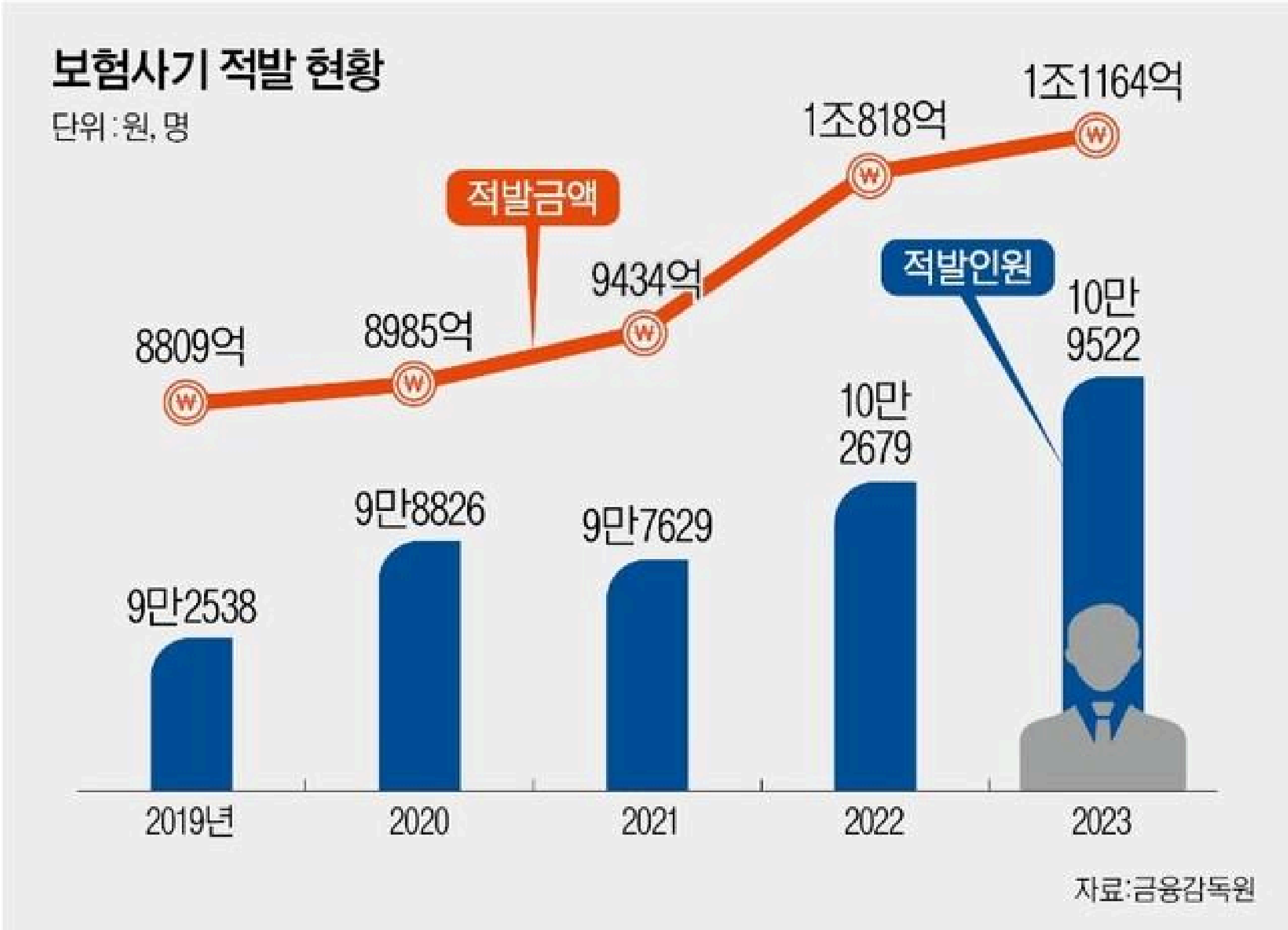


보험 사기 탐지 모델 개발

Project - 1



*해당 분석은 실제 데이터가 아닌 연습용 데이터로 제작되었습니다.

배경

- AI 시대의 변화 속도가 매우 빠르기 때문에, 이에 발맞추는 것이 중요하다고 생각합니다.
- 보험업계에서는 항상 보험 사기가 주요 이슈이며, 지속적으로 손실을 입고 있습니다.
- 데이터 분석을 활용하여 보험 사기 리스크를 효과적으로 줄이고자 하였습니다.

프로젝트 요약

- 사기 리스크 예측 모델(XGBoost) 정확도 75.6%
- 주요 변수: 사고 심각도, 피해금액, 가입기간
- 산점도·Feature Importance로 시각화
- 보험 사기 조기 탐지 기준 제시

작업 기간

2025. 02. 24 - 03. 31

사용 툴



구성원

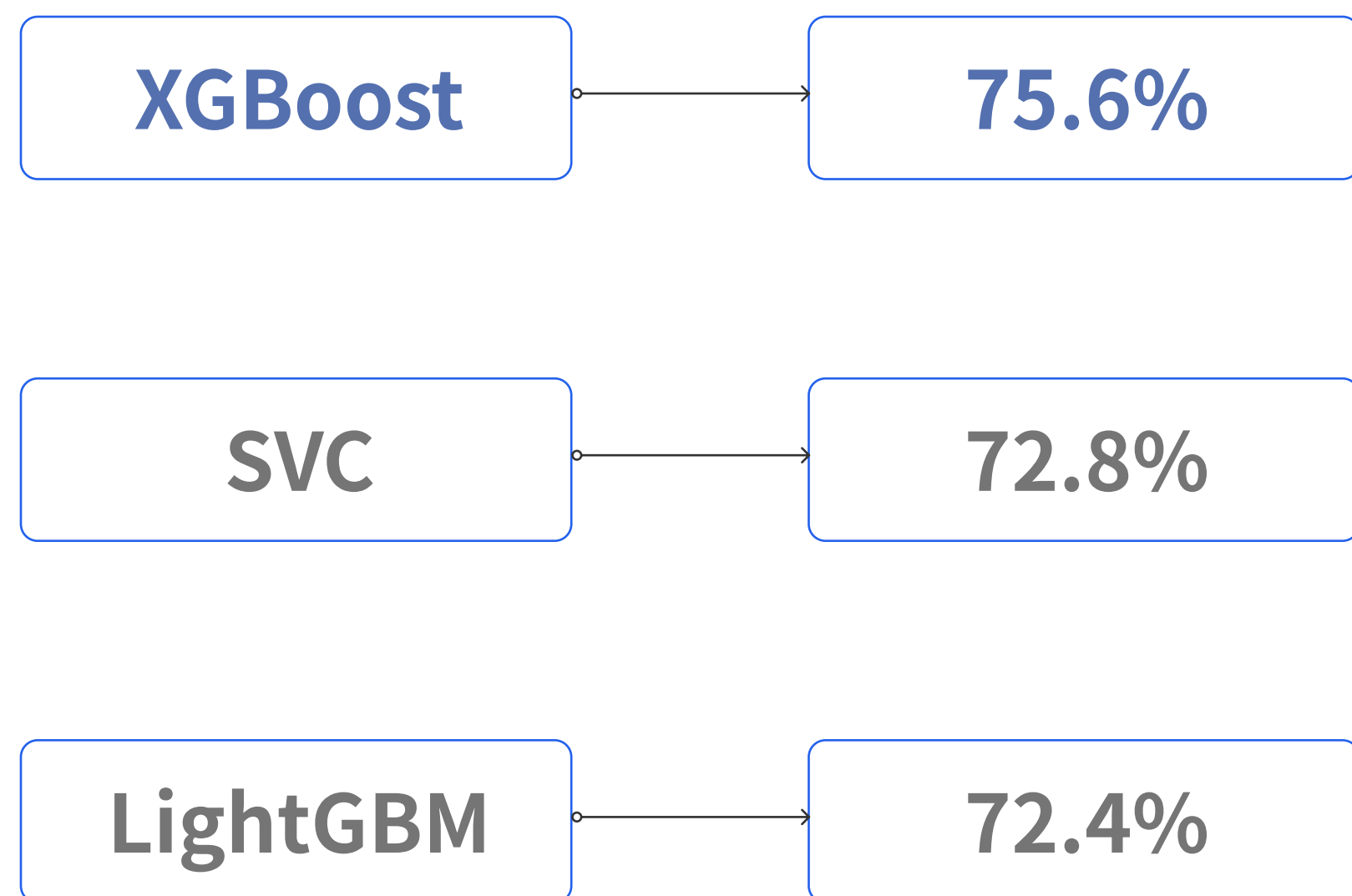
기여도 100%

Model Development

보험 사기 탐지 모델 개발

Project - 1

모델별 정확도 결과



모델별 정확도 차이 도출

- data set 1,000건인 경우 약 32건이 보험 사기로 예측 되면
리스크 비용은 수천만원 이상 차이 나기 때문

XGBoost Classifier 분석

- 부스팅 기법을 활용한 분류 모델이며, 대규모 데이터셋에서 높은 예측 및 과적합 방지하기에 해당 모델 선택
- 보험 사기 탐지 문제에서 연속형 및 범주형 변수 혼합된 데이터셋을 효과적으로 처리할 수 있음

Support Vector Classifier(SVC) 분석

- 비선형 데이터 분류에 강점, 커널 트릭을 사용해 선형적 구분할 수 없는 데이터도 고차원으로 변환하여 쉽게 분류
- 대규모 데이터 처리에 비효율 및 학습 속도가 느림

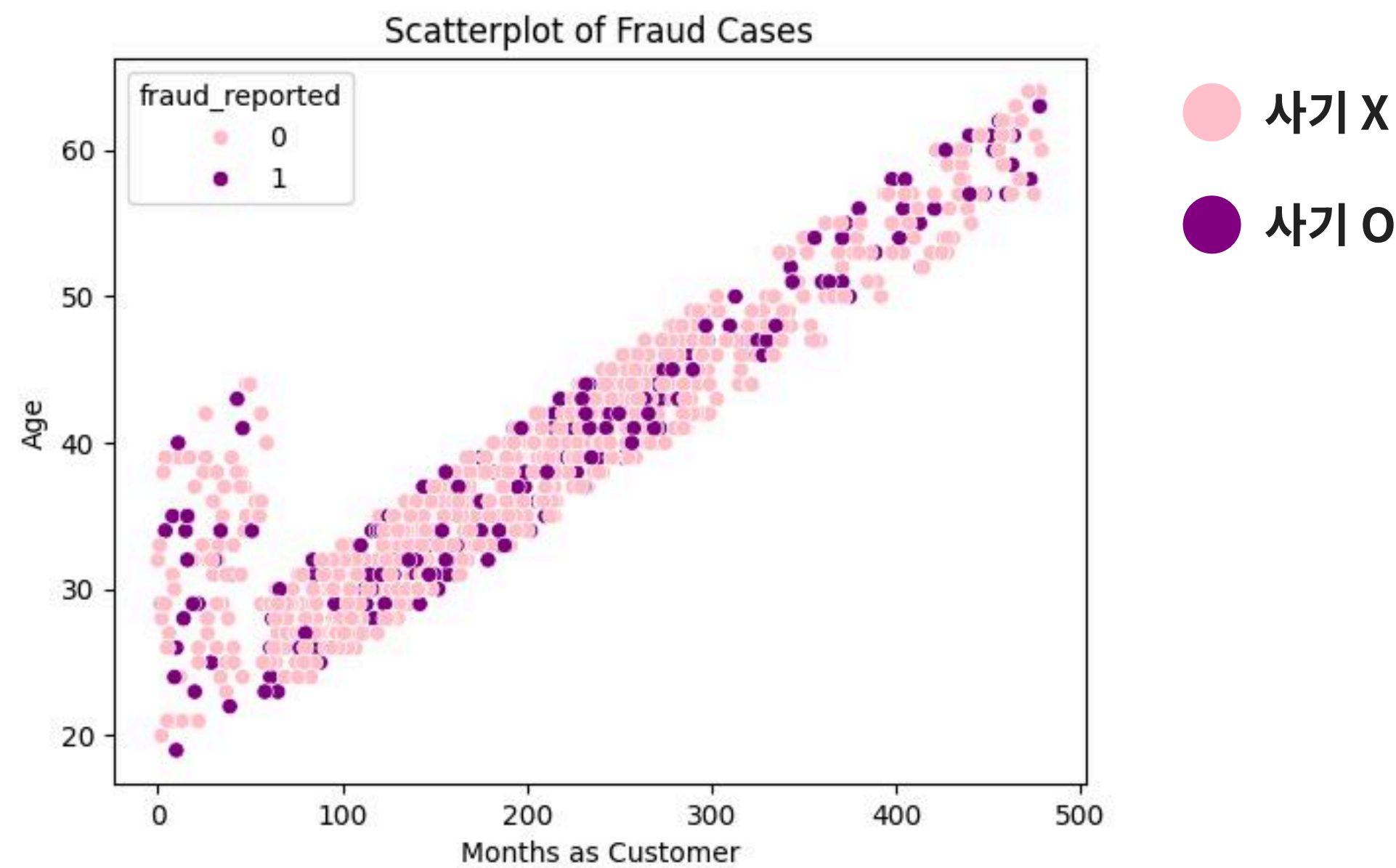
LightGBM 분석

- XGBoost와 유사한 부스팅 기법을 사용하지만, 대량의 데이터를 빠르게 학습할 수 있음
- 빠른 학습 속도, 과적합 방지 기능, 대규모 데이터셋 처리에 최적화

Model Development

보험 사기 탐지 모델 개발

사기 의심 사건 여부 산점도

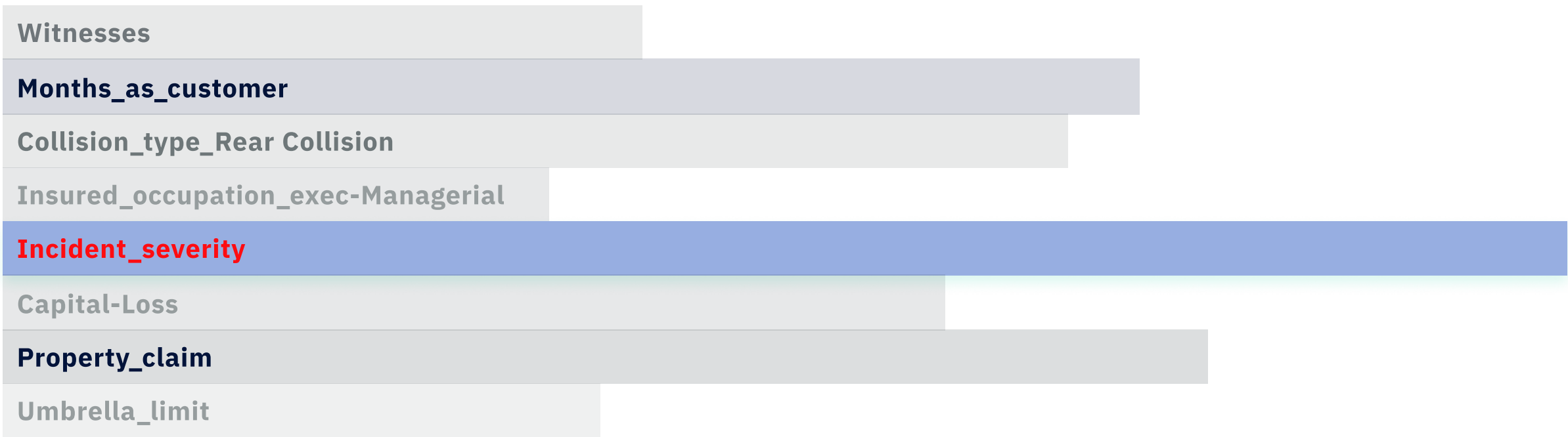


- 20-40대 사이 고객들은 주로 보험 사기 신고가 많이 발생하는 경향이 있음
→ 경제적 부담이 크고 빠르게 현금화할 수 있는 보험 청구를 악용 가능성이 큼
 - 나이가 많아지고, 가입 기간이 길어지면 보험 사기 신고 비율이 현저히 낮음
→ 보험에 대한 신뢰가 높아지면서 사기 비율 감소
- ✅ 보험 사기는 젊고 가입 기간이 짧을 때 발생 확률이 높다.

XGBoost 중요 변수

Feature Importance for XGBoost Classifier

Feature Importance



- ✅ 사고의 심각도, 피해 금액, 가입 기간이 보험 사기 예측에서 가장 중요한 변수로 작용하는 것을 도출하였습니다.
- 사고의 심각도가 클수록 청구 금액도 커지기 때문에, 사기성 청구가 발생할 확률이 높습니다.
- 이후, Kafka와 Flink를 활용해 실시간 보험 사기 탐지 프로그램을 구축하고자 합니다.