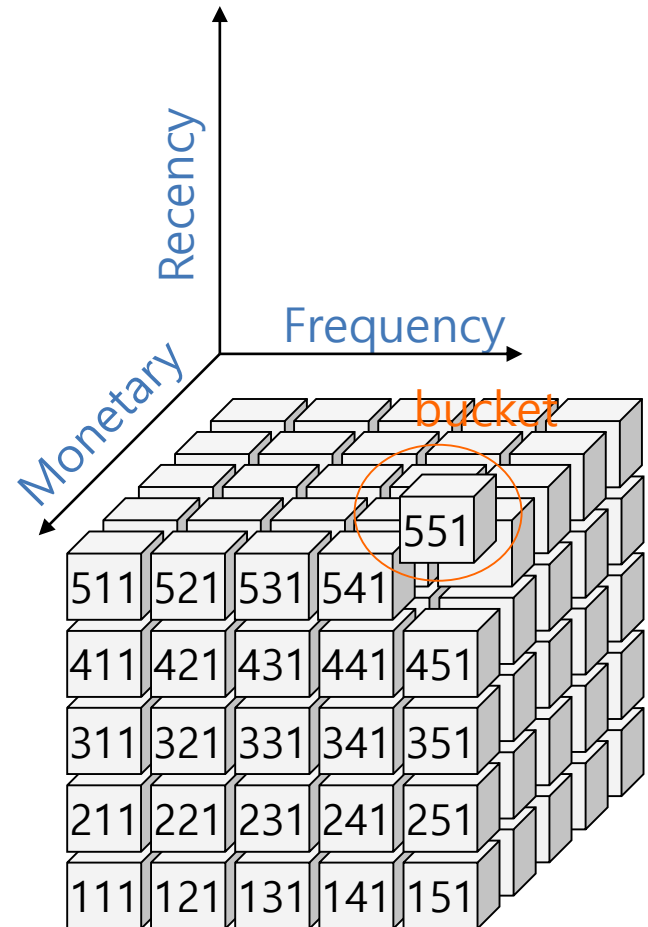
- Customer Relationship Management (CRM)
  - 신규 고객을 유치하는 하는 것보다 기존 고객을 유지하는 것이 훨씬 저비용이다



# Mini Project

## ❖ 배경이론

- RFM 데이터
  - 고객의 구매 패턴 데이터
  - Recency
  - Frequency
  - Monetary



# Mini Project

## ❖ 데이터

### ■ 학습데이터

- tmx: 입력데이터

(50,766개, 15변수)

- tmy: 반응여부

(4,786 반응, 약 9.4%)

- trndol: 반응금액

### ■ 평가데이터

- tstx: 입력데이터

- tsty: 제출물

15 input variables used in the experiments

Name	Formulation	Description
<i>Original variables</i>		
Purseas		Number of seasons with a purchase
Falord		LTD fall orders
Ordtyr		Number of orders this year
Puryear		Number of years with a purchase
Sprord		LTD spring orders derived variables
<i>Derived variables</i>		
Recency		Order days since 10/1992
Tran53	$I(180 \leq \text{recency} \leq 270)$	
Tran54	$I(270 \leq \text{recency} \leq 366)$	
Tran55	$I(366 \leq \text{recency} \leq 730)$	
Tran38	$1/\text{recency}$	
Comb2	$\sum_{m=1}^{14} \text{ProdGrp}_m$	Number of product groups purchased from this year
Tran46	$\sqrt{\text{comb2}}$	
Tran42	$\log(1 + \text{ordtyr} \times \text{falord})$	Interaction between the number of orders
Tran44	$\sqrt{\text{ordhist} \times \text{sprord}}$	Interaction between LTD orders and LTD spring orders
Tran25	$1/(1 + \text{lorditm})$	Inverse of latest-season items

# Mini Project

---

## ❖ 프로젝트 목표

- 카탈로그를 딱 1,000개만 보낼 수 있음
- 누구에게 보낼 때 가장 기대수익이 높을 것인가?

# Mini Project

## ❖ 생각할 수 있는 방법들

### ■ Classification

- $trnx-trny$ 를 모델링하여 반응확률이 가장 높은 1천명을 고르자

### ■ Regression

- $trnx-trndol$ 을 모델링하여 반응금액이 가장 큰 1천명을 고르자
- 혹시,  $trnx$  전체를 안쓰고 반응자만 모델링하면 어떨까?

### ■ 2-Stage Model

- Classification으로 반응자를 고른 후, Regression으로 반응금액을 예측하여 1천명을 고른다면?

# Mini Project

## ❖ 생각할 수 있는 방법들

- Novelty detection
  - 기존 반응자와 가장 유사한 1천명을 골라보자
  - 기존 반응하지 않은 사람과 가장 먼 1천명을 골라보자
- Ensemble
  - 모델링은 무조건 앙상블이 좋을까?



# Mini Project

## ❖ 생각할 수 있는 방법들

- Class-imbalanced Problems
  - 반응률이 9.4%인 불균형 상황에서 어떻게 해야할까?
- Cross-validation
  - 내 모델의 정확도를 어떻게 객관적으로 공정하게 평가할 것인가?
- 변수선택
  - 저 15개의 변수를 무조건 다 쓰는게 좋을까?
- Normalize는 되어있으니까 안하셔도 됩니다.

# Mini Project

## ❖ 제출

- 2022년 6월 21일(화) 23시 59분까지, 사이버캠퍼스
- 제출물
  - tsty.csv 파일에 1천명까지 1을 붙여서 제출
    - 그럼 그 사람들에 대한 반응금액의 합산이 각 팀의 점수가 됩니다
  - 모델링 과정을 간략하게 정리한 pdf(자유형식)
    - 코드 나열이 아닌 전체적인 구조에 대한 설명 필요
- 평가
  - 반응금액 점수 50% + 모델링 과정 참신성 50%

# Mini Project

## ❖ 제출

### ■ 중간 피드백

- 각 팀 별 3번까지 `tsty.csv`를 제출해주시면 점수를 계산해서 드립니다.
- Unrealistic한 상황이지만 잘 활용해보세요.
- Leader board 운영이 없는 대신 수업시간에 공개할 수도 있음

# Mini Project

## ❖ 제출

- 만약 random guess를 한다면?
  - 1,000명 중 9.4%의 반응률 = 약 94명 반응자
  - 1인당 반응금액 평균 \$47
  - $94 * 47 = 4,418$
  - 점수의 baseline이 저거라고 생각하시면 됩니다.
  - 최고 획득 가능 점수는 116,332

# Q&A

