



유추 파이프라인 만들기

8분

모델을 학습하기 위한 파이프라인을 만들고 실행한 후에는 새 데이터에 대해 동일한 데이터 변환을 수행한 다음 학습된 모델을 사용하여 해당 특징에 따라 레이블 값을 '유추'(즉, 예측)하는 두 번째 파이프라인이 필요합니다. 이를 통해 애플리케이션에서 사용하도록 게시할 수 있는 예측 서비스의 기반이 형성됩니다.

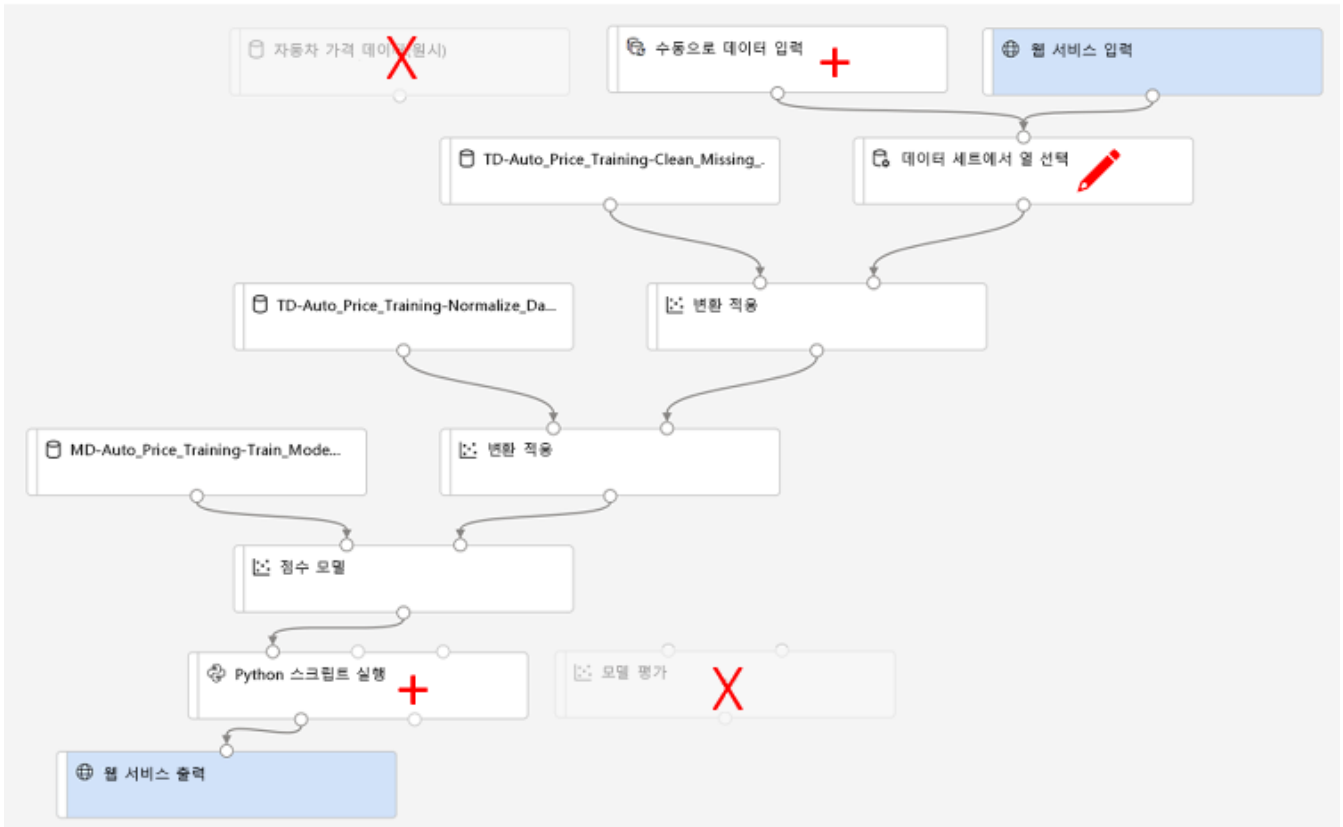
유추 파이프라인 만들기 및 실행

1. Azure Machine Learning Studio에서 **디자이너** 페이지를 클릭하여 만든 모든 파이프라인을 확인합니다. 그런 다음 이전에 만든 **자동차 가격 학습** 파이프라인을 엽니다.
2. **유추 파이프라인 만들기** 드롭다운 목록에서 **실시간 유추 파이프라인** 을 클릭합니다. 몇 초 후에 **자동차 가격 학습-실시간 유추** 라는 파이프라인의 새 버전이 열립니다.

파이프라인에 **웹 서비스 입력** 과 **웹 서비스 출력** 모듈이 포함되지 않은 경우 **디자이너** 페이지로 돌아가 **자동차 가격 학습-실시간 유추** 파이프라인을 다시 엽니다.

3. 새 파이프라인의 이름을 **자동차 가격 예측** 으로 바꾸고 새 파이프라인을 검토합니다. 여기에는 제출할 새 데이터에 대한 웹 서비스 입력과 결과를 반환하기 위한 웹 서비스 출력이 포함되어 있습니다. 학습 데이터의 통계가 새 데이터 값을 정규화하는 데 사용되고 학습된 모델이 새 데이터의 점수를 매기는 데 사용될 수 있도록 일부 변환 및 학습 단계가 해당 파이프라인에 캡슐화되었습니다.

유추 파이프라인을 다음과 같이 변경합니다.



- 자동차 가격 데이터(원시) 데이터 세트를 레이블 열(가격)을 포함하지 않은 수동으로 데이터 입력 모듈로 바꿉니다.
- 데이터 세트에서 열 선택 모듈을 수정하여 (이제 존재하지 않는) 가격 열에 대한 참조를 제거합니다.
- 모델 평가 모듈을 제거합니다.
- 예측된 레이블만 반환하도록 웹 서비스 출력 앞에 Python 스크립트 실행 모듈을 삽입합니다.

파이프라인을 수정할 때 위 이미지와 정보를 참조로 사용하여 아래의 나머지 단계를 수행합니다.

- 유추 파이프라인은 새 데이터가 원래 학습 데이터의 스키마와 일치하는 것으로 가정하므로 학습 파이프라인의 **자동차 가격 데이터(원시)** 데이터 세트가 포함됩니다. 하지만 해당 입력 데이터에는 모델이 예측하는 **가격** 레이블이 포함되어 있으므로, 아직 가격 예측이 이루어지지 않은 새 자동차 데이터에 이를 포함시키는 것은 비논리적입니다. 해당 모듈을 삭제하고 **데이터 입력 및 출력** 섹션에 있는 **수동으로 데이터 입력** 모듈로 바꾸어 다음 CSV 데이터를 포함시킵니다. 여기에는 세 대의 자동차에 대해 레이블이 없는 특징 값이 포함됩니다(텍스트 블록 전체를 복사하여 붙여넣기).

| CSV | 복사 |
|--|----|
| symboling,normalized-losses,make,fuel-type,aspiration,num-of-doors,body-style,drive-wheels,engine-location,wheel-base,length,width,height,curb-weight,engine-type,num-of-cylinders,engine-size,fuel- | |

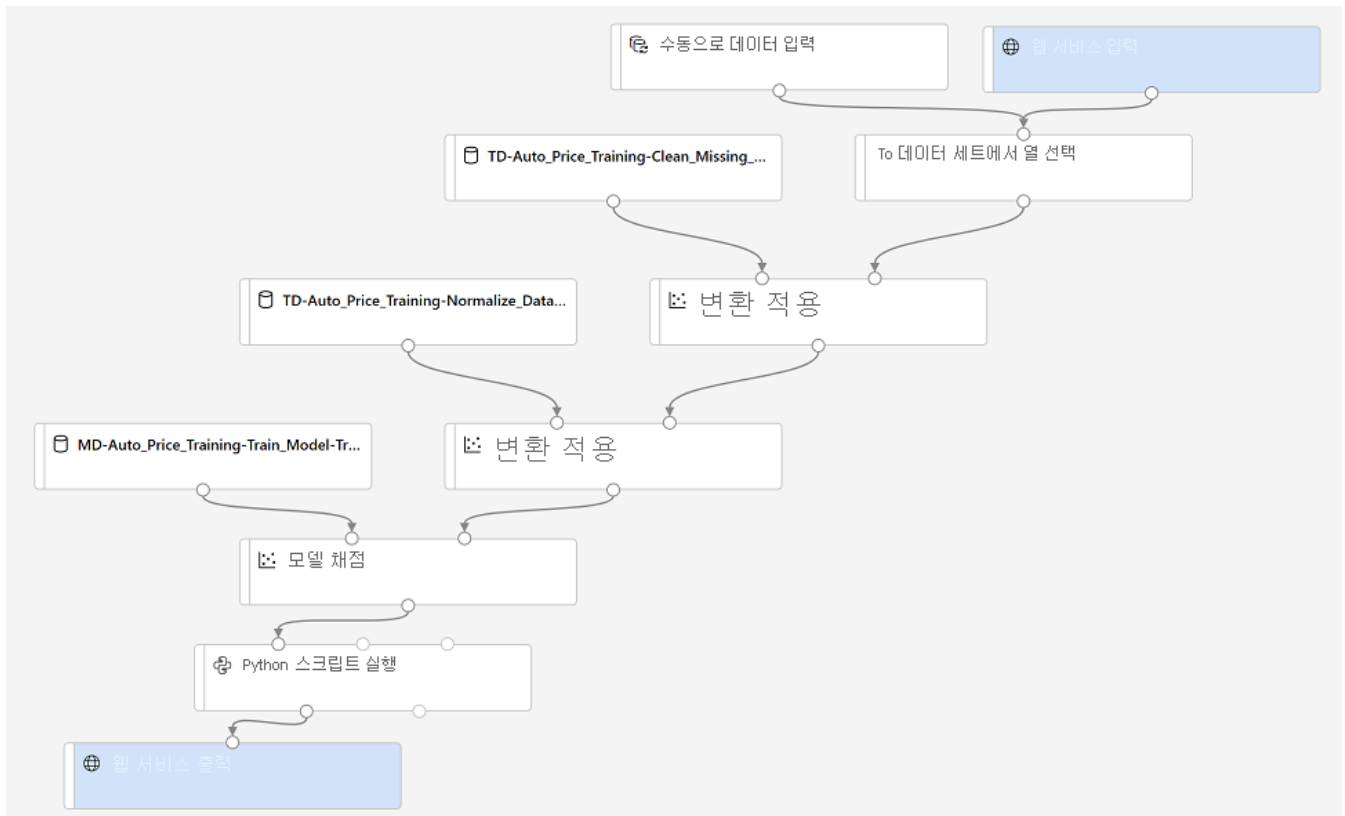
```
system,bore,stroke,compression-ratio,horsepower,peak-rpm,city-mpg,highway-mpg
3,NaN,alfa-romero,gas,std,two,convertible,rwd,front,88.6,168.8,64.1,48.8,2548,dohc,four,130,mpfi,3.47,2.68,9,111,5000,21,27
3,NaN,alfa-romero,gas,std,two,convertible,rwd,front,88.6,168.8,64.1,48.8,2548,dohc,four,130,mpfi,3.47,2.68,9,111,5000,21,27
1,NaN,alfa-romero,gas,std,two,hatchback,rwd,front,94.5,171.2,65.5,52.4,2823,ohcv,six,152,mpfi,2.68,3.47,9,154,5000,19,26
```

5. 새로운 수동으로 데이터 입력 모듈을 웹 서비스 입력 과 동일한 데이터 세트에서 열 선택 모듈의 데이터 세트 입력으로 연결합니다.
6. 들어오는 데이터의 스키마를 변경하여 가격 필드를 제외했으므로 남은 모듈에서 해당 필드를 사용할 수 없도록 제거해야 합니다. 데이터 세트에서 열 선택 모듈을 선택한 다음 설정 창에서 열을 수정하여 가격 필드를 제거합니다.
7. 유추 파이프라인에는 새 데이터로부터 예측하기에 유용하지 않은 모델 평가 모듈이 포함되어 있으므로 해당 모듈을 삭제합니다.
8. 모델 채점 모듈의 출력에는 모든 입력 특징뿐만 아니라 예측된 레이블도 포함됩니다. 예측만 포함하도록 출력을 수정하려면 다음을 수행합니다.

- 모델 채점 모듈과 웹 서비스 출력 간의 연결을 삭제합니다.
- Python 언어 섹션에서 Python 스크립트 실행 모듈을 추가하고 모든 기본 Python 스크립트를 다음 코드로 바꿉니다(점수가 매겨진 레이블 열만 선택하고 predicted_price 로 이름 변경).

| Python | 복사 |
|---|----|
| <pre>import pandas as pd def azureml_main(dataframe1 = None, dataframe2 = None): scored_results = dataframe1[['Scored Labels']] scored_results.rename(columns={'Scored Labels':'predicted_price'}, inplace=True) return scored_results</pre> | |

- 모델 채점 모듈의 출력을 Python 스크립트 실행 의 Dataset1(맨 왼쪽) 입력에 연결하고 Python 스크립트 실행 모듈의 출력을 웹 서비스 출력 에 연결합니다.
9. 파이프라인이 다음과 유사한지 확인합니다.



10. 컴퓨팅 클러스터에서 **mslearn-auto-inference** 라는 새 실험으로 파이프라인을 제출합니다. 이 작업은 다소 시간이 걸릴 수 있습니다.
11. 파이프라인이 완료되면 **Python 스크립트 실행** 모듈을 선택하고, 설정 창의 **출력 + 로그** 탭에서 **결과 데이터 세트** 를 시각화하여 입력 데이터의 3개 자동차에 대한 예측 가격을 확인합니다.
12. 시각화 창을 닫습니다.

유추 파이프라인은 해당 특징에 따라 자동차 가격을 예측합니다. 이제 클라이언트 애플리케이션에서 사용할 수 있도록 파이프라인을 게시할 준비가 되었습니다.

다음 단원: 예측 서비스 배포

계속 >