



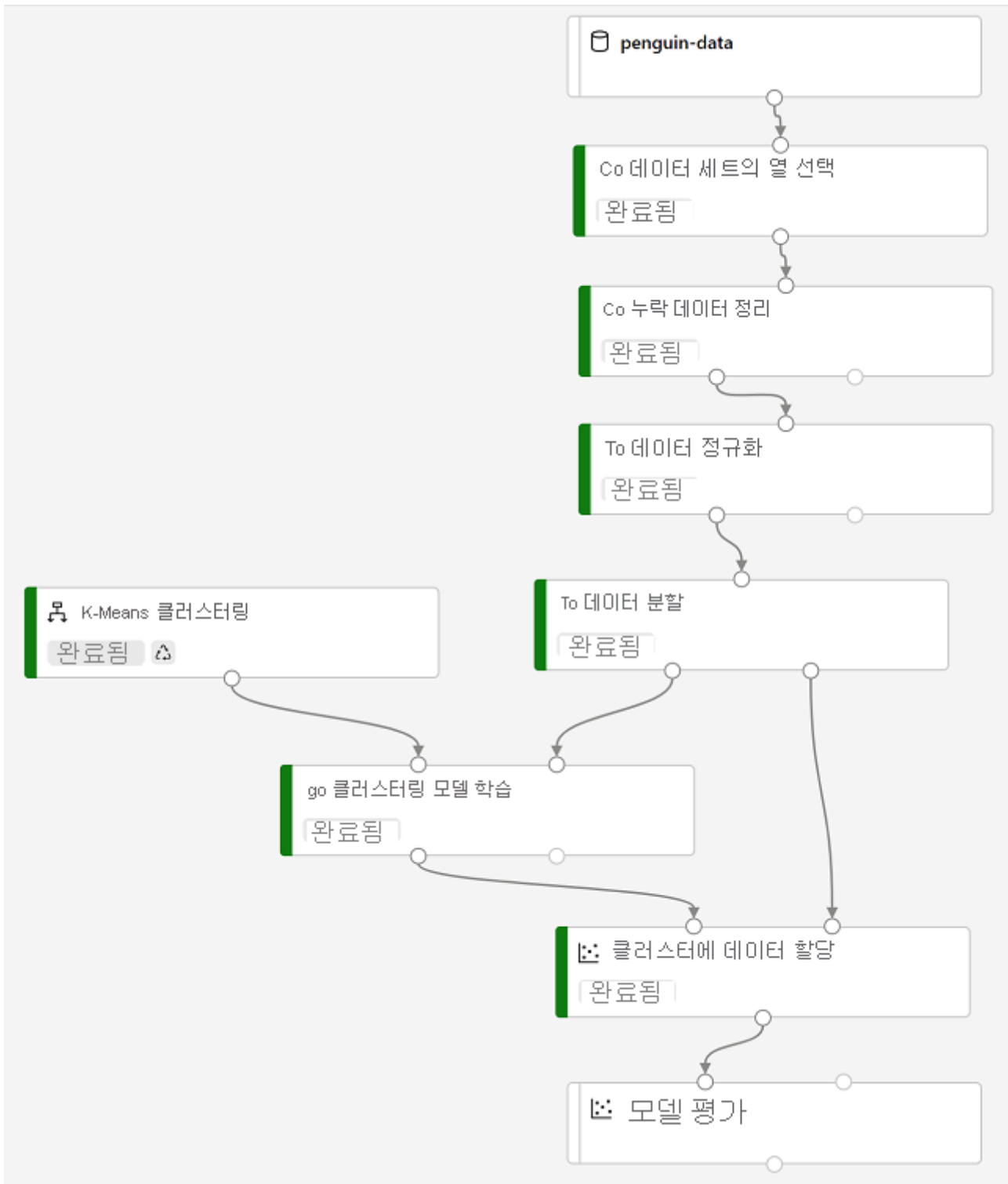
클러스터링 모델 평가

5분

클러스터링 모델을 평가하는 것은 클러스터 할당에 대해 이전에 알려진 *true* 값이 없으므로 어렵습니다. 성공적인 클러스터링 모델은 각 클러스터의 항목 간에 적절한 수준의 분리를 제공하는 모델이므로, 해당 분리를 측정하는 데 도움이 되는 메트릭이 필요합니다.

모델 평가 모듈 추가

1. 아직 열려 있지 않은 경우 이전 단원에서 만든 **Train Penguin Clustering** 파이프라인을 엽니다.
2. 왼쪽 창의 **모델 채점 및 평가** 섹션에서 **모델 평가** 모듈을 **클러스터에 데이터 할당** 모듈에 있는 캔버스로 끌어다 놓고, **클러스터에 데이터 할당** 모듈의 출력을 **모델 평가** 모듈의 **점수가 매겨진 데이터 세트**(왼쪽) 입력에 연결합니다.
3. 파이프라인이 다음과 같아야 합니다.



4. **제출** 을 선택하고, 기존 **mslearn-penguin-training** 실험을 사용하여 파이프라인을 실행합니다.
5. 실험이 완료될 때까지 기다립니다.
6. 실험이 완료되면 **모델 평가** 모듈을 선택하고 설정 창에서 **출력 + 로그** 탭의 **평가 결과** 섹션에 있는 **데이터 출력** 에서 **시각화** 아이콘을 사용하여 성능 메트릭을 봅니다. 이러한 메트릭은 데이터 과학자가 모델이 클러스터를 얼마나 잘 분리하는지 평가하는 데 도움이 될 수 있습니다. 각 클러스터에 대한 메트릭 행과 결합된 평가의 요약 행을 포함합니다. 각 행의 메트릭은 다음과 같습니다.

- **다른 중심까지의 평균 거리:** 클러스터의 각 지점과 다른 모든 클러스터의 중심이 평균적으로 얼마나 가까운지 나타냅니다.
- **클러스터 중심까지의 평균 거리:** 클러스터의 각 지점과 클러스터의 중심이 평균적으로 얼마나 가까운지 나타냅니다.
- **지점 수:** 클러스터에 할당된 지점의 수입니다.
- **클러스터 중심까지의 최대 거리:** 각 지점과 해당 지점의 클러스터 중심 간 최대 거리입니다. 이 수치가 높으면 클러스터가 광범위하게 분산되었을 수 있습니다. **클러스터 중심까지의 평균 거리**와 함께 이 통계를 활용하면 클러스터의 분산을 파악할 수 있습니다.

7. 모델 평가 결과 시각화 창을 닫습니다.

이제 작동하는 클러스터링 모델이 있으니 이를 사용하여 새 데이터를 추론 파이프라인에 있는 클러스터에 할당할 수 있습니다.

다음 단원: 유추 파이프라인 만들기

계속 >