**펄프 표백 공정 불량율 예측**

PAPER\_BLEACHING

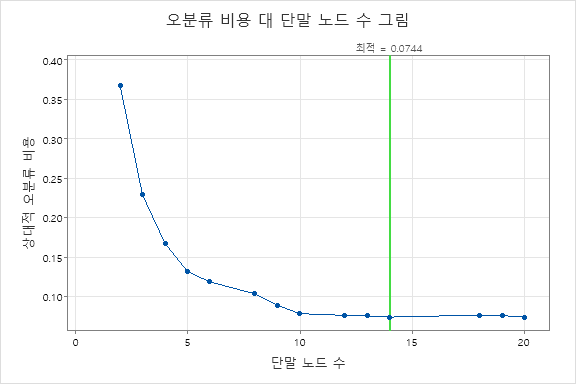
**14 노드 CART® 분류: Pulp Defects 대 Production Rate, Caustic, MgSO4, H2O2, O2, Conductivity, Discharge pH, Bleaching**

**방법**

|  |  |
| --- | --- |
| 사전 확률 | 모든 등급에 대해 같음 |
| 노드 분할 | 지니 |
| 최적 트리 | 최소 오분류 비용 |
| 모형 검증 | 10-접기 교차 검증 |
| 사용된 행 | 2906 |

**이항 반응 정보**

|  |  |  |  |
| --- | --- | --- | --- |
| **변수** | **등급** | **카운트** | **%** |
| Pulp Defects | Fail (사건) | 131 | 4.51 |
|  | Pass | 2775 | 95.49 |
|  | 모두 | 2906 | 100.00 |



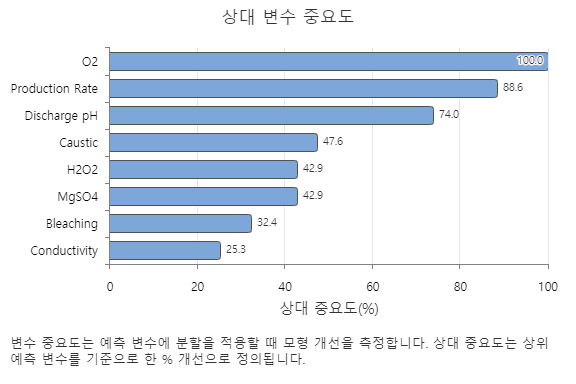
**최적 트리 다이어그램**

도표, 평면도, 지도, 기술 도면이(가) 표시된 사진

자동 생성된 설명

**모형 요약**

|  |  |
| --- | --- |
| 전체 예측 변수 | 8 |
| 중요 예측 변수 | 8 |
| 단말 노드 수 | 14 |
| 최소 단말 노드 크기 | 4 |
| **통계량** | **교육** | | **검정** |
| 평균 음수 로그 우도 | 0.0239 | | 0.1110 |
| ROC 곡선 아래 면적 | 0.9932 | | 0.9729 |
| 95% CI | (0.8669, 1) | | (0.9521, 0.9936) |
| 향상도 | 9.8585 | | 9.4415 |
| 오분류 비용 | 0.0300 | | 0.0744 |

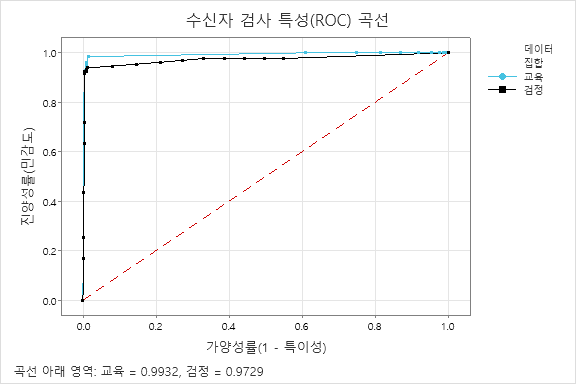


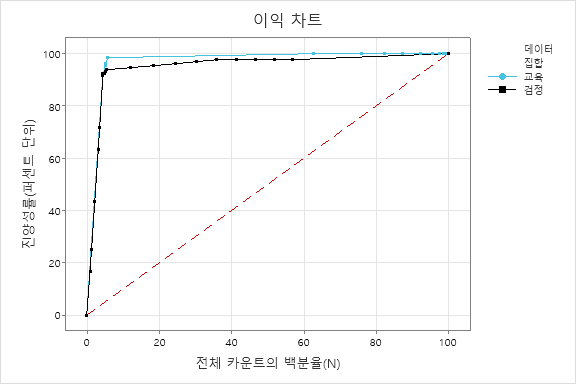
**오차 행렬**

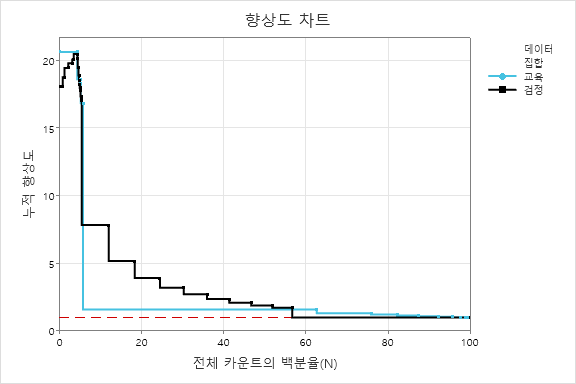
|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | **예측된 등급(교육)** | | | | | **예측된 등급(검정)** | | | |
| **실제 등급** | **카운트** | **Fail** | **Pass** | | **정답률(%)** | | **Fail** | | **Pass** | **정답률(%)** |
| Fail (사건) | 131 | 129 | 2 | | 98.5 | | 123 | | 8 | 93.9 |
| Pass | 2775 | 41 | 2734 | | 98.5 | | 37 | | 2738 | 98.7 |
| 모두 | 2906 | 170 | 2736 | | 98.5 | | 160 | | 2746 | 98.5 |
| **통계량** | | | | **교육(%)** | | **검정(%)** | |
| 진양성률(민감도 또는 검정력) | | | | 98.5 | | 93.9 | |
| 가양성률(유형 I 오차) | | | | 1.5 | | 1.3 | |
| 가음성률(유형 II 오차) | | | | 1.5 | | 6.1 | |
| 진음성률(특이성) | | | | 98.5 | | 98.7 | |

**오분류**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **입력 오분류 비용** | | **예측된 등급** | | | |
| 실제 등급 | | Fail | | Pass | |
| Fail | |  | | 1.00 | |
| Pass | | 1.00 | |  | |
|  |  | | **교육** | | | | | **검정** | | |
| **실제 등급** | **카운트** | | **오분류됨** | | **오차율(%)** | | **비용** | **오분류됨** | **오차율(%)** | **비용** |
| Fail (사건) | 131 | | 2 | | 1.5 | | 0.0153 | 8 | 6.1 | 0.0611 |
| Pass | 2775 | | 41 | | 1.5 | | 0.0148 | 37 | 1.3 | 0.0133 |
| 모두 | 2906 | | 43 | | 1.5 | | 0.0150 | 45 | 1.5 | 0.0372 |







**결과 해석**

검정데이터의 진양성률은 93.9%이고, 진음성률은 98.7% 이다. 데이터는 전체 131개의 불량품 중에 123개의 불량품을 예측하였으며, 2775개의 양품 중 2738개를 예측하였다. 가양성률(유형 1오차)는 1.3% 였으며, 양품 37개를 불량으로 잘못 분류하였으며, 가음성률(유형 2오차)는 6.1%로 불량품 8개를 양품으로 잘못 분류하였습니다.

ROC Curve는 왼쪽 상단에 가까울수록 ‘진양성률’은 1에 가깝고 ‘가양성률’은 0에 가까운 것을 의미하며 위의 모델은 상당히 양호한 것을 볼 수 있습니다.

가장 중요한 변수는 O2(산소)이며, Discharge PH의 경우 O2 대비 74% 정도 중요한 역할을 하고 있다고 해석할 수 있습니다.

**추가 분석**

**최악 5 단말 노드의 분류 효과**

**교육, 등급 = Fail(사건)**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **단말 노드** | **사건 카운트** | **전체 카운트** | **N의 비율(%)** | **사건 확률** | **비사건 확률** | **승산** |
| 8 | 5 | 20 | 0.7 | 0.250 | 0.750 | 0.333 |
| 10 | 1 | 4 | 0.1 | 0.250 | 0.750 | 0.333 |
| 12 | 3 | 17 | 0.6 | 0.176 | 0.824 | 0.214 |
| 4 | 120 | 129 | 4.4 | 0.930 | 0.070 | 13.333 |

**교육, 등급 = Pass**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **단말 노드** | **사건 카운트** | **전체 카운트** | **N의 비율(%)** | **사건 확률** | **비사건 확률** | **승산** |
| 11 | 2 | 1653 | 56.9 | 0.001 | 0.999 | 0.001 |

**피실험자를 최악 5 단말 노드로 분류하는 기준**

**등급 = Fail(사건)**

|  |  |
| --- | --- |
| **단말 노드** | **기준** |
| 8 | Production Rate > 137.257, H2O2 <= 41.5506, Discharge pH <= 7.663 |
| 10 | Production Rate > 137.257, H2O2 > 41.5506, O2 > 23.6984, Discharge pH <= 7.663 |
| 12 | Production Rate > 137.257, MgSO4 <= 3.1223, 23.1836 < O2 <= 23.2695, Discharge pH > 7.663 |
| 4 | Production Rate <= 137.257, Caustic <= 70.3891, MgSO4 > 2.90722, H2O2 > 42.664, 19.4944 < O2 <= 21.9205, Discharge pH <= 8.538 |

**등급 = Pass**

|  |  |
| --- | --- |
| **단말 노드** | **기준** |
| 11 | Production Rate > 137.257, O2 <= 23.1836, Discharge pH > 7.663 |

**결과 해석**

전체 단말 노드 14개 중 최악의 노드를 분석한 결과 단말노드 4의 불량율이 93% 였으며 가장 최고의 양품이 나온 단말노드는 11이고 양품율은 99.9% 였습니다. 각각의 조건은 다음과 같습니다.

노드 4(불량품 공식) = Production Rate <= 137.257, Caustic <= 70.3891, MgSO4 > 2.90722, H2O2 > 42.664, 19.4944 < O2 <= 21.9205

노드 11(양품 공식) = Discharge pH <= 8.538 , Production Rate > 137.257, O2 <= 23.1836, Discharge pH > 7.663

**추가 신규 데이터 예측**

신규데이터를 모델에 투입하여 결과를 예측하였다. 신규 투입 데이터는 아래와 같다.

# Production and Chemical Data

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Id | Production Rate | Caustic | MgSO4 | H2O2 | O2 | Conductivity | Discharge pH | Bleaching |
| A | 171.312 | 55.3725 | 2.84409 | 41.512 | 18.9316 | 8120.85 | 8.104 | 41.517 |
| B | 203.382 | 68.4852 | 2.954 | 57.2778 | 24.1637 | 8068.31 | 7.814 | 56.584 |
| C | 118.383 | 70.0503 | 3.0945 | 56.82 | 22.773 | 4941.51 | 9.57 | 54.342 |
| D | 195.629 | 54.9626 | 0.46815 | 42.8486 | 19.6628 | 6454.66 | 9.285 | 41.508 |
| E | 137.166 | 57.6864 | 5.36821 | 43.0847 | 20.9022 | 9243.75 | 7.223 | 63.2 |

결과는 아래와 같다.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Id | TNodeID | TClass | PEvent | PNonEvent |
| A | 11 | Pass | 0.00121 | 0.99879 |
| B | 14 | Pass | 0.0 | 1.0 |
| C | 7 | Pass | 0.0 | 1.0 |
| D | 11 | Pass | 0.00121 | 0.99879 |
| E | 4 | Fail | 0.930233 | 0.06977 |

TNodeID 는 각 데이터의 결과가 속한 단말 노드를 의미하며, TClass는 합격 여부를 의미한다. PEvent 는 불량품 확률이며, PNonEvent 는 양품 확률이다. 현재 모델은 전체 불량품의 데이터가 매우 적기 때문에 불량율 판단에 있어서 매우 높은 기준을 선택하였다.