**AUTO\_ML\_Classifier**

PAPER\_BLEACHING

**최고의 모형 검색(이항 반응): Pulp Defects 대 Production Rate, Caustic, MgSO4, H2O2, O2, Conductivity, Discharge pH, Bleaching**

**방법**

|  |
| --- |
| 선형 항 및 순서 2의 항이 포함된 로지스틱 회귀 모델을 적합합니다. |
| 2 TreeNet® 분류 모델을 적합합니다. |
| 교육 데이터 크기 2033과(와) 동일한 부트스트랩 표본 크기로 3 Random Forests® 분류 모델을 적합합니다. |
| 최적의 CART® 분류 모델을 적합합니다. |
| 데이터의 30%이(가) 포함된 별도의 검정 집합을 사용하여 최대 로그우도가 있는 모델을 선택합니다. |
| 행의 총 수: 2906 |
| 로지스틱 회귀 모형에 사용되는 행: 2893 |
| 트리 기반 모형에 사용되는 행: 2906 |

**이항 반응 정보**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | **교육** | | **검정** | |
| **변수** | **등급** | **카운트** | **%** | **카운트** | **%** |
| Pulp Defects | Fail (사건) | 99 | 4.87 | 32 | 3.67 |
|  | Pass | 1934 | 95.13 | 841 | 96.33 |
|  | 모두 | 2033 | 100.00 | 873 | 100.00 |

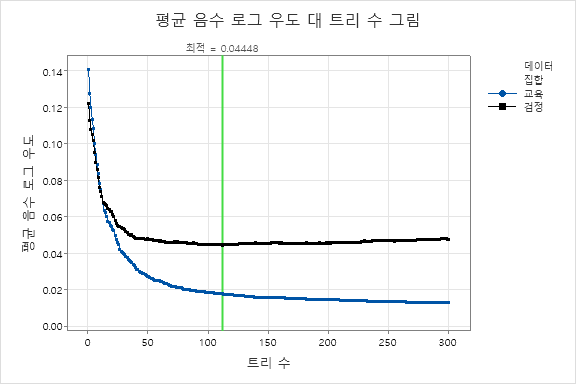
**모형 선택**

|  |  |  |  |
| --- | --- | --- | --- |
| **유형 내에서 최고의 모형** | **평균 음수 로그 우도** | **ROC 곡선 아래 면적** | **분류 율 잘못** |
| TreeNet®\* | 0.0445 | 0.9566 | 0.0126 |
| CART® | 0.0453 | 0.9715 | 0.0402 |
| 로지스틱 회귀 | 0.1050 | 0.9024 | 0.0333 |
| Random Forests® | 0.2202 | 0.9492 | 0.0137 |

\* 최소 평균 -로그 우도가 있는 모든 모형 유형에서 최고의 모형. 최고의 모형 출력이 뒤에 나옵니다.

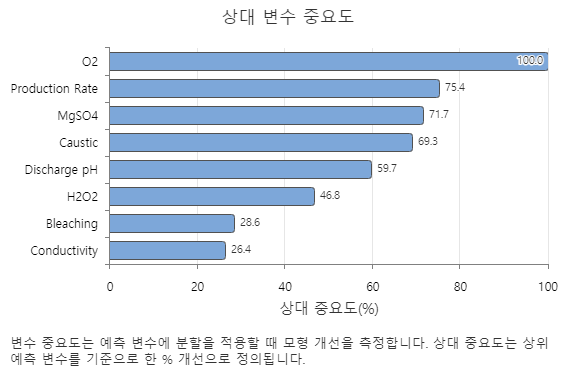
**최고의 TreeNet® 모형에 대한 하이퍼파라미터**

|  |  |
| --- | --- |
| 성장한 트리 수 | 300 |
| 최적의 트리 수 | 112 |
| 학습률 | 0.1 |
| 하위 표본 부분 | 0.7 |
| 트리당 최대 터미널 노드 수 | 6 |
| 노드 분할을 위해 선택된 예측 변수 수 | 총 예측 변수 수 = 8 |



**모형 요약**

|  |  |
| --- | --- |
| 전체 예측 변수 | 8 |
| 중요 예측 변수 | 8 |
| 성장한 트리 수 | 300 |
| 최적의 트리 수 | 112 |
| **통계량** | | **교육** | **검정** |
| 평균 음수 로그 우도 | | 0.0178 | 0.0445 |
| ROC 곡선 아래 면적 | | 0.9995 | 0.9566 |
| 95% CI | | (0.9991, 1.000) | (0.8989, 1) |
| 향상도 | | 10.0000 | 9.3750 |
| 오분류 비율 | | 0.0049 | 0.0126 |



**오차 행렬**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **예측된 등급(교육)** | | | | **예측된 등급(검정)** | | | |
| **실제 등급** | **카운트** | **Fail** | **Pass** | **정답률(%)** | **카운트** | **Fail** | **Pass** | **정답률(%)** |
| Fail (사건) | 99 | 93 | 6 | 93.94 | 32 | 26 | 6 | 81.25 |
| Pass | 1934 | 4 | 1930 | 99.79 | 841 | 5 | 836 | 99.41 |
| 모두 | 2033 | 97 | 1936 | 99.51 | 873 | 31 | 842 | 98.74 |

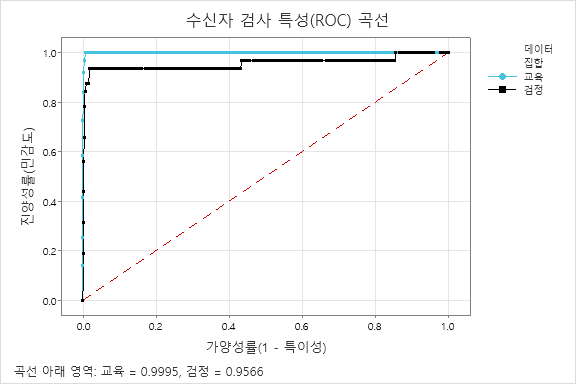
행의 사건 확률이 0.5을 초과하는 경우 사건 등급에 행을 할당합니다.

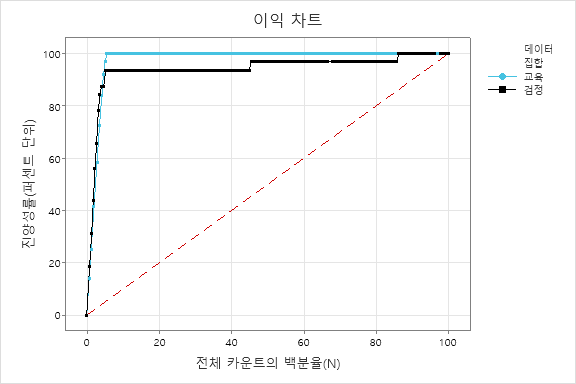
|  |  |  |
| --- | --- | --- |
| **통계량** | **교육(%)** | **검정(%)** |
| 진양성률(민감도 또는 검정력) | 93.94 | 81.25 |
| 가양성률(유형 I 오차) | 0.21 | 0.59 |
| 가음성률(유형 II 오차) | 6.06 | 18.75 |
| 진음성률(특이성) | 99.79 | 99.41 |

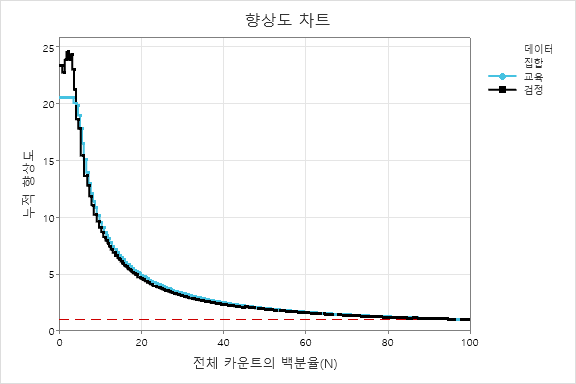
**오분류**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **교육** | | | **검정** | | |
| **실제 등급** | **카운트** | **오분류됨** | **오차율(%)** | **카운트** | **오분류됨** | **오차율(%)** |
| Fail (사건) | 99 | 6 | 6.06 | 32 | 6 | 18.75 |
| Pass | 1934 | 4 | 0.21 | 841 | 5 | 0.59 |
| 모두 | 2033 | 10 | 0.49 | 873 | 11 | 1.26 |

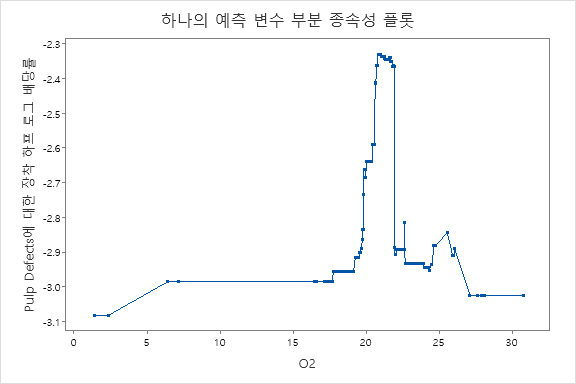
행의 사건 확률이 0.5을 초과하는 경우 사건 등급에 행을 할당합니다.

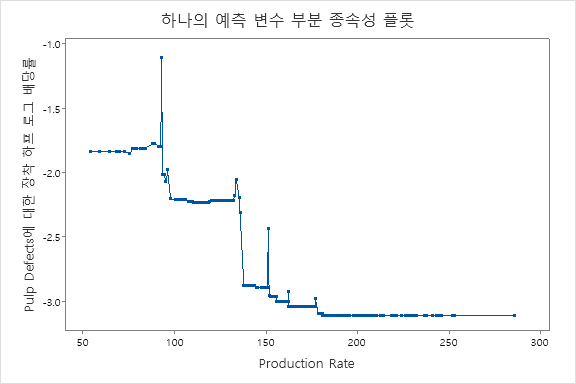


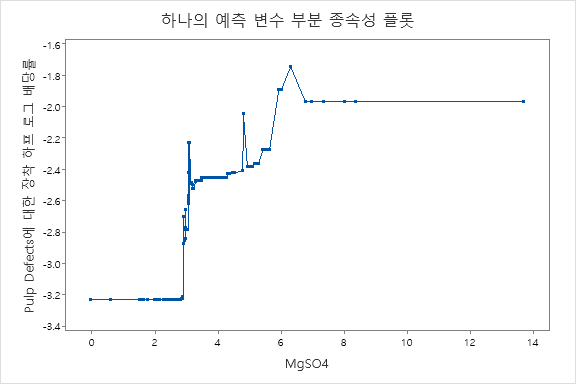


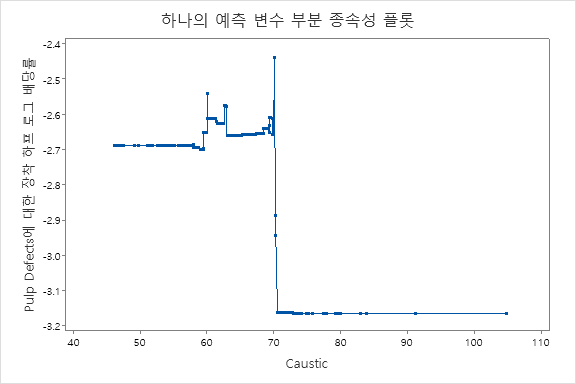


**하나의 예측 변수 부분 종속성 플롯**

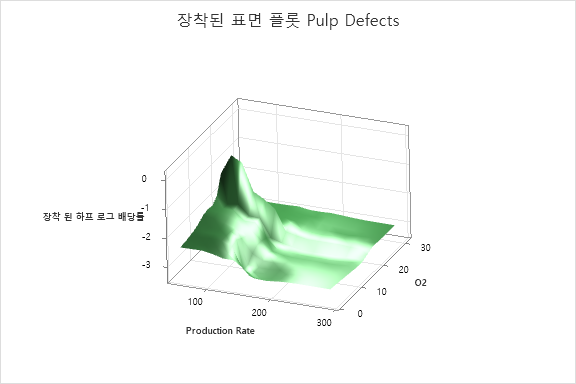


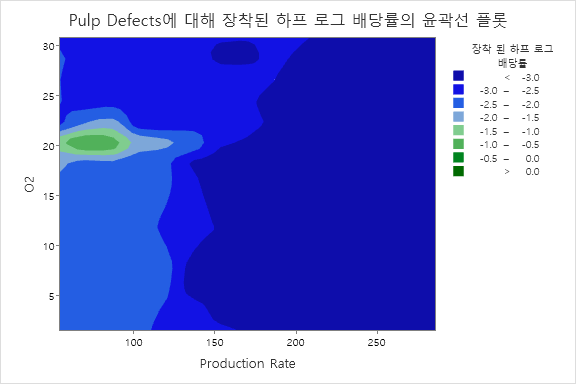






**두 개의 예측 변수 부분 종속성 플롯**





**AUTO\_ML**

minitap에서 지원하는 분류 알고리즘 4개(TreeNet, RandomForest, CART, 로지스틱 회귀)를 한번에 적용하여 표시해 주며, 사전에 파라미터를 설정할 수도 있으며, 사후에 파라미터를 조정할 수 있다.

자동적으로 가장 좋은 모델에 대한 분석 내용을 표시한다.

**결과 해석**

|  |  |  |  |
| --- | --- | --- | --- |
| **유형 내에서 최고의 모형** | **평균 음수 로그 우도** | **ROC 곡선 아래 면적** | **분류 율 잘못** |
| TreeNet®\* | 0.0445 | 0.9566 | 0.0126 |
| CART® | 0.0453 | 0.9715 | 0.0402 |
| 로지스틱 회귀 | 0.1050 | 0.9024 | 0.0333 |
| Random Forests® | 0.2202 | 0.9492 | 0.0137 |

\* 최소 평균 -로그 우도가 있는 모든 모형 유형에서 최고의 모형. 최고의 모형 출력이 뒤에 나옵니다.

기본 설정의 모형 중 평균 음수 로그 우도가 가장 낮은 TreeNet 모델이 성능이 가장 좋은 것을 확인 할 수 있다. CART 모형과 모형 성능의 차이가 별로 나지 않는 것을 확인할 수 있다.

성능이 가장 낮은 것은 RandomForest 모델인 것을 확인할 수 있다.