

2024년 K-water 대국민 물 빅데이터 공모전 아이디어 기획서

제 목	SPI6를 활용한 미래를 예측하는 가뭄 경보 시스템: 데이터 통합과 머신러닝의 결합		
성 명	팀 장	서동현	000-0000-0000
			000000@gamil.com
	팀 원	문혜선	000-0000-0000
		김준희	000-0000-0000
		노승욱	000-0000-0000
		유동현	000-0000-0000

I. 과제 목표

1. 추진 배경

- a. 급격한 기후 변화와 도시화 등 여러 요인으로 인해 세계인구의 40%가량은 만성적인 물 부족 상황이 발생하고 있다.
- b. 현재 한국의 경우 사용 가능한 수자원 대비 물 수요 비율이 높은 국가인 물 스트레스 국가이다.
- c. 한국은 계절별 강수량의 편차가 매우 크며, 국토 면적이 좁고 인구밀도가 높아 물 부족 현상이 심각해질 우려가 있다.
- d. 이를 선제 대응하기 위해 일별 강수량, 계절 그리고 댐의 저수량 데이터를 활용한 가뭄 예측이 필요하다.

2. 필요성

- a. 조기 경고 시스템 : spi6 가뭄지수를 예측함으로써 가뭄 발생을 조기에 예측하고, 선제 대응할 필요가 있다.
- b. 자원 관리 최적화 : 제한된 수자원의 효율적인 배분을 위하여 댐의 저수량 관리와 운영에 중요한 정보를 제공, 정수장과 수력발전소등의 운영을 최적화하고, 물 공급의 안정성을 보장한다.
- c. 기후 변화 대응 : 최근 급격한 기후 변화에 따라 가뭄의 발생 빈도와 강도가 증가할 수 있다. 따라서 가뭄 예측은 기후 변화에 대응하는 전략 수립에 필수적이다.

3. 목적

- a. 효율적인 물 자원 관리
다양한 데이터 분석 기법을 활용하여 가뭄의 기상학적 요소와

수문학적 요소를 체계적으로 분석함으로써 물 자원의 효율적인 분배를 도모한다.

b. 정확한 예측과 계획 수립

일별 누적강수량, 계절별, 댐별 데이터를 기반으로 가뭄 패턴을 분석하여 정확한 예측과 계획 수립을 통해 물 자원 관리 문제를 예방한다.

c. 가뭄 대응력 강화

예측된 가뭄 지수를 바탕으로 대한민국 주요 댐 부근 지역 가뭄 상황에 대한 선제 대응 능력을 강화한다.

II. 활용 데이터

기상청: 기상자료개방포털-종관기상관측(ASOS)(2019년 ~ 2022년)

기상청: 기상자료개방포털-표준강수지수 자료(2019년 ~ 2022년)

대한민국 모든 종관기상 관측 지점에서 측정한 일별 spi6 데이터를 추출한다.

대한민국 모든 종관기상관측 지점에 대한 일별 기온, 강수, 습도, 일조, 일사, 기압에 대한 데이터를 추출한다.

My Water: 기간별 자료(2018년~2022년) - 댐/보 기간별자료

다목적 댐과 용수 댐의 일별 강수량, 방류량, 저수량, 총 저수량, 유역, 유효 저수량 및 댐의 제원에 대한 데이터를 추출한다.

III. 주요 내용

1. SQL(Oracle)과 R을 활용

기후 상황과 가뭄에 대한 효율적인 물자원 분배를 위해 대한민국의 강수 특성을 반영한 주요 댐 부근 지역 SPI6 예측 모델을 개발하고 사용자 편의를 고려한 R shiny 활용 UI를 구현하여 홈페이지를 제공한다.

2. 데이터 전처리

a. 주요 다목적 댐 및 용수 댐

- My water 물 정보 포털에서 제공하는 주요 다목적 댐 및 용수 댐

의 제원과 일별 강수량, 방류량, 저수량 데이터셋을 이용한다.

- SQL을 활용하여 연도별로 구분 되어 있는 데이터셋을 통합한다.

b. 일별 기상 및 표준강수지수(SPI6) 데이터

- 기상자료개방포털에서 제공하는 종관기상관측지점에서 측정한 기상 데이터와 이를 활용하여 기상청에서 제공하는 일별 표준강수지수(SPI6) 데이터를 이용한다.
- SPI 지수를 예측을 위한 데이터로 예상된 일별 기온, 강수, 습도, 일조, 일사, 기압, SPI6 데이터를 SQL을 활용하여 분석하기 용이한 형태로 가공한다.
- 표준강수지수(SPI6)를 계산하는데 있어서 직접적인 요소라고 할 수 있는 누적강수량에 대한 데이터를 구하기 어려워 기존 일별 강수량 데이터를 따로 분리하여 6개월 단위로 누적강수량을 계산하도록 코드를 작성한다.
- 이를 토대로 표준강수지수(SPI) 예측 모델이 학습할 기상학적 요소가 담긴 훈련데이터를 생성한다.

c. SQL 활용 방안

- 앞서 가공한 기상학적 요소와 수문학적 요소와의 결합을 위한 관측소 지점명을 지역명으로 통일한다.
- 학습 데이터로써 활용하기위해 기상학적 요소와 수문학적 요소가 담긴 데이터를 통합한다.

3. 데이터 분석

a. 1차 분석

- 데이터셋 구성 : 기상 데이터와 주요 댐 데이터를 통합한다.
- 표준강수지수(SPI6) 예측을 위한 변수 특정을 위해 가뭄에 영향을 미치는 주요 요소들인 기상학적 요소와 수문학적 요소의 모든 변수들을 적용한다.
- 모델의 회귀분석 결과로 P-value와 상관관계를 확인하여 불용 변수를 제거한다.

b. 2차 분석

- 1차의 결과로 특정된 변수들을 다시 모델에 적용시켜 P-value와

상관관계를 확인한다.

- 예측에 필요한 최소한의 변수들을 특정한다.

c. 3차 분석

- 랜덤포레스트 모델의 적용을 위해 spi6의 수치를 기상청에서 지정한 수치 범위별 등급으로 분류한다.
- 해당 변수들을 R 프로그램을 활용한 분류 모델에 적용한다.

d. 모델 훈련

- 모델 설정

초기 데이터의 상관관계 분석을 위해 SQL 회귀분석 모델 사용한다.

SQL에서는 훈련데이터의 타겟 비율을 동일하게 적용할 수 없어 R을 통한 학습을 진행한다.

e. 랜덤포레스트 모델

- 모델 훈련
- 랜덤포레스트 분류 모델에 가공된 데이터셋을 적용한다.

	학습 데이터 (n)	검증 데이터 (n)
데이터 수량	42,392	4,706
비율	90%	10%

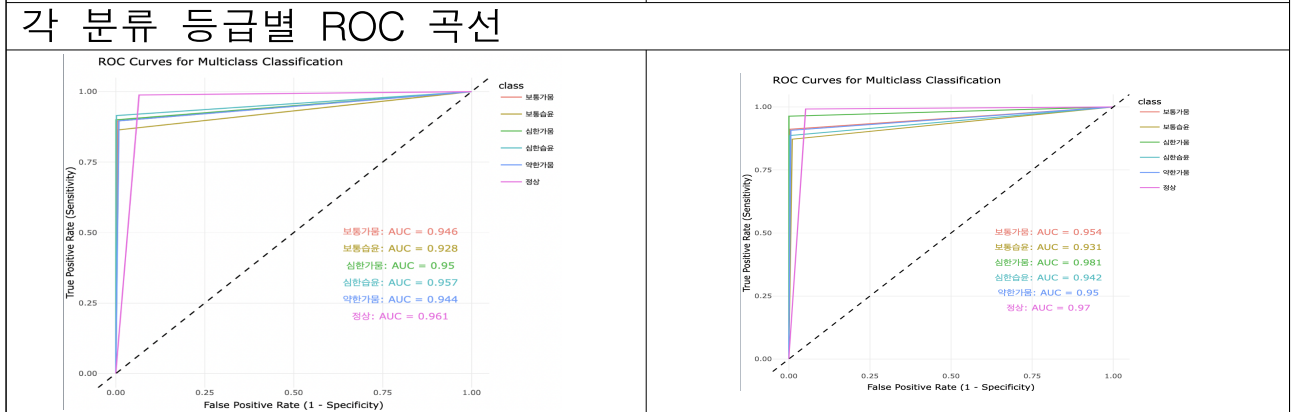
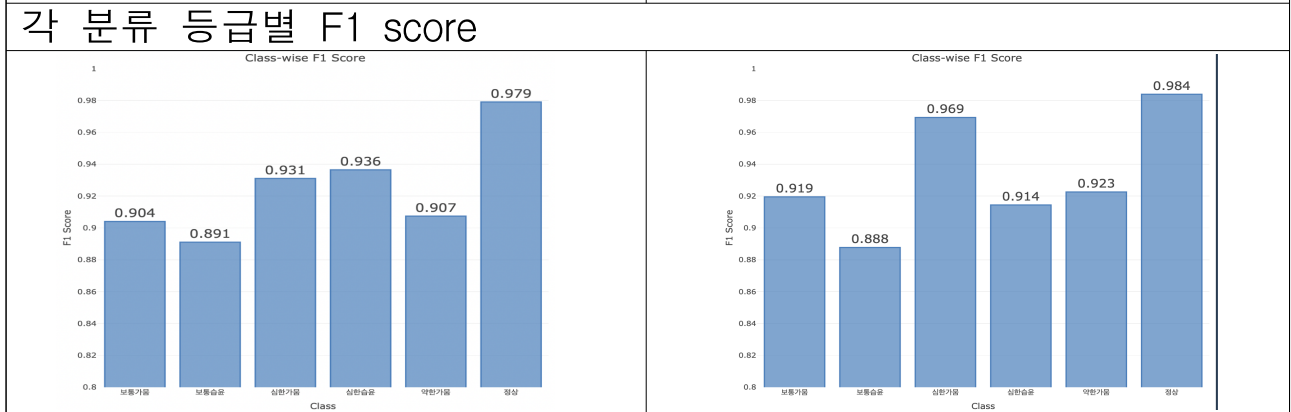
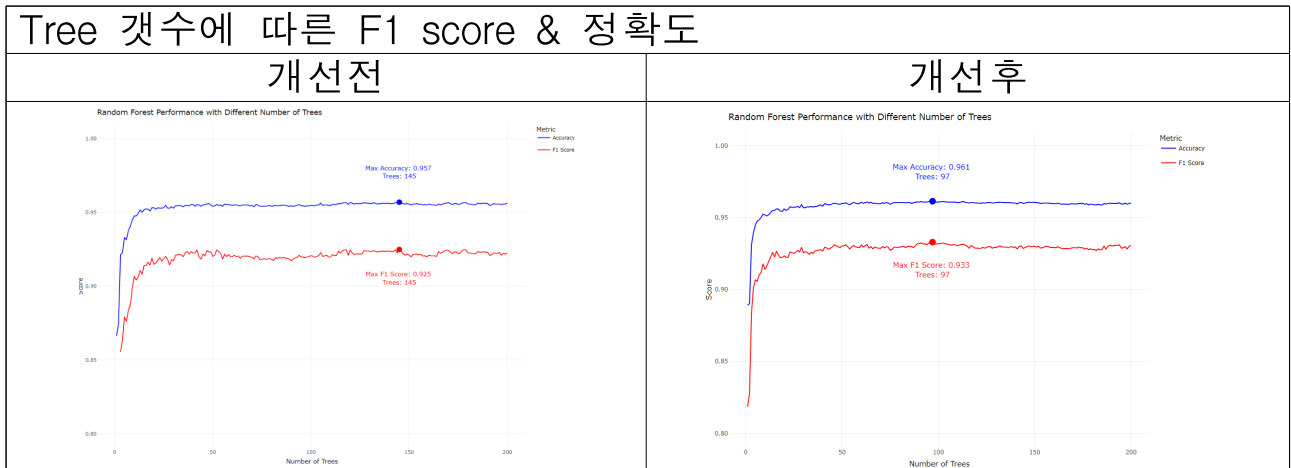
- 예측 모델의 정확도 향상을 위하여 추가 변수를 특정한다.
- 의사결정나무의 수 변화(n=1~100)에 따른 모델의 정확도를 확인한다.

f. 모델 성능 평가

- sensitivity, precision, specificity 와 F1 score를 확인한다.
- 각 분류 예측 결과에 대한 ROC 곡선 및 AUC를 계산한다.
- 결정 트리의 갯수에 따른 F1 score와 정확도를 확인한다.

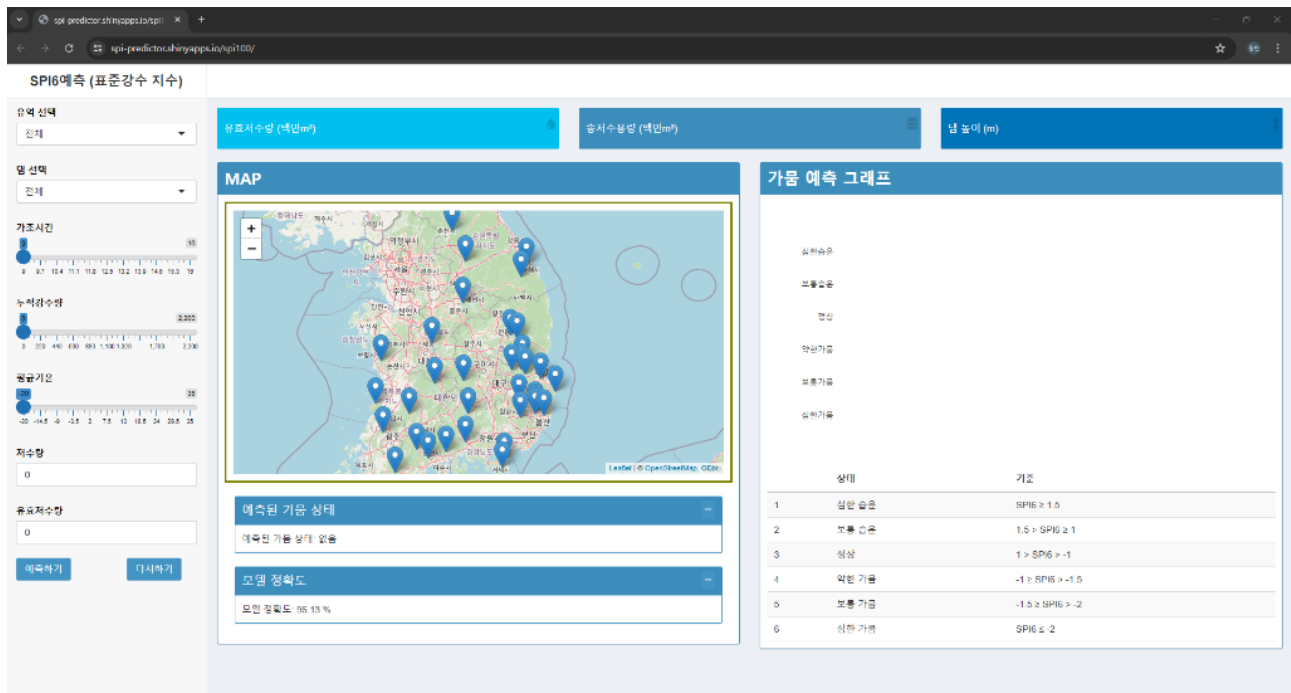
g. 모델 성능 개선

- 모델의 예측능력과 분류능력을 높이기 위해 데이터 정규화를 시행한다.
- 앞서 변수 최소화를 위해 제거했던 변수를 추가한다.



	정밀도	민감도	특이도	F1 score	정확도
개선전	0.941	0.91	0.985	0.925	95.7%
개선후	0.944	0.922	0.987	0.933	96.1%

- h. R shiny를 통한 UI를 통한 홈페이지 제작
-사용 설명 동영상 첨부한다.



- <https://spi-predictor.shinyapps.io/spi100/>

3. 참신성

a. 다중 요소 결합 분석

기상학적 요소와 수문학적 요소를 결합하여 가뭄 예측 모델을 개발한다.

b. 기존 데이터의 파생 변수 활용

기존 데이터를 심도 있게 분석하고, 누적강수량 계산 등의 정교한 데이터 가공 과정을 통해 새로운 파생 변수를 생성함으로써 예측 모델의 정확도를 획기적으로 향상시켰습니다.

4. 차별화된 특징

a. SQL과 R의 결합

SQL을 통해 데이터 전처리와 기본 분석을 수행하고, R을 통해 모델의 정밀한 분석과 평가를 수행하여 도구의 강점을 최대한 활용한다.

b. 머신러닝 분석

1차, 2차에 걸친 상관관계 분석 후 랜덤포레스트를 통해 체계적으로 평가된 변수를 적용한 머신러닝을 활용한다.

c. 사용자 편의를 고려한 UI 구현

R Shiny를 활용하여 예측 결과를 직관적으로 확인할 수 있는 사용자 인터페이스를 제공한다.

d. 종합적인 가뭄 예측 시스템

조기 경고 시스템을 구축하여 가뭄 발생을 사전에 경고하고, 선제적 대응을 가능하게 함으로써 물 자원 관리의 효율성을 극대화한다.

IV. 결과 및 기대효과

1. 결과

- a. 본 과제는 기후 변화와 도시화 등의 요인으로 인한 가뭄을 예측하는 것에 대한 필요성이 증대되는 상황에서, 다차원 데이터를 통합하고 고도화된 예측 모델을 활용하여 가뭄 지수를 예측하고, 실시간으로 대응할 수 있는 시스템을 구축하는 데 중점을 두고 있다. 데이터의 통합 및 자동화, 고도화된 머신러닝 모델의 적용, 사용자 중심의 UI 홈페이지를 제공하여 기존 방법과 차별화된 방식을 제시한다. 이를 통해 정확하고 신뢰성 있는 가뭄 예측을 가능하게 하여, 효율적인 물 자원 관리와 기후 변화 대응 전략 수립에 기여할 수 있을 것이다.

2. 기대효과

a. 효율적인 물 자원 관리

– 자원의 최적 분배

댐의 운영을 최적화하고, 수력 발전소의 효율적인 운영으로 물 공급 안정성을 보장한다.

– 농업용수 관리

가뭄 예측을 통해 농업용수의 공급을 조절하여 농작물 피해를 최소화한다.

b. 조기 경고 시스템 구축

- 조기 경고 및 대응

가뭄 발생을 조기에 경고하고 선제적 대응으로 피해를 최소화한다.

- 정책 결정 지원

실시간 데이터를 제공하여 신속한 대응책을 마련한다.

c. 기후 변화 대응

- 장기 전략 수립

기후 변화를 모니터링하고 적응 전략을 개발하여 대응력을 강화한다.

- 환경 보호

수자원을 효과적으로 관리하여 지속 가능한 물 자원 관리를 도모한다.