AUTO\_PARTS

**Random Forests® 분류: y\_faulty 대 fix\_time, a\_speed, b\_speed, separation, s\_separation, rate\_terms, mpa, load\_time, highpressure\_time**

**방법**

|  |  |
| --- | --- |
| 모형 검증 | OOB 데이터로 검증 |
| 부트스트랩 표본 수 | 300 |
| 표본 크기 | 학습 데이터 크기 21767과 동일 |
| 노드 분할을 위해 선택된 예측 변수 수 | 총 예측 변수 수의 제곱근 = 3 |
| 최소 내부 노드 크기 | 5 |
| 사용된 행 | 21767 |

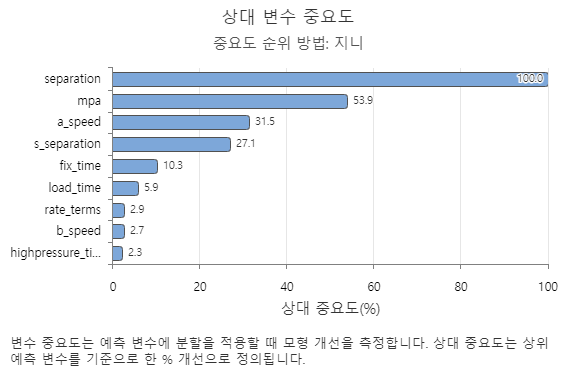
**이항 반응 정보**

|  |  |  |  |
| --- | --- | --- | --- |
| **변수** | **등급** | **카운트** | **%** |
| y\_faulty | 1 (사건) | 2842 | 13.06 |
|  | 0 | 18925 | 86.94 |
|  | 모두 | 21767 | 100.00 |



**모형 요약**

|  |  |
| --- | --- |
| 전체 예측 변수 | 9 |
| 중요 예측 변수 | 9 |
| **통계량** | | | **OOB** |
| 평균 음수 로그 우도 | | | 0.0985 |
| ROC 곡선 아래 면적 | | | 0.9916 |
| 95% CI | | | (0.9905, 0.9928) |
| 향상도 | | | 7.2825 |
| 오분류 비율 | | | 0.0333 |

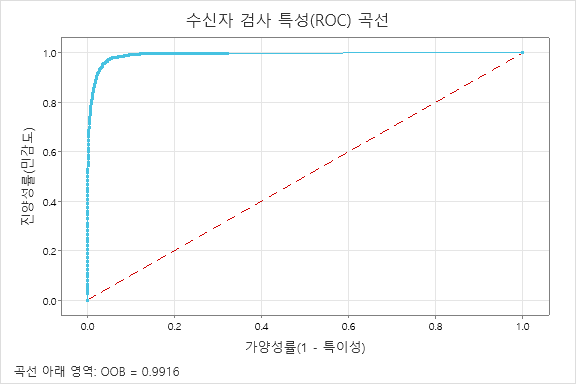


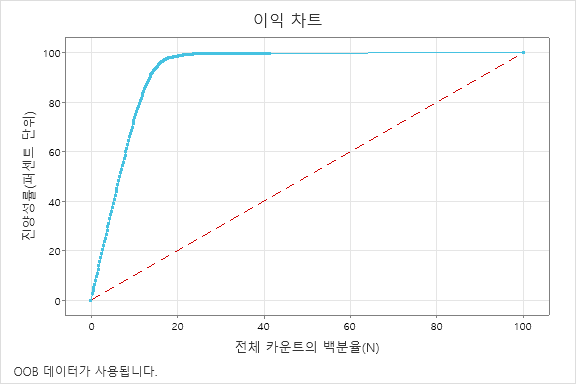
**오차 행렬**

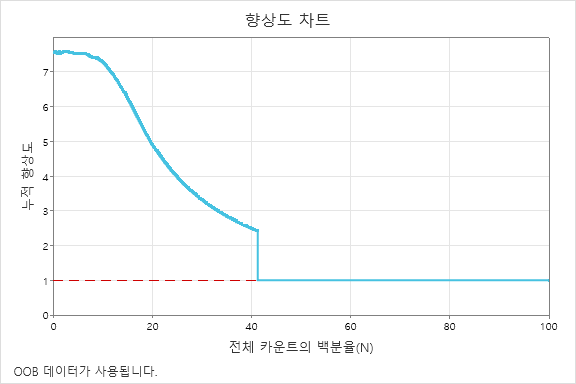
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **예측 등급(OOB)** | | | | | |
| **실제 등급** | **카운트** | **1** | **0** | | **정답률(%)** | |
| 1 (사건) | 2842 | 2337 | 505 | | 82.23 | |
| 0 | 18925 | 220 | 18705 | | 98.84 | |
| 모두 | 21767 | 2557 | 19210 | | 96.67 | |
| **통계량** | | | | **OOB(%)** | |
| 진양성률(민감도 또는 검정력) | | | | 82.23 | |
| 가양성률(유형 I 오차) | | | | 1.16 | |
| 가음성률(유형 II 오차) | | | | 17.77 | |
| 진음성률(특이성) | | | | 98.84 | |

**오분류**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **OOB** | | |
| **실제 등급** | **카운트** | **오분류됨** | **오차율(%)** |
| 1 (사건) | 2842 | 505 | 17.77 |
| 0 | 18925 | 220 | 1.16 |
| 모두 | 21767 | 725 | 3.33 |







AUTO\_PARTS

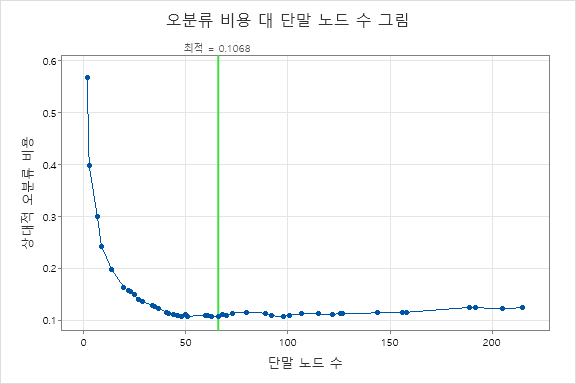
**66 노드 CART® 분류: y\_faulty 대 fix\_time, a\_speed, b\_speed, separation, s\_separation, rate\_terms, mpa, load\_time, highpressure\_time**

**방법**

|  |  |
| --- | --- |
| 사전 확률 | 모든 등급에 대해 같음 |
| 노드 분할 | 지니 |
| 최적 트리 | 최소 오분류 비용 |
| 모형 검증 | 70/30% 교육/검정 집합 |
| 사용된 행 | 21767 |

**이항 반응 정보**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | **교육** | | **검정** | |
| **변수** | **등급** | **카운트** | **%** | **카운트** | **%** |
| y\_faulty | 1 (사건) | 1976 | 12.93 | 866 | 13.35 |
|  | 0 | 13303 | 87.07 | 5622 | 86.65 |
|  | 모두 | 15279 | 100.00 | 6488 | 100.00 |



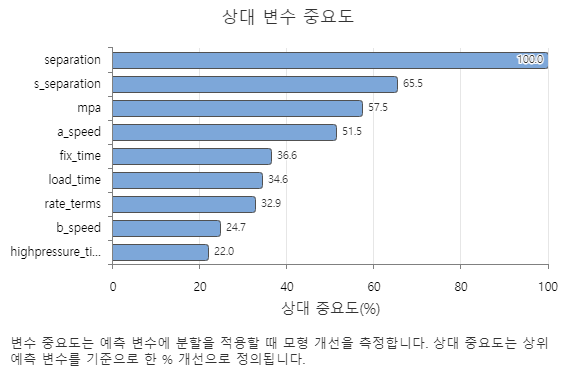
**최적 트리 다이어그램**

텍스트, 도표이(가) 표시된 사진

자동 생성된 설명

**모형 요약**

|  |  |
| --- | --- |
| 전체 예측 변수 | 9 |
| 중요 예측 변수 | 9 |
| 단말 노드 수 | 66 |
| 최소 단말 노드 크기 | 3 |
| **통계량** | **교육** | | **검정** |
| 평균 음수 로그 우도 | 0.1037 | | 0.2375 |
| ROC 곡선 아래 면적 | 0.9815 | | 0.9667 |
| 95% CI | (0.9099, 1) | | (0.8668, 1) |
| 향상도 | 6.5126 | | 6.1354 |
| 오분류 비용 | 0.0709 | | 0.1068 |

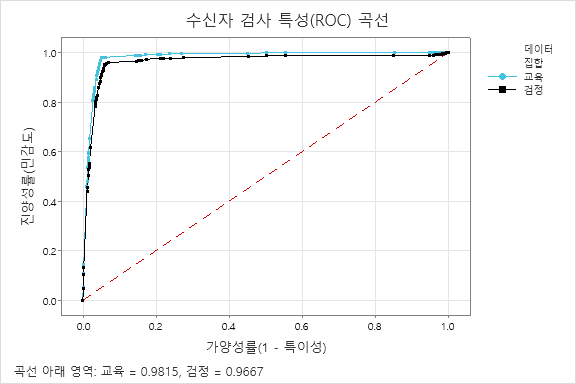


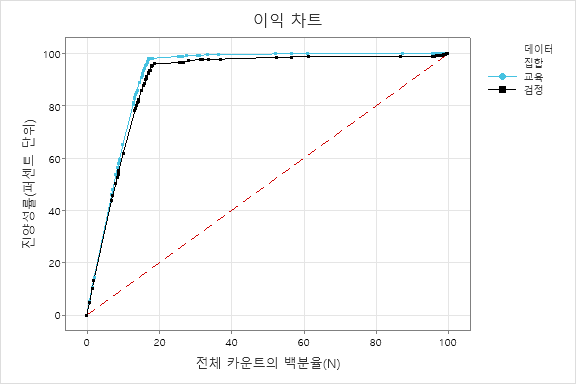
**오차 행렬**

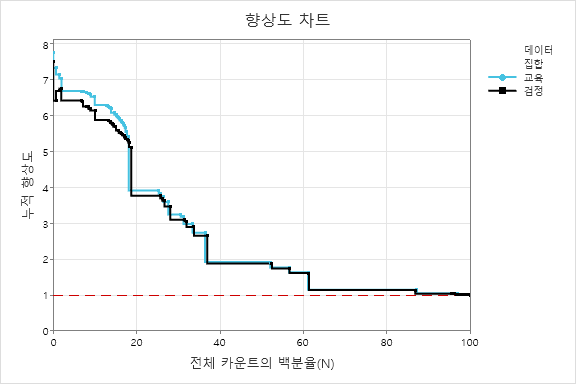
|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **예측된 등급(교육)** | | | | | | **예측된 등급(검정)** | | | | |
| **실제 등급** | **카운트** | **1** | **0** | | **정답률(%)** | | **카운트** | | **1** | **0** | **정답률(%)** |
| 1 (사건) | 1976 | 1938 | 38 | | 98.1 | | 866 | | 825 | 41 | 95.3 |
| 0 | 13303 | 687 | 12616 | | 94.8 | | 5622 | | 334 | 5288 | 94.1 |
| 모두 | 15279 | 2625 | 12654 | | 95.3 | | 6488 | | 1159 | 5329 | 94.2 |
| **통계량** | | | | **교육(%)** | | **검정(%)** | |
| 진양성률(민감도 또는 검정력) | | | | 98.1 | | 95.3 | |
| 가양성률(유형 I 오차) | | | | 5.2 | | 5.9 | |
| 가음성률(유형 II 오차) | | | | 1.9 | | 4.7 | |
| 진음성률(특이성) | | | | 94.8 | | 94.1 | |

**오분류**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **입력 오분류 비용** | | **예측된 등급** | | | |
| 실제 등급 | | 1 | | 0 | |
| 1 | |  | | 1.00 | |
| 0 | | 1.00 | |  | |
|  | **교육** | | | | | | | **검정** | | | |
| **실제 등급** | **카운트** | | **오분류됨** | | **오차율(%)** | | **비용** | **카운트** | **오분류됨** | **오차율(%)** | **비용** |
| 1 (사건) | 1976 | | 38 | | 1.9 | | 0.0192 | 866 | 41 | 4.7 | 0.0473 |
| 0 | 13303 | | 687 | | 5.2 | | 0.0516 | 5622 | 334 | 5.9 | 0.0594 |
| 모두 | 15279 | | 725 | | 4.7 | | 0.0354 | 6488 | 375 | 5.8 | 0.0534 |







**성능비교**

**랜덤포레스트**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **예측 등급(OOB)** | | | | | |
| **실제 등급** | **카운트** | **1** | **0** | | **정답률(%)** | |
| 1 (사건) | 2842 | 2337 | 505 | | 82.23 | |
| 0 | 18925 | 220 | 18705 | | 98.84 | |
| 모두 | 21767 | 2557 | 19210 | | 96.67 | |
| **통계량** | | | | **OOB(%)** | |
| 진양성률(민감도 또는 검정력) | | | | 82.23 | |
| 가양성률(유형 I 오차) | | | | 1.16 | |
| 가음성률(유형 II 오차) | | | | 17.77 | |
| 진음성률(특이성) | | | | 98.84 | |

**CART**

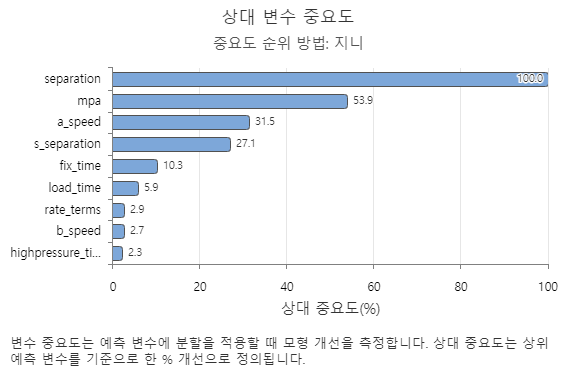
|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **예측된 등급(교육)** | | | | | | **예측된 등급(검정)** | | | | |
| **실제 등급** | **카운트** | **1** | **0** | | **정답률(%)** | | **카운트** | | **1** | **0** | **정답률(%)** |
| 1 (사건) | 1976 | 1938 | 38 | | 98.1 | | 866 | | 825 | 41 | 95.3 |
| 0 | 13303 | 687 | 12616 | | 94.8 | | 5622 | | 334 | 5288 | 94.1 |
| 모두 | 15279 | 2625 | 12654 | | 95.3 | | 6488 | | 1159 | 5329 | 94.2 |
| **통계량** | | | | **교육(%)** | | **검정(%)** | |
| 진양성률(민감도 또는 검정력) | | | | 98.1 | | 95.3 | |
| 가양성률(유형 I 오차) | | | | 5.2 | | 5.9 | |
| 가음성률(유형 II 오차) | | | | 1.9 | | 4.7 | |
| 진음성률(특이성) | | | | 94.8 | | 94.1 | |

**결과 해석**

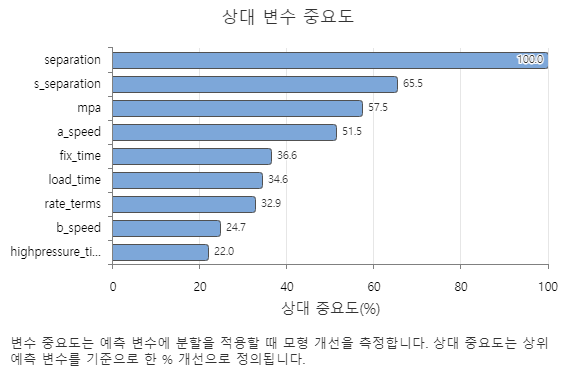
전체 모델의 정확도는 96.67%인 랜덤 포레스트가 CART 의 94.2% 보다 높지만, 진양성률은 95.3% 로 CART 가 랜덤포레스트(82.23%) 보다 더 높으며, 가음성률 또한 4.7%로 CART가 랜덤포레스트(17.77%) 보다 더 좋은 성능을 보이고 있다. 반면에 랜덤포레스트는 가양성률과 진음성률에서 CART보다 더 높은 성능을 보이고 있다. 따라서 분석의 목적과 유형에 따라 모델을 선택하는 것이 올바르다고 생각되어진다.

**변수 중요도 비교**

- 랜덤포레스트



- CART



**상대 변수 중요도**

• 랜덤포레스트분류 : Separation(100.0) > Mpa(53.9) > A\_speed(31.5)

• CART 분류 : Separation(100.0) > S\_separation(65.5) > Mpa(57.5

변수 중요도를 해석을 하는 기준은 연구자가 판단하는 것으로 전체적인 상황을 고려하여 모델과의 적합성을 고려하여 판단한다.

**예측 분석**

새로운 데이터는 다음과 같다.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| fix\_time | a\_speed | b\_speed | separation | s\_separation | rate\_terms | mpa | load\_time | highpressure\_time | 실제 y |
| 61.8 | 0.471 | 1.696 | 248.5 | 657.1 | 96 | 28.3 | 0 | 65 | 1 |
| 86.7 | 0.612 | 1.71 | 242.5 | 657.2 | 96 | 77.9 | 18.1 | 73 | 0 |
| 85.9 | 0.607 | 1.706 | 246.2 | 657.2 | 96 | 78.1 | 18.1 | 56 | 0 |
| 86.3 | 0.606 | 1.718 | 246.3 | 657 | 96 | 78.2 | 18.1 | 59 | 0 |
| 87.3 | 0.612 | 1.712 | 247.2 | 657.4 | 95 | 78 | 18.1 | 54 | 1 |

랜덤포레스트로 분류한 결과는 다음과 같다.

|  |  |  |
| --- | --- | --- |
| Class | PEvent | PNonEvent |
| 1 | 0.903333 | 0.09667 |
| 0 | 0.003333 | 0.99667 |
| 0 | 0.003333 | 0.99667 |
| 0 | 0 | 1 |
| 1 | 0.676667 | 0.32333 |

CART와 차이점은 관련 변수들이 어느 노드에 속하는지 알 수 없다는 점이다.

\*\*\*RandomForest : blackbox 모형