

1. 개요

본 보고서는 사내 인사 데이터를 바탕으로 **Random Forest** 및 튜닝된 **GBClassifier** 모델을 이용하여 퇴사 가능성이 높은 직원을 사전에 파악하기 위해 개발된 **직원 퇴사 예측 모델**에 대한 설명입니다.

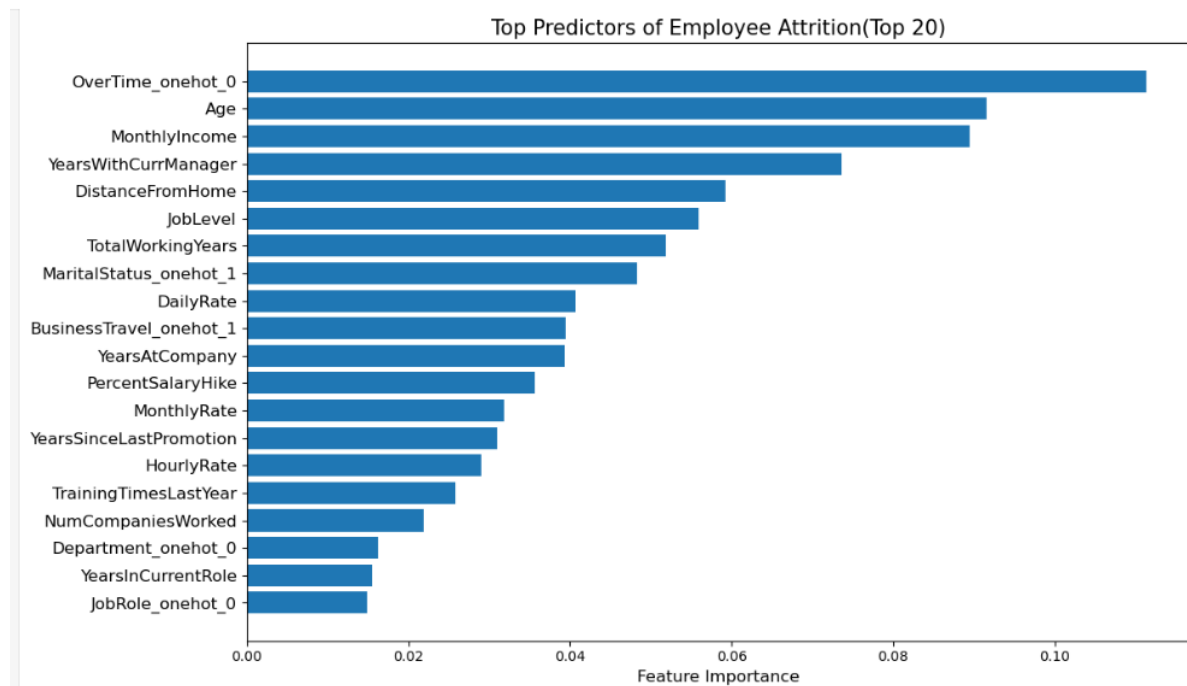
모델 개발 과정, 핵심 변수 해석, 두 모델의 주요 성능 지표, 시각화 결과, 그리고 이를 활용한 인사 전략 제안까지 포함하였습니다.

2. 모델 개발 과정

- 데이터 전처리
 - Age, MonthlyIncome, TotalWorkingYears, YearsAtCompany 등의 변수에 대한 필터링 및 이상값 제거
 - Attrition 값이 'Yes' 또는 'No'인 경우만 선택하여 이진 분류 목적에 맞게 정제
 - 범주형 변수(Gender, OverTime 등) 인덱스 부여 후 '원핫 인코딩' 진행
- 피처 엔지니어링
 - 수치형 변수는 결측치 대체(Imputer)
 - 범주형 변수는 StringIndexer와 OneHotEncoder를 통해 원-핫 인코딩 수행 (dropLast 적용)
- 벡터화 및 레이블 설정
 - 모든 피처를 VectorAssembler로 묶어 feature_vector 생성
 - Attrition을 label로 변환 (0: 재직, 1: 퇴사)
- 모델 학습 및 평가
 - RandomForestClassifier로 모델 훈련
 - 80% 학습, 20% 테스트 데이터로 성능 평가 수행

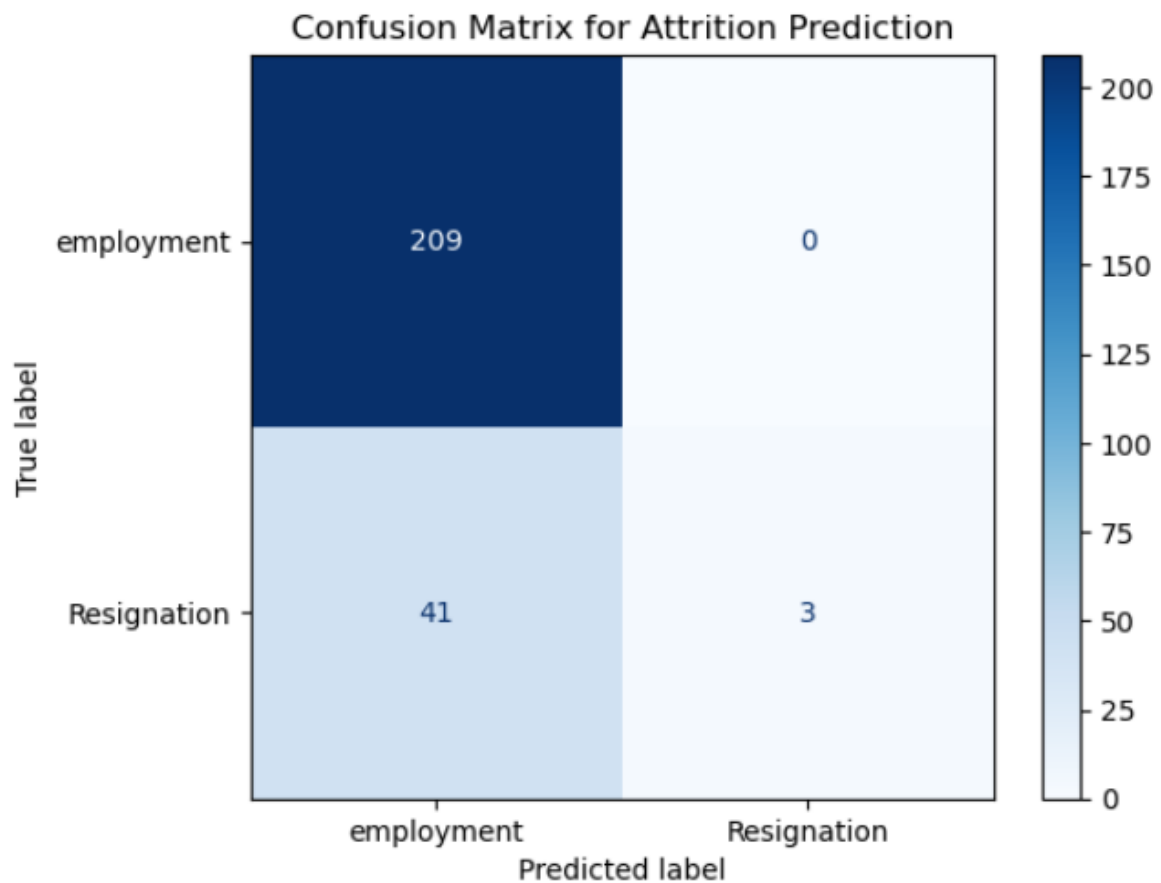
3. 시각화 결과 및 해석

3.1 주요 퇴사 예측 변수



- **OverTime_onehot_0:** 퇴사 예측 기여도 1위
 - 초과근무 여부가 가장 강력한 이탈 신호
 - **Age, MonthlyIncome:**
 - 젊은 연령, 낮은 급여일수록 퇴사 확률 ↑
 - **YearsWithCurrManager, DistanceFromHome:**
 - 매니저와의 근속 기간이 짧거나 출퇴근 거리가 멀면 이탈 위험 ↑
- 📌 전반적으로 "과중한 근무 + 낮은 보상 + 짧은 관계 유지"가 퇴사 위험 요인임을 보여줌
-

3.2 혼동행렬 분석

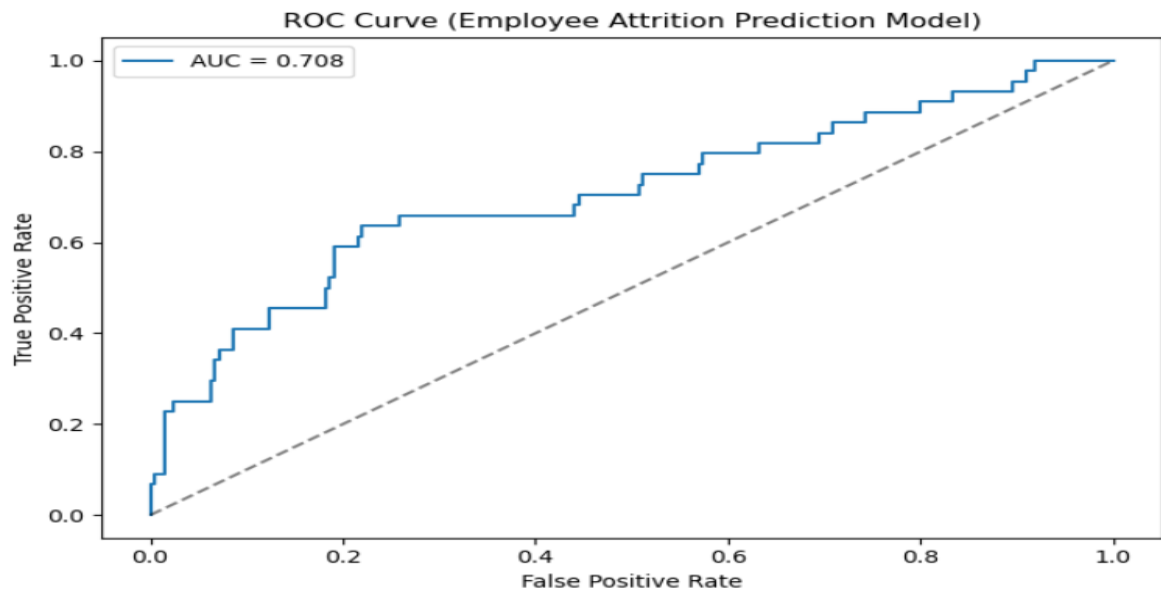


- True Negative (정확히 예측된 재직자): 209명
- True Positive (정확히 예측된 퇴사자): 3명
- False Negative (실제 퇴사자를 재직으로 오분류): 41명
- False Positive (재직자인데 퇴사자로 잘못 예측): 0명

→ 재직자 예측에는 강하지만, 퇴사자 탐지 민감도가 낮음

→ 향후 임계값 조정, 클래스 가중치 적용, 오버샘플링 필요

3.3 ROC 곡선 해석



- ROC 커브는 대각선 위에 위치하여 **무작위 예측보다 우수함**
 - **AUC (Area Under Curve): 0.708**로, 퇴사자와 재직자를 구분하는 중간 이상 수준의 성능을 나타냄
 - 이 수치는 인사 데이터에서 흔히 겪는 클래스 불균형을 감안할 때 의미 있는 결과
-

4. 주요 인사이트

1. 퇴사 가능성이 높은 직원의 공통 특징

- 초과근무 있음
- 나이가 어리고, 월급이 낮음
- 관리자와의 관계 기간 짧고, 출퇴근 거리가 길다

2. 모델의 강점

- 재직자 예측 정확도 우수 (False Positive = 0)
- 실제 데이터 기반의 신뢰도 있는 지표 확보

3. 보완할 점

- 퇴사자 탐지 민감도(Recall)가 낮음
 - 향후 이직 징후 탐지용 정성적 변수(예: 성과평가, 만족도 조사 등) 추가 필요
-

5. 활용 방안 제안

1. 고위험군 실시간 모니터링

- 월별로 퇴사 리스크가 높은 직원 식별 및 대시보드 시각화

2. 정책 실행

- 퇴사 예측 점수가 높은 직원에게 보상 재설계, 팀 이동 제안, 면담 진행 등 맞춤형 개입

3. 정책 효과 추적

- 개입 이후 실제 퇴사율 감소 여부 측정 → HR 전략 검증 수단으로 활용 가능

4. 모델 주기적 재학습

- 분기 단위로 모델 업데이트 및 성능 검토 필요
-

6. 결론

본 예측 모델은 퇴사자 탐지에 아직 개선 여지가 있으나, 전반적인 변수 영향력 분석 및 재직자 분류 정확도는 우수한 수준입니다.

단순 통계가 아닌 **데이터 기반 인사전략 수립**의 기초 도구로서 실무 적용 가치가 충분하며, 지속적인 보완을 통해 **사전적 이탈 예방 시스템**으로 발전시킬 수 있습니다.