



# 모델을 서비스로 배포

10분

자동화된 기계 학습을 사용하여 일부 모델을 학습시킨 후에는 클라이언트 애플리케이션에서 사용할 수 있는 최상의 모델을 서비스로 배포할 수 있습니다.


## 예측 서비스 배포


Azure Machine Learning에서 서비스를 ACI(Azure Container Instances) 또는 AKS(Azure Kubernetes Service) 클러스터로 배포할 수 있습니다. 프로덕션 시나리오의 경우 AKS를 배포하는 것이 좋습니다. 이 경우 유추 클러스터 컴퓨팅 대상을 만들어야 합니다. 이 연습에서는 테스트에 적합한 배포 대상인 ACI 서비스를 사용하며, 유추 클러스터를 만들 필요가 없습니다.

1. [Azure Machine Learning Studio](#) 의 **자동화된 ML** 페이지에서 자동화된 기계 학습 실험에 대해 실행을 선택하고 **세부 정보** 탭을 봅니다.
2. 최적 모델의 알고리즘 이름을 선택합니다. 그런 다음, **모델** 탭에서 **배포** 단추를 사용하여 다음 설정으로 모델을 배포합니다.
  - **이름:** predict-rentals
  - **설명:** 자전거 대여 예측
  - **컴퓨팅 형식:** ACI
  - **인증 사용:** 선택됨
3. 배포가 시작될 때까지 대기합니다. 이 작업에는 몇 초 정도 걸릴 수 있습니다. 그런 다음, **모델 요약** 섹션에서 **predict-rentals** 서비스에 대한 **배포 상태** 를 확인합니다. 이 상태는 **실행 중** 이어야 합니다. 이 상태가 **성공** 으로 변경될 때까지 기다립니다. 주기적으로 **새로 고침** 을 선택해야 할 수도 있습니다.
4. Azure Machine Learning Studio에서 **엔드포인트** 페이지를 보고 **predict-rentals** 실시간 엔드포인트를 선택합니다. 그런 다음, **사용** 탭을 선택하고 다음 정보를 확인합니다. 클라이언트 애플리케이션에서 배포된 서비스를 연결하려면 이 정보가 필요합니다.
  - 서비스에 대한 REST 엔드포인트
  - 서비스에 대한 기본 키
5. 이 값 옆에 있는 **땀 링크**를 사용하여 이를 클립보드에 복사할 수 있습니다.

## 배포된 서비스 테스트

이제 서비스를 배포했으므로 간단한 코드를 사용하여 테스트할 수 있습니다.

1. 브라우저에서 **predict-rentals** 서비스 페이지의 **사용** 페이지가 열린 상태에서 새 브라우저 탭을 열고 **Azure Machine Learning Studio** 의 두 번째 인스턴스를 엽니다. 그런 다음, 새 탭에서 **Notebooks** 페이지(작성자 아래)를 봅니다.
2. **Notebooks** 페이지의 **내 파일** 에서  단추를 사용하여 다음 설정으로 새 파일을 만듭니다.
  - **파일 위치:** Users/사용자 이름
  - **파일 이름:** Test-Bikes
  - **파일 형식:** Notebook
  - **이미 있는 경우 덮어쓰기:** 선택됨
3. 새 Notebook이 만들어지면 이전에 만든 컴퓨팅 인스턴스가 **컴퓨팅** 상자에서 선택되었는지, 그리고 상태가 **실행 중** 인지 확인합니다.
4. << 단추를 사용하여 파일 탐색기 창을 축소하고 **Test-Bikes.ipynb** Notebook 탭에 사용할 수 있는 공간을 확보합니다.
5. Notebook에서 만들어진 사각형 셀에 다음 코드를 붙여넣습니다.

Python	 복사
<pre> endpoint = 'YOUR_ENDPOINT' #Replace with your endpoint key = 'YOUR_KEY' #Replace with your key  import json import requests  #An array of features based on five-day weather forecast x = [[1,1,2022,1,0,6,0,2,0.344167,0.363625,0.805833,0.160446],       [2,1,2022,1,0,0,0,2,0.363478,0.353739,0.696087,0.248539],       [3,1,2022,1,0,1,1,1,0.196364,0.189405,0.437273,0.248309],       [4,1,2022,1,0,2,1,1,0.2,0.212122,0.590435,0.160296],       [5,1,2022,1,0,3,1,1,0.226957,0.22927,0.436957,0.1869]]  #Convert the array to JSON format input_json = json.dumps({"data": x})  #Set the content type and authentication for the request headers = {"Content-Type":"application/json",            "Authorization":"Bearer " + key}  #Send the request response = requests.post(endpoint, input_json, headers=headers)  #If we got a valid response, display the predictions if response.status_code == 200:     y = json.loads(response.json())     print("Predictions:")     for i in range(len(x)):         print (" Day: {}. Predicted rentals: {}".format(i+1, max(0, round(y["result"][i])))) </pre>	

```
else:  
    print(response)
```

### ❗ 참고

코드의 세부 정보에 대해 너무 걱정하지 마세요. 가상 날씨 예측 데이터를 사용하여 5일 간의 특징을 정의하고, 만든 **predict-rentals** 서비스를 사용하여 5일 간의 자전거 대여 수를 예측합니다.

6. **predict-rentals** 서비스의 **사용** 페이지가 포함된 브라우저 탭으로 전환한 다음 서비스에 대한 REST 엔드포인트를 복사합니다. Notebook을 포함하는 탭으로 돌아가 키를 코드에 복사하고 YOUR\_ENDPOINT를 대체합니다.
7. **predict-rentals** 서비스의 **사용** 페이지가 포함된 브라우저 탭으로 전환한 다음 서비스에 대한 기본 키를 복사합니다. Notebook을 포함하는 탭으로 돌아가 키를 코드에 복사하고 YOUR\_KEY를 대체합니다.
8. Notebook을 저장한 다음 셀 옆에 있는 ▶ 단추를 사용하여 코드를 실행합니다.
9. 5일 기간에 포함된 각 날의 예상 대여 수가 반환되는지 확인합니다.

## 다음 단원: 지식 점검

[계속 >](#)