

## 농업 생산 데이터 분석 주제 후보

주제: 기후 변화와 작물 생산량의 상관관계 분석:

목표: 다양한 기후 및 농업 데이터 수집하여 기후변화가 작물 생산량에 미치는 영향을 파악.  
데이터를 기반으로 생산량 변화 예측 및 기후 변화 대응 전략을 도출.

### 1) 데이터 수집

KOSIS 채소생산량(엽채류)(1980~2023):

[https://kosis.kr/statHtml/statHtml.do?conn\\_path=l2&orgId=101&tblId=DT\\_1ET0021](https://kosis.kr/statHtml/statHtml.do?conn_path=l2&orgId=101&tblId=DT_1ET0021)

KASS 농축산물 단위당 생산량(2017~2021)

[https://kass.mafra.go.kr/statHtml/statHtml.do?mode=tab&orgId=114&tblId=DT\\_114985\\_A001](https://kass.mafra.go.kr/statHtml/statHtml.do?mode=tab&orgId=114&tblId=DT_114985_A001)

2) 데이터 구성: 작물 생산량 데이터와 기후 데이터 (강수량, 기온 등)를 통합.

3) 데이터 저장: 데이터의 지속적인 확장성을 고려한 HDFS 기반의 저장 구조 설계.

4) 데이터 처리 및 분석: Spark와 Python을 활용.

5) 분석 및 모델링: Hadoop 기반의 분산 환경에서 데이터를 저장, 처리하며, 분석 성능 최적화. Spark MLlib로 생산량 예측 모델 개발.

6) 시각화: Tableau를 활용하여 분석 결과를 시각화.

### Comment:

- 1) 농작물 데이터가 종류별, 년도별, 지역별로 크게 구성되어 있어 데이터 정제가 어려움
- 2) 강수량, 기온과의 유의미한 값이 도출이 될지 모르겠음(작물은 년도별이지만 날씨데이터는 더욱 세부적이고 자연재해 등 변수가 많음)
- 3) 실시간 데이터 스트리밍 기술(Kafka)을 활용한 분석이 가능한지 모름.