



유추 파이프라인 만들기

8분

모델을 학습하기 위해 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. 이 파이프라인을 통해 사용할 애플리케이션에 게시할 수 있는 예측 서비스의 기반이 형성됩니다.

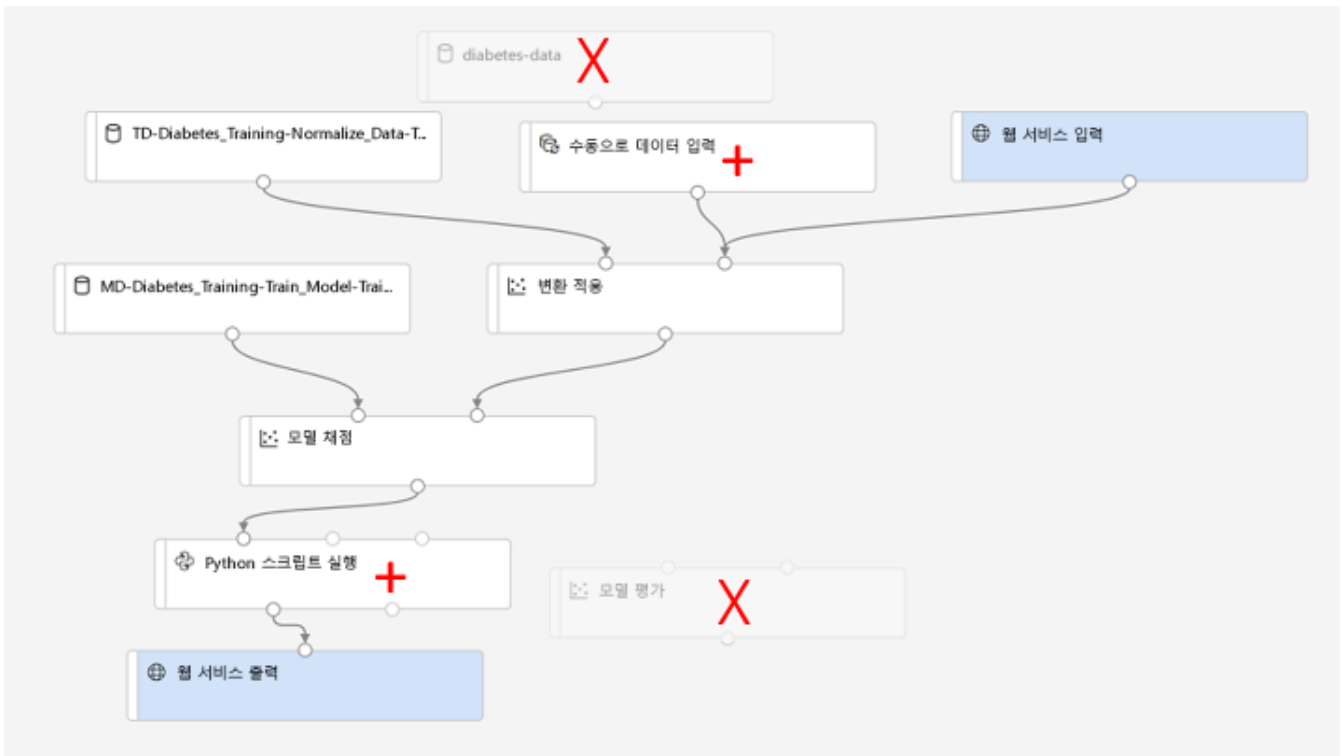
유추 파이프라인 만들기

1. Azure Machine Learning Studio에서 **디자이너** 페이지를 클릭하여 만든 모든 파이프라인을 표시합니다. 그런 다음 이전에 만든 **당뇨병 학습** 파이프라인을 엽니다.
2. **유추 파이프라인 만들기** 드롭다운 목록에서 **실시간 유추 파이프라인** 을 클릭합니다. 몇 초 후에 **당뇨병 학습-실시간 유추** 라는 새 버전의 파이프라인이 열립니다.

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

3. 새 파이프라인의 이름을 **당뇨병 예측** 으로 바꾸고 새 파이프라인을 검토합니다. 제출할 새 데이터에 대한 웹 서비스 입력이 포함되어 있고, 웹 서비스 출력이 포함되어 결과를 반환합니다. 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.



- **diabetes-data** 데이터 세트를 레이블 열(Diabetic)을 포함하지 않은 **수동으로 데이터 입력** 모듈로 바꿉니다.
- **모델 평가** 모듈을 제거합니다.
- 환자 ID, 예측된 레이블 값, 확률만 반환하도록 웹 서비스 출력 앞에 **Python 스크립트 실행** 모듈을 삽입합니다.

파이프라인을 수정할 때 아래의 단계를 따르고 위 이미지와 정보를 참조합니다.

- 유추 파이프라인은 새 데이터가 원래 학습 데이터의 스키마와 일치하는 것으로 가정하므로 학습 파이프라인의 **diabetes-data** 데이터 세트가 포함됩니다. 해당 입력 데이터에는 모델이 예측하는 **Diabetic** 레이블이 포함되어 있지만, 아직 당뇨병 예측이 이루어지지 않은 새 환자 데이터에 이를 포함시키는 것은 비논리적입니다. 해당 모듈을 삭제하고 **데이터 입력 및 출력** 섹션에서 **수동으로 데이터 입력** 모듈을 사용하여 세 명의 신규 환자 관찰에 대해 레이블을 제외한 특징 값이 포함된 다음 CSV 데이터가 포함되도록 바꿉니다(텍스트 블록 전체 복사 및 붙여넣기).

CSV	복사
PatientID,Pregnancies,PlasmaGlucose,DiastolicBloodPressure,TricepsThickness,SerumInsulin,BMI,DiabetesPedigree,Age 1882185,9,104,51,7,24,27.36983156,1.350472047,43 1662484,6,73,61,35,24,18.74367404,1.074147566,75 1228510,4,115,50,29,243,34.69215364,0.741159926,59	

- 새 **수동으로 데이터 입력** 모듈을 **변환 적용** 모듈의 동일한 **데이터 세트 입력**에 **웹 서비스 입력** 으로 연결합니다.

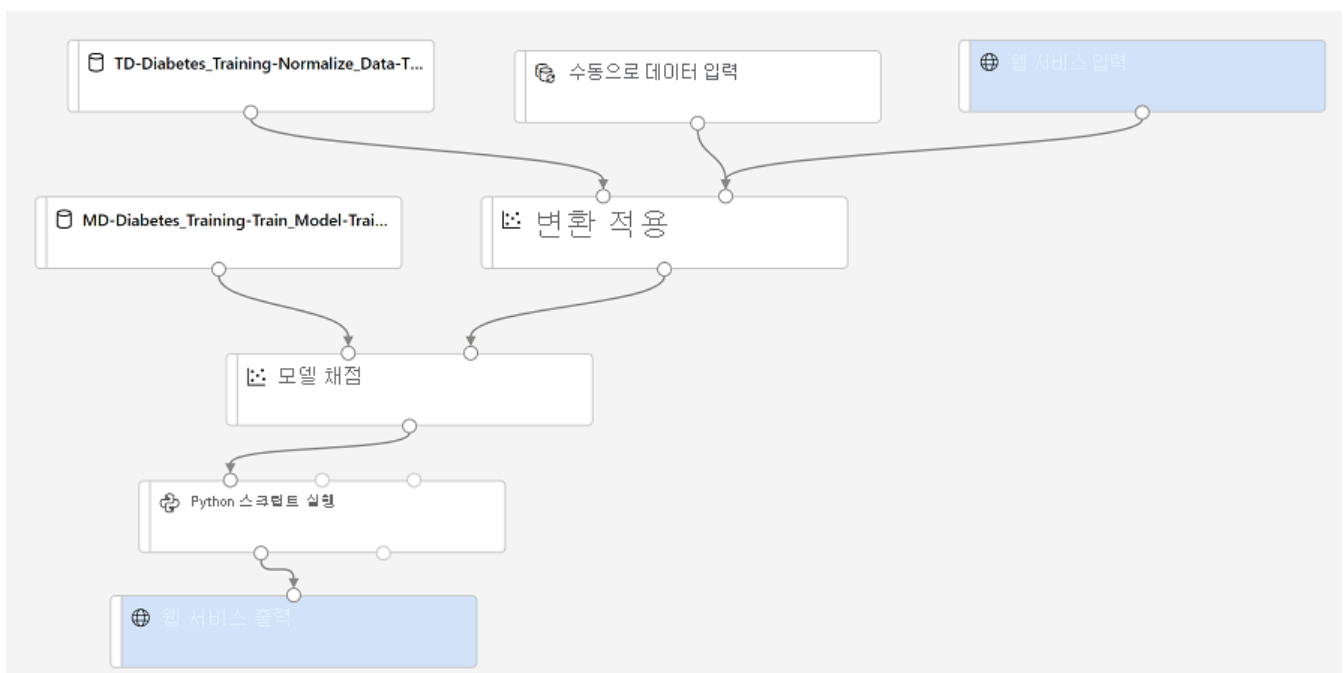
6. 유추 파이프라인에는 새 데이터를 예측할 때 유용하지 않은 **모델 평가** 모듈이 포함되어 있으므로 해당 모듈을 삭제합니다.
7. **모델 채점** 모듈의 출력에는 모든 입력 특징뿐만 아니라 예측된 레이블과 확률 점수도 포함됩니다. 출력을 예측과 확률로만 제한하려면 다음을 수행합니다.

- **모델 채점** 모듈과 **웹 서비스 출력** 간의 연결을 삭제합니다.
- **Python 언어** 섹션에서 **Python 스크립트 실행** 모듈을 추가하고 모든 기본 Python 스크립트를 다음 코드로 바꿉니다(**PatientID**, **Scored Labels**(점수가 매겨진 레이블), **Scored Probabilities**(점수가 매겨진 확률) 열만 선택하고 적절하게 이름 변경).

Python	복사
<pre>import pandas as pd def azureml_main(dataframe1 = None, dataframe2 = None): scored_results = dataframe1[['PatientID', 'Scored Labels', 'Scored Probabilities']] scored_results.rename(columns={'Scored Labels':'DiabetesPrediction', 'Scored Probabilities':'Probability'}, 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.

8. 파이프라인이 다음과 유사한지 확인합니다.



9. 컴퓨팅 클러스터에서 **mslearn-diabetes-inference** 라는 새 실험으로 파이프라인을 실행합니다. 다소 시간이 걸릴 수 있습니다.
10. 파이프라인이 완료되면 **Python 스크립트 실행** 모듈을 선택하고, 설정 창의 **출력 + 로그** 탭에서 **결과 데이터 세트** 를 시각화하여 입력 데이터의 세 환자 관찰값에 대한 예측 레이블과 확률을 확인합니다.

유추 파이프라인은 특징을 기준으로 환자에게 당뇨병 위험이 있는지 여부를 예측합니다. 이제 클라이언트 애플리케이션에서 사용할 수 있도록 파이프라인을 게시할 준비가 되었습니다.

다음 단원: 예측 서비스 배포

계속 >