

# 이탈리안 라이그라스 생산량 예측 회귀 모델 비교 연구

윤홍상, 황민우, 신문선, 양승학\*

건국대학교 컴퓨터공학과, \*축산과학원

e-mail:hsyoon27@naver.com /ghkdalsdn312@naver.com

/msshin@kku.ac.kr/y64h@korea.kr

## A Comparative Study of Regression Models for Predicting Italian Ryegrass Yield

Hong-Sang Yoon, Min-Woo Hwang, Moon-Sun Shin, Seung-Hak Yang\*

Dept. of Computer Engineering, Konkuk University

\*National Institute of Animal Science

### 요약

본 논문에서는 이상기후로 인한 생산량 변화와 재배적지 변동이 발생하고 있는 사료작물의 생산량 예측을 위한 회귀모델을 비교 분석하였다. 동계사료작물인 이탈리아라이그라스의 생산량에 영향을 미치는 기후요소들인 유효적산온도, 강수량, 최저기온 등에 대한 데이터를 수집하여 IRG 생산량과의 상관분석을 수행하였으며 특징추출을 통한 기후요소 선정을 수행하여 LASSO 모델, RIDGE 모델, ElasticNet 모델 기반 생산량 예측모델을 구축하여 기후영향 취약성에 대한 비교 분석을 수행하였다.

### 1. 서론

최근 빈번하게 발생하고 있는 이상기후 현상은 농업 생태계에 중대한 영향을 미치고 있다[1]. 지난 한 세기 동안 평균 기온 상승과 계절 길이의 변화를 지속적으로 발생하고 있으며, 이러한 변화는 축산농가의 사료작물 생산성의 불안정을 심화시키고 있다. 본 논문에서는 동계사료작물인 이탈리아 라이그라스(IRG)의 기후변화에 따른 생산성 변화를 예측하기 위하여 IRG 생산량 데이터와 지역 기후 데이터를 통합하여 생산량 예측모델에 대한 비교분석을 수행하였다. 분석을 위한 기초데이터셋을 준비하고 IRG 생산량을 예측하는 모델을 구축하기 위해 머신러닝 기반 회귀기법인 LASSO, RIDGE, ElasticNET을 적용하여 IRG 생산량 예측 모델을 도출하였으며 이를 비교분석하였다. 본 연구의 결과는 향후 동계사료작물 재배 적지 설정과 전자기후도 구축에 활용될 수 있을 것이다.

### 2. 기초데이터수집 및 전처리

본 연구에서는 기후변화에 따른 IRG 사료작물의 기후 영양성을 확인하고 생산량 예측 모델 개발을 위해 두 가지 주요 데이터를 수집하였다. 국립축산원에서 수집한 시험 데이터인 IRG생산량(건물 수량, kg/ha) 데이터는 건물 수량에 따라 해당 연도와 지역을 구분하여 구성되어

있다. 기존 연구에서 활용되었던 304개의 데이터에 추가로 수집, 공유된 데이터를 통합해 총 1037개의 데이터를 활용해 기초데이터셋을 구축하였다. 그림1은 이번 연구에서 구축한 기초데이터 DB를 보여주는 그림으로 일부 수집 과정에서 누락된 값들은 이후 전처리과정에서 NA로 처리하여 모델에 적용하였다.

	1기(kg/ha)	2기(kg/ha)	합계(kg/ha)	GDD											
				Reference	GDDTOTAL	GDDOTA	9월	10월	11월	12월	1월	2월	3월	4월	
2012년	1235	412	1647	1108	1627.35	239.9	27.5	33.4	67.7	6.1	7.3	0	32.4	121.5	
2011년	705	1315	2020	100	1417.15	147.8	43.9	25.5	14.3	27.2	9	28.5	8.3	93.3	
2012년	705	1315	2020	1112	1627.35	239.9	27.5	33.4	67.7	6.1	7.3	0	32.4	121.5	
2012년	869	1197	2065	1124	1627.35	239.9	27.5	33.4	67.7	6.1	7.3	0	32.4	121.5	
2009년	2263		2263	196	1465.6	238.45	23.3	4.7	77	20.5	2.5	23.8	7.1	68.1	
2012년	854	1516	2369	1120	1627.35	239.9	27.5	33.4	67.7	6.1	7.3	0	32.4	121.5	
2012년	990	1484	2474	1128	1627.35	239.9	27.5	33.4	67.7	6.1	7.3	0	32.4	121.5	
2012년	958	1516	2474	1116	1627.35	239.9	27.5	33.4	67.7	6.1	7.3	0	32.4	121.5	
2011년	1083	1598	2681	412	1417.15	147.8	43.9	25.5	14.3	27.2	9	28.5	8.3	93.3	
2012년	1083	1598	2681	1104	1627.35	239.9	27.5	33.4	67.7	6.1	7.3	0	32.4	121.5	
2016년	755	1978	2733	228	1746.9	349.6	31.8	74.1	111.6	21.2	1	65.9	57.4	89.3	
2012년	2871		2871	1212	1989	331	71.2	34.6	118.9	10.5	20	5.2	78	51.5	
2011년	2978		2978	416	1809.55	287.1	90.5	41.5	15.9	45.7	4.5	58.4	23.5	80.8	
2012년	2978		2978	1188	1989	331	71.2	34.6	118.9	10.5	20	5.2	78	51.5	

[그림 1] IRG생산량예측 기초데이터셋

동계사료작물 IRG의 기후 영향성을 확인하기 위해 수집된 기상데이터들은 월별 최고기온 평균, 최저기온 평균, 평균기온 평균, 강수량, 강수일수 등 기후데이터들이다. 그리고 겨울 사료작물의 월동과 생육에 영향을 미치는 유효적산온도 데이터들도 추가적으로 수집하여 유효적산온도(GDD) 계산 방식인 ((일일 최고기온+일일 최저기온)/2)-4를 계산한 값이 4보다 클 경우 해당 값을, 그보다 작을 경우 0을 부여하도록 하여 유효적산온도를 계산

하였다. 이후 1월부터 4월까지의 유효적산온도(GDDFJTA) 데이터와 1월부터 12월까지 유효적산온도(GDDTOTAL) 데이터를 수집 및 계산하여 기초데이터 DB를 구축하였다.

### 3. IRG생산량 예측 회귀모델 비교

#### 3.1 탐색적 데이터 분석(EDA)

본 연구에서는 1차 수확량(반응 변수)과 기후 변수(설명 변수) 간의 통계적 관계를 사전 파악하여 향후 예측모델 설계에 있어 유의미한 변수 선정을 돕고자 하였다. 이를 위해 상관분석, 통계적 중요도 분석, 다중공선성 진단을 병행하여 변수 간의 상관성 및 중복성을 체계적으로 검토하였다.

##### (1) 상관분석 (Pearson Correlation Analysis)

IRG 1차 수확량과 각 기후 변수 간의 상관관계를 피어슨 상관 분석한 결과를 표1로 정리하였다. 3월 평균기온( $r = 0.548$ )과 3월 최저기온( $r = 0.527$ )이 가장 높은 양의 상관성을 나타냈다. 이는 월동 이후 생육이 재개되는 시점의 기온 조건이 수확량 변동에 핵심적으로 작용함을 의미한다. 또한, 1~2월 평균 및 최저기온( $r \approx 0.51 \sim 0.52$ ) 역시 상위권에 포함되어, 겨울철 저온 환경이 초기 생육 안정성에 중요한 영향을 미친다는 점을 시사한다. 누적 유효적산온도(GDDFJTA,  $r = 0.513$ )와 강수량 합계( $r = 0.508$ )도 주요한 양의 상관성을 보여 누적된 온도와 토양 수분 공급이 IRG 생산성에 긴밀히 연결되어 있음을 확인하였다.

[표 1] 1차 수확량과 기후 변수 간 상관계수 (상위/하위 주요 항목)

순위	기후 변수	상관계수 (r)
[1]	3월 평균 기온	0.548
[2]	3월 최저 기온	0.527
[3]...[5]	1 ~ 2월 최저, 평균 기온	0.52 ~ 0.51
[6]	GDDFJTA	0.513
[7]	강수량 합계	0.508
...(중략)...		
[12]	GDDTOTAL	0.468
...(중략)...		
[31]	10월 강수량 합계	0.288
...(중략)...		
[56]	12월 일조시간	-0.391
[57]	1월 일조시간	-0.452

##### (2) 변수 중요도 분석 (F-통계량 기반 Feature Selection)

사이킷런의 SelectKBest 모듈과 f\_regression 함수를 활용하여 기후 변수의 통계적 중요도를 평가한 결과를 표2로 정리하였다. 1차 수확량에 대한 기후 변수의 특성 중요도를 평가한 결과 12월~3월 평균, 최저기온 변수가  $F = 444.26 \sim 352.07$ 로 상위 점수로 평가하고 있다. 1차 수확량에 있어 월동 및 생육 기간인

12~3월까지의 최저, 평균 기온이 큰 영향을 미치고 있는 것으로 분석된다. 이어서 GDDFJTA( $F = 369.56$ )와 강수량 합계( $F = 360.35$ )가 높은 점수를 보여, 누적 온도 및 수분 조건이 IRG 생산성 예측에서 핵심적 역할을 하고 있음을 확인하였다. 이외에도 12월 및 2월 최고기온, 12월 강수량 등 겨울철 기후 변수들이 일정 수준 이상의 설명력을 나타냈다. 이러한 결과는 IRG 생육 과정에서 겨울철 기후 조건이 다음 해 수확량 예측에 있어 중요한 기반을 제공함을 도출해 낼 수 있다.

[표 2] 1차 수확량에 대한 기후 변수의 F-통계량 (상위 15개)

순위	특성 이름	F-통계량 점수
[1]	3월 평균 기온	444.26
[2]	3월 최저 기온	397.61
[3]	1월 최저 기온	383.83
[4]	2월 평균 기온	381.87
[5]	2월 최저 기온	371.40
[6]	GDDFJTA	369.56
[7]	강수량합계	360.35
[8]	12월 최저 기온	358.91
[9]	1월 평균 기온	352.07
[10]	12월 평균 기온	336.40
[11]	10월 최저 기온	298.80
[12]	GDDTOTAL	290.53
[13]	12월 최고 기온	286.47
[14]	2월 최고 기온	276.67
[15]	12월 강수량	271.00

##### (3) 다중공선성 분석 (Variance Inflation Factor)

기후 변수 간 다중공선성을 평가하기 위해 VIF를 산출하였고 이를 표3으로 정리하였다. 일반적으로  $VIF \geq 10$  이상일 경우 다중공선성을 의심할 수 있는데 분석 결과 다수의 변수에서 매우 높은 VIF 값이 확인되었다.

[표 3] 기후 변수 구분별 최고, 최저 VIF

구분	최고-최저	기후 변수	VIF
강수량	최고	9월 강수량	Inf
	최저	강수량 합계	Inf
최고기온	최고	1월 최고	2196.01
	최저	5월 최고	120.45
최저기온	최고	11월 최저	949.52
	최저	5월 최저	73.60
GDD	최고	GDDTOTAL	107.26
	최저	GDDFJTA	76.67
평균기온	최고	12월 평균	21.11
	최저	4월	7.06

특히 강수량 관련 변수는  $VIF = \infty$ 로 평가되어 다른 변수와의 완전한 선형 종속성을 보였다. 또한 기온 변수(평균·최고·최저기온) 역시 70~2000 이상의 수치를 기록하여 월별 기온 간 강한 중복성을 드러냈다. 특히 기후데이터는 시간적 연속성과 누적성을 지니고 있어, 월별 기온, 강수량 지표 간 강한 상관성이 불가피하다. 따라서 본 연구에서는 기후데이터의 특성으로 인해 높은 VIF 값이 도출되었음을 감안하여, 유사 변수 중 대표 변수를 선별하여 모델 안정성과 예측력을

높이고자 하였다.

### 3.2 IRG 생산량 예측 모델 구축

본 연구에서는 IRG 사료작물의 생산량과 기후데이터 간의 영향성을 분석하기 위해 1차 건물 수량을 종속변수로 설정하고 예측에 적합한 기후요소들을 독립변수로 설정하여 머신러닝 기반 회귀분석을 수행한다. 회귀분석에 사용된 변수 구성에는 다음 표4와 같다.

[표 4] IRG 생산성 예측 모델 선택 변수

구분	변수	정의
생산성	DMY1	1차 예측시기 건물 수량
	PREOCT	10월 강수량 합
강우	PRESUM	연간 내린 비의 양
	PREAFTOVWIN	월동 후 강우의 합
평균기온	MINTDEC	12월 최저 평균기온
	MINTJAN	1월 최저 평균기온
	MINTFEB	2월 최저 평균기온
	MINTMAR	3월 최저 평균기온
유효적산온도	GDDFJTA	유효적산온도(1~4월)

본 연구에서는 IRG 1차 수확량을 예측하기 위해 정규화 회귀 기법에 기반한 세 가지 모델(Lasso 회귀, Ridge 회귀, ElasticNet 회귀)을 적용하였다. 이들 기법은 모두 선형 회귀를 확장한 기법으로 정규화(regularization)를 수행하여 과적합과 일반화 성능 향상을 수행하는 기법이다. 그러나 본 연구에서는 변수 선택 방식에 있어서 앞서 선별한 변수들을 특정해 사용한다는 차이가 있으며, 이러한 차별성은 기후 변수 간 다중공선성이 존재하고 변수가 다수 포함된 본 연구의 데이터 특성을 반영한 조치이다.

세 가지 회귀 모델은 Python의 scikit-learn 패키지를 활용하여 구현하였다. 데이터는 훈련 세트와 테스트 세트를 8:2로 분할하였으며, 훈련 세트는 K-Fold 교차검증을 통해 모델 학습 및 하이퍼파라미터 최적화를 수행하였다. 각 모델은 Pipeline 구조를 사용하여 데이터 표준화(StandardScaler)를 거친 후 회귀 모델 학습을 진행하는 형태로 구성하였다. 이를 통해 연구를 진행함에 있어 항상 일관되는 데이터 전처리와 학습 과정을 수행하도록 하였다. 하이퍼파라미터의 탐색은 GridSearchCV를 통해 수행되었으며, 총 2단계로 탐색을 진행하였다. 1차에서는 로그 스케일로 설정된 넓은 범위의 정규화 계수( $\alpha$ )를 탐색하였고, ElasticNet의 경우 정규화 계수( $\alpha$ )와 L1 비율을 동시에 탐색하였다. 이어서 2차 정밀 탐색에서는 1차에서 얻은 최적값 주변을 선형 분할하여 추가로 탐색함으로써 더 안정적인 조합을 조율하였다. 최적의 파라미터 조합은 훈련 세트 교차 검증 성능을 기준으로 결정되었으

며, 이후 테스트 세트를 통해 모델 성능을 평가하였다. Ridge, Lasso, ElasticNet 회귀 모델을 대상으로 K-폴드 교차검증을 수행하여 각 모델의 성능을 사전에 검증하고, 이를 통해 모델의 일반화 가능성을 평가하였다. 이어서 GridSearchCV를 활용하여 하이퍼파라미터를 탐색하고, 이를 기반으로 각 모델의 예측 성능을 비교하였다.

[표 5] 하이퍼파라미터 최적화 모델 교차 검증 성능표

Model-Fold	RMSE	MAE	R <sup>2</sup>
Ridge-1	3269.36	2720.61	0.37
Ridge-2	3605.54	2862.58	0.06
Ridge-3	3523.74	2882.71	0.43
Ridge-4	3307.87	2706.02	0.28
Ridge-5	3230.36	2512.00	0.42
평균	3387.37	2736.79	0.31
Lasso-1	3277.44	2720.40	0.36
Lasso-2	3603.44	2863.75	0.06
Lasso-3	3512.38	2881.84	0.43
Lasso-4	3294.41	2691.56	0.28
Lasso-5	3241.43	2514.90	0.41
평균	3385.82	2734.49	0.31
ElasticNet-1	3269.66	2720.57	0.37
ElasticNet-2	3605.62	2862.65	0.06
ElasticNet-3	3523.19	2882.73	0.43
ElasticNet-4	3307.20	2705.22	0.28
ElasticNet-5	3230.92	2512.05	0.42
평균	3387.32	2736.64	0.31

5-폴드 교차 검증 결과(표5 참조), 세 모델의 평균 RMSE는 3385.82 ~ 3387.37로 거의 동일한 수준이었으며, 해당 교차 검증의 평균  $R^2$  역시 약 0.31수준으로 유사하였다. 이는 기후 데이터의 구조적 특성상, 정규화 방식의 차이가 예측 성능에 미치는 영향이 제한적임을 보여준다. 그럼에도 ElasticNet은 L1과 L2 규제를 동시에 반영함으로써 가장 적은 오차 폭과 안정적인 성능을 보여줬으며, 향후 변화된 환경 조건에서도 우수할 가능성이 높은 모델임을 도출할 수 있다.

최적화된 하이퍼파라미터를 적용하여 테스트 세트에서 성능을 평가한 결과(표6 참조), 세 모델의 성능은 전반적으로 유사한 수준을 보였다. Ridge( $\alpha = 2.111$ ) 모델은 Test RMSE=3405.16, Test  $R^2 = 0.41$ 을 기록하여 안정적인 예측력을 나타냈다. Lasso( $\alpha = 9.259$ ) 모델은 Test RMSE=3409.91, Test  $R^2 = 0.40$ 으로 Ridge 모델 대비 소폭 높은 오차를 기록했는데, 이는 기후 변수 간 강한 상관성에 대응하기 위해 대표 변수 8개를 사전에 선별하여 학습한 결과, Lasso 특유의 L1 기반 변수 선택의 추가적 축소 여지가 제한되었기 때문으로 해석된다. ElasticNet( $\alpha = 0.084$ , l1\_ratio = 0.96) 모델은 L1 및 L2 규제를 동시에 반영하여 변수 선택과 다중공선성 완화를 동시에 달성하였으며, Test RMSE=3405.49으로 소폭 낮은 오차와 Test  $R^2 = 0.41$ 로 세 모델 중 가장 균형 잡힌 성능을 나타

냈다. 전반적으로 RMSE와  $R^2$ 의 차이는 모델 간 격차가  $\leq 4.75$ 로 극히 미미하여 본 데이터 특성상 정규화 방식의 차이가 총체적 성능에 큰 영향을 미치지 않았음을 보여준다. 다만 ElasticNet이 상대적으로 안정적이고 해석 가능한 모델임을 시사한다.

[표 6] 하이퍼파라미터 최적화 모델 단일 훈련 성능표

모델	$\alpha$ (alpha)	L1 ratio	Train RMSE	Test RMSE	Train $R^2$	Test $R^2$
Ridge	2.111	-	3324.86	3405.16	0.36	0.41
Lasso	9.259	-	3326.63	3409.91	0.35	0.40
ElasticNet	0.084	0.96	3325.14	3405.49	0.36	0.41

#### 4. 결론

본 논문에서는 이상기후로 인한 생산량 변화와 재배적지 변동이 발생하고 있는 사료작물의 생산량 예측을 위한 회귀모델을 비교 분석하였다. 동계사료작물인 이탈리아 라이그라스(IRG)의 생산량에 영향을 미치는 유효적산온도, 강수량, 최저기온 등 기후요소 데이터를 수집하여 IRG 생산량과의 상관분석을 수행하였으며 특징추출을 통해 대표 변수를 선정한 뒤 Lasso, Ridge, ElasticNet 기반의 생산량 예측모델을 구축하여 기후 영향 취약성에 대한 비교 분석을 수행하였다. 비교 분석한 결과 Lasso는 대표 변수를 사전에 압축한 조건에서 L1 기반 변수선택의 추가적 효과가 제한되어 예측력은 Ridge, ElasticNet과 유사하나 약간 높은 오차를 보였다. Ridge는 다중공선성 하에서 계수 안정성이 높고, 가장 낮은 Test RMSE으로 미세하게 우수한 일반화 성능을 보였다. ElasticNet은 L1, L2 규제를 병행하여 해석성과 안정성의 균형을 확보하였으며, Test RMSE와  $R^2$ 가 두 모델보다 일부 지표에서 근소하게 우수한 결과를 보여주고 있어 향후 변화된 환경 조건에서도 생산량 예측이 우수할 가능성이 높은 모델임을 확인하였다. 본 연구의 결과는 IRG 재배 적지 선정과 사료작물 전자기후도 구축의 기초데이터로 활용될 수 있다.

#### Acknowledgement

본 연구는 농촌진흥청 신농업기후변화대응체계구축(R&D)사업(과제번호 RS-2024-00400758)의 지원으로 수행되었음.

#### 참고문헌

[1] Hatfield, J.L., Morton, L.W. and Hall, B. 2018. Vulnerability of grain crops and croplands in the Midwest to climatic variability and adaptation strategies. *Climatic Change* January. 146:263-275.

[2] Jeong Sung Jung, Hyung Soo Park, Hee Jung Ji, Ki Yoong Kim, Se Young Lee and Bae Hun Lee, Predicting Changes in the Suitable Agro-climate Zone of Italian Ryegrass Cultivars with RCP 8.5 Climate Change Scenario, *Journal of the Korean Society of Grassland and Forage Science* 2020 40(4): pp. 196 ~ 204, <https://doi.org/10.5333/KGFS.2020.40.4.196>

[3] Impact of Climate Change on Agriculture and Its Mitigation Strategies: A Review. *Sustainability* 2021, 13, 1318. <https://doi.org/10.3390/su13031318>

[4] Stöckle, C.O., Higgins, S., Nelson, R., Abatzoglou, J., Huggins, D., Pan, W., Karimi, T., Antle, J., Eigenbrode, S.D. and Brooks, E. 2018. Evaluating opportunities for an increased role of winter crops as adaptation to climate change in dryland cropping systems of the US Inland Pacific Northwest. *Climatic change* 146: 247-261.

[5] M.S.Shin, S.W.Lee, S.M.Hwang and S.H.Moon, "Correlation Analysis of Production Changes of Feed Crop Data according to Climate Change" , in *Proceeding of the 14th Annual International Conference on Future Information and Communication Engineering* , Jeju: pp. 149-152, 2022.

[6] Moon-Sun Shin (2025). A Study on Predicting Models for Pasture Production according to Climate Change using Lasso/Ridge Regression Method. *Journal of the Korea Academia-Industrial cooperation Society*, 26(4), 179-184. 10.5762/KAIS.2025.26.4.179

[7] Jae-Seok Lim, Moon-Sun Shin, & Sang-Ho Moon (2024-10-24). 기후 변화에 따른 목초 생산량 영향 분석 연구. 한국정보통신학회 종합학술대회 논문집, 경남.