

# 곰소만 해역의 기계학습을 이용한 분변성대장균 발생 예측

김종구\* · 강훈\*\* · 권민선\*\* · 장효상\*\*\*

\*, \*\* 군산대학교

## Occurrence Prediction of Fecal Coliform Using Machine Learning In Gomso-Bay of Korea

Jong-Gu Kim\* · Hoon Kang\*\* · Min-Sun Kwon\*\* · Hyo-Sang Jang\*\*\*

\*, \*\* Kunsan National University

**핵심용어** : 기계학습, 분변성대장균, 발생 예측, 곰소만

**Key Words** : Machine Learning, Fecal Coliform, Occurrence Prediction, Gomso-Bay

### 1. 개요 및 연구목적

우리나라의 경우 안전한 수산물 생산과 관련한 연안해역의 미생물학적 위생관리는 미흡한 실정이다. 이런 미생물학적 위생요소로부터 수산물 생산해역 관리 및 안전성을 확보할 수 있는 기술의 개발 및 방안의 마련이 절실한 시점이다.

본 연구에서는 패류생산해역인 곰소만 해역에서 11년(2005~2015년)간 조사한 분변성 대장균 자료를 이용하여 기계학습의 일반적인 과정(데이터 수집, 가공, 모델 훈련, 모델 성능 평가, 모델 성능개선)을 통하여 해역의 위생요소발생을 예측하고자 하였다.

### 2. 연구방법

본 연구에서는 매일 곰소만 해역에서 11년간(2005-2015년) 조사한 수온, 염분, pH, 분변성대장균 자료와 기상 자료(UV, 강수량)를 이용하였다.

기존의 연구들은 선형회귀분석이나 비선형회귀분석 등을 사용하여 대장균의 정량적 수치를 예측하였다. 본 연구에서는 예측성능의 일관성을 확보하기 위하여 정형화된 통계모델이 아닌 기계학습의 기법을 이용하여 분변성 대장균의 발생 여부를 예측하였다.

기계학습 알고리즘 선택 시 정확도, 학습시간, 선형성, 매개변수 등을 고려하여 본 연구에서는 로지스틱 회귀분석(Logistic Regression), 나이브 베이즈(Naive-Bayes), 결정나무(Decision Tree), 서포트 벡터 머신(Support Vector Machine)의 4가지 알고리즘을 선택하였다. 각 적합 알고리즘은 10년간 자료를 훈련데

이터를 사용하여 1년(2015년)의 분변성 대장균 발생여부를 얼마나 정확하게 예측하였는지 평가하였다.

### 3. 결과 및 고찰

수온, 염분, pH, 분변성대장균 자료와 기상 자료(UV, 강수량)를 활용하여 기계학습 기법을 이용해 곰소만 해역의 분변성 대장균 발생 가능성을 예측하였다. 전통적인 방식으로 예측하였을 때, 4가지 알고리즘 모두 좋은 성능을 보였으며, 특히 로지스틱 회귀모델과 서포트 벡터 머신이 가장 좋은 성능을 보였다. 하지만 선택한 알고리즘의 성능이 미래에도 유사한 결과를 가져올 수 있는지 추정하는 미래성능 추정이라는 검증과정을 통해 성능향상을 할 수 있었다. 이러한 미래성능 추정기법으로 K-fold 교차검증(k=10)을 사용한 결과는 결정나무, 나이브 베이즈 순으로 높은 성능을 보였다.

### 4. 결론

기계학습을 이용하여 곰소만 해역의 10년간 자료를 훈련데이터로 사용하여 1년(2015년)의 분변성 대장균 발생여부를 예측해 본 결과, 일반적인 결과는 로지스틱 회귀모델과 서포트 벡터 머신 알고리즘이 가장 좋은 성능을 보였다. 하지만 미래성능 추정기법을 사용한 결과는 결정나무, 나이브 베이즈 알고리즘 순으로 높은 성능을 보였다.

본 연구결과는 수산물 생산해역에서 분변성 대장균과 유사한 질병을 야기하는 바이러스 또는 병원성균에 의한 질병발생 예보를 위한 예측모델 개발에 활용되고, 수산물 생산해역을 관리하는데 유용한 자료가 될 것으로 판단된다.

\* First Author : kjg466@kunsan.ac.kr, 010-3456-3947

† Corresponding Author : hyosaeng@nate.com, 010-9224-4628