

연구논문

생산성 영향인자를 활용한 기계학습 기반 셰일가스 생산량 예측모델 분석

이다영¹ · 신효진² · 임중세^{2*}¹한국해양대학교 해양에너지자원공학과, ²한국해양대학교 에너지자원공학과

Analysis of the Shale Gas Production Prediction Model Based on Machine Learning Using Productivity Factors

Da-Yeong Lee¹, Hyo-Jin Shin² and Jong-Se Lim^{2*}¹Department of Ocean Energy & Resources Engineering, National Korea Maritime & Ocean University, Busan, Korea²Department of Energy & Resources Engineering, National Korea Maritime & Ocean University, Busan, Korea

*Corresponding Author. Jong-Se Lim, jslim@kmou.ac.kr, Department of Energy & Resources Engineering, National Korea Maritime & Ocean University, Busan, Korea

Received

7 December 2023

Final version Received

28 February 2024

Accepted

26 March 2024

Abstract

Owing to the uncertainties caused by complex flow mechanisms in shale gas reservoirs, production prediction using machine learning has recently been performed. However, efficiently collecting and analyzing production-related factors is difficult. In this study, the productivity factors of different criteria were identified using feature selection, and gas production was predicted using the selected factors in various machine learning models. The error rate improved when productivity factors such as total fluid, measured depth, and lower perforation based on the filter method were applied to the random forest model. In addition, the potential for extending the proposed feature selection process and machine learning algorithms in Marcellus Shale was identified, and a reliable production prediction was expected to be possible by utilizing limited variables even at the sites where production information was difficult to obtain.

Key words : shale gas, feature selection, productivity factor, machine learning, production prediction

요약

셰일가스 저류층에서의 복잡한 유동 메커니즘으로 불확실성이 존재하여 최근 기계학습을 활용한 생산량 예측을 수행하고 있으나, 생산관련인자를 효율적으로 수집하고 분석하는 데는 어려움이 있다. 이 연구에서는 특징 선택을 활용하여 서로 다른 기준의 생산성 영향인자를 파악하고, 이를 다양한 기계학습 알고리즘에 적용하여 Barnett Shale의 가스 생산량을 예측하였다. 이에 Filter 기법 기반의 총주입유체량, 측정 깊이, 하부 천공 위치 등 생산성 영향인자를 random forest 모델에 적용할 경우 오차율을 개선할 수 있음을 확인하였다. 이와 더불어 Marcellus Shale에서 제안된 특징 선택 과정 및 기계학습 알고리즘의 확장 가능성을 파악하였으며, 이 연구를 통해 생산 정보 취득이 어려운 경우에도 제한된 변수를 활용하여 신뢰성 있는 생산량 예측이 가능할 것으로 사료된다.

주요어 : 셰일가스, 특징 선택, 생산성 영향인자, 기계학습, 생산량 예측

서 론

셰일가스는 기존의 천연가스와는 달리 미세한 공극 특성, 복잡한 유동 메커니즘으로 인하여 생산 및 생산량 예측에 불확실성을 지닌다(Niu *et al.*, 2022). 생산량 예측을 위해서는 수치 시뮬레이션, 생산감퇴곡선분석법(decline curve analysis, DCA) 등을 활용하고 있다. 수치 시뮬레이션을 활용한 연구에서는 셰일가스 생산량 예측 시 저류층 모델에서의 공극 크기 분포, 자연 균열, 수압파쇄(hydraulic fracturing) 매개변수의 민감도를 분석하여 생산량 예측 정확도를 향상시켰으며(Geng *et al.*, 2018), 가스 흡착, 가스 점도를 활용하여 셰일가스 저류층의 생산성을 파악하였다(Guo *et al.*, 2014; Ekundayo and Rezaee, 2019). 그러나 수치 시뮬레이션의 경우 저류층 모델 구축에 있어 생산정 정보와 물성 정보와 같은 다양한 인자를 필요로 하며, 생산이력 일치에 많은 시간이 소요된다(Niu *et al.*, 2021). 이에 비해, DCA는 생산이력만을 활용하며 빠른 계산이 가능하다는 장점을 기반으로 셰일가스 생산량 예측에 DCA 기반 기술을 적용하는 다양한 연구가 수행되었다(Tan *et al.*, 2018). Kim *et al.*(2014)은 DCA를 활용한 궁극가채량 예측 시 몬테카를로 시뮬레이션을 적용하여 셰일가스 생산 시 변동성을 고려하였으며, Zhang *et al.*(2016)은 셰일유정에서의 생산 시 발생하는 배유부피 증가와 연관된 매개변수를 활용한 방정식인 extended exponential DCA를 제안함으로써 경계 조건에서의 유동을 반영한 바 있다. Duong (2010)은 셰일가스의 균열 유동(fracture flow)을 고려한 새로운 DCA 모델을 제안하였으며, Wang *et al.*(2017)은 이를 기반으로 시간에 따라 변화하는 파쇄시간지수(fracture time exponent)를 통합한 모델을 제시하여 궁극가채량 예측성능을 개선할 수 있음을 확인하였다.

최근에는 기계학습을 활용하여 유·가스 생산량을 예측하고 있으나, 입·출력 자료의 특성이나 수에 따라 예측 정확도에 차이를 보였다(Luo *et al.*, 2019; Shin *et al.*, 2021). 이와 더불어 예측 시 모든 생산관련인자를 활용하는 것은 자료 수집 및 분석에 있어 많은 시간과 비용이 들어 효율성이 저하될 뿐만 아니라 모델의 복잡도가 증가할 수 있다. 따라서 다양한 생산관련인자 중 셰일가스의 생산성에 비교적 큰 영향을 미치는 주요인자를 분석하고, 이를 입력자료로 활용하여 생산량을 예측하는 연구가 수행되고 있다.

Panja *et al.*(2018)은 초기 gas-oil rate(GOR), 초기 저류층 압력 등의 8개 입력자료를 활용하여 기계학습 모델을 구축하였으며, 결정계수(coefficient of determination, R^2)와 normalized root mean square error(NRMSE)를 통해 3가지 기계학습 모델에 대한 성능을 비교하였다. Oh *et al.*(2021)은 Barnett Shale 가스정의 저류층 물성, 유정 완결 조건, 월

별 생산량 등의 자료에 피어슨 상관분석을 수행하여 입력인자를 구성하고, 다양한 case의 RF 기반 궁극가채량 예측 모델을 구축하여 성능을 개선하였다. 최근에는 ChangNing 가스정의 생산량을 예측하기 위하여 저류층의 지질학적 특성, 시추관련 인자, 생산정 완결 인자에 대한 중요도 분석 연구가 수행되었으며, 이를 random forest(RF), extreme gradient boosting(XGB), artificial neural network(ANN), support vector machine(SVM)에 적용하여 SVM의 예측 오차가 가장 낮음을 확인한 바 있다(Zhai *et al.*, 2022). 이러한 기존 연구들은 단일 기법을 활용하여 주요인자를 분석하거나 단일 알고리즘 기반의 생산량 예측을 수행하였다. 이에 다수의 기준에 따른 생산성 영향인자를 분석하고 각기 다른 기계학습 모델과 조합하여 생산량 예측모델의 신뢰성과 정확성을 확보하여야 한다. 따라서 이 연구에서는 셰일가스 생산에 영향을 미치는 인자와 기계학습 알고리즘의 다양한 조합을 통해 가스 생산량을 예측하고자 하였으며, 모든 생산관련인자를 활용한 예측결과와 비교하여 예측모델의 개선 여부를 파악하고자 한다.

특징 선택 및 기계학습 알고리즘 개요

특징 선택(feature selection)

특징 선택은 중복되거나 관련이 있는 인자들을 기반으로 하위집합을 형성하여 모델에 활용되는 인자의 수를 축소하는 과정을 의미한다. 과도한 수의 입력인자(input feature)를 사용할 경우 기계학습 모델은 과적합되어 새로운 자료에 대한 적용이 어려울 수 있다. 이에 불필요한 인자들을 제거하여 연산 부담을 감소시킴으로써 모델의 학습 및 예측 속도를 향상시킬 수 있으며(Jain and Zongker, 1997), 이러한 특징 선택은 Embedded, Filter, Wrapper 기법의 3가지 유형으로 구분된다. Embedded 기법은 모델 내부에서 인자를 선택하며, 인자에 대한 적절한 가중치를 주어 중요도를 결정한다(Pudjihartono *et al.*, 2022). 그 중 least absolute shrinkage and selection operator(LASSO)는 모델 입력인자의 절대값 합에 대한 제약조건을 설정하며, 이를 위해 일부 인자를 0으로 축소한다(Fonti and Belitser, 2017). LASSO는 n 개의 샘플과 각 샘플의 p 개의 인자에 대해 식 (1)과 같이 표현된다(Jo and Yoon, 2017).

$$(\hat{\alpha}, \hat{\beta}) = \underset{\alpha, \beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^n (y_i - \alpha - \sum_{j=1}^p \beta_j x_{ij})^2 + \lambda \sum_{j=1}^p |\beta_j| \right\}, \quad (1)$$

$$\text{subject to } \sum_{j=1}^p |\beta_j| \leq s,$$

여기서 x 는 입력인자, y 는 예측변수를 의미하며, α 는 오

차, β 는 회귀계수를 나타낸다. 또한, λ 는 정규화항의 강도를 조절하며 제약조건에 해당하는 s 는 값이 작아질수록 더 많은 인자를 0으로 축소시켜 모델을 단순화한다.

Filter 기법은 특징 선택 과정을 생산량 예측을 위한 회귀 모델과는 별도로 독립적으로 수행하는 방법으로, 인자 사이의 통계적인 특성이나 관계를 분석하여 중요도를 파악한다(Abusamra, 2013). 해당 기법에는 chi square test, analysis of variance(ANOVA) 등이 포함되며, 그 중 식 (2)와 같은 피어슨 상관계수는 인자 간의 선형적인 강도와 방향성을 측정하는 통계적 지표이다(Hall, 1999).

$$\rho(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}, \quad (2)$$

여기서 ρ 는 피어슨 상관계수에 해당하며, x 와 y 는 상관관계 분석 변수이다. 또한, \bar{x} , \bar{y} 는 두 변수의 평균을 나타낸다.

이와 더불어 특징 선택 시 기계학습 모델의 성능을 최적화하는 Wrapper 기법은 선택된 인자의 부분 집합을 활용하여 모델을 학습시킨다(Peng *et al.*, 2005). 해당 기법에는 sequential forward selection, sequential backward selection, genetic algorithms, recursive feature elimination(RFE) 등의 모델이 포함된다. RFE는 RF와 같은 훈련된 모델의 트리 내에서 노드 분할 시 사용되는 불순도 감소량을 기반으로 각 인자의 중요도 평균을 구하며, 중요도가 낮은 순으로 인자를 제거하여 기계학습 모델의 성능을 개선한다(Guyon and Elisseeff, 2003).

기계학습 알고리즘

트리 기반의 앙상블 기법은 블랙박스로 취급된 다른 알고리즘과 달리 예측값에 대한 입력인자의 중요도를 설명할 수 있으며, 데이터 전처리가 거의 필요하지 않는 장점이 있다. 또한, 다양한 유형의 변수를 취급하며 복잡한 비선형 관계에도 적용 가능하다(Zhang and Haghani, 2015). 그 중 서로 독립적이고 병렬적인 여러 결정트리로 구성된 앙상블 알고리즘인 RF는 classification and regression trees(CART) 결정트리와 배깅(bagging)을 기반으로 한다(Zhai *et al.*, 2022). 배깅은 bootstrap sampling을 통하여 여러개의 CART 결정트리를 구성하고, 문제에 대한 개별 결정트리의 예측 평균을 도출하여 편향성을 유지하고 분산을 줄인다(Breiman, 1994). 이와 더불어 gradient boosting machine(GBM)의 한계를 극복하기 위해 개발된 알고리즘인 XGB는 단계적으로 모델을 구축하고 특성 손실 함수의 예상값을 최소화하여 모델을 업데이트한다. 결정트리를 순차적으로 학습

하고 보완하는 잔차(residual) 처리과정을 수행함으로써 강력한 앙상블 모델을 생성하며, 정규화 및 복잡성 제어로 과적합을 방지하고 불순도를 감소시킨다. 이에 따라 대규모 데이터 처리 및 이상치 대응과 같은 환경에서의 뛰어난 예측성능을 제공한다(Chen and Guestrin, 2016; Yeom *et al.*, 2023). Light GBM(LGB)은 gradient 값이 큰 표본들을 일정 비율로 추출하고, gradient 값이 작은 표본들은 무작위로 제거하여 표본의 수를 줄이는 gradient based one side sampling과 공간에서 상호배타적인 특징들을 효율적으로 묶는 exclusive feature bundling을 적용한 GBM 기반의 알고리즘이다(Ke *et al.*, 2017). Al Daoud(2019)에 따르면, 해당 알고리즘은 예측 정확도를 유지하며 속도를 높이고 메모리 사용률을 줄일 수 있다.

이러한 기계학습의 성능은 모델에 사용되는 하이퍼파라미터(hyperparameter)에 의존하기 때문에 알고리즘 별 하이퍼파라미터 최적화를 통한 높은 성능의 모델을 구현하여야 한다(Won *et al.*, 2023). RF의 경우, 결정트리의 개수에 해당하는 number of estimators, 결정트리의 최대 깊이인 max depth, 리프노드가 가져야 할 최소 샘플 수인 min samples leaf, 내부 노드 분할 시 최소 샘플 수를 뜻하는 min samples split 등의 하이퍼파라미터를 통해 모델을 구성한다. XGB는 number of estimators, max depth와 더불어 모델의 가중치와 연관된 learning rate 및 과적합 방지 매개변수인 gamma 등을 활용하여야 하며, LGB의 경우 모델의 복잡도를 제어하는 변수인 number of leaves에 대한 고려가 필요하다.

생산성 영향인자를 활용한 세일가스 생산량 예측

이 연구에서는 생산정 자료로부터 생산성 영향인자를 분석하고, 이를 다양한 기계학습 알고리즘에 적용하여 세일가스의 생산량을 예측하기 위해 Fig. 1과 같은 절차를 따라 분석을 수행하였다. 예측모델의 성능평가를 위한 지표로는 식 (3)의 결정계수(R^2)와 더불어 식 (4), (5)와 같이 root mean square error(RMSE), mean absolute percentage error(MAPE)를 활용하였다.

$$R^2 = 1 - \frac{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2} \quad (3)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (4)$$

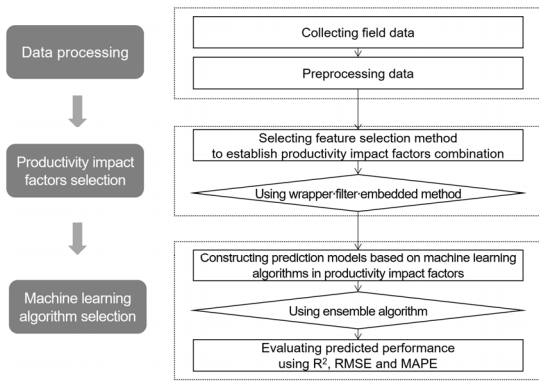


Fig. 1. Work flow of this study.

$$MAPE(\%) = \frac{100}{n} \sum_{i=1}^n \frac{|\hat{y}_i - y_i|}{y_i} \quad (5)$$

여기서 y_i 는 실제 생산량, \hat{y}_i 는 예측된 생산량, \bar{y} 는 실제 생산량의 평균에 해당한다.

현장자료 취득

Barnett Shale은 미국 텍사스 Forth Worth 분지에 위치하고 있으며, 주로 고생대 석탄기인 미시시피기(Mississippian)와 펜실베이니아기(Pennsylvanian)에 형성되었다(New Mexico Bureau of Geology and Mineral Resources, 2007). 25개의 카운티로 구성된 지질구조 가운데 Denton, Johnson, Tarrant, Wise의 4개 핵심 카운티는 1999년부터 2014년까지 Barnett Shale 지역 전체 가스 생산량의 83%를 차지하였다(Federal Reserve Bank of Dallas, 2023). 이 연구에서는 Enverus의 DrillingInfo를 활용하여 미국 텍사스 주에 위치한 Barnett 셰일가스 수평정을 대상으로 한 생산량 자료를 사용하였으며(Fig. 2), 아래와 같은 조건을 바탕으로 자료를 수집하였다. 취득된 자료의 한계로 인하여 공극률(porosity), 유체 투과도(permeability)와 같은 저류층 물성 및 인공채유(artificial lift) 적용여부와 같은 생산조건에 따른 영향분석은 고려하지 않았다.

1. 100개월 이하의 생산이력 제외
2. 누적가스 생산량이 1 Bcf 이하인 생산정 제외
3. 생산관련인자 중 결측치가 존재하는 생산정 제외

위의 조건을 통해 총 475개의 생산정 자료를 선정하였으며, 입력자료는 Barnett Shale 현장자료의 생산량 정보를 제외한 well information, completion design, stimulation design, gas property와 관련한 22개의 생산관련인자로 구성하였다(Fig. 3). Fig. 3에서 well information은 생산정의

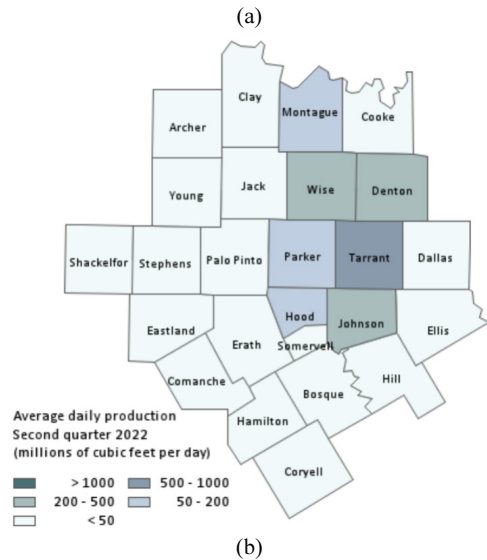


Fig. 2. Barnett Shale basin: (a) location of Barnett Shale and (b) average daily production in the second quarter of 2022 (Enverus, 2023; Federal Reserve Bank of Dallas, 2023).

위치와 연관된 인자, 시추공의 길이(measured depth, MD), 수직깊이(true vertical depth, TVD), 해수면 기준 생산정의 고도(elevation) 등과 같은 정보를 포함하며, 유정완결 설계 조건과 연관된 completion design은 수평정의 heel과 toe에 해당하는 상·하부 천공 위치(upper-lower perforation), 실제 수평정 길이(lateral length), 측정 수평정 길이(horizontal length) 및 수압파쇄 단계 수(stage count) 등으로 구성된다. 또한, 수압파쇄 시 활용된 주입유체(fluid) 및 프로판트(proppant)의 총량, 주입유체 단위 부피 당 프로판트 양(proppant concentration) 등이 stimulation design에 해당되며, gas gravity는 생산된 가스의 비중을 뜻한다.

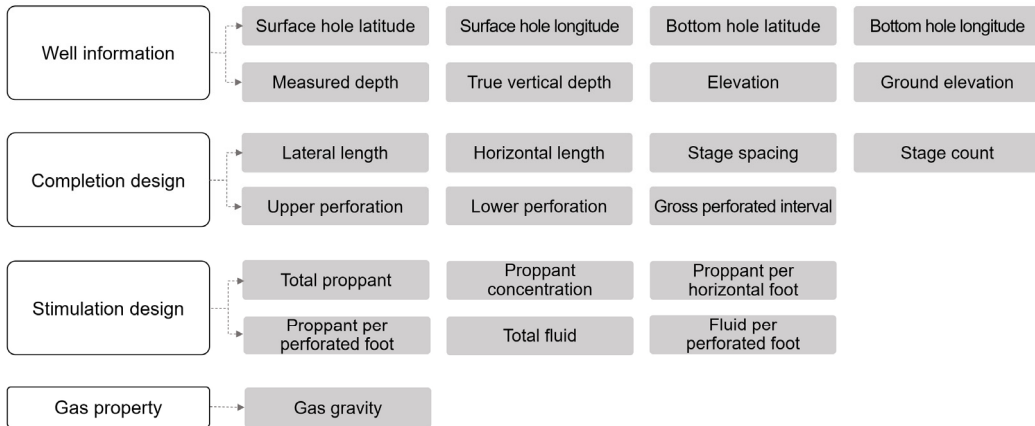


Fig. 3. Production-related factors of Barnett Shale.

Elevation	1.00	-0.09	-0.41	-0.42	-0.10	0.16	0.10	0.16	0.19	-0.06	-0.21	0.02	-0.24	-0.28	-0.09	-0.26	0.44	-0.05	-0.71	-0.04	-0.70	-0.11
1.00	Ground Elevation	-0.10	-0.41	-0.42	-0.11	0.15	0.09	0.15	0.18	-0.05	-0.21	0.02	-0.23	-0.28	-0.09	-0.26	0.44	-0.05	-0.71	-0.05	-0.70	-0.11
-0.09	-0.10	MD	0.44	0.39	0.99	0.80	0.84	0.80	0.52	-0.10	0.29	0.51	-0.18	-0.21	0.02	-0.25	-0.12	0.26	0.26	0.26	0.26	0.57
-0.41	-0.41	0.44	MD	0.76	0.43	-0.02	-0.03	-0.02	-0.03	0.00	0.16	0.17	0.15	0.16	0.18	0.04	-0.31	0.35	0.81	0.35	0.81	0.26
-0.42	-0.42	0.39	0.76	Upper Perforation	0.39	0.81	0.83	0.81	0.62	-0.10	0.30	0.62	-0.17	-0.21	0.01	-0.24	-0.12	0.24	0.27	0.24	0.27	0.57
-0.10	-0.11	0.99	0.43	0.39	Lower perforation	0.81	0.83	0.81	0.62	-0.10	0.30	0.62	-0.17	-0.21	0.01	-0.24	-0.12	0.24	0.27	0.24	0.27	0.57
0.16	0.15	0.80	-0.02	-0.22	0.81	0.85	0.85	1.00	0.70	0.01	0.28	0.51	-0.21	-0.33	-0.10	-0.29	0.04	0.09	-0.14	0.09	-0.14	0.47
0.10	0.09	0.84	-0.03	0.04	0.83	0.85	Horizontal Length	0.85	0.67	-0.10	0.20	0.54	-0.36	-0.33	-0.10	-0.31	0.02	0.14	-0.15	0.14	-0.14	0.48
0.16	0.15	0.80	-0.02	-0.22	0.81	1.00	0.85	DI Lateral Length	0.70	0.01	0.28	0.51	-0.21	-0.33	-0.10	-0.29	0.04	0.09	-0.14	0.09	-0.14	0.47
0.19	0.18	0.62	-0.03	-0.06	0.62	0.70	0.67	0.70	Stage Count	-0.55	0.10	0.46	-0.26	-0.35	-0.08	-0.33	0.10	0.18	-0.18	0.18	-0.17	0.36
-0.06	-0.05	-0.10	0.00	-0.18	-0.10	0.01	-0.10	0.01	-0.55	Stage Spacing	0.30	-0.01	0.33	0.35	0.09	0.35	-0.04	-0.04	0.02	-0.04	0.02	-0.07
-0.21	-0.21	0.29	0.16	0.06	0.30	0.28	0.20	0.28	0.10	0.30	Total Proppant	0.55	0.76	0.74	0.41	0.45	-0.09	0.12	0.18	0.12	0.19	0.20
0.02	0.02	0.61	0.17	0.08	0.62	0.61	0.54	0.61	0.46	-0.01	0.55	Total Fluid	0.26	0.14	0.68	-0.36	-0.10	0.11	0.09	0.11	0.10	0.48
-0.24	-0.23	-0.18	0.15	0.06	-0.17	-0.21	-0.36	-0.21	-0.26	0.33	0.76	0.26	0.93	0.56	0.57	-0.07	0.02	0.23	0.02	0.23	-0.09	-0.09
-0.28	-0.28	-0.21	0.16	0.17	-0.10	-0.33	-0.33	-0.33	-0.35	0.35	0.74	0.14	0.93	0.51	0.65	-0.07	0.06	0.24	0.06	0.24	0.06	0.42
-0.09	-0.09	0.02	0.18	0.17	0.01	-0.10	-0.10	-0.10	-0.08	0.09	0.41	0.68	0.56	0.51	0.51	-0.17	-0.11	0.07	0.18	0.07	0.18	0.12
-0.26	-0.26	-0.25	0.04	0.05	-0.24	-0.29	-0.31	-0.29	-0.33	0.35	0.45	-0.36	0.57	0.65	-0.17	-0.03	-0.03	-0.05	0.16	-0.05	0.16	-0.24
0.44	0.44	-0.12	-0.31	-0.27	-0.12	0.04	0.02	0.04	0.10	-0.04	-0.09	-0.10	-0.07	-0.07	-0.11	-0.03	Gas Gravity	0.32	-0.60	0.32	-0.60	-0.21
-0.05	-0.05	0.26	0.35	0.26	0.24	0.09	0.14	0.09	0.18	-0.04	0.12	0.11	0.02	0.06	0.07	-0.05	0.32	Surface Hole Latitude	-0.07	1.00	-0.07	0.11
-0.71	-0.71	0.26	0.81	0.67	0.27	-0.14	-0.15	-0.14	-0.18	0.02	0.18	0.09	0.23	0.24	0.18	0.16	-0.60	-0.07	Surface Hole Longitude	-0.06	1.00	0.21
-0.04	-0.05	0.26	0.35	0.26	0.24	0.09	0.14	0.09	0.18	-0.04	0.12	0.11	0.02	0.06	0.07	-0.05	0.32	1.00	Bottom Hole Latitude	-0.07	-0.07	0.11
-0.70	-0.70	0.26	0.81	0.67	0.27	-0.14	-0.14	-0.14	-0.17	0.02	0.19	0.10	0.23	0.24	0.18	0.16	-0.60	-0.07	1.00	-0.07	Bottom Hole Longitude	0.22
-0.11	-0.11	0.57	0.26	0.20	0.57	0.47	0.48	0.47	0.36	-0.07	0.20	0.48	-0.09	-0.12	0.12	-0.24	-0.21	0.11	0.21	0.11	0.22	Cum_Gas

Fig. 4. Correlation coefficients among production-related factors.

특징 선택을 활용한 생산성 영향인자 파악

이 연구에서는 Embedded 기법의 LASSO 회귀, Filter 기법의 피어슨 상관계수, Wrapper 기법의 RFE를 활용하여 22개의 인자 중 셰일가스의 생산성에 비교적 큰 영향을 미치는 생산성 영향인자를 분석하였다. LASSO 회귀분석 시에는 생산관련인자 중 값의 크기가 큰 일부 인자의 과도한 영향을 방지하기 위하여 표준화(standardization)를 통한 전처리를 수행하였다. 또한, 5회의 교차검증(cross validation, CV)을 기반으로 모델의 정규화 강도를 조절하는 하이퍼파라미터 λ에 대하여 0.001에서 10 사이의 조건 내 최적화를 진행하였다. 이에 일부 입력인자의 회귀계수가 0으

로 축소되었으며, 해당 인자를 제외한 surface hole longitude, lateral length, proppant per horizontal foot, gas gravity의 4개 생산성 영향인자를 파악하였다.

피어슨 상관계수를 활용하여 생산관련인자와 가스 생산량 간의 선형적인 관계를 분석하였다. lower perforation과 MD는 생산량에 대하여 0.57의 가장 높은 상관계수를 가졌으며, gross perforation interval, horizontal length, total fluid, lateral length의 경우 생산량에 대해 약 0.47의 상관성을 나타내었다. Fig. 4는 다양한 인자의 가스 생산량에 대한 상관계수를 나타낸 것으로서 상관계수가 0.3 이상인 7개의 인자를 생산성 영향인자로 선정하였다.

RF 기반의 RFE를 활용하기 위해 생산관련인자의 중요도에 따른 예측모델을 구축하고, 성능을 파악하여 최적의 예측성능을 지니는 생산성 영향인자를 분류하고자 하였다. 예측모델 구축 시 학습 및 테스트 자료는 각각 80%, 20%로 분할하였으며, number of estimators, max depth, min samples leaf, min samples split의 하이퍼파라미터를 사용하였다. CV는 5회 수행하였으며, 중요도 순위가 낮은 인자를 순차적으로 1개씩 제거한 후 예측성능지표인 MAPE와

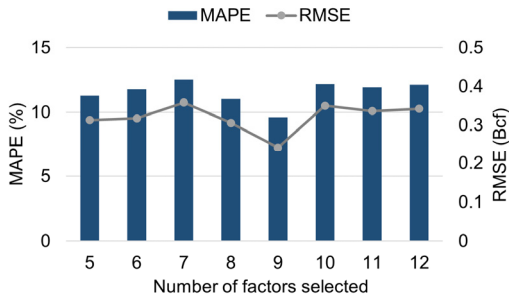


Fig. 5. Recursive feature elimination (RFE) results based on the number of factors.

RMSE를 분석한 결과 영향인자 9개 활용 시 예측 오차가 가장 낮은 것을 확인하였다(Fig. 5). 해당 9개 영향인자는 well information, completion design, stimulation design으로부터 각각 3개씩 도출되었으며, 이는 생산성 정보를 통한 지리적 조건, 수평정의 길이, 주입유체 및 프로판트에 대한 특성을 모두 포함하고 있어 생산량 예측 시 수반되는 물리적 현상을 반영가능할 것으로 사료된다. 이러한 분석을 통해 가스 생산량에 영향을 미치는 주요한 인자는 포함하고 관련성이 없는 인자는 제외하여 생산량 예측을 위한 최적의 입력인자를 구성하였다.

Table 1은 LASSO 회귀, 피어슨 상관계수, RFE를 통해 선정한 생산성 영향인자를 정리한 표이다. Embedded 기법의 경우, well information, completion design, stimulation design, gas property 중 각각 1개의 인자를 생산성 영향인자로 선정하였다. 이는 규제를 통하여 생산관련인자 간 다중공선성(multicollinearity)을 방지하는 LASSO의 특성으로 판단된다. Filter 기법의 경우 대부분의 생산성 영향인자가 completion design과 관련되었으며, 이를 통하여 유정 완결 설계조건이 생산량에 대하여 비교적 높은 상관관계가

Table 1. Classification results of productivity factors based on feature selection

	Parameter	Embedded	Filter	Wrapper
Well information	Surface hole latitude			
	Surface hole longitude	○		
	Bottom hole latitude			
	Bottom hole longitude			○
	Measured depth		○	○
	True vertical depth			○
Completion design	Elevation			
	Ground elevation			
	Lateral length	○	○	
	Horizontal length		○	○
	Stage spacing			
	Stage count		○	
Stimulation design	Upper perforation			○
	Lower perforation		○	○
	Gross perforated interval		○	
	Total proppant			
Gas property	Proppant concentration			○
	Proppant per horizontal foot	○		
	Proppant per perforated foot			
Gas property	Total fluid		○	○
	Fluid per perforated foot			○
	Gas gravity	○		

Table 2. Optimization of hyperparameters

	RF	XGB	LGB
Learning rate	-	[0.0005 : 0.1]	
Number of estimators	[100 : 1000]		
Max depth	[1 : 10]		
Min samples leaf	[1 : 3]	-	
Min samples split	[2 : 10]	-	
Gamma	-	[1 : 3]	-
Number of leaves	-	[5 : 35]	

있음을 파악하였다. MD, horizontal length, lower perforation, total fluid는 Wrapper와 Filter 기법에서 공통적인 생산성 영향인자로 활용되었으며, 생산량에 대한 선형적인 상관관계가 비교적 큰 인자는 중요도 측면에서도 주요 인자임을 확인하였다.

기계학습 알고리즘 별 세일가스 생산량 예측모델 설계

3가지로 분류된 생산성 영향인자에 대하여 RF, XGB, LGB 알고리즘을 적용한 9가지 세일가스 생산량 예측모델을 구축하였다. 생산량 예측 시, 각 생산정의 100개월 시점의 누적생산량을 고려하였다. 총 475개의 생산량 자료에 대하여 80%를 학습자료로 나머지 20%는 테스트자료로 분할하였으며, 표준화를 통해 입력인자의 평균은 0이고 표준편차는 1인 정규분포로 스케일을 조정하였다. 이와 더불어 알고리즘에 따른 기계학습 모델 구축 시 모델을 구성하는 하이퍼파라미터에 대한 최적화를 수행하여 오차를 최소화하고자 하였으며, 이에 교차검증 기반의 그리드 탐색을 활용하였다(Table 2).

기계학습 알고리즘 별 세일가스 생산량 예측결과

입력인자에 따른 기계학습 알고리즘 별 세일가스의 생산량 예측결과를 도출하였으며, 생산관련인자 기반 예측모델에 대하여 생산성 영향인자 기반 예측모델의 성능 개선 여부를 파악하였다(Fig. 6). 22개의 생산관련인자를 활용하여 생산량 예측 시, RF의 R²은 약 0.75이며, XGB와 LGB에 비하여 상대적으로 높은 값을 가졌다. Embedded 기법 기반 생산성 영향인자를 적용한 3가지 예측모델의 경우, 22개의 생산관련인자를 활용한 모델과 비교하면 예측성능이 전반적으로 감소하였다. 이는 LASSO 회귀로 인하여 4개의 생산성 영향인자를 제외한 생산관련인자를 0으로 축소함에 따라 영향인자에 대한 편향이 발생하여 모델의 예측성능이 저하한 것으로 판단된다.

Filter 기법을 적용한 결과, RF 모델의 R² 값은 약 0.91으로 향상되었으며 MAPE는 약 13%에서 약 8.6%로 감소하여 성능 개선을 확인하였다. 반면, XGB와 LGB의 경우, 성능이 향상되지 않음을 확인하였다(Fig. 7). 또한, Wrapper 기법을 활용한 경우, RF 모델의 R² 값이 약 0.90으로 개선되었다. 이는 생산량 예측에 영향을 주는 주요한 인자만을 사용함으로써 예측모델이 자료 간 패턴을 효과적으로 학습하였으며, 이상치에 대한 영향을 비교적 적게 받은 결과로 분석된다.

Table 3은 생산성 영향인자의 구성에 따른 각 모델의 예측성능지표를 정리한 것이다. RF 알고리즘은 XGB와 LGB에 비하여 전반적으로 낮은 오차율을 나타냈다. 그 중 Filter, Wrapper 기법을 기반으로 한 RF 모델은 생산관련인자를 활용한 경우보다 예측성능이 개선되어 세일가스 생산 자료에 적합한 예측모델임을 확인하였다. 또한, Filter 기법 기반 7개의 인자를 활용할 경우 RMSE와 MAPE는 최대로 낮아지는 경향을 보였다. 그에 반하여 XGB 모델의 경우 성능의 개선을 확인할 수 없어 세일가스 생산량 예측에서의 활용이 어려울 것으로 보인다. LGB 모델은 XGB 모델과 동일하게 예측성능 개선에 어려움이 있었으나 각각 4, 7, 9개의 생산성 영향인자를 활용함에도 불구하고 오차율이 유사하였다. 이러한 결과는 여러개의 결정트리를 결합하여 작동하는 RF 예측모델의 특성으로 인하여 인자 간 복잡한 상호작용 및 중요도를 고려할 수 있기 때문으로 사료된다. 그에 반하여 XGB와 LGB의 경우, gradient boosting을 사용하여 결정트리를 순차적으로 구축하게 되며, 해당 과정은 비교적 인자의 중요도를 기반으로 이루어지기에 RF에 비하여 비교적 낮은 예측성능을 지니는 것으로 추측된다.

결과적으로 세일가스 생산량 예측 시 Filter와 Wrapper 기법 기반의 RF 모델은 모든 생산관련인자를 활용한 경우보다 예측오차를 더 크게 감소시켰다. 이에 Filter와 Wrapper 기법에서 공통적으로 주요하다고 판단된 인자인 total fluid, horizontal length, MD, lower perforation에 대해 피어슨 상관분석을 수행하였으며, 모든 인자는 서로 0.5 이상의 높은 상관성을 보였다(Fig. 8). 각 인자는 수평정의 길이 및 파쇄 시 주입유체와 연관되며, 수평정의 길이에 비례하여 균열 표면적이 증가함에 따라 세일가스 저류층의 생산성에 영향을 미친다. 따라서, 생산정의 유정완결 설계 및 주입유체 조건은 생산량 예측 모델의 성능향상을 위해 필수적인 입력인자라 할 수 있다. 그러나 예측성능 향상을 위해 반복적으로 학습을 수행하며 알고리즘에 의존하는 Wrapper 기법에 비하여 Filter 기법은 계산량이 적고 학습 알고리즘과 독립적이기에 예측모델 중 가장 효율적인 모델로 판단하였다.

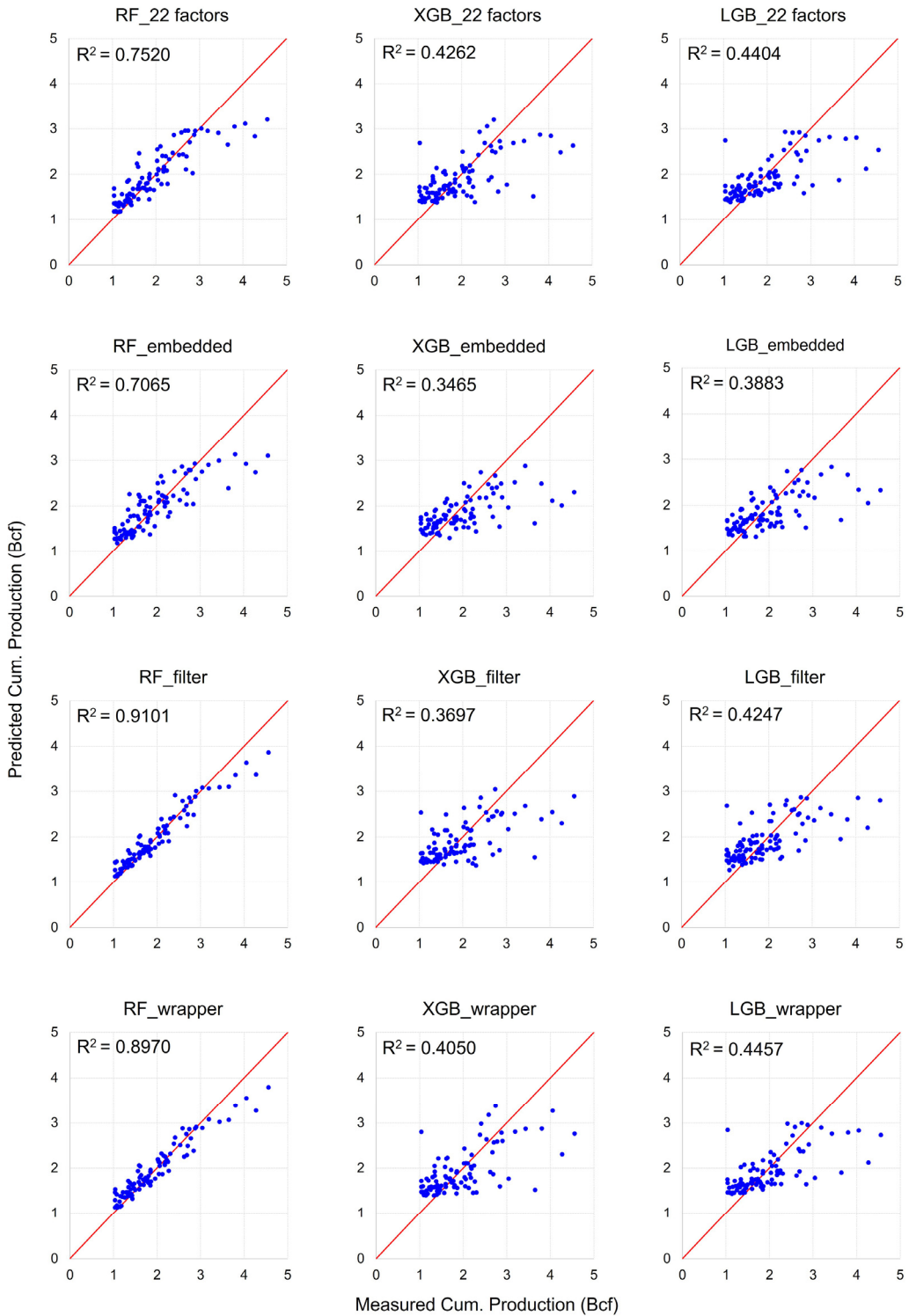


Fig. 6. Comparison of production prediction based on productivity impact factors.

Marcellus Shale에서의 생산량 예측모델 적용성 검토

Filter 기법을 활용한 생산성 영향인자 분석

Barnett Shale의 자료를 활용한 생산량 예측모델의 경우, 3 Bcf 이상의 생산량 자료가 비교적 적어 생산량이 증가할수록 예측성능이 감소하는 경향에 대한 추가적인 분석이 필요하다. 이에 셰일가스 저류층은 지역에 따라 생산성이 상이하므로 다른 Shale 분지에서의 활용성을 확인하고자 Marcellus Shale에 대하여 생산량 예측모델의 적용성을 검토하였으며, EIA(2023)에 따르면 해당 셰일층은 2021년 기준 가장 높은 천연가스 매장량을 기록한 바 있다. 미국 Appalachia 분지에 위치한 Marcellus Shale은 중기 Devonian 시대에 형성되었으며, 2040년까지 하루에 약 21 Bcf의 가스를 생산할 것으로 예상된다(Syed *et al.*, 2022). Barnett Shale의 생산정 자료와 동일한 조건 및 과정으로 전 처리를 수행하여 총 958개의 생산정 자료를 분석에 활용하였으며, 취득 가능한 현장자료를 고려하여 gas property를 제외한 well information, completion design, stimulation design의 21개 생산관련인자로 구성하였다.

Filter 기법의 피어슨 상관계수를 활용한 결과, 누적 가스 생산량에 대하여 MD, lower perforation, total proppant, total fluid, surface hole longitude, bottom hole longitude의 6개 인자가 0.4 이상의 상관관계를 보였다(Fig. 9). 그중 Barnett Shale의 가스 생산량에 대해 낮은 선형 상관관계를 지녔던 surface hole longitude, bottom hole longitude가 0.5의 상관계수를 가지는 영향인자로 파악되었다. 이는 북쪽의 뉴욕주에서 남쪽의 켄터키주까지 넓게 위치한 해당 Shale의 지리적 분포로 인하여 각 생산정의 위치정보가 상이하기에 지질학적 조건의 변동성이 존재하며(Fig. 10), 이러한 특성이 생산성에 비교적 큰 영향을 준 것으로 사료된다. 또한, Barnett Shale에서 생산성 영향인자로 판단한 MD, lower perforation, total fluid가 다른 Shale 분지에서도 영향인자로 활용됨을 확인하였다.

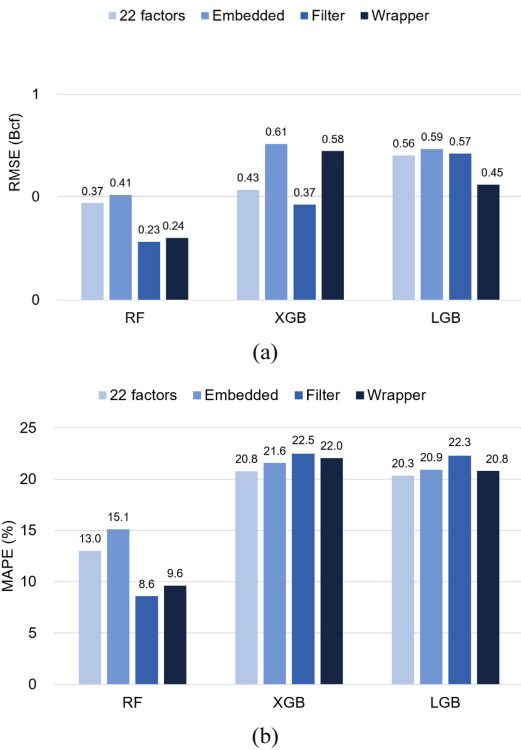


Fig. 7. Evaluation indices of various algorithms using different productivity impact factors: (a) RMSE and (b) MAPE.

Total Fluid	0.54	0.61	0.62
0.54	Horizontal Length	0.84	0.83
0.61	0.84	MD	0.99
0.62	0.83	0.99	Lower Perforation

Fig. 8. Correlation coefficients of productivity impact factors.

Table 3. Evaluation indices of various algorithms using different productivity impact factors

	RF		XGB		LGB	
	RMSE (Bcf)	MAPE (%)	RMSE (Bcf)	MAPE (%)	RMSE (Bcf)	MAPE (%)
22 factors	0.3744	13.02	0.4262	20.77	0.5625	20.34
Embedded	0.4074	15.11	0.6078	21.57	0.5881	20.92
Filter	0.2254	8.584	0.3697	22.49	0.5703	22.27
Wrapper	0.2413	9.610	0.5800	22.05	0.4457	20.79

Elevation	1.00	0.32	0.29	0.24	0.31	0.19	0.20	0.19	0.11	-0.03	-0.04	-0.11	-0.15	-0.21	-0.30	0.19	0.35	0.25	0.35	0.25	-0.08	
1.00	Ground Elevation	0.31	0.28	0.23	0.31	0.19	0.20	0.19	0.11	-0.03	-0.04	-0.11	-0.15	-0.21	-0.30	0.20	0.36	0.25	0.35	0.25	-0.08	
0.32	0.31	MD	0.59	0.55	0.95	0.73	0.80	0.73	0.40	0.08	0.35	0.36	-0.27	-0.24	-0.23	0.02	0.07	0.19	0.07	0.19	0.41	
0.29	0.28	0.59	TVD	0.80	0.59	0.10	0.08	0.10	0.07	0.02	-0.11	-0.03	-0.19	-0.24	-0.14	-0.12	-0.11	0.11	-0.11	0.11	0.21	
0.24	0.23	0.55	0.80	Upper Perforation	0.57	-0.08	0.08	-0.08	-0.03	0.04	-0.19	-0.11	-0.18	-0.21	-0.12	-0.14	-0.14	0.06	-0.14	0.06	0.18	
0.31	0.31	0.95	0.59	0.57	Lower Perforation	0.77	0.74	0.77	0.43	0.07	0.37	0.38	-0.09	-0.26	-0.24	0.02	0.06	0.19	0.06	0.19	0.43	
0.19	0.19	0.73	0.10	-0.08	0.77	Gross Perforated Interval	0.84	1.00	0.54	0.06	0.60	0.55	0.03	-0.15	-0.20	0.13	0.18	0.18	0.18	0.18	0.39	
0.20	0.20	0.80	0.08	0.08	0.74	0.84	Horizontal Length	0.84	0.42	0.13	0.48	0.42	-0.26	-0.18	-0.24	0.15	0.16	0.14	0.16	0.14	0.30	
0.19	0.19	0.73	0.10	-0.08	0.77	1.00	DI Lateral Length	0.94	0.54	0.06	0.60	0.55	0.03	-0.15	-0.20	0.13	0.18	0.18	0.18	0.18	0.39	
0.11	0.11	0.40	0.07	-0.03	0.43	0.54	0.42	0.54	Stage Count	-0.51	0.62	0.75	0.29	0.26	0.39	-0.24	-0.15	-0.16	-0.15	-0.16	0.28	
-0.04	-0.03	0.08	0.02	0.04	0.07	0.06	0.13	0.06	-0.51	Stage Spacing	-0.20	-0.31	-0.24	-0.29	-0.43	0.27	0.16	0.18	0.16	0.18	-0.03	
-0.04	-0.04	0.35	-0.11	-0.19	0.37	0.60	0.48	0.60	0.62	-0.20	Total Proppant	0.84	0.56	0.66	0.47	0.18	0.13	0.14	0.13	0.14	0.48	
-0.11	-0.11	0.36	-0.03	-0.11	0.38	0.55	0.42	0.55	0.75	-0.31	0.84	Total Fluid	0.45	0.52	0.67	-0.31	-0.09	-0.05	-0.09	-0.05	0.48	
-0.15	-0.15	-0.27	-0.19	-0.18	-0.09	0.03	-0.26	0.03	0.29	-0.24	0.56	0.45	Proppant Per Horizontal Foot	0.67	0.54	0.08	0.01	0.04	0.01	0.04	0.24	
-0.21	-0.21	-0.24	-0.24	-0.21	-0.26	-0.15	-0.18	-0.15	0.26	-0.29	0.66	0.52	0.67	Proppant per Perforated Foot	0.80	0.11	0.00	0.00	0.00	0.00	0.22	
-0.30	-0.30	-0.23	-0.14	-0.12	-0.24	-0.20	-0.24	-0.20	0.39	-0.43	0.47	0.67	0.54	0.80	Proppant Concentration	-0.46	-0.26	-0.22	-0.26	-0.22	0.21	
0.19	0.20	0.02	-0.12	-0.14	0.02	0.13	0.15	0.13	-0.24	0.27	0.18	-0.31	0.08	0.11	-0.46	Proppant	0.42	0.35	0.42	0.35	-0.03	
0.35	0.36	0.07	-0.11	-0.14	0.06	0.18	0.16	0.18	-0.15	0.16	0.13	-0.09	0.01	0.00	-0.26	0.42	Surface Hole Latitude	0.93	1.00	0.93	0.36	
0.25	0.25	0.19	0.11	0.06	0.19	0.18	0.14	0.18	-0.16	0.18	0.14	-0.05	0.04	0.00	-0.22	0.35	0.93	Surface Hole Longitude	0.93	1.00	0.50	
0.35	0.35	0.07	-0.11	-0.14	0.06	0.18	0.16	0.18	-0.15	0.16	0.13	-0.09	0.01	0.00	-0.26	0.42	1.00	0.93	Bottom Hole Latitude	0.93	0.35	
0.25	0.25	0.19	0.11	0.06	0.19	0.18	0.14	0.18	-0.16	0.18	0.14	-0.05	0.04	0.00	-0.22	0.35	0.93	1.00	0.93	Bottom Hole Longitude	0.93	0.50
-0.08	-0.08	0.41	0.21	0.18	0.43	0.39	0.30	0.39	0.28	-0.03	0.48	0.48	0.24	0.22	0.21	-0.03	0.38	0.50	0.35	0.50	Cum_Gas	

Fig. 9. Correlation coefficients of production-related factors.

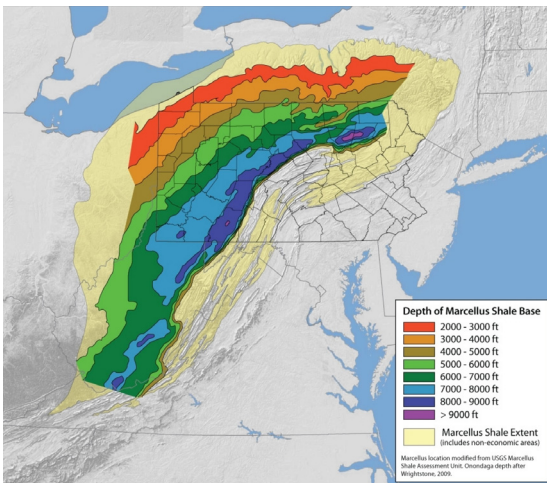


Fig. 10. Marcellus Shale basin (PHYS, 2023).

생산성 영향인자를 활용한 RF 모델의 생산량 예측

RF 모델의 생산량 예측결과, 생산관련인자를 사용한 경우 Barnett Shale의 R²인 0.7보다 향상된 0.9 이상의 값을 확인하였다. 이는 Barnett Shale 대비 높은 생산량을 가지는 자료를 포함하여 약 2배의 생산정 자료를 활용한 기계학습 모델의 학습성능 향상으로 인함을 알 수 있다(Fig. 11). 이에 따라 생산량의 크기에 따른 충분한 학습이 이루어진다면 모델의 성능이 개선될 수 있음을 파악하였다. 6개의 생산성 영향인자를 기반으로 한 RF 모델은 모든 생산관련인

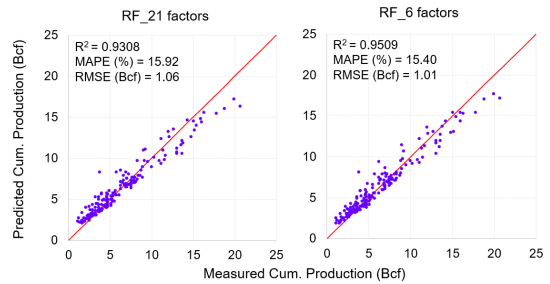


Fig. 11. Production prediction results based on productivity impact factors.

자를 활용한 예측모델보다 적은 개수의 인자를 활용하였음에도 불구하고 R² 값은 0.95 이상의 높은 성능을 지녔으며, 예측성능 지표인 MAPE와 RMSE를 통해 오차율의 안정성을 확인하였다. 이로써 Barnett Shale 외의 Shale 분지에서도 예측모델의 적용성을 검증하였다.

결론

이 연구에서는 셰일가스 생산량 예측의 불확실성을 개선하기 위해 특징 선택 및 기계학습 알고리즘을 활용하여 생산성 영향인자를 분석하고 예측모델을 구축하였다. 생산량 정보를 제외한 Barnett Shale의 입력자료에 대하여 LASSO 회귀, 피어슨 상관분석, RFE를 적용하였으며, 그 중 total fluid, horizontal length, MD, lower perforation 등

을 영향인자로써 활용하였다. 이후 생산성 영향인자를 입력자료로 하여 RF, XGB, LGB 기반 예측모델의 성능을 평가한 결과 Filter 기법을 기반으로 한 RF 모델이 가장 우수한 예측성능 및 효율성을 보였다. 또한, Marcellus Shale의 생산정 자료를 바탕으로 한 Filter 기법 기반 RF 모델을 구축하여 다른 Shale 분지에서의 적용 가능성을 확인하였다. 해당 분지의 경우 지역적으로 넓게 분포되어 있는 특성으로 인하여 생산정 위치의 중요도가 비교적 높음을 파악하였으며, 해당 정보를 활용하지 않을 경우 신뢰성 있는 생산량 예측에 어려움이 있을 것으로 사료된다. 이와 더불어 Barnett Shale과 동일하게 total fluid, horizontal length, MD, lower perforation이 생산성 영향인자로 활용되어 향후 현장에서의 생산량 예측 시 해당 인자를 고려해야 할 것으로 판단된다.

이 연구에서는 현장에서 사용 가능한 자료만을 입력자료로 활용하여 셰일가스 저류층의 물성으로 인한 물리적 특성을 반영할 수 없었기에 이에 대한 추가적인 분석이 필요할 것으로 보인다. 제한된 방식을 통해 생산 정보 확보가 어려운 경우에서도 제한된 변수를 활용하여 신뢰성 있는 생산량 예측이 가능할 것으로 사료된다.

사 사

이 논문은 2021년도 산업통상자원부의 재원으로 해외자원개발협회의 지원(NO. 2021060002, 디지털 오일필드 전문인력 양성)과 한국에너지기술평가원의 지원을 받아 수행된 연구임(NO. 20216110100050, 소형발전 연계 중소규모 가스전 개발/생산 현장 운영기술 개발).

References

Abusamra, H., 2013. A comparative study of feature selection and classification methods for gene expression data of glioma, *Procedia Computer Science*, 23, p.5-14.
 Al Daoud, E., 2019. Comparison between XGBoost, LightGBM and CatBoost using a home credit dataset, *International Journal of Computer and Information Engineering*, 13(1), p.6-10.
 Breiman, L., 1994. *Machine learning*, 24, Kluwer academic publishers, Amsterdam, The Netherlands, p.123-140.
 Chen, T. and Guestrin, C., 2016. XGBoost: A scalable tree boosting system, *Proceedings of the 22nd ACM SICKDD International Conference on Knowledge Discovery and Data Mining*, KDD, California, p.785-794.
 Duong, A.N., 2010. An unconventional rate decline approach for tight and fracture-dominated gas wells, *Canadian*

Unconventional Resources and International Petroleum Conference, Canadian Society for Unconventional Gas, Calgary, Canada, 15p.
 Ekundayo, J.M. and Rezaee, R., 2019. Numerical simulation of gas production from gas Shale reservoirs-Influence of gas sorption hysteresis, *Energies* 2019, 12(18), 12p.
 Enverus, 2023.08.30., <https://app.drillinginfo.com/gallery/>
 Federal Reserve Bank of Dallas, 2023.08.30., <https://www.dallasfed.org/research/energy11/barnett#region>
 Fonti, V. and Belitser, E., 2017. Feature selection using LASSO, *Amsterdam Research Paper in Business Analytics*, 30, p.1-25.
 Geng, L., Li, G., Wang, M., Li, Y., Tian, S., Pang, W., and Lyu, Z., 2018. A fractal production prediction model for Shale gas reservoirs, *Journal of Natural Gas Science and Engineering*, 55, p.354-367.
 Guo, C., Wei, M., Chen, H., He, X., and Bai, B., 2014. Improved numerical simulation for Shale gas reservoirs, *Offshore Technology Conference-Asia*, Kuala Lumpur, Malaysia, 17p.
 Guyon, I. and Elisseeff, A., 2003. An introduction to variable and feature selection, *Journal of Machine Learning Research*, 3, p.1157-1182.
 Hall, M.A., 1999. *Correlation-based Feature Selection for Machine Learning*, PhD Thesis, Waikato University, New-Zealand, 199p.
 Jain, A. and Zongker, D., 1997. Feature selection: evaluation, application, and small sample performance, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(2), p.153-158.
 Jo, J.H. and Yoon, S.R., 2017. Lasso regression of RNA-Seq data based on bootstrapping for robust feature selection, *KIISE Transactions on Computing Practice*, 23(9), p.557-563.
 Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q. and Liu, T.Y., 2017. LightGBM: A highly efficient gradient boosting decision tree, *31st Conference on Neural Information Processing Systems*, Neural Information Processing Systems, California, USA, 9p.
 Kim, J.S., Shin, H.J., and Lim, J.S., 2014. Probabilistic decline curve analysis for forecasting estimated ultimate recovery in Shale gas play, *Journal of The Korean Society of Mineral and Energy Resources Engineers*, 51(6), p.808-819.
 Luo, G., Tian, Y., Bychina, M., and Ehlig-Economides, C., 2019. Production-strategy insights using machine learning: Application for Bakken Shale, *SPE Reservoir Evaluation & Engineering*, 22(3), 17p.
 New Mexico Bureau of Geology and Mineral Resources, 2007. *The Barnett Shale in southeastern New Mexico: distribution, thickness, and source rock characterization*, 502, New Mexico tech socorro, , New Mexico, 57p.
 Niu, W., Lu, J., and Sun, Y., 2021. A production prediction

- method for Shale gas wells based on multiple regression, *Energies*, 14(5), 11p.
- Niu, W., Lu, J., and Sun, Y., 2022. Development of Shale gas production prediction models based on machine learning using early data, *Energy Reports*, 8, p.1229-1237.
- Oh, H.T., Ki, S.I., Park, C.H., and Jang, I.S., 2021. Analysis of uncertainty trend for estimated ultimate recovery prediction of Shale gas with various production periods based on machine learning, *Journal of The Korean Society of Mineral and Energy Resources Engineers*, 58(5), p.475-490.
- Panja, P., Velasco, R., Pathak, M., and Deo, M., 2018. Application of artificial intelligence to forecast hydrocarbon production from Shales, *Petroleum*, 4(1), p.75-89.
- Peng, H., Long, F., and Ding, C., 2005. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8), p.1226-1238.
- PHYS, 2023.08.30., <https://phys.org/news/2010-10-animations-extent-marcellus-Shale.html>
- Pudjihartono, N., Fadason, T., Kempa-Liehr, A.W., and O'sullivan, J.M., 2022. A review of feature selection methods for machine learning-based disease risk prediction, *Frontiers in Bioinformatics*, 2, 17p.
- Shin, H.S., Nguyen-Le, V., Kim, M., Shin, H.D., and Little, E., 2021. Development of production-forecasting model based on the characteristics of production decline analysis using the reservoir and hydraulic fracture parameters in Montney Shale Gas reservoir, Canada, *Geofluids*, 2021, 12p.
- Syed, F.I., Alnaqbi, S., Muther, T., Dahaghi, A.K., and Negahban, S., 2022. Smart Shale gas production performance analysis using machine learning applications, *Petroleum Research*, 7(1), p.21-31.
- Tan, L., Zuo, L., and Wang, B., 2018. Methods of decline curve analysis for Shale Gas reservoirs, *Energies* 2018, 11(3), 18p.
- U.S. Energy Information Administration (EIA), 2023.08.30., <https://www.eia.gov/naturalgas/crudeoilreserves/>
- Wang, K., Li, H., Wang, J., Jiang, B., Bu, C., Zhang, Q., and Luo, W., 2017. Predicting production and estimated ultimate recoveries for Shale gas wells: A new methodology approach, *Applied Energy*, 206, p.1416-1431.
- Won, J.H., Shin, J.M., Kim, J.H., and Lee, J.W., 2023. A survey on hyperparameter optimization in machine learning, *Journal of Korean Institute of Communications and Information Sciences*, 48(6), 15p.
- Yeom, J.Y., Kim, H.Y., Lee, K.B., Chang, C.D., and Jo, Y.G., 2023. Detection of borehole breakout depth in image logs using machine learning algorithms, *Journal of The Korean Society of Mineral and Energy Resources Engineers*, 60(4), p.223-230.
- Zhai, S., Geng, S., Li, C., Gong, Y., Jing, M., and Li, Y., 2022. Prediction of gas production potential based on machine learning in Shale gas field: a case study, *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, 44(3), p.6581-6601.
- Zhang, H., Rietz, D., Cagle, A., Cocco, M., and Lee, J., 2016. Extended exponential decline curve analysis, *Journal of Natural Gas Science and Engineering*, 36, p.402-413.
- Zhang, Y. and Haghani, A. 2015. A gradient boosting method to improve travel time prediction, *Transportation Research Part C: Emerging Technologies*, 58, p.308-324.



이 다 영

2024년 국립한국해양대학교 대학원 해양
에너지자원공학과 공학석사

현재 국립한국해양대학교 해양에너지자원공학과 공학석사
(E-mail; dylee@g.kmou.ac.kr)



신 효 진

2020년 한국해양대학교 대학원 해양에너지
자원공학과 공학박사

현재 국립한국해양대학교 에너지자원공학과 박사후연구원
(E-mail; hjshin@kmou.ac.kr)



임 종 세

1998년 서울대학교 대학원 자원공학과 공
학박사

현재 국립한국해양대학교 에너지자원공학과 교수
(E-mail; jslim@kmou.ac.kr)
