

## 천수만 해역의 분변성대장균 발생 예측 모델 적용

김종구\* · 강훈\*\* · 권민선\*\* · 장효상\*\*\*

\*, \*\* 군산대학교

### Application of Predictive Model for Generation of Fecal Coliform In Chunsu-Bay of Korea

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

\*, \*\* Kunsan National University

**핵심용어** : 기계학습, 분변성대장균, 발생 예측, 천수만

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

#### 1. 개요 및 연구목적

연안해역의 미생물학적 위해소로부터 수산물의 안전성을 확보하기 위해서는, 생산되어 유통되는 해양생물에 대한 전수 조사 및 생산되는 해역의 관리가 필요하다. 그래서 수산물 생산해역을 관리하고 생산단계에서 수산물 안전성을 확보할 수 있는 기술의 개발이 절실한 시점이다.

기존 곰소만 해역에서 기계학습을 이용하여 해역의 위해요소인 분변성대장균 발생예측을 평가하였다. 곰소만 해역에 적용하였던 기계학습의 4가지 알고리즘에 대해 미래 성능 추정까지 완료했지만, 적합 모델을 실제로 적용하고 검증하는 과정이 남아있다. 그래서 적합한 모델이 해역특성이 다른 천수만 해역에 적용하여 좋은 성능을 보이는지 확인해 보고자 한다.

#### 2. 연구방법

모델에 입력되는 자료는 매월 천수만 해역에서 8년간(2008~2015년) 조사한 수온, 염분, pH, 분변성대장균 자료와 기상 자료(UV, 강수량)를 이용하였다.

4가지 각 적합 알고리즘은 곰소만 해역의 10년간 자료를 훈련데이터를 사용하여, 천수만 해역의 8년(2008~2015년)의 분변성 대장균 발생여부를 얼마나 정확하게 예측하였는지 평가하였다.

#### 3. 결과 및 고찰

곰소만 해역에 적용하였던 4가지 알고리즘을 천수만 해역의 관측된 데이터에 적용하였으며, 예측 정확도와 Kappa 수를 구

하여 모델을 검증하였다. 그 결과 검출을 검출로 예측한 정분율로 고려하였을 때 로지스틱 회귀모델의 정분류율이 가장 높았다. 적합 모델의 정확도와 Kappa값에서는 나이브 베이즈와 결정트리 알고리즘이 좋은 성능을 보였고, K-fold 교차검증 결과까지 함께 참조했을 때 분변성 대장균 예측에 가장 좋은 성능을 보이는 알고리즘은 나이브 베이즈로 확인되었다. 하지만, 실제 데이터의 불검출/검출 비율이 약 6배 정도로 높다보니 성능평가 하는데 신중해야 하며, 지속적인 모니터링을 통해 자료의 축적이 요구되고, 모델의 학습 또한 계속되어야 할 것이다.

#### 4. 결론

분변성 대장균 발생예측을 위하여 곰소만 해역에 적합된 4가지 알고리즘을 해역 특성이 다른 천수만 해역에 적용하여 검증하였다. 예측 정확도 및 Kappa 수를 구하여 모델을 검증한 결과, 나이브 베이즈 및 결정트리 알고리즘 순으로 안정적이고 좋은 성능을 보였다. 천수만 해역에 적용한 검증 결과와 K-fold 교차검증의 성능 안전성을 함께 검토하였을 때, 수산물생산해역에서 분변성 대장균 예측에 가장 적절한 알고리즘은 나이브-베이지스로 판단되었다.

본 연구결과는 단순히 해역의 분변성 대장균의 발생 예측하는데 그치지 않고, 수산물의 안전성을 위협하는 비브리오 등의 치명적인 위해요소를 예측하는데 크게 기여할 것으로 판단된다.

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

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