**인공지능 학습 결과서**

**1 . Logistic Regression**

* **선정 이유**

해석 용이성: 모델의 계수를 통해 각 피처가 종속 변수에 미치는 영향을 직관적으로 이해할 수 있음.

범주형 데이터 처리: 범주형 데이터를 적절히 인코딩하여 사용하면, 상관 관계가 적은 데이터에서도 기본적인 성능을 평가할 수 있음.

* **최적 파라미터** : {'C': 0.01, 'fit\_intercept': True, 'penalty': 'l2', 'solver': 'sag'}
* **성과 지표**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| train\_accuracy | test\_accuracy | AUC | Precision | Recall | F1 Score |
| 0.8150 | 0.8133 | 0.9034 | 0.8231 | 0.8100 | 0.8165 |

텍스트, 라인, 그래프, 스크린샷이(가) 표시된 사진

자동 생성된 설명

**2. RandomForest**

* **선정 이유**

비선형성과 복잡성 처리: Random Forest는 여러 개의 의사결정 나무를 앙상블하여 예측을 수행하기 때문에 비선형성과 복잡한 데이터 구조를 잘 처리할 수 있음.

범주형 데이터 처리: Random Forest는 범주형 데이터를 원-핫 인코딩 같은 전처리 과정을 거쳐 효과적으로 처리할 수 있음.

안정성과 해석력: 다수의 트리를 사용하여 안정적인 예측 결과를 제공하며, 각 피처의 중요도를 해석할 수 있는 장점이 있음.

* **최적 파라미터** : {'n\_estimators': 400, 'min\_samples\_split': 10, 'min\_samples\_leaf': 2, 'max\_features': 'auto', 'max\_depth': 30}
* **성과 지표**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| train\_accuracy | test\_accuracy | AUC | Precision | Recall | F1 Score |
| 0.9665 | 0.8455 | 0.9350 | 0.8467 | 0.8534 | 0.8500 |

텍스트, 라인, 도표, 그래프이(가) 표시된 사진

자동 생성된 설명

**3. XGB**

* **선정 이유**

강력한 성능: XGBoost는 부스팅 알고리즘을 사용하여 매우 강력한 성능을 발휘하는 모델로, 특히 데이터의 상관 관계가 낮을 때도 좋은 성능을 보임.

범주형 데이터 처리: 원-핫 인코딩을 통해 범주형 데이터를 효과적으로 처리할 수 있음.

* **최적 파라미터** : {'subsample': 1.0, 'n\_estimators': 300, 'min\_child\_weight', 'max\_depth': 3, 'learning\_rate': 0.2, 'gamma': 0.1, 'colsample\_bytree': 1.0}
* **성과 지표**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| train\_accuracy | test\_accuracy | AUC | Precision | Recall | F1 Score |
| 0.8301 | 0.8300 | 0.9416 | 0.8616 | 0.8371 | 0.8492 |

텍스트, 라인, 그래프, 스크린샷이(가) 표시된 사진

자동 생성된 설명

**4. CatBoost**

* **선정 이유**

범주형 데이터 자동 처리: CatBoost는 범주형 데이터를 자동으로 처리할 수 있는 특수한 알고리즘을 제공하여, 원-핫 인코딩 없이도 범주형 데이터를 효과적으로 다룰 수 있음.

높은 성능: 상관 관계가 낮은 데이터에서도 강력한 성능을 발휘하며, 과적합을 방지하는 다양한 기법이 포함되어 있음.

사용 편의성: 데이터 전처리 과정이 간소화되어 사용이 매우 편리함.

* **최적 파라미터** : {'depth': 4, 'grow\_policy': 'Depthwise', 'learning\_rate': 0.0860, 'min\_child\_samples': 21, 'n\_estimators': 261}
* **성과지표**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| train\_accuracy | test\_accuracy | AUC | Precision | Recall | F1 Score |
| 0.8588 | 0.8565 | 0.9466 | 0.8584 | 0.8625 | 0.865 |

텍스트, 라인, 그래프, 스크린샷이(가) 표시된 사진

자동 생성된 설명

**5. 신경망**

* **선정 이유**

자동적인 특성 학습: 신경망은 입력 데이터로부터 특성을 자동으로 학습할 수 있음. 범주형 변수의 경우 원-핫 인코딩 등의 방법으로 변환하더라도 신경망은 이러한 범주형 특성의 패턴을 스스로 인식하고 활용할 수 있음.

다중 레벨 추상화: 신경망은 여러 은닉층을 통해 데이터의 다양한 추상적인 특성을 학습할 수 있음. 이는 복잡한 데이터셋에서 중요한 특성을 추출하고 분류하는 데 유리함.

대규모 데이터 처리: 대규모의 데이터셋에서도 효과적으로 작동할 수 있는 능력이 있음. 많은 양의 범주형 변수가 포함된 데이터도 신경망은 처리할 수 있으며, 이를 통해 정확도를 높일 수 있음.

* **성과지표**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| train\_accuracy | test\_accuracy | AUC | Precision | Recall | F1 Score |
| 0.8301 | 0.8300 | 0.9163 | 0.8326 | 0.8369 | 0.8347 |

라인, 텍스트, 그래프, 도표이(가) 표시된 사진

자동 생성된 설명