

| 연구 단계 | 세부 프로세스 | 구체적인 기술 구현 및 내용 |
|----------------------|-----------------------------|---|
| 1. 데이터 수집 | Depresjon 데이터셋 확보 | 노르웨이의 'Depresjon' 데이터셋 사용. 액티그래피(Actigraphy) 시계를 통해 분 단위로 기록된 활동량 데이터와 인구통계학적 정보 포함. |
| 2. 데이터 전처리 | 클래스 불균형 해소 | 특정 클래스가 치우친 문제를 해결하기 위해 ADASYN 기법 적용. |
| | 데이터 정규화 | 데이터의 척도를 맞추기 위해 Z-score 정규화 기법을 사용하여 표준화 수행. |
| 3. 피처 엔지니어링 | 다차원 특징 추출 | 로우 데이터를 5개 범주(시간 도메인, 주파수 도메인, 수면 주기, 전이 특징, 인구통계)의 통계적 특징으로 변환. |
| | 시간/주파수 도메인 | 평균, 표준편차, 왜도, 첨도, 자기상관(Autocorrelation), 전력 스펙트럼 밀도(PSD) 평균 등을 추출. |
| 4. 모델 학습 및 분류 | 알고리즘 구현 | 5종의 모델 구현 : 로지스틱 회귀(LR), SVM, 랜덤 포레스트(RF), XGBoost , 신경망(NN). |
| | 분류 작업 수행 | 1. 이진 분류: 우울증 유무(Yes/No) 판별. 2. 다중 분류: MADRS 점수에 따른 심각도(정상, 경도, 중등도) 분류. |
| 5. 모델 평가 | 성능 지표 검증 | 정확도(Accuracy), 정밀도(Precision), 재현율(Recall), F1-score, MCC 등을 통해 모델 성능 평가. |
| | 최적 모델 선정 | XGBoost 가 이진 분류(84.94%) 및 다중 분류(85.91%)에서 가장 높은 정확도를 기록함. |
| 6. 설명 가능한 AI(XAI) | 결과 해석 및 투명성 | SHAP 과 LIME 기법을 적용하여 모델이 왜 해당 예측을 했는지, 어떤 특징(예: PSD 평균, 나이 등)이 중요하게 작용했는지 시각화 및 설명. |