

이 제품, 수상한데요?

1등팀 김성연



목차

01

AI 모델
핵심 요약

02

변수 전처리

03

모델링

04

모델 활용 방안

01

AI 모델 핵심 요약

모델을 간략히 소개합니다.



다양한 코드 입력받기

○ 어느 국가에서 누가 보냈는지?

적출국가코드, 원산지국가코드, 신고인부호, 수입자 부호, 해외거래처부호, 특송업체부호,

○ 어떤 절차를 거쳤는지?

통관지 세관부호, 수입거래 구분코드, 수입종류 코드, 징수형태 코드, 운송수단 유형코드, 원산지 표기 유무 코드

○ 무슨 제품군인지?

신고중량, 과세가격 원화금액, HS10단위부호, 관세율, 관세율구분코드

AI 모델 사용하기

AI 모델은 대부분이 범주형 부호인 입력을 받고 우범여부를 다음과 같이 판단합니다.

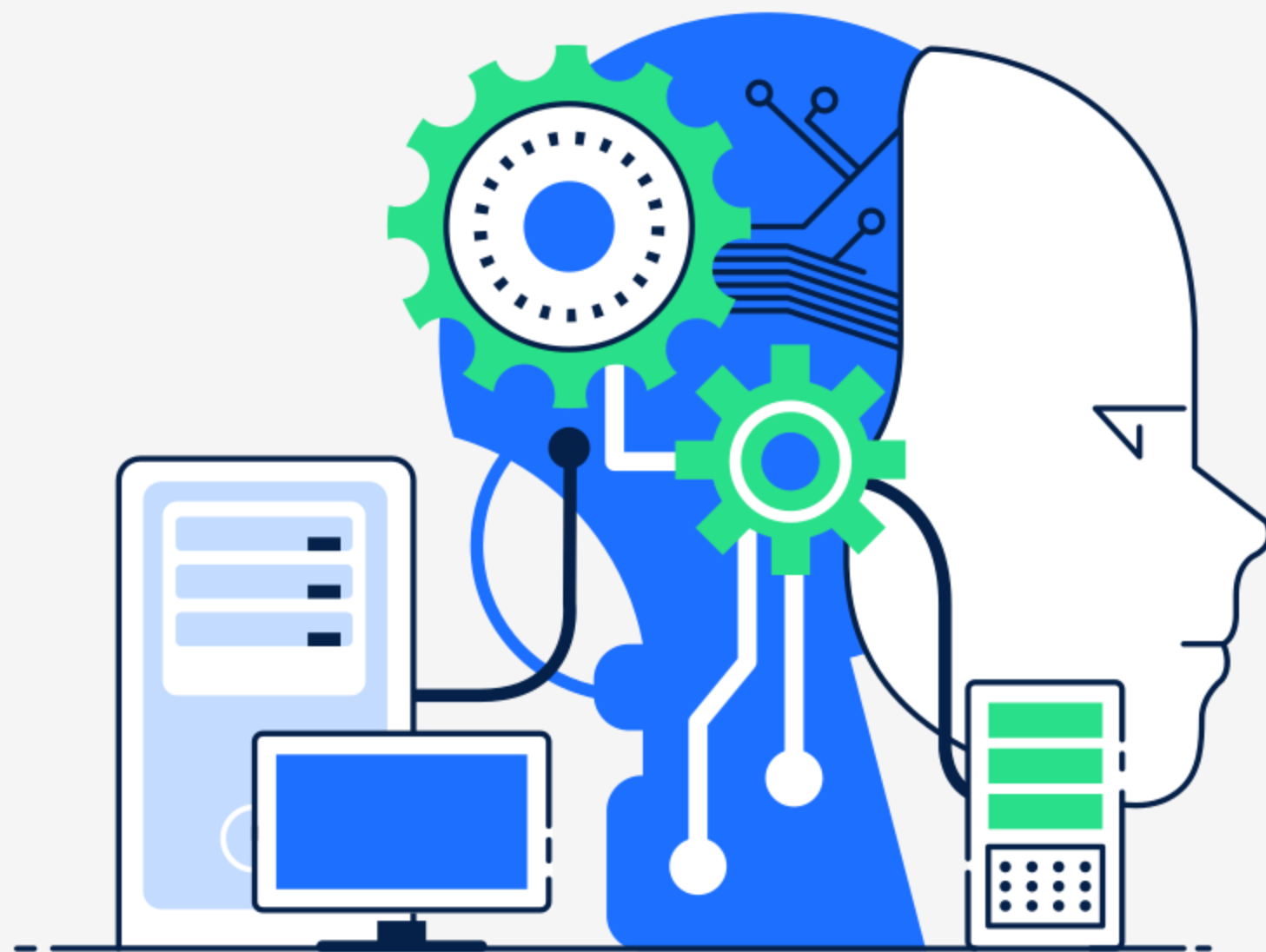
" 'HK'국가에서 'FWVNMKK' 사람이
'RZM2Z77' 거래처를 사용했는데 12
(성남) 세관을 이용했다고?..
이 제품, 수상한데요? "



AI 모델이란?

AI 모델은 창의력 있는 학습을 하지 않습니다. 쌓인 데이터에 충실해 결과를 도출할 뿐입니다.

AI 모델의 우범화물 탐지 메커니즘은 경력이 오래된 사람의 행동과 비슷합니다. 다만 사람 대비 복잡한 계산 능력이 뛰어난 것이 장점입니다.



02

변수 전처리

모델링 전 변수 전처리 과정을 설명합니다.

범주형 코드 전처리

01

표본이 적은 항목
기타 항목으로 묶기

70번(완도) 세관 같이 표
본이 적은 항목을 기타
항목으로 묶었습니다.

02

일부 범주의 NULL값
하나의 항목으로 취급

해외거래처와 특송업체
부호 범주의 NULL(미입
력)값을 하나의 항목으로
취급했습니다.

03

모든 범주형 변수
One-Hot Encoding

범주 내 항목별로 해당한
다면 1, 아니라면 0을 갖
는 One-Hot Encoding을
진행했습니다.

변수 추가



신고 일자 변수를 이용해
요일과 공휴일 변수를 추가
했습니다.



데이터 내 우범여부가 1인 비
율이 약 21.7% 로 낮아 랜덤
오버샘플링을 통해 비율을
50% 로 맞췄습니다.



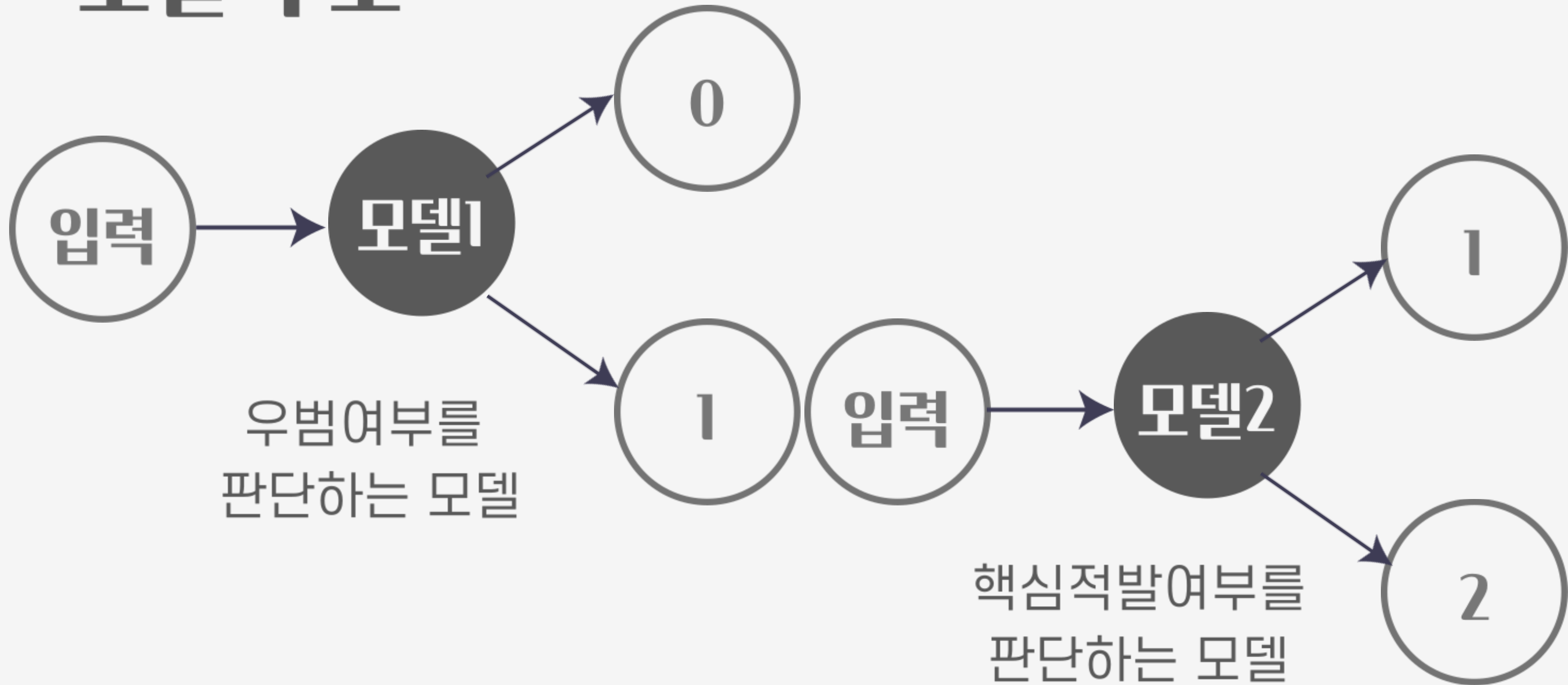


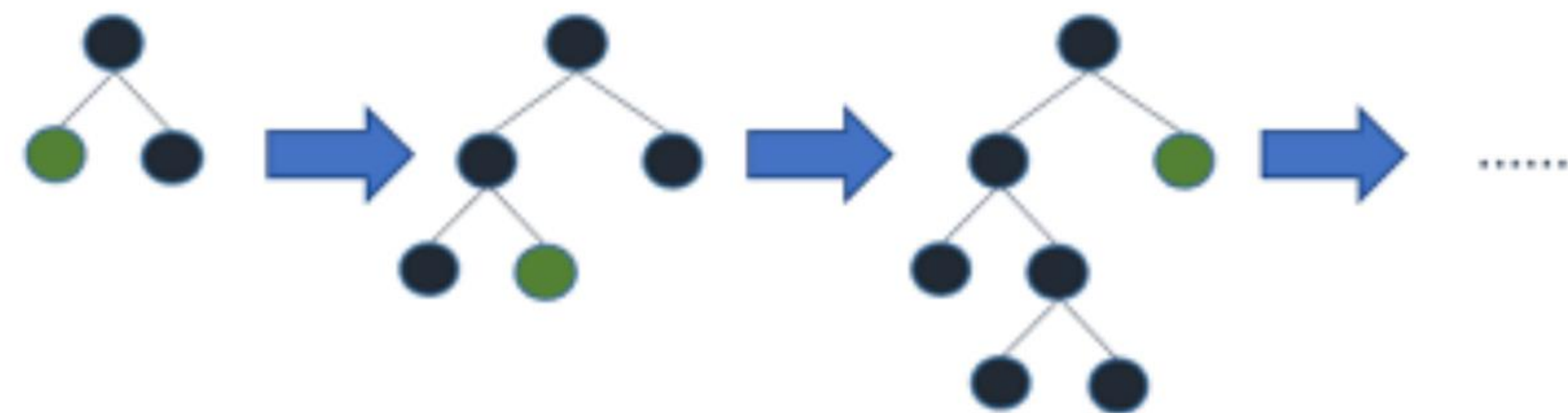
03

모델링

전반적인 모델 구조를 소개합니다.

모델 구조





LightGBM

사용한 모델

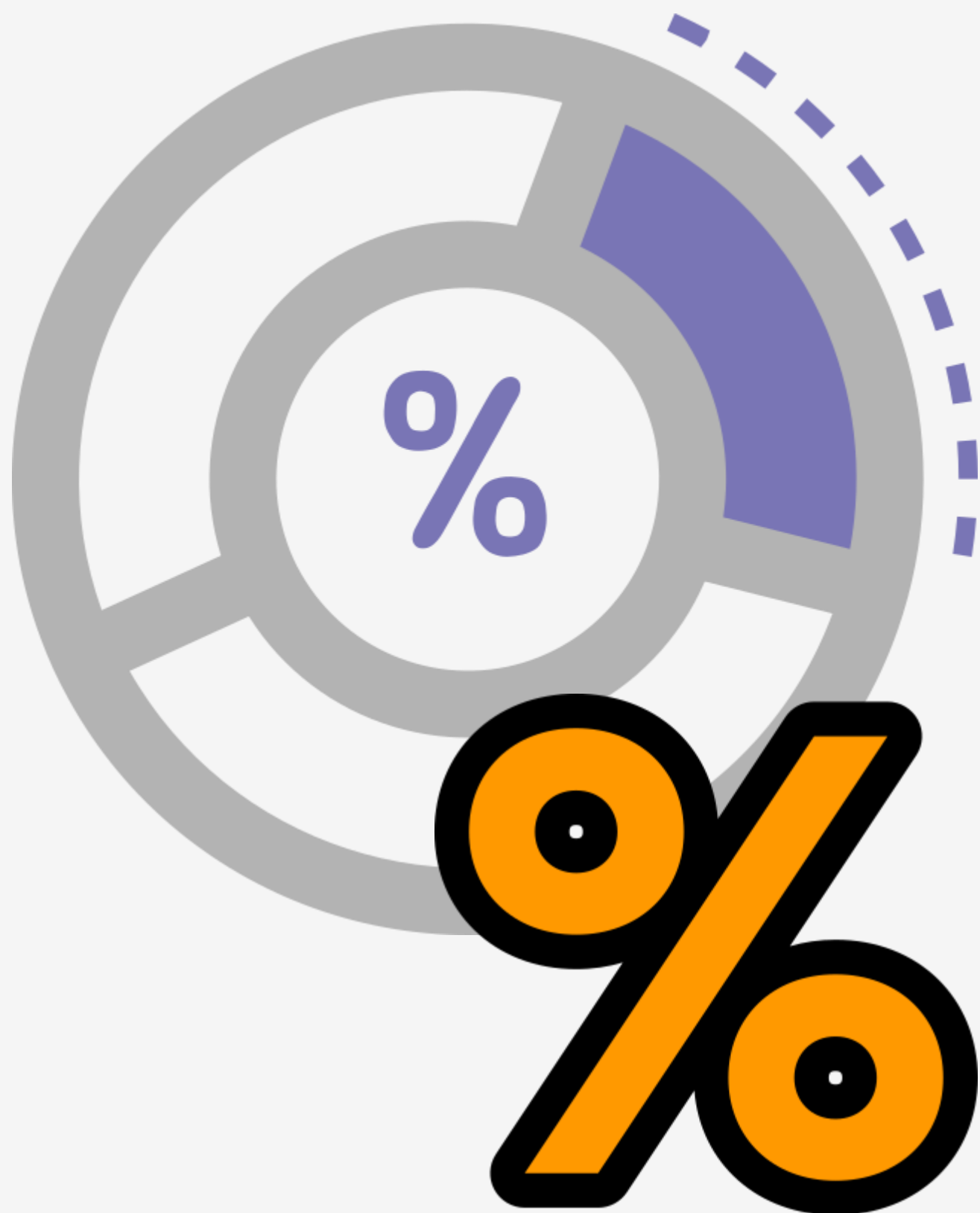
Light GBM은 트리기반
그래디언트 부스팅 모델
입니다.

속도가 빠르고, 캐글 내
수상경력이 많은 성능 좋
은 모델입니다.

04

모델 활용 방안

실제 모델이 어떻게 활용되는지 알아봅니다.



얼마나 수상한가요?

"AI 모델은 우범화물일 확률을 알려줍니다."

*모델 적용시
고려해야하는 사항*

- 우범화물 적발율
- 모델의 우범화물 예측 정확도
- 의심되는 화물 확인 비용



모델을 더 잘 활용하려면? +

○ 목표 우선순위 정하기

목표 우선순위에 따라 우범화물 판단 기준점을 정한다면 모델을 원하는 목적에 맞게 사용할 수 있습니다.

○ 주기적인 모델 유지보수

입력값이 대부분 범주형 부호이므로 지속적인 관찰을 통해 일정한 주기로 업데이트 해야합니다.

발표를 들어주셔서
감사합니다 :))

1등팀 김성연