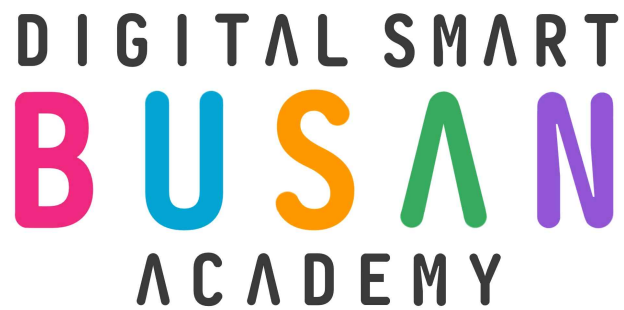


# 프로젝트명



팀명: GENOM

팀원: 김영휘, 장진원, 박수빈

디지털스마트부산 아카데미

# 1. 주제 선정 배경

## 가) 환경분석

### ☐ AI 모델 필요성

- 시장에서 세 품종이 동시에 유통될 때, 각 품종의 고유한 생산 품목(우유 및 식육)의 가치 및 가격 산정에 부정 유통이 차단되기 위해 현장에서 사용할 수 있다.
- 즉, 많은 SNP 정보를 통해 분류하는 것보다, 더욱더 적은 SNP 정보로 높은 분류 성능을 내는 것이 중요합니다.

### ☐ 품종 분류 모델을 개발

- 개체 정보와 사전에 구성된 15개의 SNP 정보를 바탕으로 해야 한다.

## 나) 시장분석

### ☐ 15개의 SNP 정보

- 염색체 정보, Genetic distance, 각 마커의 유전체상 위치 정보를 바탕으로 데이터를 사용하였다.
- 즉, 국내 이외 해외에서 여러 가지 지역에서 다방면으로 활용할 수 있습니다.

## 2. 개발 목적

### ☐ 배경

- 동일 개체를 확인하거나, 동일 품종을 구분하는 데 활용이 가능
  - 개체 정보와 SNP 정보를 이용하여 A, B, C 품종을 분류

### ☐ 주제

- 개체와 SNP 정보를 이용하여 품종 분류 AI 모델 개발
  - 최고의 품종 구분 정확도를 획득

### 3. 아이디어 소개 및 차별성

#### 가) 아이디어 소개

☐ 품종의 다양성 혹은 품종 부정 유통을 방지하기 위함

- 농축 수산 현장에서는 유전체 변이정보를 이용해서 품종을 구분하는 연구를 합니다.

#### 나) 기존 시스템 분석

☐ 데이터 분석

- 많은 SNP 정보를 통해 분류

#### 다) 차별화 기술 방안

☐ 기존 데이터 분석과 차별화

- 보다 더 적은 SNP 정보로 높은 분류 성능을 내는 것이 중요합니다.
  - 따라서, 개체 정보와 15개의 SNP 정보를 바탕으로 품종 분류 모델을 개발하였습니다.

## 4. 개발환경 및 사용 알고리즘

### ☐ 클라우드 기반 개발환경과 사용한 언어

- CoLaboratory(Google)
- Python(PSF)

### ☐ 머신러닝 모델링을 하기 전에 기계학습

- 데이터를 전처리
  - 범주형 데이터이기에 레이블 인코딩(Label encoding)을 적용했습니다.

### ☐ 머신러닝 모델

- Random Forest
  - 과대 적합을 방지하기 위해, 최적의 기준 변수를 랜덤하게 선택하는 모델
- Decision Tree
  - 일련의 분류 규칙을 통해 데이터를 분류, 회귀하는 지도 학습 모델이며, 결과 모델이 Tree 구조로 되어 있기 때문에 Decision Tree라고 불림
- Logistic Regression
  - 이벤트가 발생할 확률을 결정하는 데 사용되는 통계 모델
- KNN
  - 지도 학습에 한 종류로 거리 기반 분류분석 모델

### ■ 모델 비교

- 각 모델을 동일한 Train 데이터를 통해 학습시켜 정확도를 비교해본 결과 KNeighborsClassifier 모델이 가장 높은 성능을 보였습니다.

### ■ 교차 검증 이후의 모델 비교

- 교차 검증 이후에 모델의 정확도를 모두 비교해본 결과 RandomForest 모

델이 가장 높은 성능을 보였습니다.

- 교차 검증 후의 모든 모델의 정확도가 교차 검증 전보다 낮은 수치를 보였습니다, 따라서 데이터의 수가 부족하면 교차 검증을 수행하는 것이 모델 성능 개선에 좋은 효과를 주지 못하는 것을 파악하였습니다.

## 5. 비즈니스 모델 분석

가) 시장규모

☐ 글로벌 유전체 시장 규모

- 글로벌 시장조사업체 글로벌 게노믹스에 따르면 162억 달러(약 19조 원)
- 그에 비해, 국내 유전체 분석시장은 1,500억 원

나) 비즈니스 모델

☐ 빅데이터 기반 서비스 비즈니스 모델

- 빅데이터 시장의 핵심으로 농업 생산

☐ 수익 모델

- 교육 : 데이터 분석가, 관리자
- ETC: 농부, 제품 개발 컨설팅

## 6. 기대효과

### ☐ 경제적 효과

- 사람이 아닌 기계를 통해 시간과 인건비를 줄일 수 있고, 휴먼에러를 줄이고 높은 분류 성능이 가능해집니다.

### ☐ 사회적 효과

- 품종별 범죄를 예방할 수 있습니다.