

이전 찜,구매 기록 기반 찜유저 구매 확률 예측

번개장터 데이터팀 인턴 프로젝트

목차

01

서론

- 프로젝트 기본 설명

02

본론

- 분석 과정 세부 설명

03

결론

- 모델 적용 결과
- 발전 방향
- 활용 방안



이

서론

프로젝트 기본 설명

“

특정 상품을 찜한 유저들의 구매 확률은 어떠할까?

- 유저의 이전 찜, 구매 기록과 판매자 정보에 기반한, "현재 찜 상품 구매확률 예측 모델"
- 분석 대상: 20201126~20201205 찜 유저 전체

+

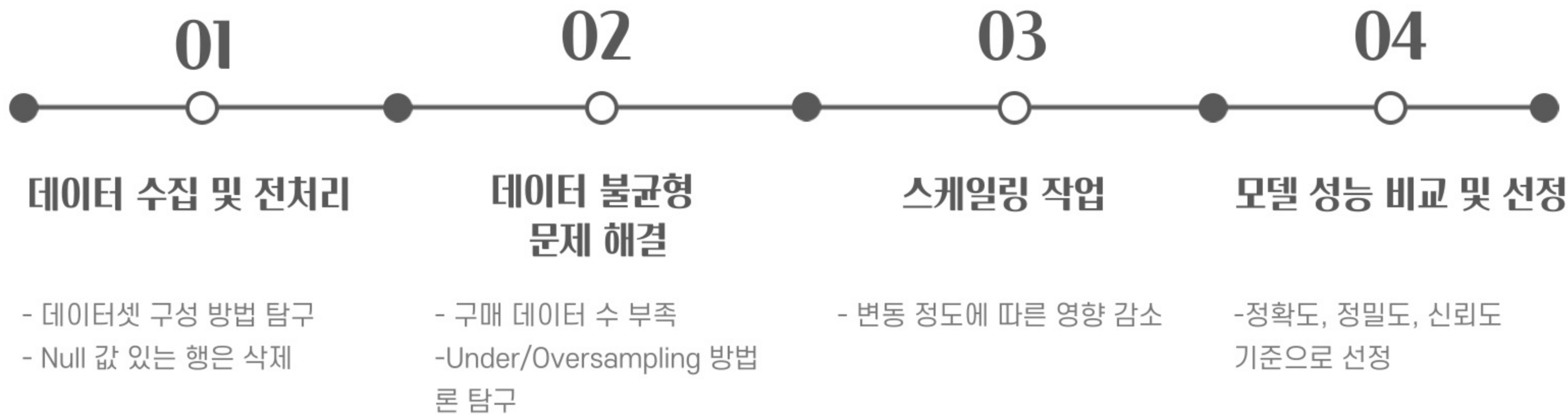
02

본론

분석과정 세부 설명

분석 과정 세부 설명

분석 절차 시간순 정리



테이터 수집 및 전처리

테이터셋 구성 방법 탐구

- 활용 테이블

테이터 수집 및 전처리

테이터셋 구성 방법 탐구

- 테이터셋 예시

데이터 수집 및 전처리

데이터 분포 확인 및 전처리

| ID | 과일 |
|----|-----|
| 1 | 사과 |
| 2 | 바나나 |
| 3 | 체리 |

One-Hot Encoding

| ID | 사과 | 바나나 | 체리 |
|----|----|-----|----|
| 1 | 1 | 0 | 0 |
| 2 | 0 | 1 | 0 |
| 3 | 0 | 0 | 1 |

LabelEncoder

| ID | 과일 |
|----|----|
| 1 | 0 |
| 2 | 1 |
| 3 | 2 |

<https://mizykk.tistory.com/13>

데이터 불균형 문제 해결

Oversampling 방법을 중심으로...

스케일링

변수별 변동 정도 차이에서 오는 영향 감소

모델 성능 비교 및 선정

정확도, 정밀도, 재현율 기준으로.

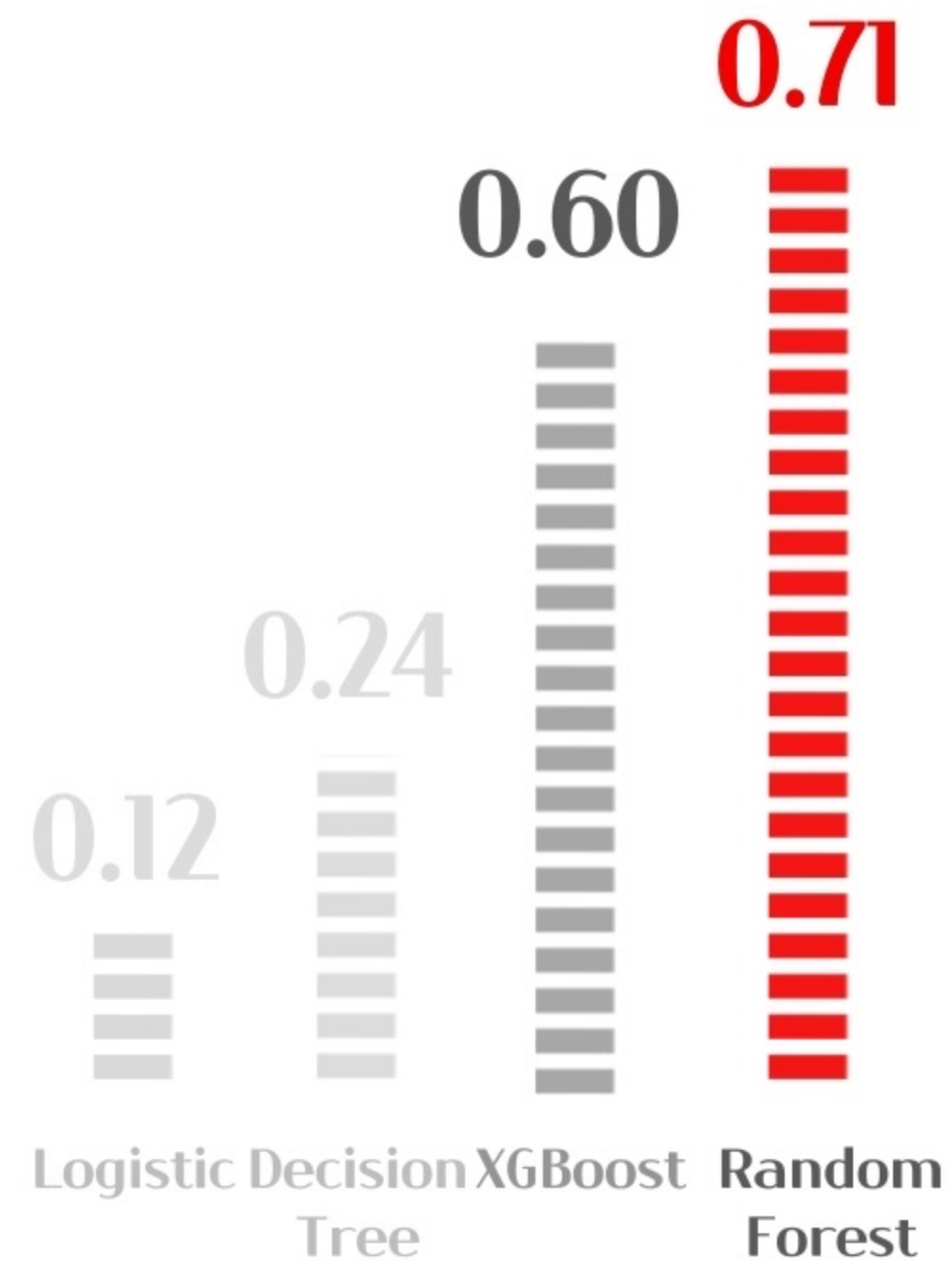
03

결론

모델적용결과, 한계점 및 발전방향

최적 분석기법 선정

분석기법 별 Precisin 비교



한계점 및 발전방향

01 데이터 불균형 정도 컸음

번장 프로미스, 번개페이 기록만을 가지고 구매 여부를 판단해, 실제 구매가 일어난 것 보다 적게 기록됐을 것. 번개톡 기록 뿐만 아니라 실제 구매 여부를 더 알 수 있다면 구매(1) row 수가 늘어날 것.

03 샘플링 방법 대안 탐구 필요

현재 smote 이외에도 효과적인 샘플링 방법이 존재.(adasyn 등). 과적합은 피하면서 조금 더 효율적인 오버샘플링 방법을 찾아야 한다.

02 주관적인 기간 설정

며칠간의 찜을 통한 탐색 후에 구매하는 유저들을 고려하기 위해 찜 기간과 구매 기간에 차이를 두었는데 이 과정이 주관적으로 진행됐음. 추가 통계자료를 활용하면 더 정확할 것이다.



활용 방안

번개장터 앱에 도입 시...

○ 예상 활용 모습

좌측 이미지와 같이 구매 확률이 상대적으로 높은 상점에 표시

○ 기대효과

구매확률 상위 상점 집중 공략으로 제품 판매에 투자되는 시간 및 피로도 감소 -> 번개장터 긍정적 인식 -> 활용률 높아짐

발표를 들어주셔서
감사합니다 :))