|  |  |  |
| --- | --- | --- |
| **| 학습주제** | | * Python에서 랜덤 포레스트 분류기를 사용한 교통 혼잡 여부 예측 |
| **| 학습목표** | | * Python에서 랜덤 포레스트 분류기를 학습하고 평가하는 방법을 이해한다. * 시간대별 교통량 데이터를 기반으로 교통 혼잡 여부를 예측하는 과정을 학습한다. * train\_test\_split 함수를 사용하여 데이터를 훈련 세트와 테스트 세트로 분리하고, RandomForestClassifier로 모델을 학습하는 방법을 학습한다. * classification\_report를 사용하여 모델 성능을 평가하는 방법을 학습한다. |
| **| 학습 개념** | 랜덤 포레스트(Random Forest)는 여러 개의 의사결정 나무(Decision Tree)를 모아 앙상블 학습을 수행하는 방법입니다. 각 의사결정 나무는 데이터를 학습한 후 독립적으로 예측을 수행하며, 그 결과들을 모아 최종 예측을 도출합니다. 이 모델은 일반적으로 과적합을 방지하고, 높은 예측 성능을 제공합니다.  이 코드에서는 시간대별 교통량 데이터를 독립 변수로 사용하고, 교통 혼잡 여부를 종속 변수로 설정하여 예측 모델을 학습합니다. 데이터는 훈련 세트와 테스트 세트로 분리되어 모델을 학습시키고, 테스트 셋에서 선택된 샘플의 혼잡 여부를 예측합니다. | |
| **| 학습 방향** | 이번 예제에서는 시간대별 교통량 데이터를 기반으로 혼잡 여부를 예측하는 과정을 다룹니다. 이를 위해 먼저 데이터를 불러와 독립 변수(시간대별 교통량)와 종속 변수(혼잡 여부)를 설정한 후, train\_test\_split 함수를 사용해 데이터를 훈련 데이터와 테스트 데이터로 나누고, 랜덤 포레스트 분류기를 통해 모델을 학습합니다. 이후, 테스트 셋에서 임의로 선택된 샘플에 대해 예측을 수행하고, 실제 혼잡 여부와 비교하여 모델 성능을 확인합니다. | |
| **| 요구 사항** | 1. 데이터 준비 및 전처리  - 데이터를 불러와 시간대별 교통량 데이터를 독립 변수로 설정하고, 혼잡 여부를 종속 변수로 설정하세요.  2. 데이터 분리  - train\_test\_split 함수를 사용하여 데이터를 훈련 세트(70%)와 테스트 세트(30%)로 분리하세요. stratify=y를 사용하여 종속 변수 비율을 유지하세요.  3. 모델 생성 및 학습  - RandomForestClassifier 객체를 생성하고, 훈련 데이터로 모델을 학습하세요.  4. 예측 및 비교  - 테스트 데이터에서 한 샘플을 선택한 후, 그 샘플에 대해 모델이 예측한 혼잡 여부와 실제 혼잡 여부를 비교하여 모델 성능을 확인하세요. | |