

기업맞춤형 AI-X융복합 인재 양성 교육

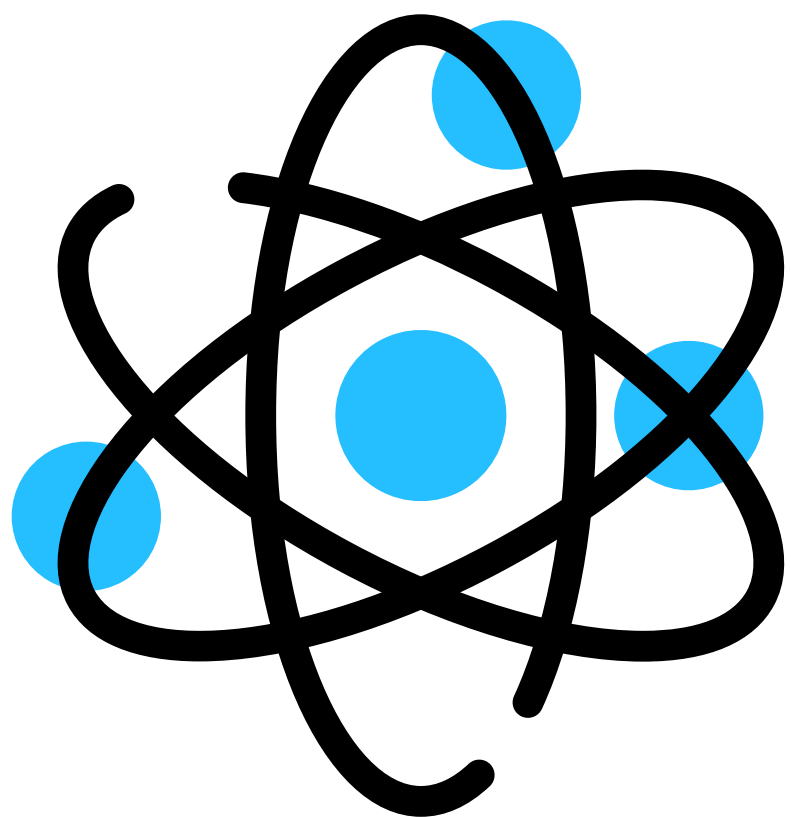
장마철 침수 위험 예측 시스템

팀명 : SOOM(숨)

정서희 고정현 김영신 장슬찬

목차

PART 01	프로젝트 개요 및 배경
PART 02	데이터 수집
PART 03	다중 AI 모델 개발 (시간별/일별/통합)
PART 04	모델 성능 평가
PART 05	웹 애플리케이션 시스템 구현
PART 06	주요 기능 시연 및 사용자 시나리오
PART 07	기대효과 및 확장 계획



프로젝트 개요 및 배경

주제 선정 및 배경

- 최근 수년간 장마철 집중호우와 극한 기상 현상 증가
- 도심 침수로 인해 기업 운영 마비, 자산 손실, 고객 서비스 중단 발생
- 기상 정보는 제공되지만, 기업에 특화된 "사전 대응 시스템"은 부재

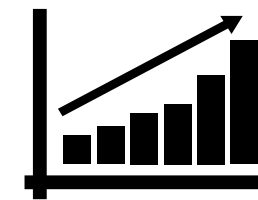
목적 및 필요성



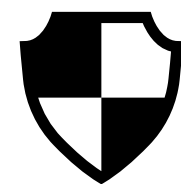
기업 운영 연속성 보장



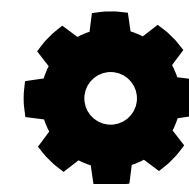
경제적 손실 예방



경쟁우위 확보 및 신시장 창출



금융 리스크 관리 고도화



디지털 전환 및 스마트화

업무분장



정서희

- 모델링 및 데이터 수집
- 데이터 전처리
- 일정관리
- 웹 구현



고정현

- 모델링 및 데이터 수집
- 데이터 전처리
- 시각화
- 웹 구현



김영신

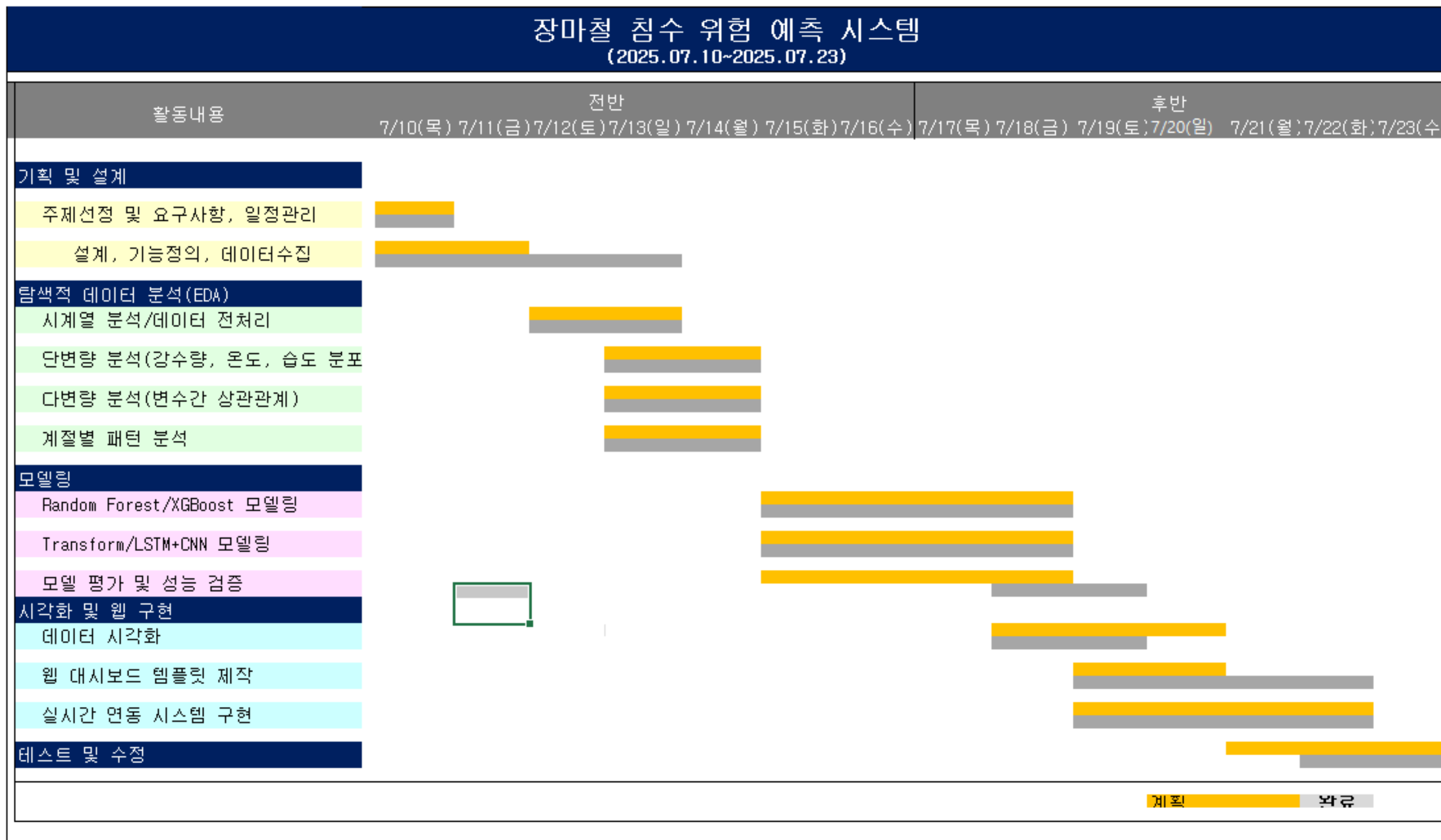
- 모델링 및 데이터 수집
- 데이터 전처리
- 웹 구현 구성
- 웹 구현



장슬찬

- 모델링 및 데이터 수집
- 데이터 전처리
- 시각화
- 웹 구현

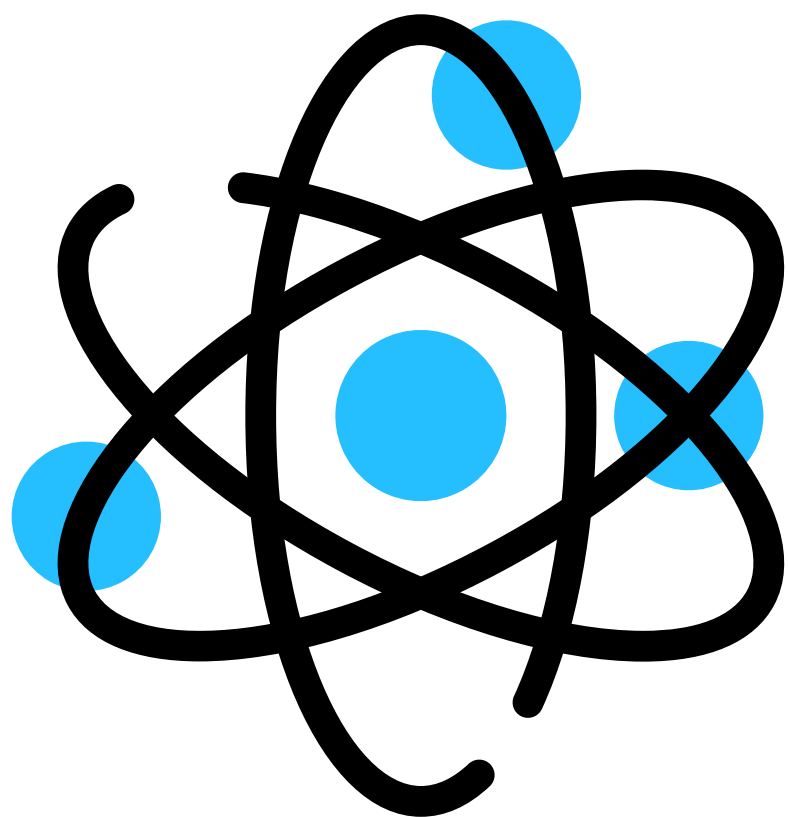
Gantt Chart를 이용한 일정관리



개발환경

(Resource)

OS	Windows 10	Open Source Numpy 1.21.6 Pandas 1.5.3 Matplotlib 3.6.3 Seaborn 0.12.2 Scikit-learn 1.2.1 Scipy 1.10.0 Requests 2.28.2 Openpyxl 3.1.1 Xlrd 2.0.1
Language	Python 3.10.13	
IDE	Vscode Anaconda jupyter notebook	Framework Flask 2.2.3



데이터 수집

참조 데이터

- 01 기상청 ASOS(지상) 시간별 데이터
- 02 기상청 ASOS(지상) 일별 데이터
- 03 물정보포탈 실제 침수 사례 데이터
- 04 네이버 날씨 웹크롤링 데이터
- 05 네이버 뉴스 웹크롤링 데이터

기상청 ASOS 데이터 (2000-2025)

데이터 규모

- 총 데이터 포인트: 약 227,000개
- 수집 기간: 25년간 (2000-2025년)
- 지역: 서울시 기상관측소 (ASOS)
- 데이터 타입: 시간별 218,015개 + 일별 9,250개

데이터셋 구성

- 시간별 기상 데이터: 실시간 예측용 고해상도 데이터
- 일별 기상 데이터: 중기 예보 및 트렌드 분석용
- 통합 데이터: 시간적 특성과 침수 위험도가 추가된 최종 분석용 데이터

기상청 ASOS 데이터 (2000-2025)

데이터 특징

- 다중 시간 스케일: 시간별/일별 분석으로 다양한 예측 환경 지원
- 계절성 반영: 25년간 장기 데이터로 계절별 패턴 학습
- 실무 적용성: 실제 기상청 관측 데이터 기반의 신뢰성 높은 데이터

핵심 변수

- 기상 변수 (9개): 기온, 강수량, 풍속, 풍향, 습도 등
- 시간 변수 (6개): 연도, 월, 일, 요일, 평일/주말 구분
- 강우 패턴 (4개): 강우 여부, 강우 시간, 최대 강수량 등
- 타겟 변수: 침수 위험도 (0: 안전, 1: 위험)

실제 침수 사건 데이터 -물정보포털

[https://www.water.or.kr/kor/menu/
sub.do?menuId=16_166_168_268](https://www.water.or.kr/kor/menu/sub.do?menuId=16_166_168_268)

*서울 실제 침수 사례

2000.8.23~9.1

2002.8.30~9.1

2005.8.2~8.11

2006.7.9~7.29

2007.9.13

2011.7.26~7.29

2013.7.11~7.15, 7.18

2018.8.23~8.24

2018.8.26~9.1

2019.9.28~10.3

2020.7.28~8.11

2020.8.28~9.3

2020.9.1~9.7

2022.8.8~8.17

2022.8.28~9.6

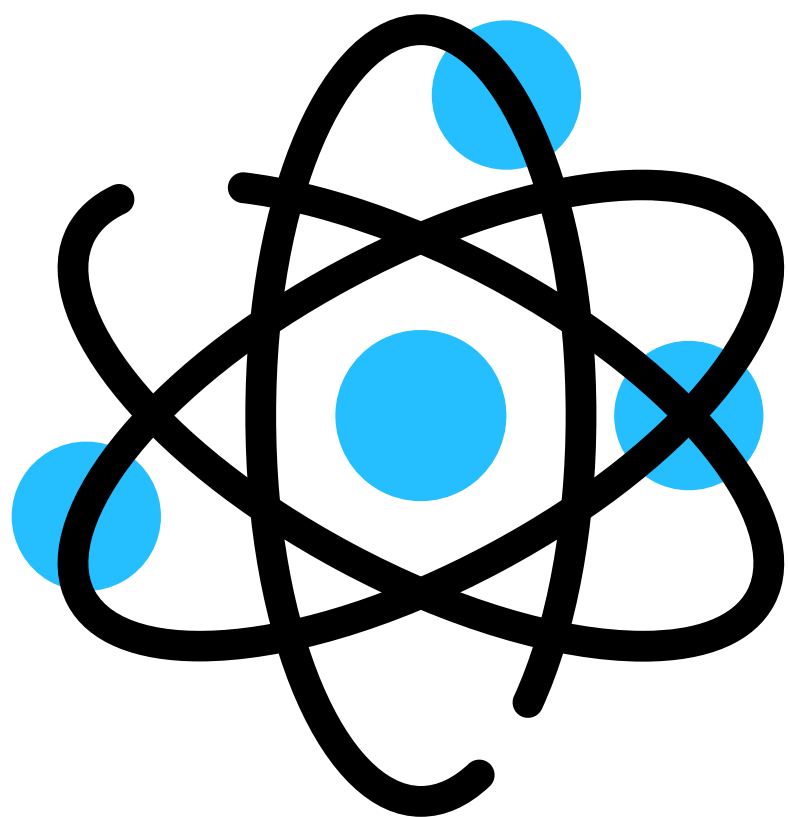
실시간 수집 데이터

네이버 뉴스 API 활용

- 서울 날씨 관련 뉴스 실시간 수집
- 날씨 키워드 필터링 (날씨, 기온, 비, 폭염 등)
- 중복 제거 및 자동 업데이트
- Excel 파일로 체계적 관리

네이버 오늘날씨 크롤링

- 서울 지역 실시간 기상 정보
- 온도, 날씨상태, 미세먼지 등
- 매일 자동 갱신 시스템
- 웹 대시보드 연동 준비



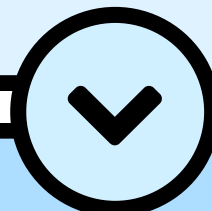
다중 SI 모델 개발 (시간별/일별/통합)

데이터 파이프라인

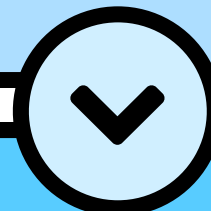
01 수집: 기상청 API + 뉴스 크롤링 (1시간마다)



02 전처리: 결측치 처리, 특성 추출, 지역별 가중치 적용



03 예측: 4모델 앙상블 → 지역별 위험도 계산



04 시각화: 웹 대시보드 실시간 업데이트

전처리 파이프라인

01 결측치 처리

중앙값 대체,
0값 처리

02 특성 추출

시간, 계절, 강수
패턴 변수 생성

03 라벨링

실제 침수 사건 +
강수량 10mm 기준

04 정규화

모델별 최적
스케일링 적용

4가지 ML/DL 모델

Random Forest

설계 목적: 안정적 기본 예측

핵심 특징: 300개 트리 앙상블, 특성 중요도 해석 가능

주요 역할: 전체 시스템의 신뢰성 담보

최적화: 클래스 불균형 자동 처리

LSTM+CNN

설계 목적: 시계열 패턴 전문 분석

핵심 특징: 7일 시퀀스 윈도우, 시간 의존성 학습

주요 역할: 기상 변화 추세 예측

최적화: CNN으로 지역 패턴, LSTM으로 시간 패턴

XGBoost

설계 목적: 최고 정확도 추구

핵심 특징: 그래디언트 부스팅, 정규화 적용

주요 역할: 정밀한 침수 위험도 계산

최적화: 불균형 데이터 가중치 10배 적용

Transformer

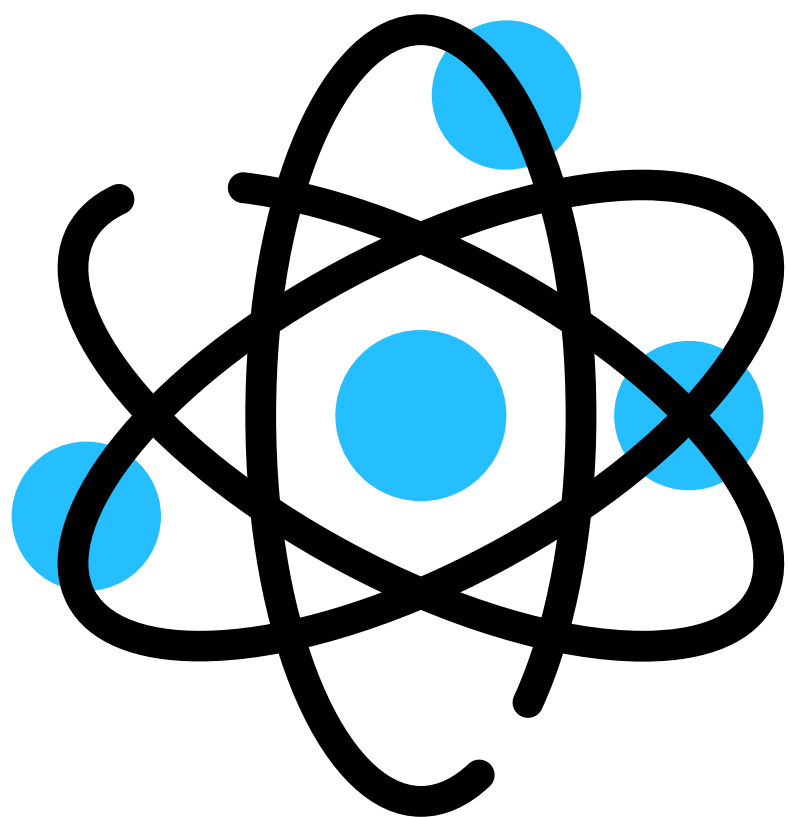
설계 목적: 높은 재현율로 침수 예측 실패 방지

핵심 특징: 어텐션 메커니즘, SMOTE 데이터 증강

주요 역할: 침수 감지 민감도 극대화

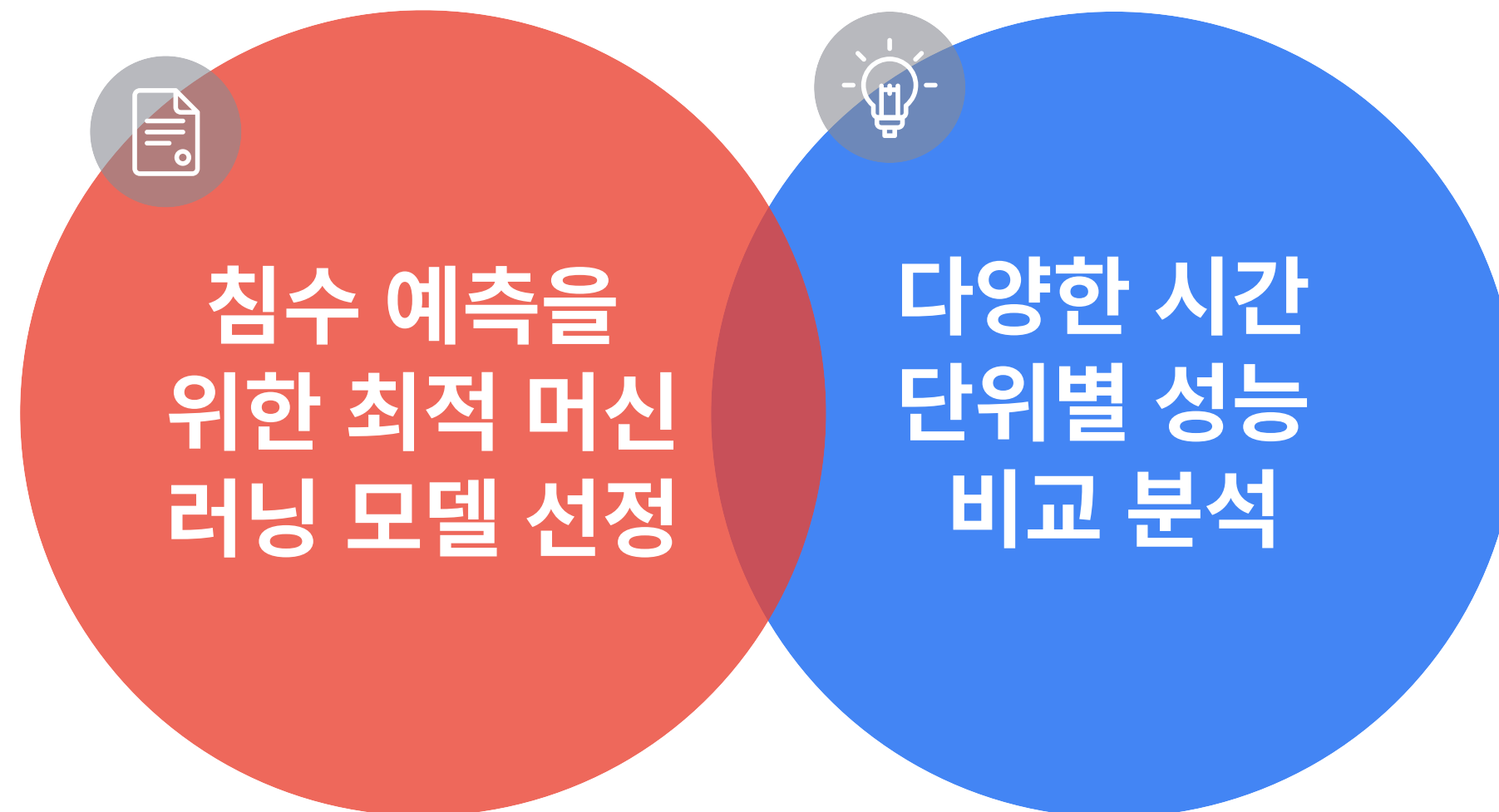
최적화: 91.0% 재현율로 false negative 최소화

4가지 모델



모델 성능 평가

평가 목적



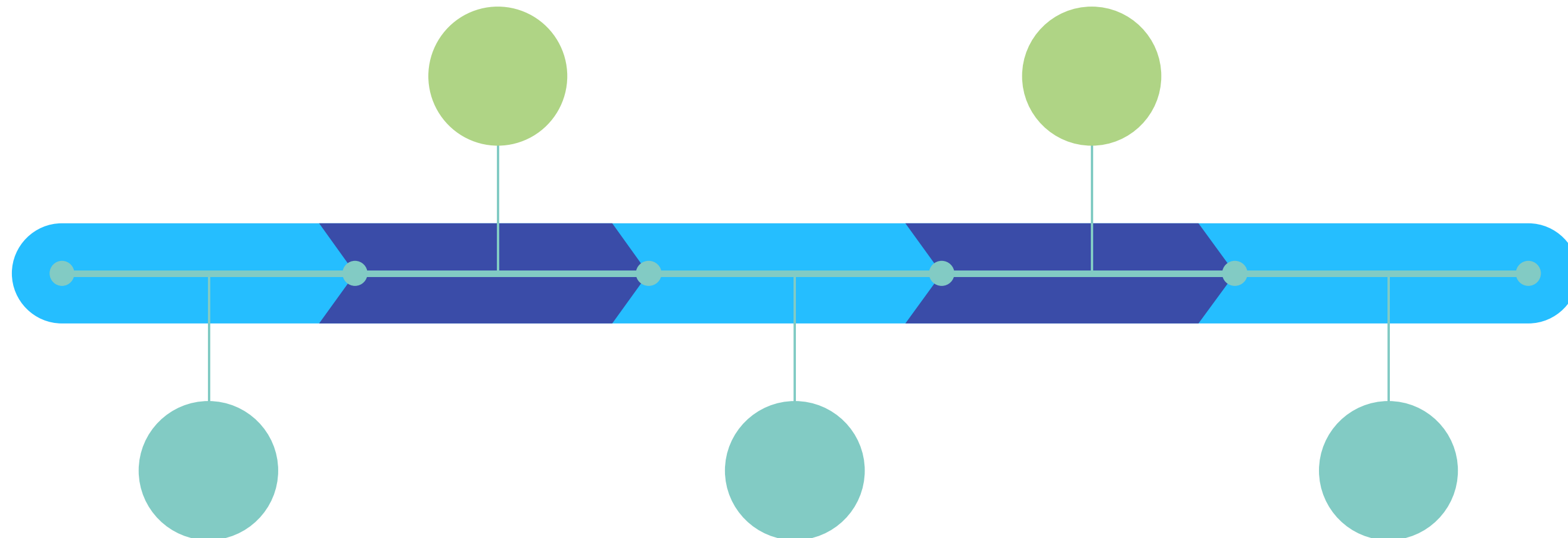
평가 지표

Accuracy

전체 정확도

Recall

실제 침수 탐지율
(재현율)



Precision

홍수 예측 정밀도

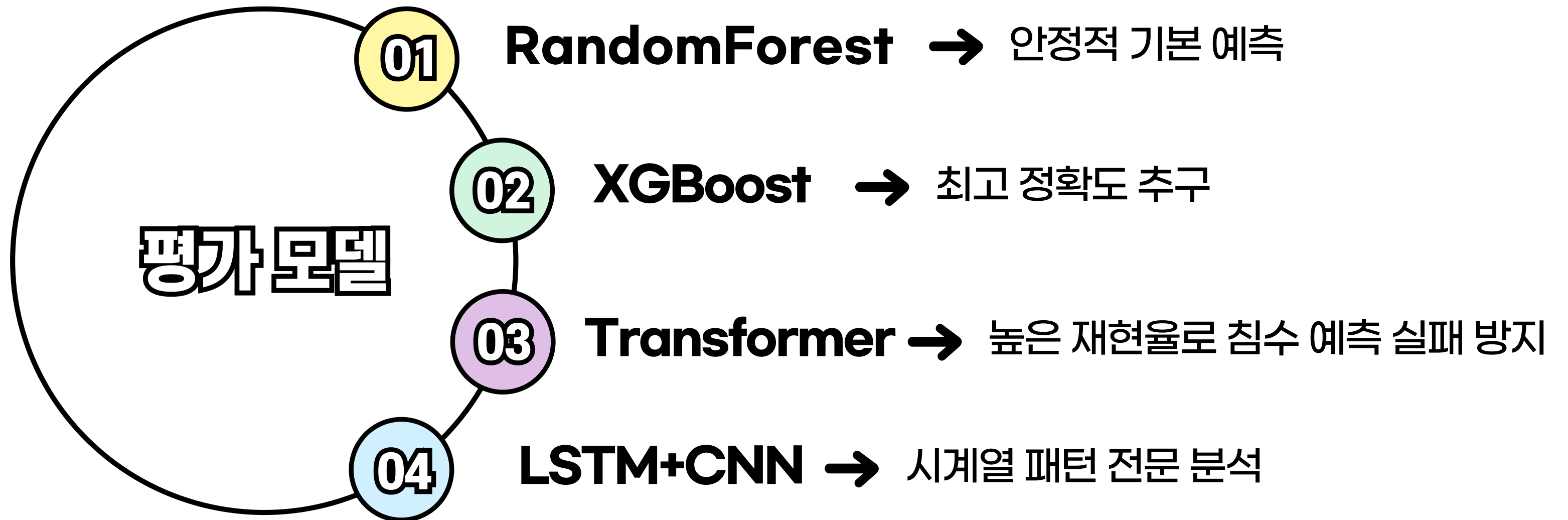
F1-Score

정밀도와 재현율의 조화평균

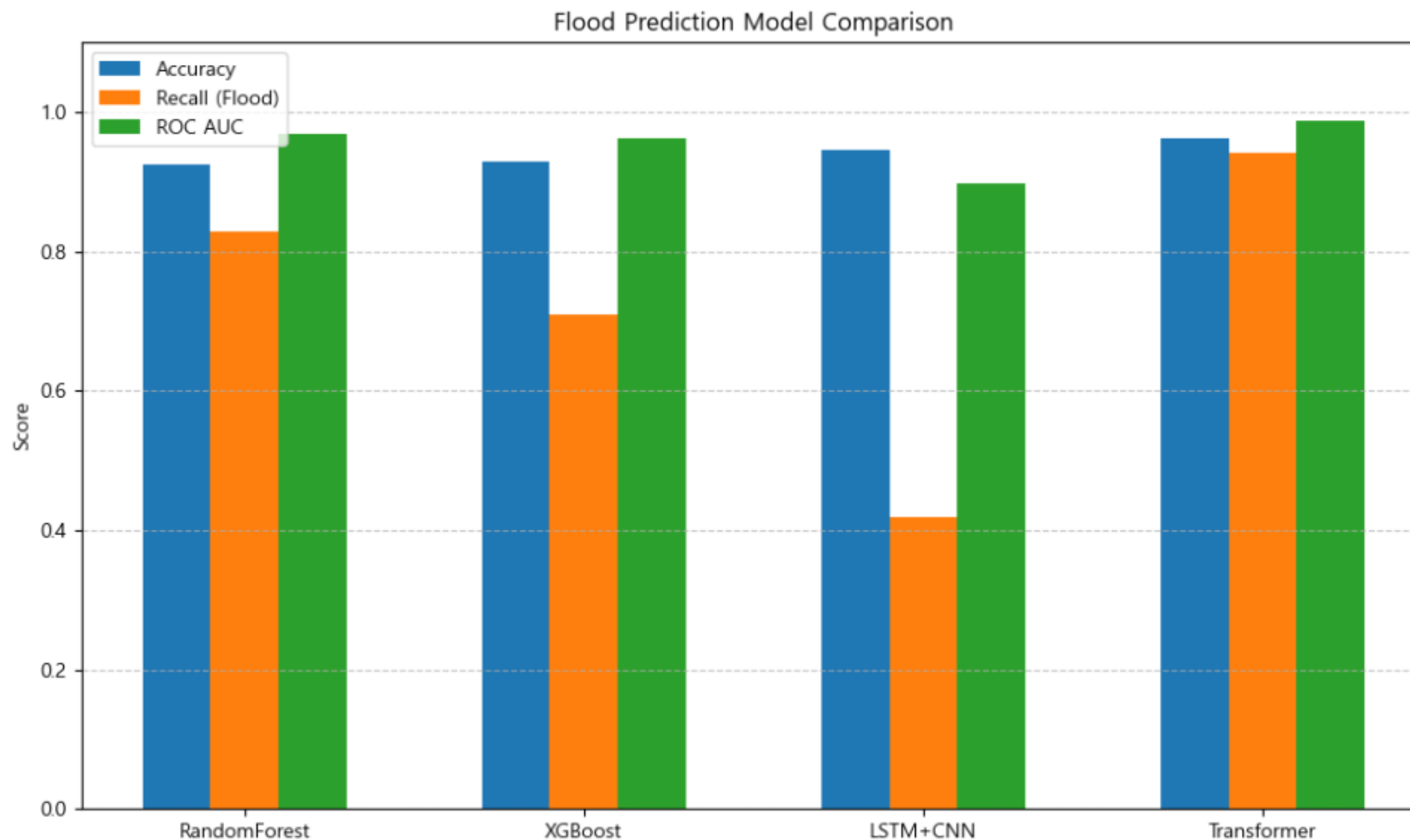
ROC-AUC

분류 성능 종합 평가

평가 모델

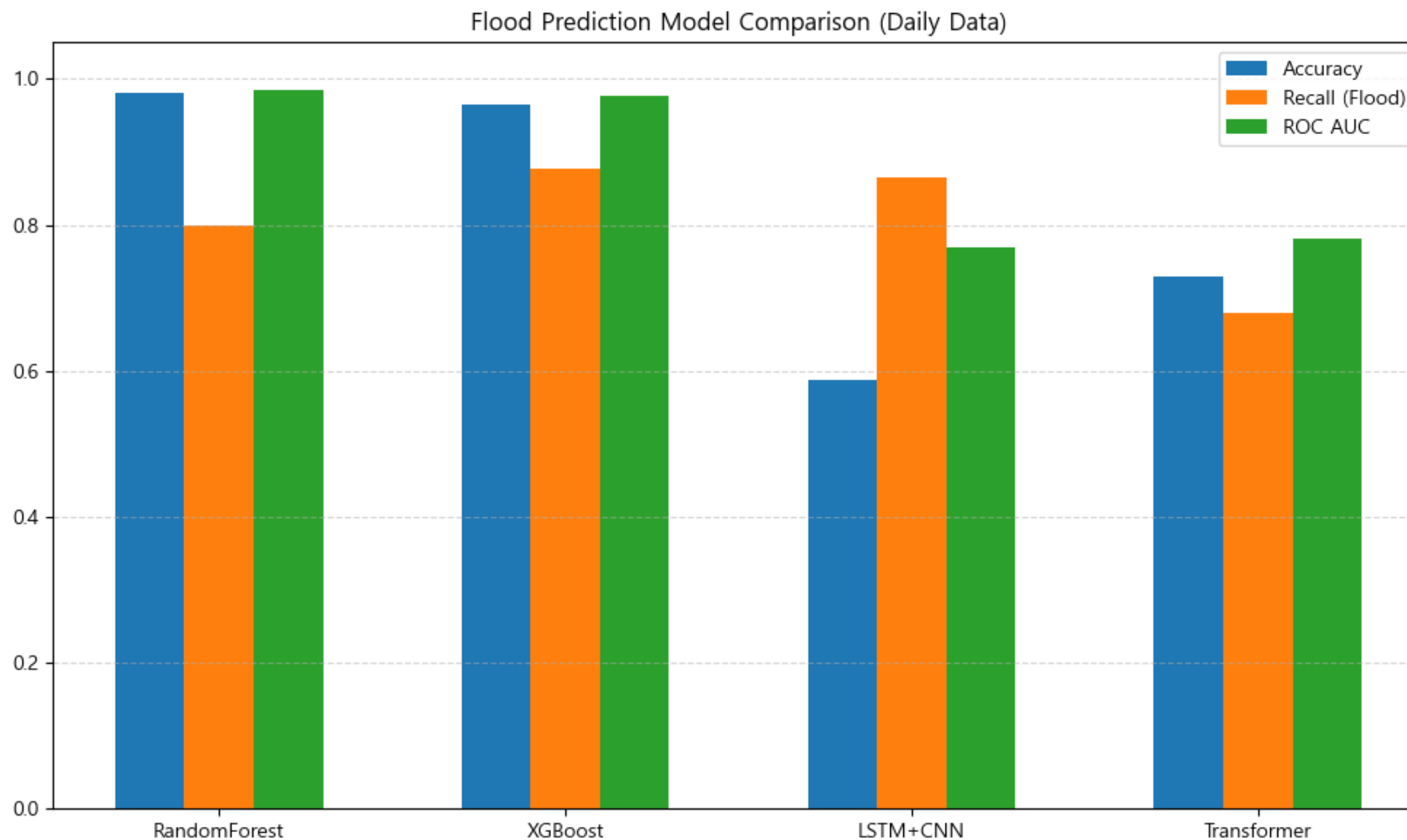


시간별 데이터



Transformer: 압도적 재현율(94.1%)과 ROC AUC(98.7%)로 침수 농침 최소화
RandomForest: 정확도와 재현율의 좋은 균형, 실시간 대응에 최적
XGBoost: 높은 정확도와 안정적 성능
LSTM+CNN: 높은 정확도지만 침수 탐지 성능 낮음

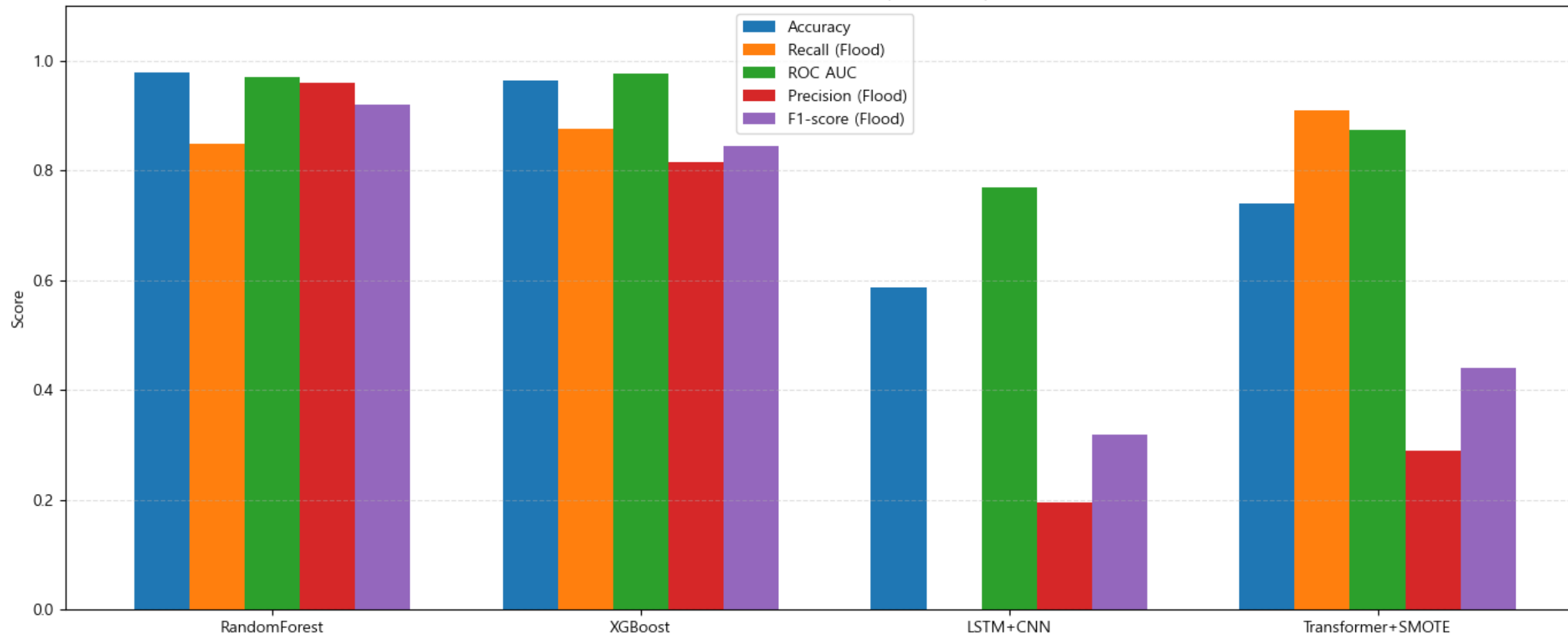
일별 데이터



RandomForest: 전체적으로 가장 안정적, 모든 지표에서 우수한 성능
XGBoost: 침수 예측(재현율 87.7%)에서 가장 뛰어남, 성능 균형 우수
LSTM+CNN: 정확도는 낮지만 재현율 높아 침수 감지에 민감
Transformer: 재현율은 양호하지만 정밀도가 낮아 오탐 발생 가능

통합 데이터

침수 예측 모델 성능 비교 (5가지 지표)



RandomForest: 압도적 정확도 (98.0%)와 정밀도(96.0%)로 신뢰성 최고, 전체 시스템 안정성 담보

XGBoost: 최고 ROC AUC(97.7%)와 우수한 재현율(87.6%)로 가장 균형 잡힌 성능, 실무 활용도 최적

LSTM+CNN: 심각한 재현율 문제 (0.0%)로 침수를 전혀 감지 못함, 시계열 학습 실패로 개선 시급

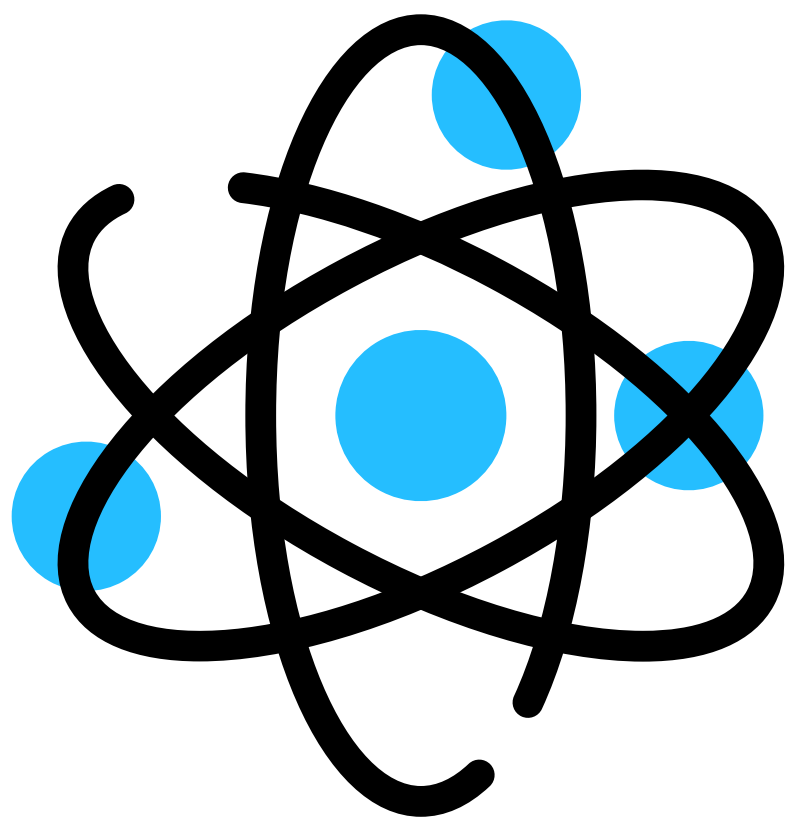
Transformer: 최고 재현율(91.0%)로 침수 놓침 최소화, SMOTE 효과로 민감도 극대화하지만 정밀도 희생

통합 데이터 지표

모델	Accuracy	Recall	ROC AUC	Precision	F1-score
RandomForest	0.980	0.850	0.970	0.960	0.920
XGBoost	0.964	0.876	0.977	0.816	0.845
LSTM+CNN	0.587	0.000	0.769	0.196	0.320
Transformer+SMOTE	0.740	0.910	0.874	0.290	0.440

데이터 타입별 성능 비교





웹 애플리케이션 시스템 구현

웹 페이지 구성(Flask)

메인 대시보드

뉴스 모니터링 페이지

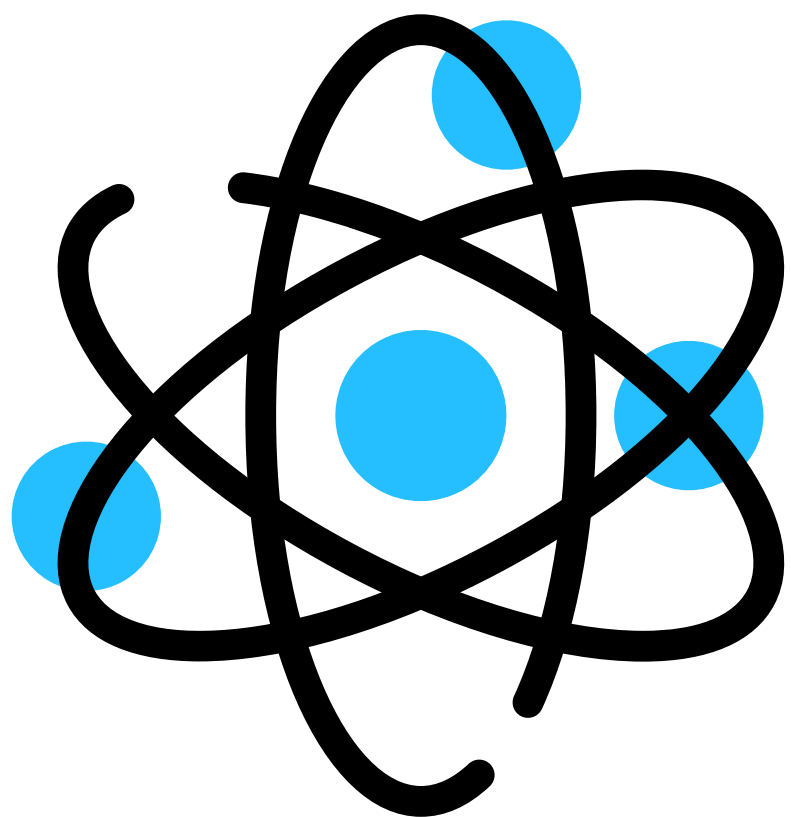
실시간 지도 페이지

사용자 모델 선택 페이지

모델 비교 페이지

사용자 인증

<https://rainflood.onrender.com>



주요 기능 시연 및 사용자 시나리오

사용자 시나리오

지방자치단체 & 공공기관

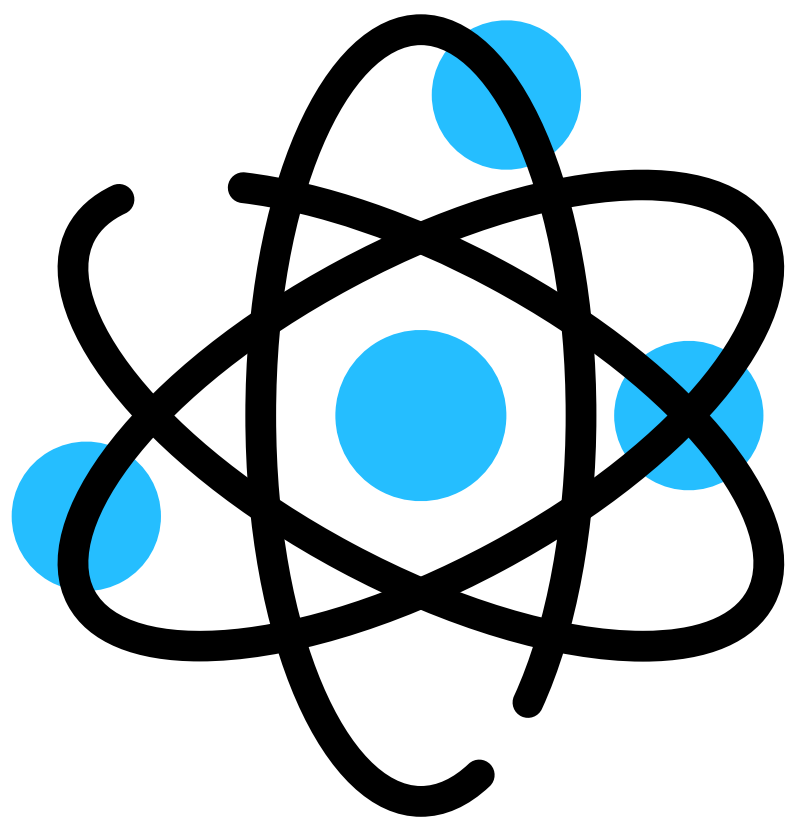
"강남구청 재난안전과에서 오전 9시 집중호우 경보 발령
→ 시스템에서 강남구 침수 위험도 85점 (매우높음) 알림
→ 즉시 응급대응팀 출동, 저지대 차량 통제 조치 실행"

건설사&부동산 개발업체

"현대건설 신규 아파트 단지 개발 계획 수립
→ 마포구 예정지 침수 위험도 95점 (최고 위험) 확인
→ 배수 시설 강화 설계 + 1층 상가 배치 변경으로 리스크 완화"

물류/유통회사

"쿠팡 로지스틱스 센터에서 강남/서초 배송 계획 수립
→ 오후 집중호우 예보 시 침수 위험 경로 자동 감지
→ 안전 경로로 자동 재배치, 배송 지연 최소화"



기대효과 및 확장 계획

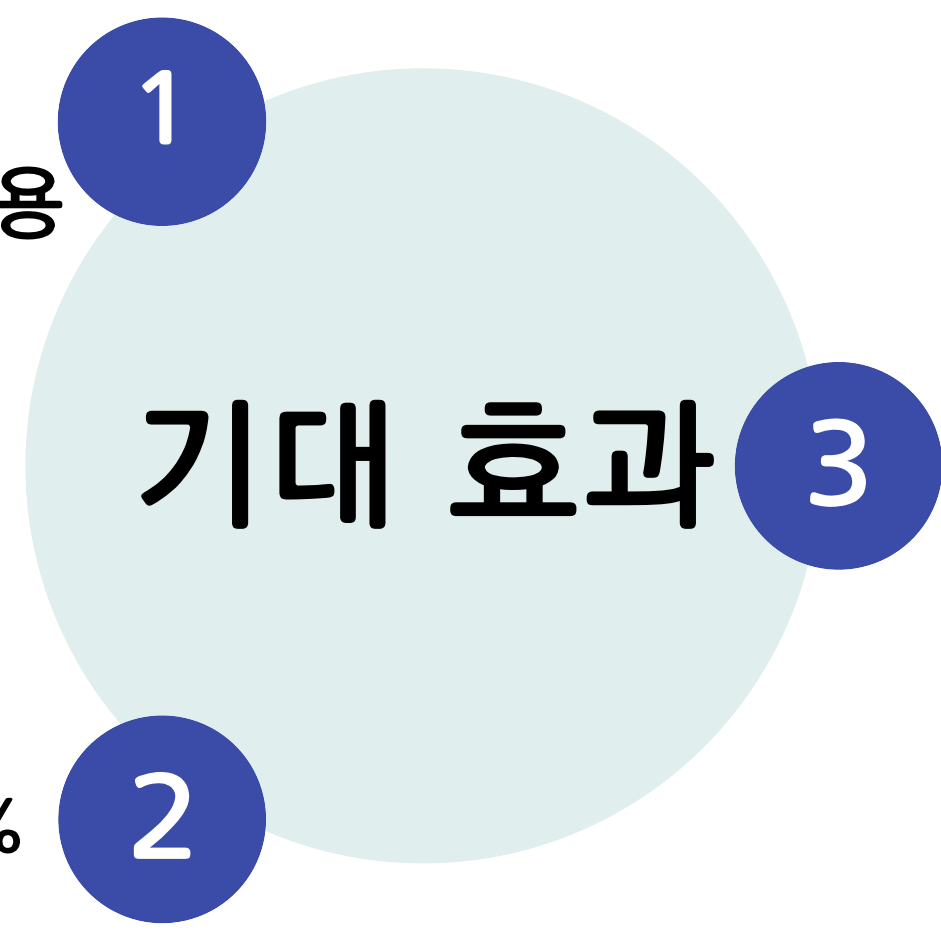
경제적&사회적 기대 효과

경제적 효과: 건설업계

사전 리스크 분석으로 설계 변경 비용
30% 절감

경제적 효과: 물류업계

배송 중단 방지로 운영 효율성 25%



사회적 가치 창출

공공 안전: 시민 생명 보호, 사회
안전망 강화
도시 회복력: 기후변화 대응,
스마트시티 구현
데이터 기반 의사결정: 경험 의존
→ 과학적 근거 기반 정책 수립

기술 확장 로드맵

AI 모델 고도화

현재: 4개 전통 ML/DL 모델
→ 확장: 생성형 AI +
강화학습 + 멀티모달 AI

서비스 확장

현재: 침수 예측
→ 확장: 산사태, 지진, 태풍
등 종합 재난 예측 플랫폼



Thank You