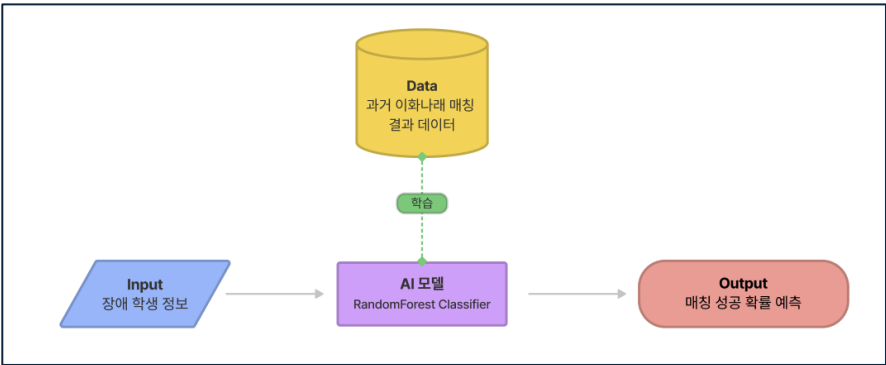


1. 목적
- 누적된 과거 이화나래 신청 장애학생의 정보와, 매칭 성공 결과에 대한 데이터를 바탕으로 AI 모델을 구축함. 학습시킨 모델에 장애 학생의 정보를 입력하면, 해당 학생이 도우미와 매칭될 확률을 예측함.



전체 모델 플로우차트

2. 모델 구축
- 2.1. 학습 데이터셋
 - 이화여대 장애학생지원센터에서 제공하는 기본 데이터셋을 활용함.

순번	장애 유형	학수 번호	분반	교과목명	교수명	시간	학점	요일/교시/교실	지원 종류	비고
----	-------	-------	----	------	-----	----	----	----------	-------	----

 - 모델 구축 과정에서 다양한 특징(feature)를 추가함에 따라 모델이 더 많은 유형의 패턴을 감지하고 그에 따라 성능이 향상될 것으로 기대되어, 아래와 같이 더 다양한 feature를 추가함.

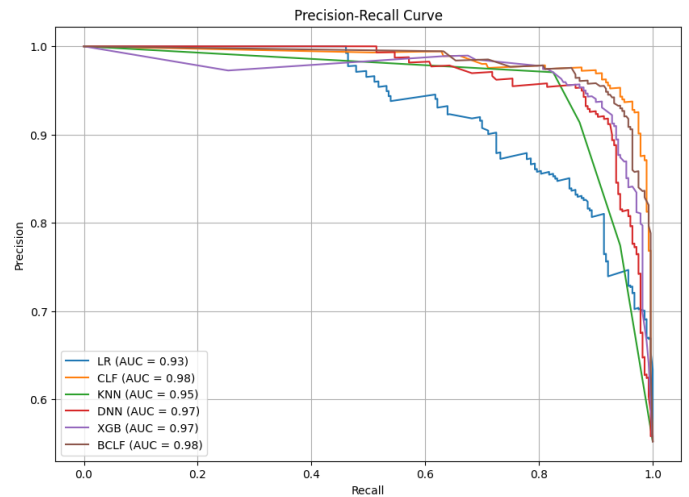
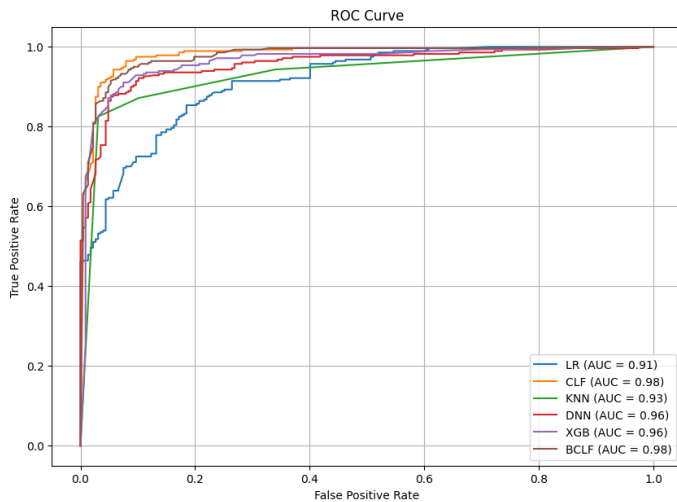
장애 유형	학수 번호	분반	교과목명	개설학과/전공	교과영역	교과목 구분	교수명	정원
학점	요일/교시	지원종류	원격강의	이동지원	교생실습지원	수업참여지원	알림사항지원	시간
- 2.2. 데이터 전처리
 - 사이킷런의 Ordinal Encoder로 범주형 데이터를 처리함.
- 2.3. 데이터 Augmentation
 - 장애학생지원센터에서 제공한 기본 데이터셋의 개수가 26개로 매우 한정적임. 효과적인 모델을 구축하기 위하여 CTGAN(Conditional Tabular Generative Adversarial Network)를 활용하여 데이터를 기존 26개에서 2026개로 증강시키는 작업을 수행함.
 - CTGAN: 단일 테이블 데이터에 대한 딥러닝 기반 GAN으로, 실제 데이터에서 학습하며 데이터를 생성함.

2.4. 모델 성능 비교 및 선택

- 머신러닝 모델 5개, 딥러닝 모델 1개 총 6개의 모델을 학습시켜 성능을 비교 분석한 후 가장 효과적인 모델을 선택함.

- 1) Logistic Regression
- 2) RandomForest Classifier
- 3) kNN
- 4) XGBoost
- 5) Bagging Classifier
- 6) DNN

- **Accuracy: RandomForest Classifier(0.9408) > Bagging Classifier(0.9289) > XGBoost(0.9151) > DNN(0.9053) > kNN(0.8836) > Logistic Regression(0.8303)**
- **ROC Curve, PR Curve:** 아래 그래프와 같이 나타남.



- 종합적인 성능을 비교 분석하였을 때, 6개의 모델 중 **RandomForest Classifier**가 가장 우수한 성능을 보임.

3. 모델 활용

- 4명의 가상 데이터를 활용하여 모델의 작동 방식을 확인함. 엑셀에 기록된 장애 학생 정보를 AI 모델에 입력하면, 아래와 같이 **해당 학생이 매칭될 확률과, 확률이 가장 낮은 순으로 학생들을 랭킹함.**

[매칭 성공 확률이 낮은 순으로 정렬]

****매칭 성공 확률: 0.39207229437229424 %**

학생 Index: 5

학생 정보:

장애유형	뇌병변
학수번호	20007
분반	1
교과목명	교육심리
개설학과/전공	교육학과
교과영역	전기
교과목 구분	전기
교수명	이승은
정원	30
시간	2
학점	2
요일/교시	수 4-5
교실	학
지원종류	노트필기
원격강의	X
이동지원	0
교생실습지원	X
수업참여지원	X
알림사항지원	0

Name: 5, dtype: object

****매칭 성공 확률: 0.4655 %**

학생 Index: 0

학생 정보:

장애유형	시각
학수번호	38640
분반	2
교과목명	북한과인권문제
개설학과/전공	북한학
교과영역	전선
교과목 구분	전선
교수명	박민주
정원	100
시간	3
학점	3
요일/교시	화 4~4nn금 5~5
교실	포
지원종류	노트북속타
원격강의	X
이동지원	0
교생실습지원	X
수업참여지원	X
알림사항지원	X

Name: 0, dtype: object

****매칭 성공 확률: 0.6686333333333333 %**

학생 Index: 2

학생 정보:

장애유형	시각
학수번호	36500
분반	1
교과목명	디지털노리설계
개설학과/전공	컴퓨터공학
교과영역	전선
교과목 구분	전선
교수명	박상수
정원	50
시간	3
학점	3
요일/교시	월 3~3nn수 2~2
교실	공A
지원종류	노트필기
원격강의	X
이동지원	X
교생실습지원	X
수업참여지원	0
알림사항지원	X

Name: 2, dtype: object

****매칭 성공 확률: 0.728263492063492 %**

학생 Index: 4

학생 정보:

장애유형	청각
학수번호	34947
분반	2
교과목명	교육철학및교육사
개설학과/전공	교육학과
교과영역	전기
교과목 구분	전기
교수명	이은주
정원	30
시간	2
학점	2
요일/교시	화 2~3
교실	교
지원종류	기타
원격강의	X
이동지원	X
교생실습지원	X
수업참여지원	0
알림사항지원	X

Name: 4, dtype: object