



데이터솔루션



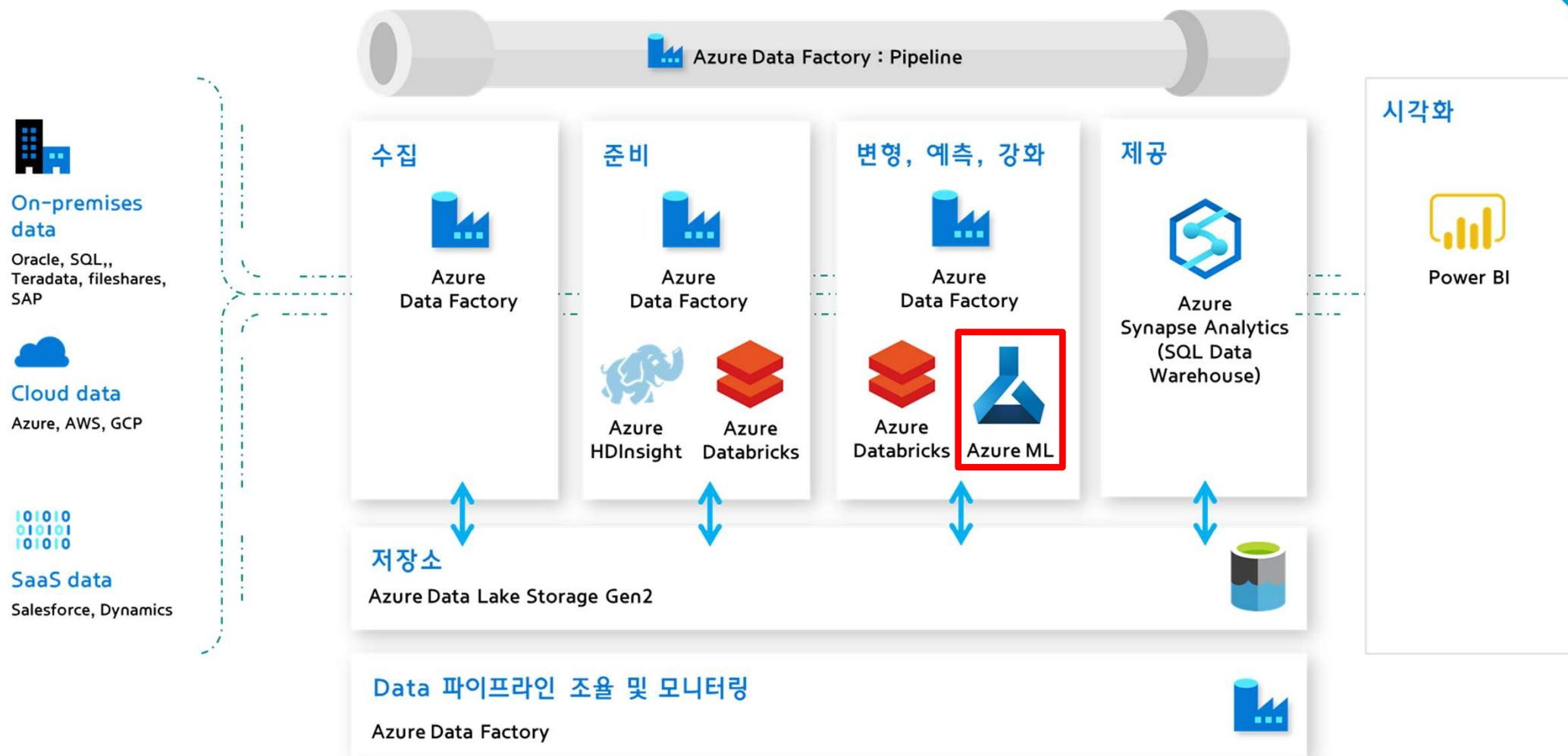
Microsoft

ASO Azure ML

01 | Azure Big Data 분석 요소



01 | Azure Big Data 분석 요소



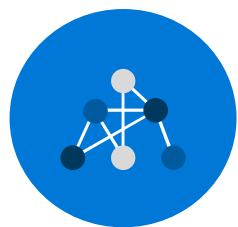
02 | Azure AI

AI apps & agents



Azure Bot Service
Azure Cognitive Services

Machine learning



Azure Machine Learning

Knowledge mining



Azure Cognitive Search

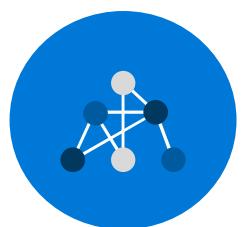
02 | Azure AI

AI apps & agents



Azure Bot Service
Azure Cognitive Services

Machine learning



Azure Machine Learning

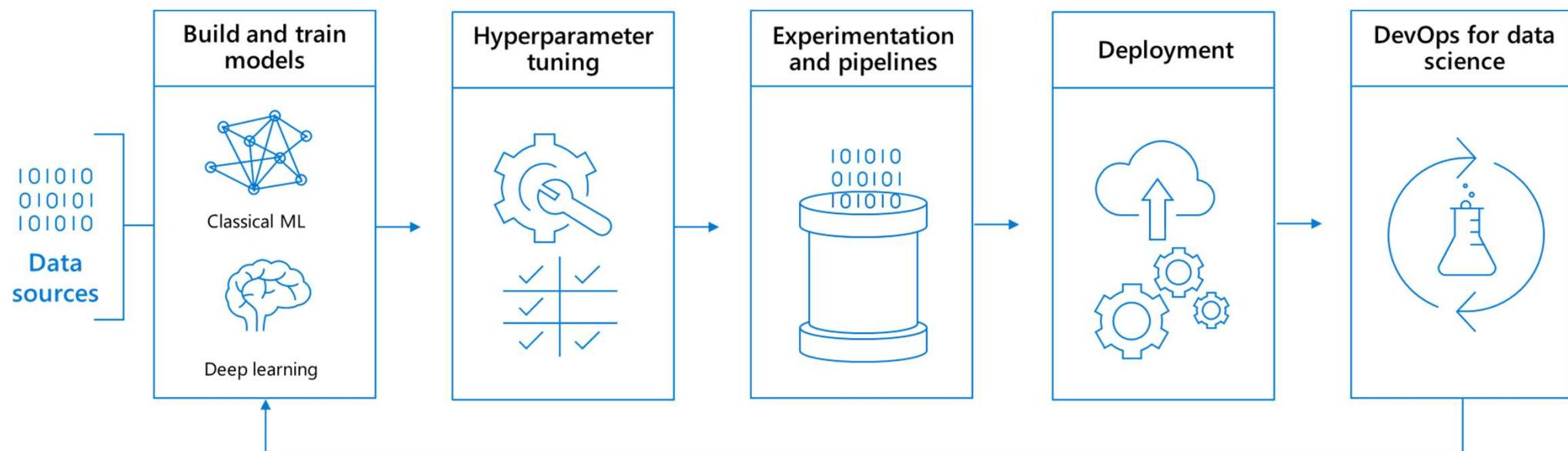
- 빠른 모델 구축 및 배포
- 엔터프라이즈급 머신러닝 서비스 제공
- 머신러닝을 위한 DevOps로 업계 최고의 MLOps를 사용하여 출시 시간을 가속화

Knowledge mining



Azure Cognitive Search

03 | Machine Learning Process



✓ 머싱러닝,
딥러닝 모델
구축

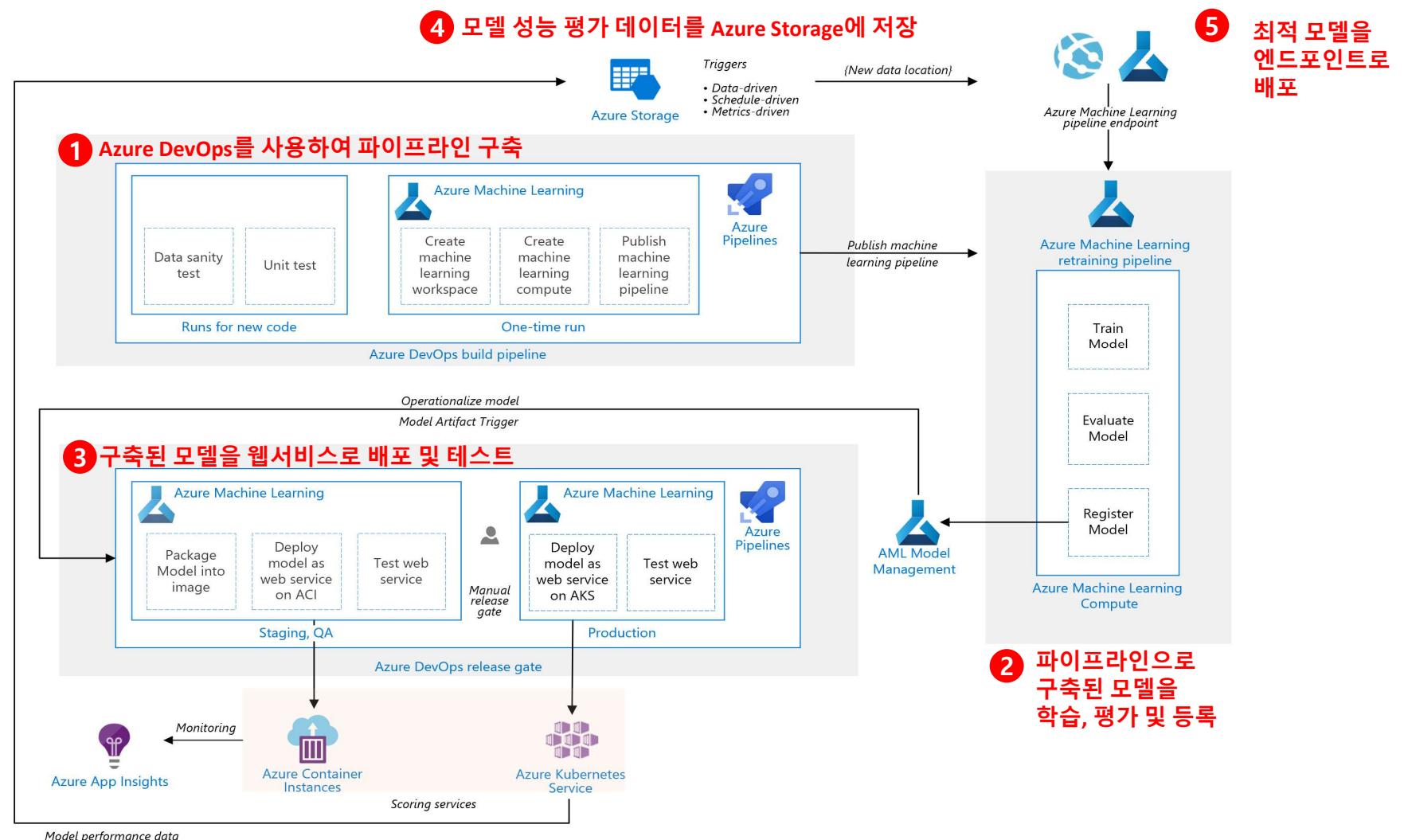
✓ 하이퍼
파라미터
튜닝 및
최적화

✓ 머싱 러닝
파이프라인
을 사용하여
반복 가능한
워크플로 및
실험 구축

✓ 실험(분석)
결과를 운영
시스템에
배포

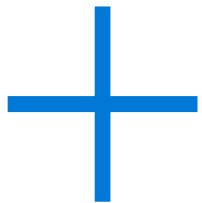
✓ 기존 DevOps
프로세스를
사용하여 ML
모델 운영

04 | Azure Machine Learning Architecture



05 | Azure Machine Learning Service

**Set of Azure
ML Services**



Python
SDK

R
SDK

JAVA
SDK
...

That enables you to:

- ✓ Prepare Data
- ✓ Build Models
- ✓ Train Models
- ✓ Manage Models
- ✓ Track Experiments
- ✓ Deploy Models

06 | Azure Machine Learning Service

End To End, 확장 가능하고 신뢰할 수 있는 플랫폼으로 모든 사람에게 AI 제공



데이터 과학 생산성 향상



실험 속도 향상



손 쉽게 모델 배포 및 관리



고객 니즈 기반 구축 가능

- 자동화된 머신 러닝
- 컴퓨팅 관리
- 간단한 배포
- 머신 러닝을 위한 DevOps
- 오픈 소스 프레임워크 지원
- Python, R SDK 활용

2

Chapter

Azure Machine Learning Automated ML

01 | Introduction of Automated ML

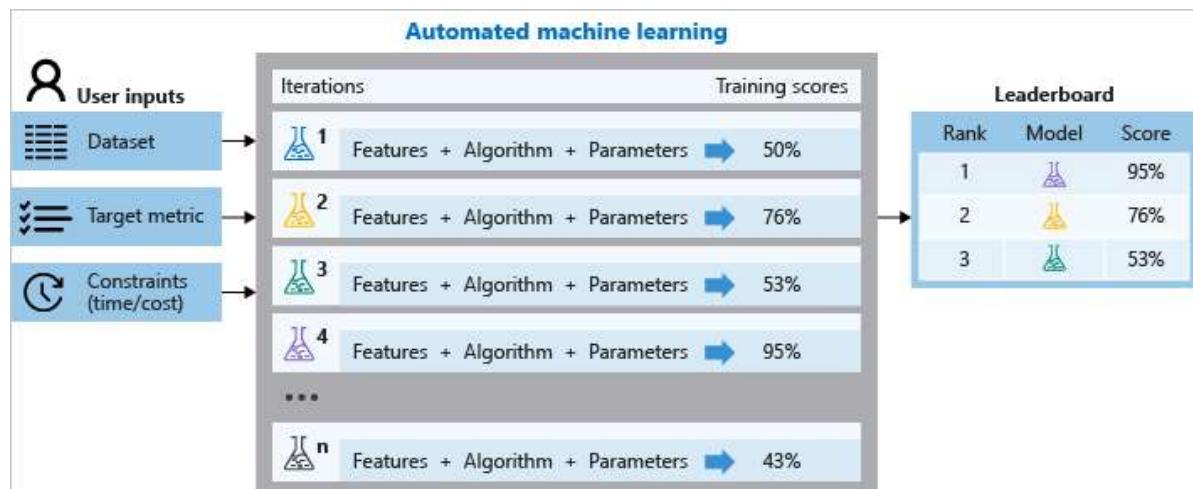
Automated Machine Learning 이란?

- 머신 러닝 모델 개발의 시간이 많이 걸리는 반복적인 작업을 자동화 하는 프로세스
- 모델 품질을 유지하면서 확장성, 효율성 및 생산성이 높은 ML 모델을 구축할 수 있음

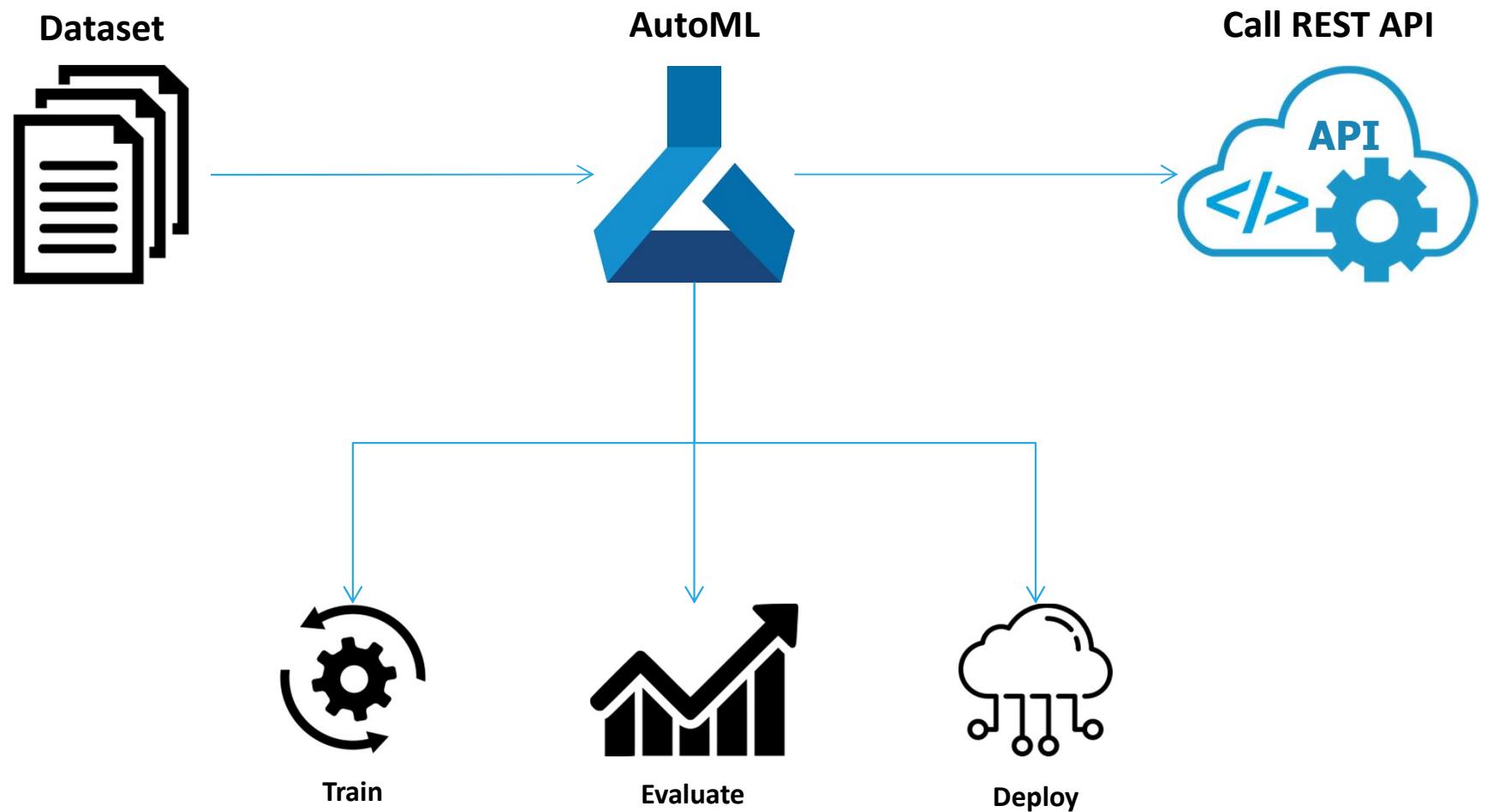
데이터 과학자, 분석가 및 개발자 관점에서 다음 수행할 수 있음

- 광범위한 프로그래밍 지식 없이 머신 러닝 솔루션 구현
- 시간 및 리소스 절약
- 데이터 과학 모범 사례 활용
- Agile 문제 해결 제공

Automated ML 예시



02 | How AutoML work?



03 | 실습 데이터

Columns	Description
transactionID	Unique transaction Id
accountID	Unique account Id
transactionAmountUSD	Transaction amount in USD e.g., 12345.00
transactionAmount	Transaction amount in currency expressed in transactionCurrencyCode e.g., 12345.00
transactionCurrencyCode	Currency code of the transaction. 3 alphabet letters, e.g., USD
transactionDate	Date when transaction occurred. Typically in the time zone of the processor, Format: yyyyymmdd, e.g., 20000101
transactionTime	Time when transaction occurred. Typically in the time zone of processing end. Format: hhmmss, eg. 153059
localHour	The hour in local time. Value of 0-23
transactionIPaddress	Full IP Address for IPv4: 000.000.000.000
ipState	State of IP address originated from 2 alphabet letters
ipPostcode	Postal Code of IP address originated from
ipCountryCode	Country code of IP address originated from
isProxyIP	Whether the IP address is a proxy or not
browserLanguage	Similar to country code
paymentInstrumentType	Type of payments: C -- Credit Card D -- Debit Card P -- Paypal K -- Check H -- Cash O -- Other
cardType	Type of cards M -- Magnetic C -- Chip
paymentBillingPostalCode	payment billing postal code
paymentBillingState	payment billing state
paymentBillingCountryCode	payment billing country code
cvvVerifyResult	M-- CVV2 Match N-- CVV2 No Match P--Not Processed S--Issuer indicates that CVV2 data should be present on the card, but the merchant has indicated data is not present on the card U--Issuer has not certified for CVV2 or Issuer has not provided Visa with the CVV2 encryption keys Empty--Transaction failed because wrong CVV2 number was entered or no CVV2 number was entered
digitalItemCount	Number of digital items purchased. (e.g. music, ebook, software, etc, that can be directly downloaded online)
physicalItemCount	Number of physical items purchased (that needs to be shipped)
label	True/False for fraud transaction

온라인 거래 사기 데이터

- Woodgrove 은행에서 제공하는 온라인 거래 내역
- 사기 데이터: 1,151건
- 비사기 데이터: 198,849건
- Oversampling을 통해 비사기 데이터 1,151건으로 축소
- 총 2,302건의 온라인 거래 데이터를 활용

실습 목적

- 온라인 거래 데이터 기반으로 사기 데이터를 식별하기 위한 최적 분류 모델 구축
- Auto ML 실습이 원활하게 진행하기 위해 실습 데이터 및 스토리지 이미 생성됨

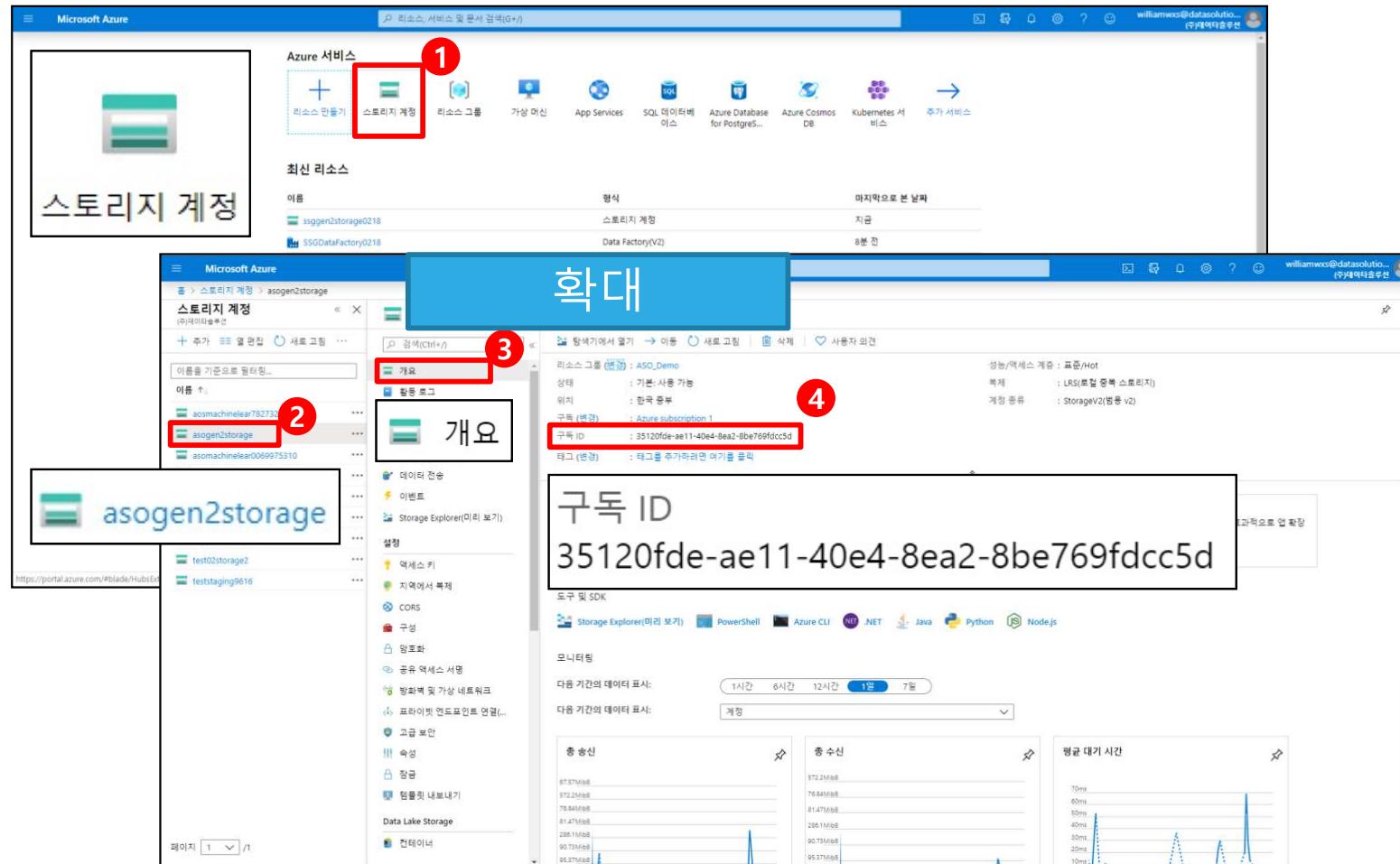
04 | MS Azure 로그인

The screenshot shows the Microsoft Azure homepage. At the top right, there is a red box highlighting the '로그인' (Login) button in the '내 계정' (My Account) dropdown menu. Below the header, a large banner features the text 'Azure 목표를 기반으로 개발하세요.' (Develop based on Azure goals) and '클라우드, 온-프레미스, 에지에서 원하는 도구와 프레임워크를 사용하여 애플리케이션을 빌드, 배포, 관리할 수 있는 100개 이상의 서비스를 통해 아이디어를 솔루션으로 전환하세요.' (Build, deploy, and manage applications using 100+ services with the tools and frameworks you need, from cloud, on-premises, and edge). A green button labeled '체험 계정 만들기 >' (Create a trial account) is visible. To the right of the banner, there is a callout box with the text 'Microsoft Learn을 통해 클라우드 기술을 빠르게 익히고 인증서를 받으세요. >' (Get certified quickly by learning cloud technologies through Microsoft Learn). The main content area below the banner is titled 'Azure를 통해 획기적인 아이디어를 솔루션으로 구현하기' (Implement revolutionary ideas into solutions using Azure) and features four sections: '미래를 위한 준비' (Preparing for the future), '원하는 조건에 따른 빌드' (Build according to requirements), '원활하게 하이브리드 운영' (Smooth hybrid operation), and '신뢰할 수 있는 클라우드' (Trusted cloud).

MS Azure 로그인 화면

- 계정 로그인 진행
- 로그인 진행한 후에 포털 클릭

05 | 계정 및 스토리지 정보 확인



계정 및 스토리지 정보 확인

1. Azure 포털 화면에서 스토리지 계정 클릭
2. 스토리지 asogen2storage 선택 (실습 데이터 스토리지)
3. <개요>탭을 클릭하여 리소스 그룹 정보 확인
4. 구독 ID 확인

06 | 계정 및 스토리지 정보 확인

The screenshot shows the Azure Storage account configuration page for 'asogen2storage'. The left sidebar lists various storage accounts. The 'Access Keys' tab is selected, highlighted by a red box and labeled '5'. In the main pane, under the 'key1' section, the key value is displayed as a long string of characters, also highlighted by a red box and labeled '6'. The URL of the page is visible at the top: /responsible/service-and-document-search(G+/-) asogen2storage - 액세스 키.

계정 및 스토리지 정보 확인

5. <액세스 키>탭을 클릭
6. 스토리지 계정 ID 확인 및 복사
7. 복사된 스토리지 키를 메모장에 붙여넣기

❖ 데이터 저장소를 만들 때
스토리지 키를 입력 필요

07 | 리소스 그룹 접속

The screenshot shows the Microsoft Azure portal interface. At the top, there's a navigation bar with various service icons like Storage, App Services, SQL Database, etc. Below it, a main menu bar has '리소스 그룹' (Resource Group) highlighted with a red box and a red number '1'. Underneath, a '최신 리소스' (Latest Resources) section lists 'SSG_Demo' and 'ssgen2storage0218'. On the left, a sidebar titled '리소스 그룹' shows a list of resource groups: 'ASOS-com', 'ASO_Databricks', 'ASO_Demo', 'ASO_ML' (which is selected and highlighted with a red box and red number '2'), and 'cloud-shell-storage-southeastasia'. In the center, a detailed view of the 'ASO_ML' resource group is shown. It includes sections for '구독 (Subscription)', '태그 (Tags)', '설정 (Settings)', and '모니터링 (Monitoring)'. A specific resource named 'ASO_MachineLearning' is highlighted with a red box and red number '3'.

리소스 그룹 접속

1. 리소스 그룹 클릭
2. 리소스 그룹 목록 확인
 - ASO_ML 선택
3. ASO_ML 목록 확인
 - ASO_MachineLearning 선택

08 | Azure ML Studio 접속



스튜디오를 시작합니다.

자동화된 ML ②

실행 번호	실험	업데이트된 시간	상태
35	ASO_AUTOML	19 Feb 2020 10:55	완료
4	Designer_test_v1	19 Feb 2020 10:53	완료
1	Designer_test_v1	19 Feb 2020 10:35	실패
36	ASO_AUTOML	19 Feb 2020 10:34	완료
1	ASO_AUTOML	19 Feb 2020 10:30	취소됨

모든 실행 보기 →

이름	유형	프로비전 상태	만든 날짜:
notebook-compute	컴퓨팅 인스턴스	성공	19 Feb 2020 10:40
compute-instance-ds3	컴퓨팅 인스턴스	성공	19 Feb 2020 10:34
ml-compute-ds13	Machine Learning 컴퓨...	크기 조정(노드 37) > 2	19 Feb 2020 10:27

모든 컴퓨팅 보기 →

자동화된 ML 실험을 만들고 템플릿을 사용해 편리하게 배포합니다.

AOS-MachineLearning 화면

1. Azure Machine Learning Studio 시작하기 클릭

Azure Machine Learning Studio 화면

2. Automated ML 선택

09 | Automated ML 접속

Automated ML 접속

- Automated ML 새로 만들기
- 수행했던 실험 확인할 수 있음
- 설명서
 - Automated ML 관련 컨셉 이해, 자습서, 블로그 등을 확인 할 수 있음

실험	실행 ID	상태	만든 날짜:
ASO_AUTOML	AutoML_cae6cdaa-a938-4d14-9b84-54c0d580f85d	완료	19 Feb 2020 10:31
ASO_AUTOML	AutoML_8c1ce8c2-5587-4d97-b395-000354fa06b7	취소됨	19 Feb 2020 10:13

10 | 데이터 세트 생성

Microsoft Azure Machine Learning

ASO_MachineLearning > 자동화된 ML > 실행 시작

새 자동화된 ML 실행 만들기

데이터 세트 선택

+ 데이터 세트 만들기

로컬 파일에서

데이터 저장소에서

웹 파일에서

열린 데이터 세트에서

데이터 세트 형식

만든 날짜:

modifiedTime

데이터 세트 형식	만든 날짜:	modifiedTime
표 형식	19 Feb 2020 10:17	19 Feb 2020 10:17
표 형식	19 Feb 2020 10:10	19 Feb 2020 10:10

+ 데이터 세트 만들기

로컬 파일에서

데이터 저장소에서

웹 파일에서

열린 데이터 세트에서

다음 >

뒤로

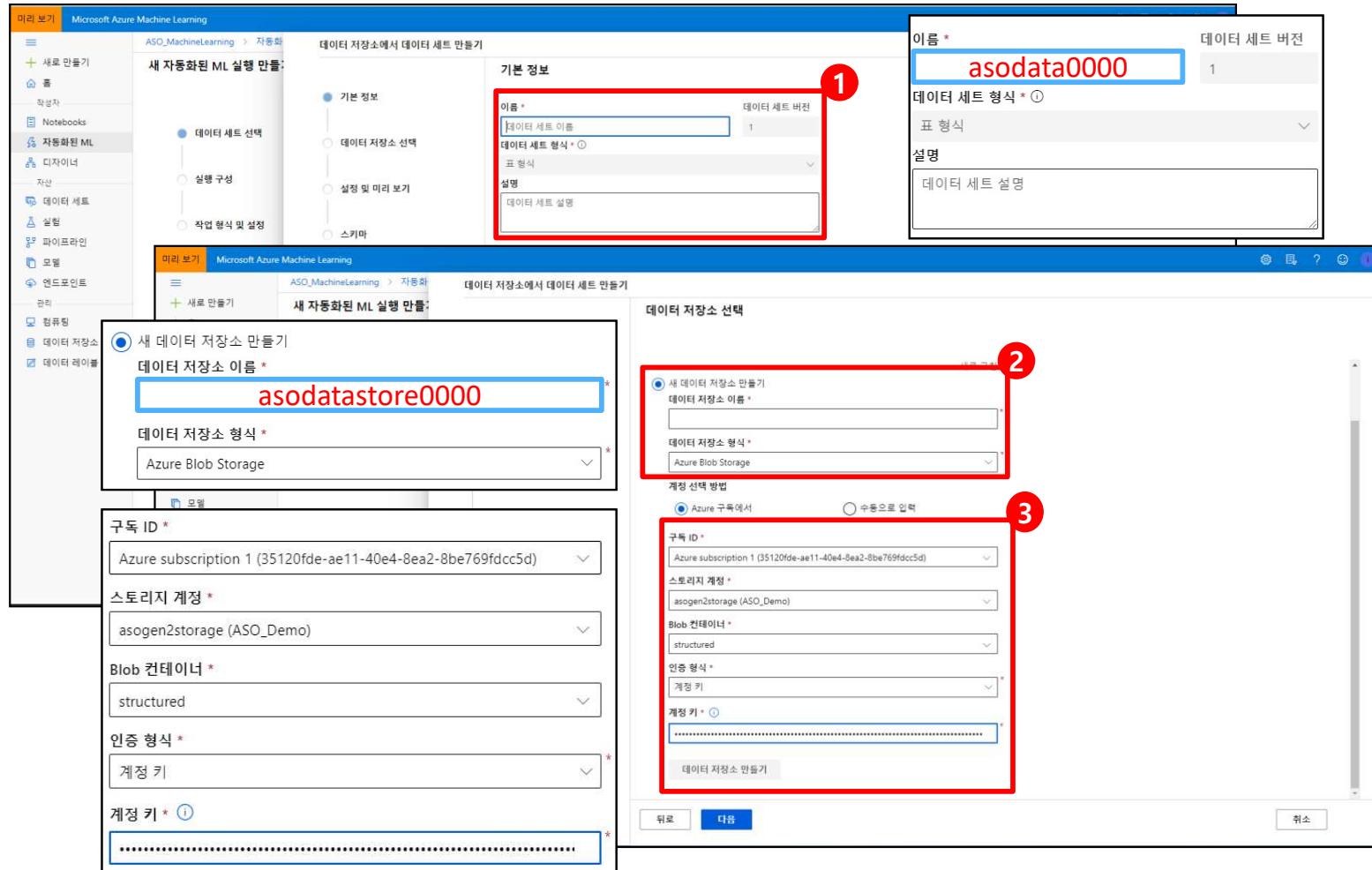
다음

취소

데이터 세트 만들기

- **로컬 파일**
 - 로컬환경에서 데이터 선택
- **데이터 저장소**
 - 기 구축된 저장소 활용
- **웹 파일**
 - Directory URL 연결
 - 온라인 데이터 수집
- **오픈 데이터 세트**
 - 무료 제공 데이터

11 | 실습 데이터 불러오기



데이터 세트 만들기

- 1. 데이터 기본 정보 작성**
 - 데이터 세트 이름
➤ **asodata0000**
 - 데이터 형식
 - 데이터 설명
- 2. 새 데이터 저장소 만들기**
 - 데이터 저장소 이름 입력
➤ **asodatastore0000**
 - 데이터 저장소 형식 선택
➤ **Azure Blob Storage**
- 3. Azure 계정 정보 입력**
 - 구독 ID
 - 스토리지 계정
➤ **asogen2storage**
 - Blob 컨테이너
➤ **structured**
 - 인증 형식
➤ 계정 키
 - 계정 키
➤ 복사된 스토리지 키 참고

12 | 실습 데이터 불러오기

The screenshot shows the 'Data from storage' step configuration in the Azure Machine Learning Studio. The 'Path' input field contains the value 'dbo.balanced_transactionstest.csv/*.*.csv'. A red box highlights the 'Find' button in this field, and the number '4' is placed next to it.

데이터 저장소에서 불러오기

- 파일 경로 찾아보기
 - 데이터 파일 경로 지정
 - csv 형식 파일만 불러오기

13 | 데이터 세트 설정

Microsoft Azure Machine Learning

ASQ_MachineLearning > 자동화

새 자동화된 ML 실행 만들기

데이터 저장소에서 데이터 세트 만들기

설정 및 미리 보기

파일 형식

구분 기호로 분리됨

구분 기호 예제

쉼표 Field1,Field2,Field3

인코딩

UTF-8

열 머리글

모든 파일의 머리글이 동일함

행 건너뛰기

없음

transactionID	accountID	transactionAmountUSD	transactionAmount	transactionCurrencyCode	transactionDate	transactionTime	local
706FC255-09E1-4	A91480087	130.00	130.00	USD	20130830	123302	7.00
						175654	-99.00
						165401	11.00
						192523	14.00
						110621	6.00
						224938	17.00
						40915	20.00
						23681	?? ??

데이터 세트 설정

1. 파일 형식 선택
 - 구분 기호로 분리됨
2. 구분 기호 선택
 - 쉼표
3. 인코딩
 - UTF-8
4. 열 머리글
 - 모든 파일의 머리글이 동일함
5. 행 건너뛰기
 - 없음
6. 미리보기
 - 데이터 맞게 들어오는지 확인

14 | 스키마

The screenshot shows the Azure Machine Learning Studio interface with the following details:

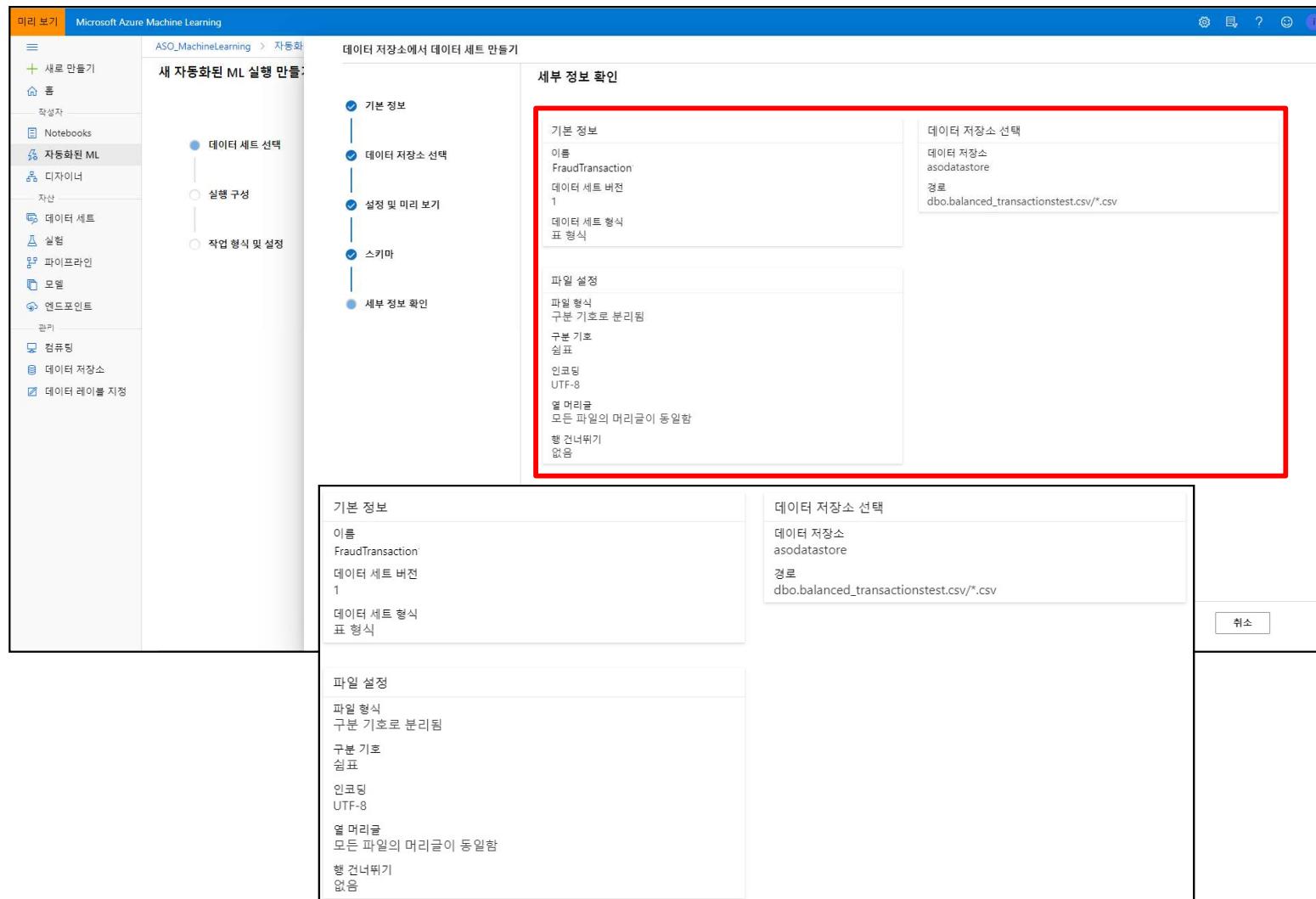
- Left Sidebar:** Includes '미리 보기' (Preview), '새로 만들기' (New), '품' (Items), '작성자' (Author), 'Notebooks', '자동화된 ML' (Automated ML), '디자이너' (Designer), '자산' (Assets), '데이터 세트' (Data Set), '실험' (Experiment), '파이프라인' (Pipeline), '모델' (Model), '연드포인트' (Endpoint), '관리' (Management), '컴퓨팅' (Computing), and '데이터 저장소' (Data Storage).
- Central Pipeline View:** Shows a pipeline titled '새 자동화된 ML 실행 만들기' (Create a new automated ML run). The '스키마' (Schema) step is currently selected.
- Schema Configuration Table:** This table maps columns from a CSV file to database tables. The columns are:

포함	열 이름	속성	유형	형식 설정 및 예제
Path	선택한 유형에 적용할 수 없습니다...	문자열	dbo.balanced_transactiontest.csv/part-00000-tid-20092...	
transactionID	선택한 유형에 적용할 수 없습니다...	문자열	206EC255-09E1-4601-9B2D-D2C0B668ED3F, E37A10BB-6...	
accountID	선택한 유형에 적용할 수 없습니다...	문자열	A914800878614266, A985157013380235, A84442812012...	
transactionAmountUSD	선택한 유형에 적용할 수 없습니다...	10진수	139.99, 84.99, 85.79	
transactionAmount	선택한 유형에 적용할 수 없습니다...	10진수	139.99, 84.99, 85.79	
transactionCurrencyCode	선택한 유형에 적용할 수 없습니다...	문자열	USD, USD, USD	
transactionDate	선택한 유형에 적용할 수 없습니다...	전송	20130830, 20130523, 20130903	
- Bottom Navigation:** Buttons for '이전' (Previous), '다음' (Next), and '취소' (Cancel).

스키마

- 각 열의 데이터 유형을 확인 및 변경

15 | 세부 정보 확인



데이터 세부 정보 확인

- 생성된 데이터 세트 정보 확인

16 | 데이터 세트 정보 확인

1. 데이터 미리 보기 확인

• 사용할 데이터 정보 확인

2. 데이터 통계 확인

• 데이터 세트 관련 통계 확인 가능

데이터 세트 정보 확인

1. 데이터 미리 보기 확인
 - 사용할 데이터 정보 확인
2. 데이터 통계 확인
 - 데이터 세트 관련 통계 확인 가능

17 | 데이터 세트 선택

The screenshot shows the Azure Machine Learning Studio interface. On the left, there's a sidebar with various options like '새로 만들기' (Create New), 'Notebooks', '자동화된 ML', etc. The main area is titled '새 자동화된 ML 실행 만들기' (Create New Automated ML Run). A sub-section titled '데이터 세트 선택' (Data Set Selection) is active. It displays a table of data sets with columns: '데이터 세트 이름' (Data Set Name), '데이터 세트 형식' (Data Set Format), '만든 날짜:' (Created Date), and 'modifiedTime'. The first row, 'asodata0000', is highlighted with a red box. Below the table, there's a large button with the text 'asodata0000' and a circular icon. At the bottom, there are '이전' (Previous) and '다음' (Next) buttons.

데이터 세트 이름	데이터 세트 형식	만든 날짜:	modifiedTime
asodata0000	표 형식	27 Feb 2020 10:46	27 Feb 2020 10:46
store1	표 형식	25 Feb 2020 16:54	25 Feb 2020 16:54
movie_review_total	표 형식	24 Feb 2020 21:31	24 Feb 2020 21:31
movie_review_sample	표 형식	19 Feb 2020 10:17	19 Feb 2020 10:17
FraudTransaction	표 형식	19 Feb 2020 10:10	19 Feb 2020 10:10

데이터 세트 선택

- 데이터 목록에서 생성된 데이터 세트를 선택

18 | 실험 구성

Microsoft Azure Machine Learning

ASO_MachineLearning > 자동화된 ML > 실행 시작

새 자동화된 ML 실행 만들기

실험 이름 *

asotest0000

대상 열 *

label

학습 클러스터 선택 *

컴퓨팅 선택...

새 컴퓨팅 만들기

실험 구성

데이터 세트
FraudTransaction (데이터 세트 보기)

실험 이름 *

asotest0000

대상 열 *

label

학습 컴퓨팅 선택 *

컴퓨팅 선택...

새 컴퓨팅 만들기

새 학습 클러스터

컴퓨팅 이름 *

mi-compute-ds

Machine Learning 컴퓨팅은 하나 이상의 노드로 구성된 관리형 고속 환경입니다. 자세히 알아보기

지역 *

koreacentral

가상 머신 크기 *

Standard_DS12_v2

가상 머신 우선 순위 *

전용

최소 노드 수 *

0

최대 노드 수 *

6

규모 축소 전 유휴 시간(초) *

120

고급 설정

다음

만들기

취소

실험 구성

1. 실험 이름 입력
➤ asotest0000
2. 데이터 분석 대상 선택
 - 분석 대상 칼럼 선택
3. 학습 컴퓨팅 생성 및 선택
 - 분석 과제에 따라 적합한 학습 컴퓨팅 선택
➤ ml-compute-ds13
4. “새 컴퓨팅 만들기” 클릭하여 새 학습 컴퓨팅을 생성

❖ 실험용 컴퓨팅 이미 생성됨.
컴퓨팅 생성과정만 참고

19 | 새 컴퓨팅 생성

Microsoft Azure Machine Learning

ASO_MachineLearning > 자동화된 ML > 실행 시작

컴퓨팅 이름 * ml-compute-d

Machine Learning 컴퓨팅은 하나 이상의 노드로 구성된 관리형 교육 환경입니다. 자세히 알아보기

지역 * koreacentral

가상 머신 크기 * Standard_DS12_v2

가상 머신 우선 순위 * 전용

최소 노드 수 * 0

최대 노드 수 * 6

규모 축소 전 유형 시간(초) * 120

새 학습 클러스터

Machine Learning 컴퓨팅은 하나 이상의 노드로 구성된 관리형 교육 환경입니다. 자세히 알아보기

지역 * koreacentral

가상 머신 크기 * Standard_DS12_v2

가상 머신 우선 순위 * 전용

최소 노드 수 * 0

최대 노드 수 * 6

규모 축소 전 유형 시간(초) * 120

고급 설정

만들기 취소

새 컴퓨팅 만들기

1. 컴퓨팅 이름 입력
2. 지역 선택
3. 가상 머신 크기 선택
4. 최대/최소 노드 수 선택
5. 규모 축소 전 유형 시간

20 | 작업 형식 선택

미리 보기 Microsoft Azure Machine Learning

ASO_MachineLearning > 자동화된 ML > 실행 시작

새 자동화된 ML 실행 만들기

작업 형식 선택

1. 데이터 세트 선택

2. 실행 구성

3. 작업 형식 및 설정

분류

대상 열에서 예/아니요, 파랑, 빨강, 녹색 등의 여러 범주 중 하나를 예측하려는 경우

딥 러닝(미리 보기) 사용

회귀

연속 숫자 값을 예측하려는 경우

시계열 예측

시간을 기준으로 값을 예측하려는 경우

추가 구성 설정 보기 1 **기능 선택 설정 보기**

2

추가 구성

주 메트릭 AUC 가중치

자동 기능화

최적 모델 설명

차단된 알고리즘

학습 작업 시간(분)

메트릭 점수 임계값

유효성 검사 유형

데이터 유효성 검사 백분율

최대 동시 반복 횟수

저장

작업 형식 선택

- 분류 ✓
- 회귀
- 시계열 예측

작업 추가 설정

1. “추가 구성 설정 보기” 클릭
2. 추가 설정 입력 화면

21 | 작업 추가 설정

주 메트릭 ⓘ

AUC 가중치

자동 기능화 ⓘ

최적 모델 설명 ⓘ

차단된 알고리즘 ⓘ

자동화된 ML에서 학습 중 사용하지 않을 알고리즘 목록입니다.

종료 기준

학습 작업 시간(분) ⓘ 1

메트릭 점수 임계값 ⓘ

메트릭 점수 임계값

유효성 검사

유효성 검사 유형 ⓘ 학습-유효성 검사 분할

데이터 유효성 검사 백분율 * ⓘ 30

동시성

최대 동시 반복 횟수 ⓘ 6

차단된 알고리즘 ⓘ

자동화된 ML에서 학습 중 사용하지 않을 알고리즘 목록입니다.

LogisticRegression
SGD
MultinomialNaiveBayes
BernoulliNaiveBayes
SVM
LinearSVM
KNN
DecisionTree
RandomForest

분류관련 추가 설정

1. 주 메트릭 선택
 - 정확도
 - AUC 가중치
 - 일반 매크로 회수
 - 평균 정밀도 점수 가중치
 - 정밀도 점수 가중치
2. 자동 기능화 체크
3. 최적 모델 설명 체크
4. 차단된 알고리즘 선택
5. 종료 기준 설정
6. 유효성 검사 유형 선택
 - 자동
 - k-fold cross validation
 - Monte Carlo cross validation
 - Train-Validation Split
7. 동시성 설정

- ❖ 학습 작업 시간(분) - X
- ❖ 학습 작업 시간(시) - O

22 | 실험 결과 확인

The screenshot shows the Azure Machine Learning studio interface. On the left, there's a sidebar with various options like Notebooks, New Experiment, and Datasets. A red box labeled '1' highlights the 'Automated ML' option under the 'Experiment' section. The main area is titled 'Automated ML' and displays two experiments:

실험	실행 ID	상태	만든 날짜:	기간
ASO_AUTOML	AutoML_cae6cdaa-a938-4d14-9b84-54c0d580f85d	완료	19 Feb 2020 10:31	00:20:54
ASO_AUTOML	AutoML_8c1ce8c2-5587-4d97-b395-000354fa06b7	취소됨	19 Feb 2020 10:13	00:11:39

A red box labeled '2' highlights the 'AutoML_cae6cdaa-a938-4d14-9b84-54c0d580f85d' link in the first row. Below this, a large callout box is labeled '실행 ID' and contains the text 'AutoML_cae6cdaa-a938-4d14-9b84-54c0d580f85d'. The bottom section is titled '설명서' and lists three items: '개념: 자동화된 ML이란 무엇인가요?', '자습서: 자동화된 ML을 사용하여 첫 번째 분류 모델 만들기', and '로그: 자동화된 ML의 새로운 기능을 사용하여 더욱 정확한 예측 결과 만들기'.

실험 결과 확인

- <자동화된 ML>탭을 클릭
- 실행 ID를 클릭하여 실험 결과 확인 가능

23 | 실험 결과 확인

세부 정보

모델 이름	VotingEnsemble
메트릭 값	0.8317033003965854
시작 시간	19 Feb 2020 10:53
기간	00:01:43
SDK 버전	1.1.0rc0
배포 상태	아직 배포 없음

모델 세부 정보 보기

실험 결과 확인

3. <세부 정보>탭을 클릭
4. 최적 모델 정보 확인
 - 모델이름
 - 메트릭 값
5. 모델 세부 정보 및 상세한 결과 값들을 확인할 수 있음

결과 해석

- 측정 메트릭은 AUC 가중치이며 83.2%의 예측 정확도를 보였음
- 이에 따라 다른 모델을 비해서 Voting Ensemble은 성능 좋은 모델(챔피언 모델)임을 확인할 수 있음
- AUC 평가 기준
 - 0.90-1 = excellent
 - 0.80-0.90 = good
 - 0.70-0.80 = fair
 - 0.60-0.70 = poor
 - 0.50-0.60 = fail

24 | 실험 결과 확인

실험 결과 확인

6. <모델 세부 정보>탭을 클릭

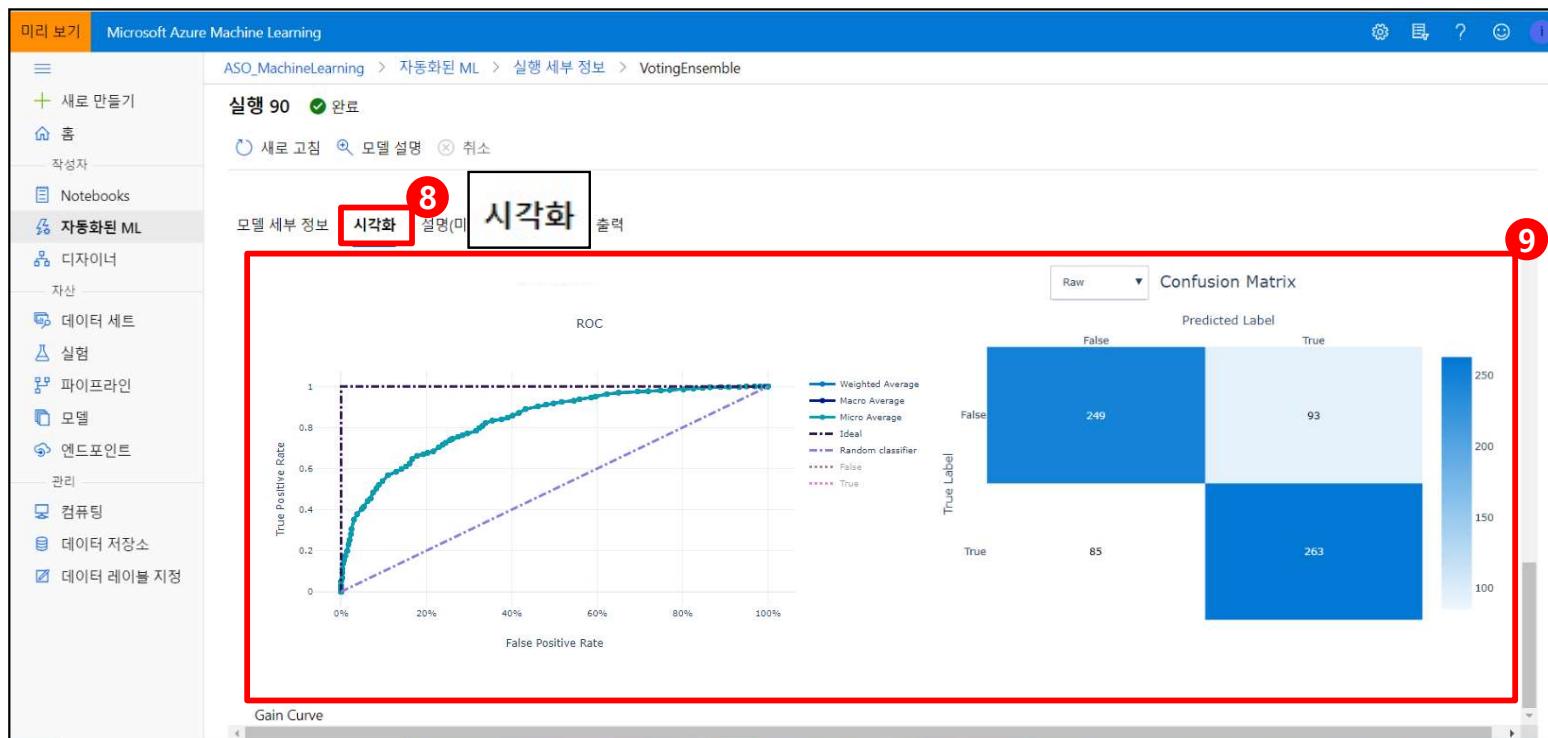
7. 주 메트릭 값 외에 다른 메트릭 값들도 확인할 수 있음

결과 해석

- AUC가중치 외에는 74.2%의 정확도를 확인할 수 있음
- 83%의 평균 정밀도 가중치도 확인했음
- 매크로 평균은 평균들의 평균 개념임
- 마이크로 평균은 전체 값들의 평균 개념임

메트릭	값
정확도	0.74203
AUC 매크로	0.83170
AUC 마이크로	0.83146
AUC 가중치	0.83170
평균 정밀도 점수 매크로	0.83015
평균 정밀도 점수 마이크로	0.83228
평균 정밀도 점수 가중치	0.83036

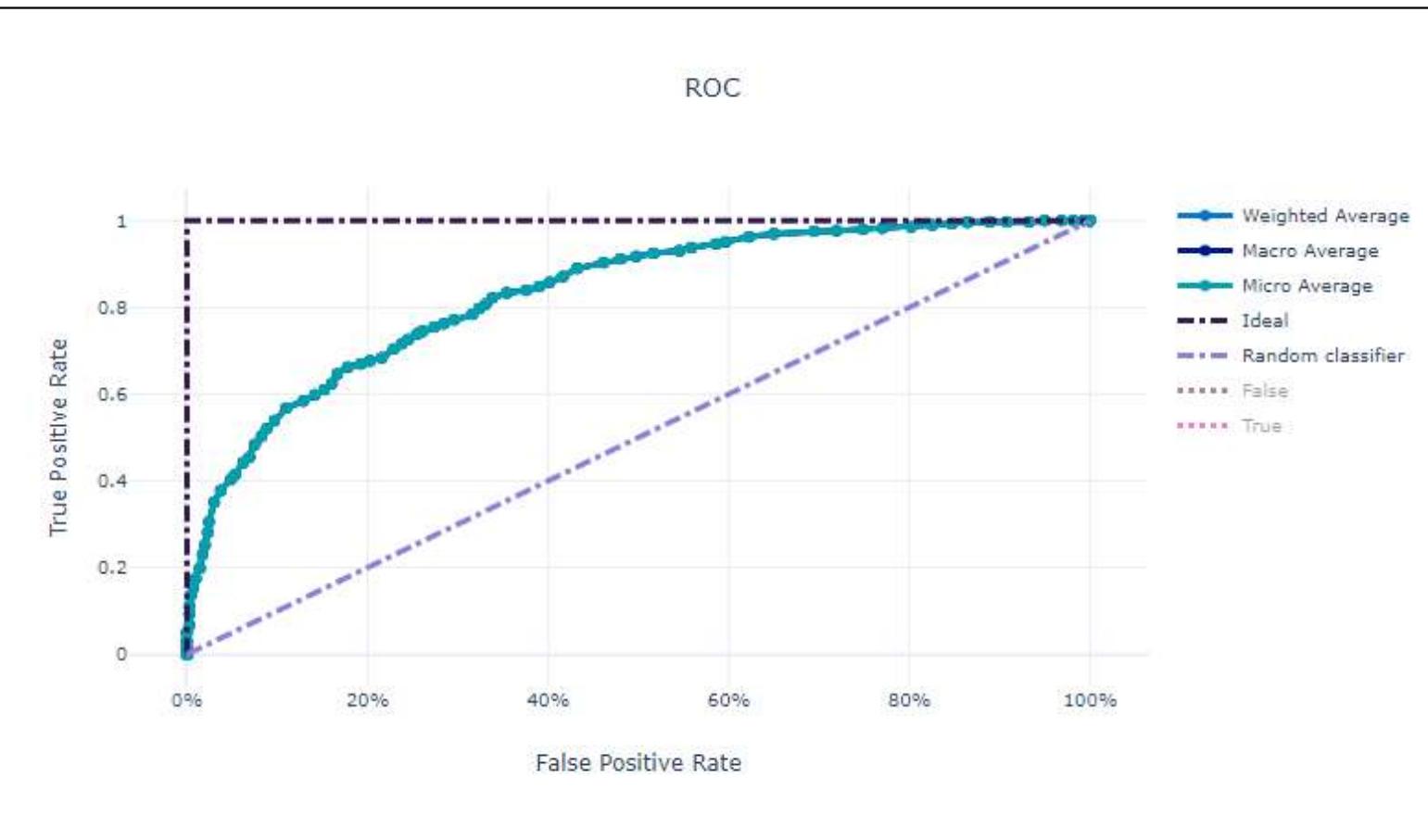
25 | 실험 결과 확인



실험 결과 확인

8. <시각화>탭을 클릭
9. 다양한 시각화 형식을 확인할 수 있음

26 | 실험 결과 확인

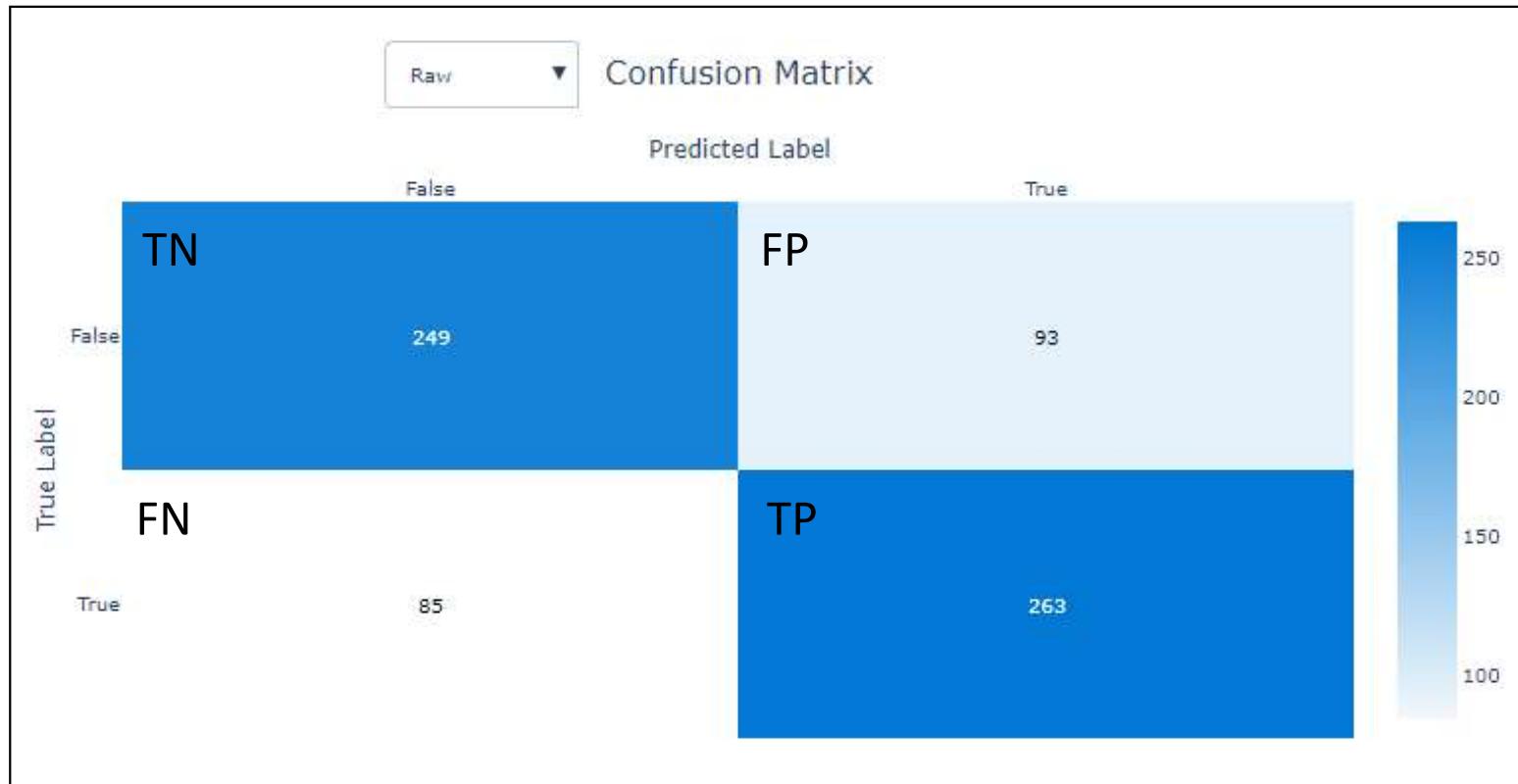


- ❖ TPR: True Positive Rate
- ❖ FPR: False Positive Rate

ROC 곡선

- ROC Curve (Receiver Operating Characteristic Curve)
- ROC는 다양한 thresholds 설정의 분류 문제 성능
- Thresholds settings을 낮추면 더 많은 항목이 양성으로 분류되므로 FP와 TP가 모두 증가
- 추가로 ROC 곡선은 거짓 양성 비율(FPR)/진실 양성 비율(TPR)의 곡선
- 값이 1에 가까울수록 모델 성능이 좋음을 나타냄
- AUC (Area Under Curve)는 ROC 곡선 아래부분의 넓이
- AUC 면적이 넓을수록 모델의 성능이 좋음

27 | 실험 결과 확인



Confusion Matrix

- 모델 예측 결과 10건을 예측 했을 때 7.4건의 정확도를 가짐
- 이에 기반으로 Precision(정밀도), Recall(재현율), Accuracy(정확성), F1-Score를 구할 수 있음
- 계산 결과에 따라 이 모델의 정확성은 74.2%로 보였음
- 이 모델에서 어떤 데이터가 사기일 것이라고 평가했을 때, 이 평가가 정확할 확률은 73.9%로 확인됨
- 이 모델에서는 모든 사기 데이터 중 75.6%가 정확하게 식별됨
- F1-Score가 높아야 성능이 좋다고 보면 되며, 이 모델의 F1-Score는 0.747이것으로 보였음

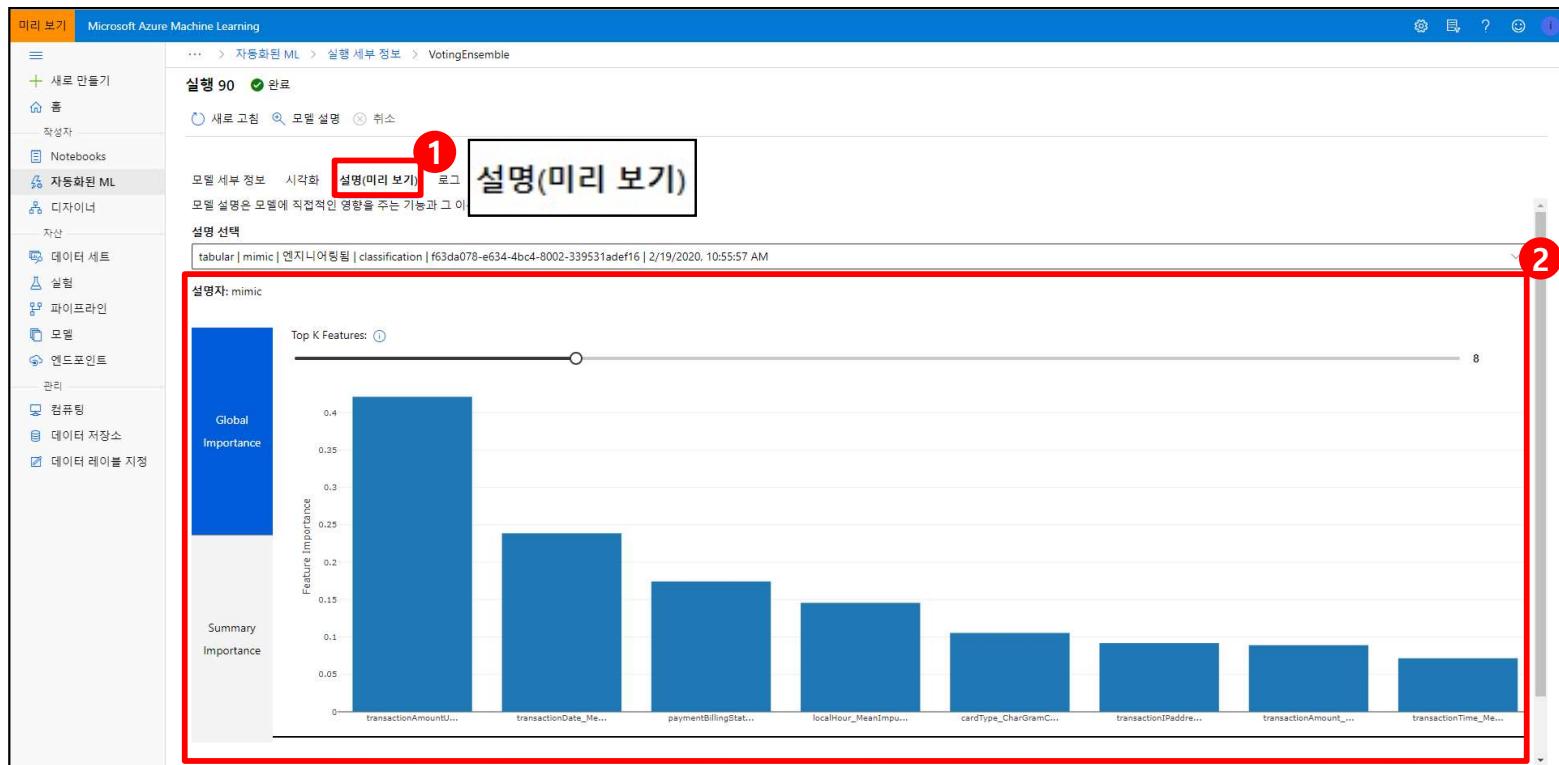
- Precision
 $= \frac{TP}{TP+FP}$
 $= \frac{263}{263+93}$
 $= 73.9\%$

- Recall
 $= \frac{TP}{TP+FN}$
 $= \frac{263}{263+85}$
 $= 75.6\%$

- Accuracy
 $= \frac{(TP+TN)}{(TP+TN+FP+FN)}$
 $= \frac{(263+249)}{690}$
 $= 74.2\%$

- F1-Score
 $= \frac{2(Precision*Recall)}{(Precision+Recall)}$
 $= \frac{2(0.739*0.756)}{(0.739+0.756)}$
 $= 0.747$

28 | 실험 결과 확인



모델에 영향 미치는 변수(피쳐)

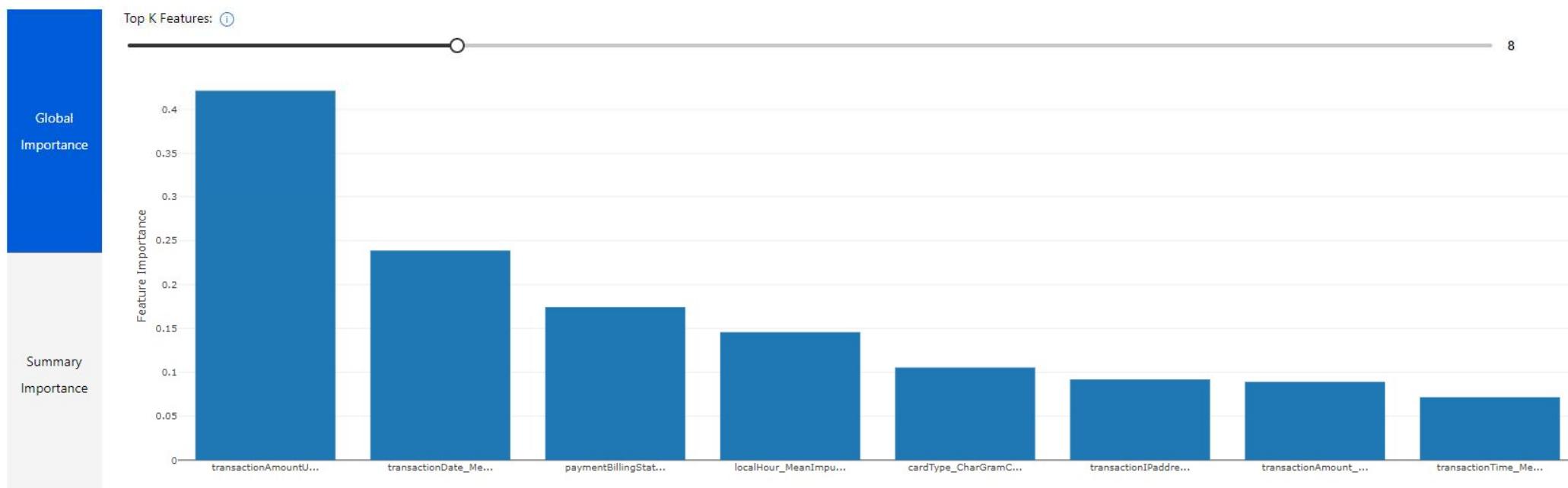
1. <설명(미리 보기)>탭을 클릭
2. 모델에 직접적인 영향을 주는 피쳐들의 중요도를 확인할 수 있음

29 | 실험 결과 확인

설명 선택

tabular | mimic | 엔지니어링됨 | classification | f63da078-e634-4bc4-8002-339531aef16 | 2/19/2020, 10:55:57 AM

설명자: mimic



글로벌 중요도

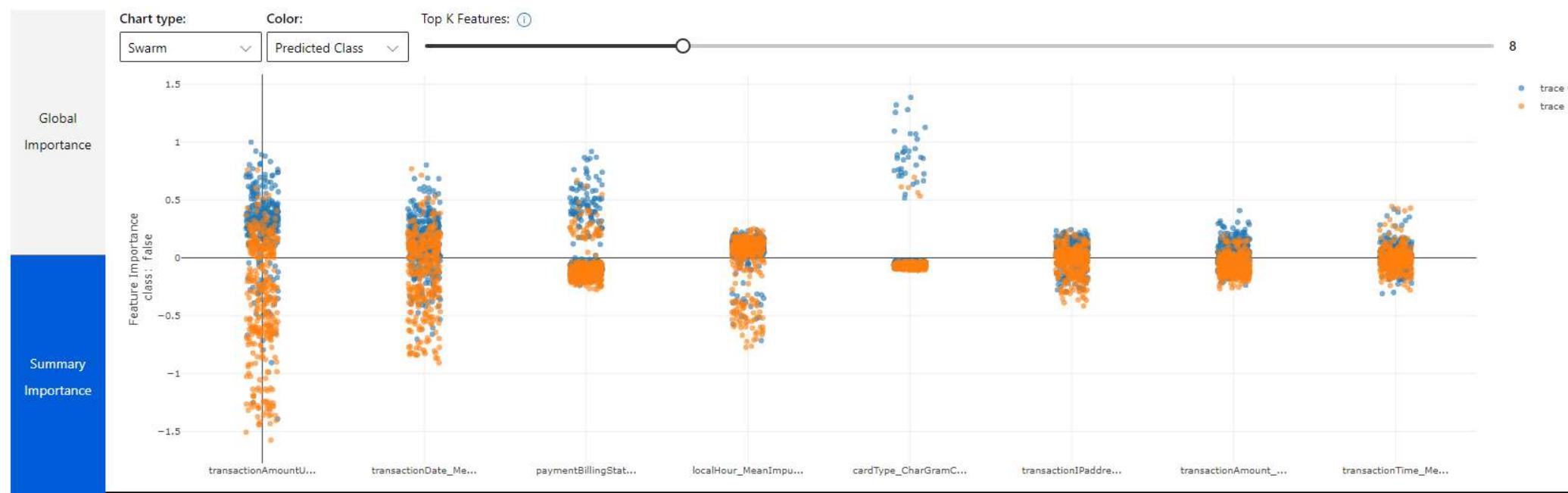
1. 최상위 K(구성 가능 K) 중요 기능을 전역적으로 보여줌
2. 그림에 따라 transactionAmountUSD이라는 변수가 0.421의 글로벌 중요도로 보였으며 이 변수가 중요한 역할로 모델에 영향을 많이 미침

30 | 실험 결과 확인

설명 선택

tabular | mimic | 엔지니어링됨 | classification | f63da078-e634-4bc4-8002-339531adef16 | 2/19/2020, 10:55:57 AM

설명자: mimic



요약 중요도

- 모든 데이터 요소에서 로컬 기능 중요도 값을 사용하여 예측 값에 대한 각 기능 영향의 분포를 표시함.
- 사기 데이터는 주황색 점으로 표시되고 비사기 데이터는 파란색 점으로 표시됨
- 분포의 모양과 점의 밀도에 따라 transactionAmountUSD 변수가 이 모델에 가장 중요한 역할로 보였음

31 | 모델 배포

The screenshot shows the Azure Machine Learning Studio interface. On the left, there's a sidebar with various options like '새로 만들기' (Create New), 'Notebooks', '자동화된 ML', etc. The main area shows a '실행 세부 정보' (Run Details) panel for 'ASO_MachineLearning > 자동화된 ML > 실행 세부 정보'. It displays details about a run, including the '권장 모델' (Recommended Model) which is a 'Voting Ensemble' with an AUC metric of 0.8317033003965854. Below this, there's a large blue button labeled '최적 모델 배포' (Deploy Best Model). To the right, there's a detailed '모델 배포' (Model Deployment) configuration pane. This pane has three numbered callouts:

- 1.** A red box highlights the '최적 모델 배포' (Deploy Best Model) button at the bottom left of the deployment pane.
- 2.** A red box highlights the '모델 배포' (Model Deployment) configuration area, which includes fields for '이름' (Name), '설명' (Description), '실행 요약' (Run Summary), '컴퓨팅 형식' (Compute Type), and '인증 사용' (Authentication Usage).
- 3.** A red box highlights the '배포' (Deploy) button at the bottom right of the deployment pane.

모델 배포

1. <최적 모델 배포>버튼을 클릭
2. 모델 배포 정보 입력
 - 모델이름
 - 모델관련 설명
 - 컴퓨팅 형식 선택
 - Azure Kubernetes Service(AKS)
 - Azure Container Instances(ACI)
 - “인증 사용” 체크
3. <배포>버튼을 클릭하여 배포 작업을 진행

32 | 모델 배포

The screenshot shows two main windows from the Azure Machine Learning Studio:

- Left Window (Azure Machine Learning Studio Home):**
 - Left sidebar: 모델 (Model) is selected, indicated by a red box labeled **4**.
 - Main area: 엔드포인트 (Endpoint) is selected, indicated by a red box labeled **5**.
- Right Window (Endpoint Details):**
 - Header: ASO_MachineLearning > 엔드포인트
 - Table: Shows deployed endpoints. One row for 'aso-championmodel...' is highlighted with a red box and labeled **5**. The table columns are: 이름 (Name), 설명 (Description), 만든 날짜 (Created Date), 업데이트한 날짜 (Last Updated Date), 업데이트한 사람 (Last Updated By).
 - Details for 'aso-championmodel':
 - 세부 정보 (Detailed Information): Includes endpoint status (Healthy), service ID (aso-championmodel), and creation date (2/20/2020 2:12:06 PM).
 - 사용 (Usage): Includes resource details like CPU (1), memory (2 GB), and auto-scaling settings.
 - REST Endpoint: Shows the URL `http://ce80a50b-01ab-4a29-8af7-84bb0b44b0a8.westus.azurecontainer.io/score`.

모델 배포

4. <엔드포인트>탭을 클릭
5. 배포된 모델을 확인하여 클릭
6. REST 엔드포인트에서 scoring uri를 확인

33 | 모델 호출

기본 사용 정보

REST 앤드포인트
http://ce80a50b-01ab-4a29-8af7-84bb0b44b0a8.westus.azurecontainer.io/score

키 사용 토큰 사용

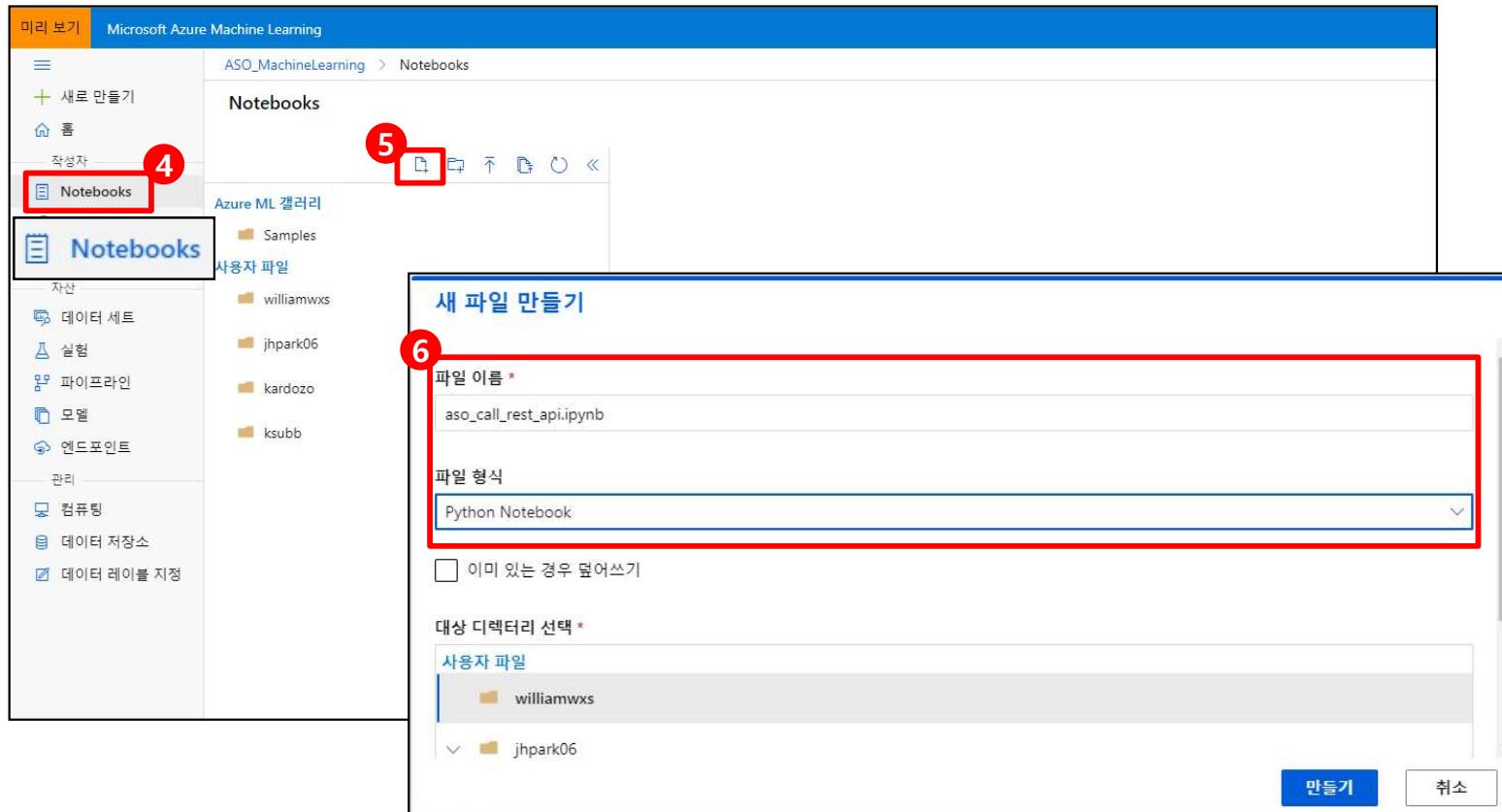
기본 키
QWpOOJHS6nrd8aQQpFj65vakmJ9OxX6A [다시 생성](#)

보조 키
hdtNRfwbbAwZPbOOx9Tlu5MHjTmvPb [다시 생성](#)

모델 호출

- <사용>탭을 클릭
- REST 앤드포인트를 확인 및 URI 복사
- “키 사용” 체크하여 기본 키를 복사

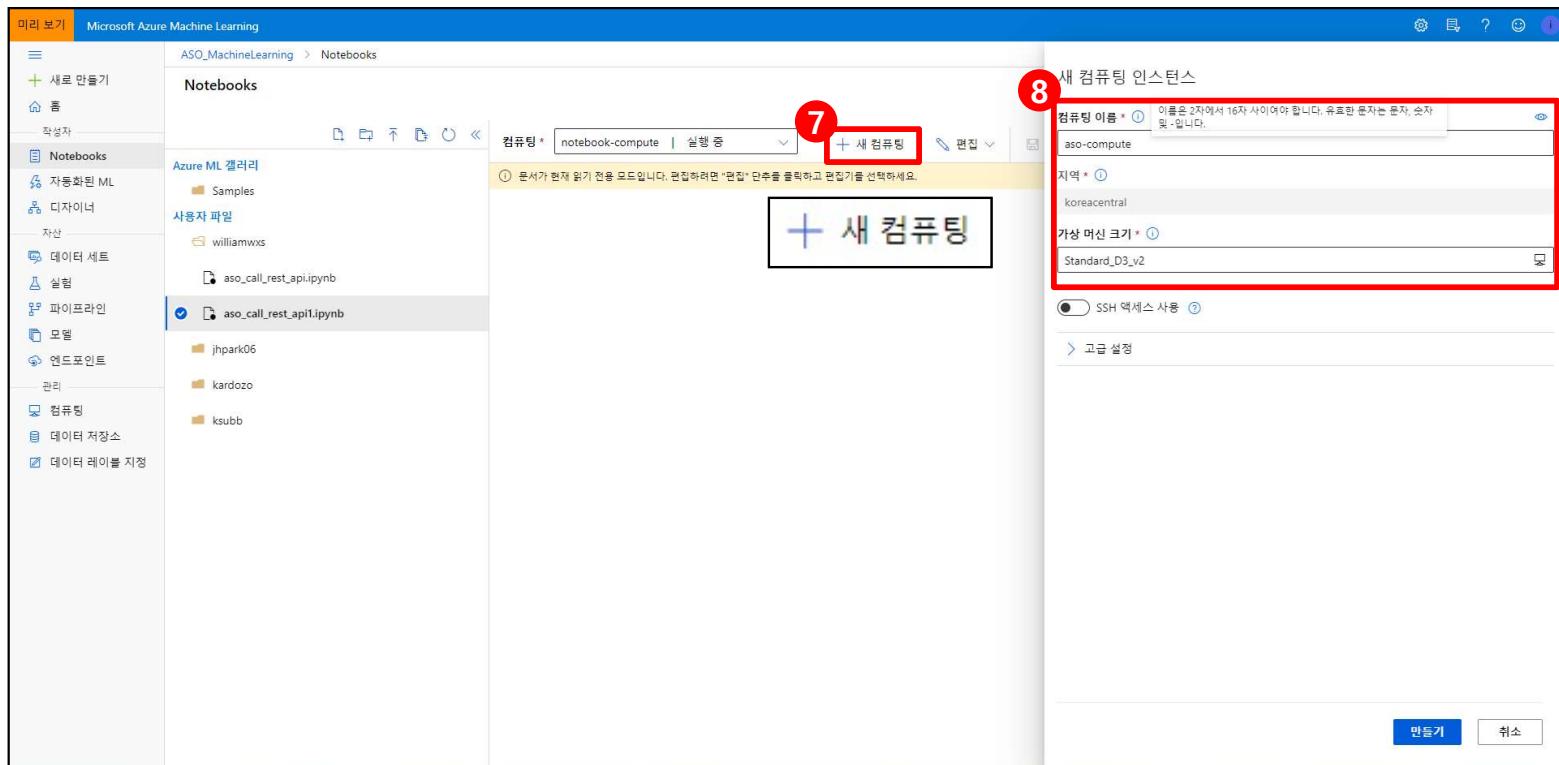
34 | 모델 호출



모델 호출

4. <Notebooks>탭을 클릭
5. <새 파일 만들기>탭을 클릭
6. 파일 이름 입력 및 파일 형식 "Python Notebook" 선택

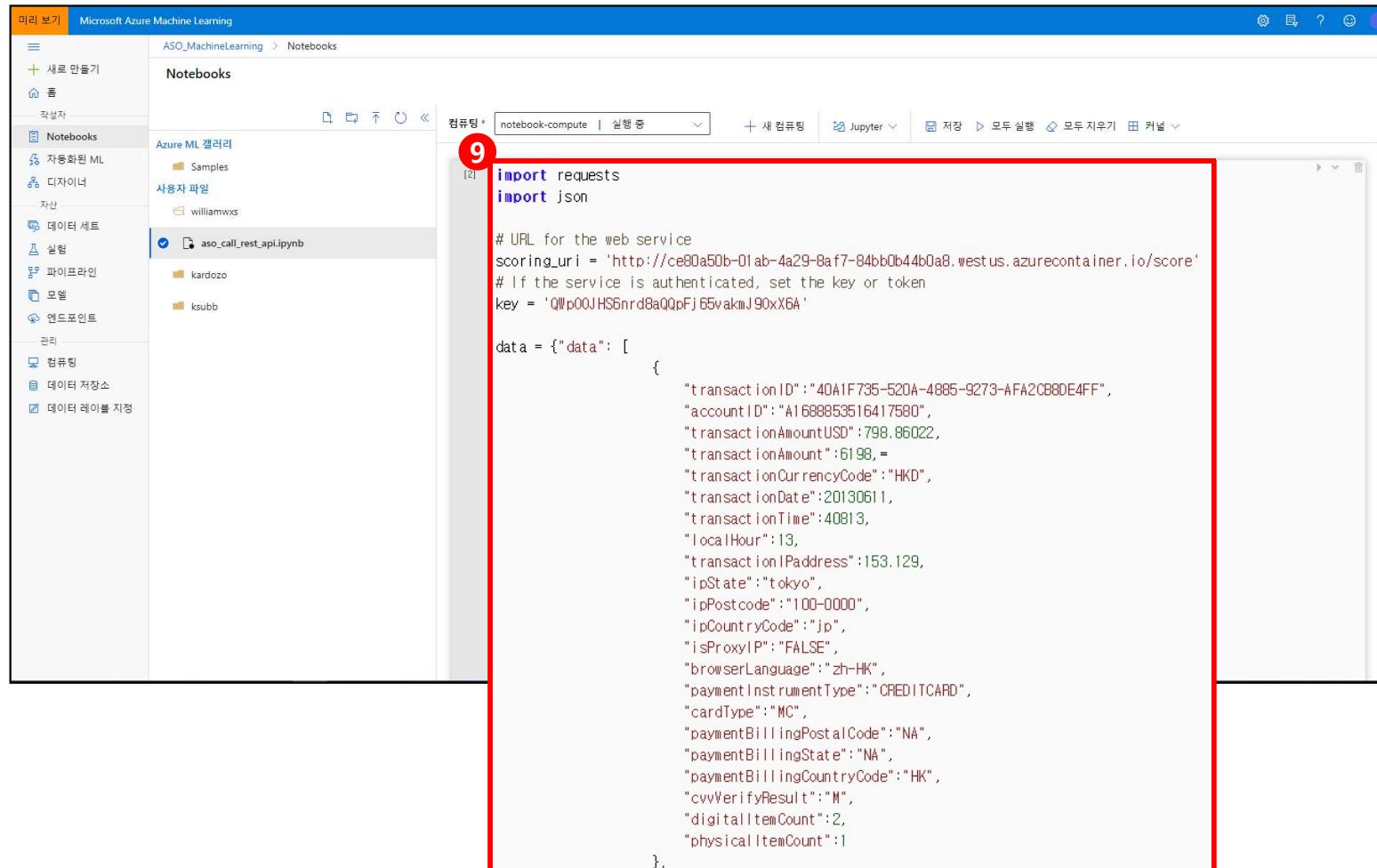
35 | 모델 호출



모델 호출

7. <새 컴퓨팅>탭을 클릭
8. 컴퓨팅 이름, 가상 머신 크기 등 정보 입력

36 | 모델 호출



The screenshot shows a Microsoft Azure Machine Learning interface. On the left, there's a sidebar with various options like '새로 만들기', 'Notebooks', '자동화된 ML', etc. The main area is a Jupyter Notebook titled 'aso_call_rest_api.ipynb'. The code in the notebook is:

```

9 import requests
import json

# URL for the web service
scoring_uri = 'http://ce80a50b-01ab-4a29-8af7-84bb0b44b0a8.westus.azurecontainer.io/score'
# If the service is authenticated, set the key or token
key = 'QWp00JHS6nrd8aQQpFj65vakmJ90xX6A'

data = {"data": [
    {
        "transactionID": "40A1F735-520A-4885-9273-AFA2CB8DE4FF",
        "accountId": "A1688853516417580",
        "transactionAmountUSD": 798.86022,
        "transactionAmount": 6198,
        "transactionCurrencyCode": "HKD",
        "transactionDate": 20130611,
        "transactionTime": 40813,
        "localHour": 13,
        "transactionIPaddress": 153.129,
        "ipState": "tokyo",
        "ipPostcode": "100-0000",
        "ipCountryCode": "jp",
        "isProxyIP": "FALSE",
        "browserLanguage": "zh-HK",
        "paymentInstrumentType": "CREDITCARD",
        "cardType": "MC",
        "paymentBillingPostalCode": "NA",
        "paymentBillingState": "NA",
        "paymentBillingCountryCode": "HK",
        "cvvVerifyResult": "M",
        "digitalItemCount": 2,
        "physicalItemCount": 1
    }
]}

```

모델 호출

9. 메모장에 있는 코드를 복사하고 실행창에 붙여넣기

37 | 호출 결과 확인

```

import requests
import json

# URL for the web service
scoring_uri = 'http://ce80a50b-01ab-4a29-8af7-84bb0b44b0a8.westus.azurecontainer.io/score'
# If the service is authenticated, set the key or token
key = 'QWp00JHSnrd8aQQpFj65vakmJ90xX6A'

data = {"data": [
    {
        "transactionID": "40A1F735-520A-4885-9273-AFA2CB8DE4FF",
        "accountID": "A1688853516417580",
        "transactionAmountUSD": 798.86022,
        "transactionAmount": 6198,
        "transactionCurrencyCode": "HKD",
        "transactionDate": 20130611,
        "transactionTime": 40813,
        "localHour": 13,
        "transactionIPaddress": 153.129,
        "ipState": "tokyo",
        "ipPostcode": "100-0000",
        "ipCountryCode": "jp",
        "isProxyIP": "FALSE",
        "browserLanguage": "zh-HK",
        "paymentInstrumentType": "CREDITCARD",
        "cardType": "MC",
        "paymentBillingPostalCode": "NA",
        "paymentBillingState": "NA",
        "paymentBillingCountryCode": "HK",
        "cvvVerifyResult": "M",
        "digitalItemCount": 2,
        "physicalItemCount": 1
    }
]}

테스트 결과: {"result": [true, false]}

```

모델 호출 코드 실행

- 복사된 REST 엔드포인트 URI와 키를 해당 위치에 맞게 붙여넣기
- 테스트할 데이터를 json파일 형식으로 입력
- <모두 실행>탭을 클릭
- 모델 호출 코드를 실행하여 테스트 결과 확인

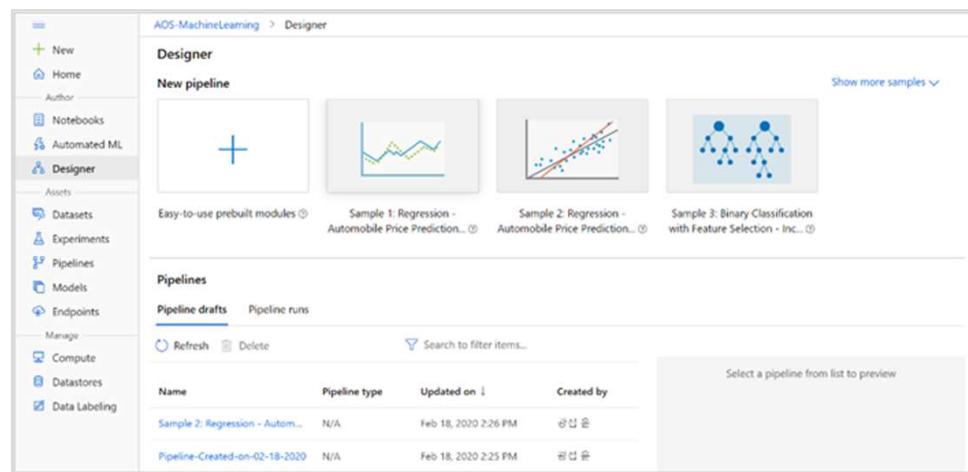
3

Chapter

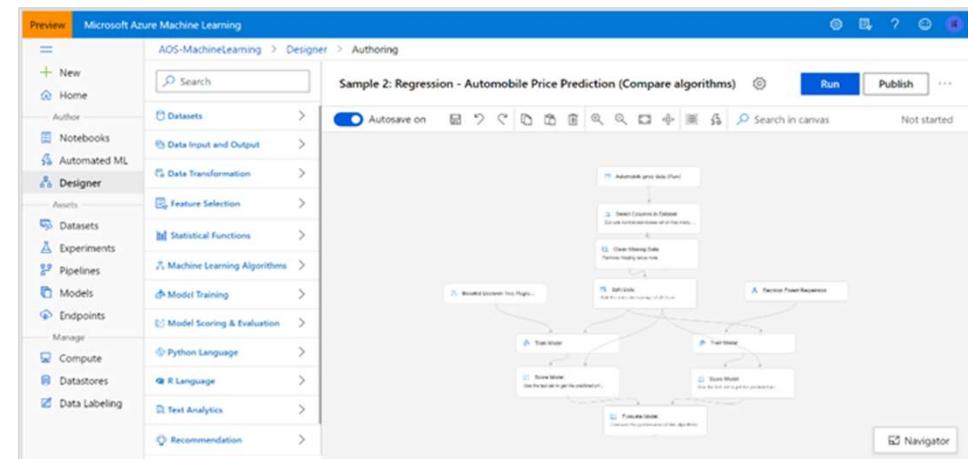
Azure Machine Learning Designer

01 | Azure ML Designer 소개

- Azure ML Designer는 Drag&Drop 방식의 Machine Learning 기반 분석 서비스
- Azure ML Designer는 분석가가 구상한 분석 프로세스를 손쉽게 구현
 - Drag & Drop방식으로 각 Module을 연결하여 누구나 손쉽게 분석 프로세스를 구현
 - Feature Selection, hyperparameter tuning 등의 복잡한 작업도 Module 하나로 손쉽게 수행
- Azure ML Designer는 구축된 분석 모델을 다양한 형태로 배포 가능
 - Pipeline 구성을 통해 DataFactory, Databricks 등 다른 Azure 서비스와 연계하여 분석 모델 활용
 - Rest API 구성을 통해 실시간 분석 모델 배포 가능



▲ 다양한 분석 예제를 통한 손쉬운 활용



▲ Drag&Drop 방식을 통한 직관적인 분석 프로세스

02 | Azure ML Designer 홈화면

The screenshot shows the Azure Machine Learning Studio home page for the workspace 'ASO_MachineLearning'. On the left, there is a sidebar with various menu items: New, Home, Author, Notebooks, Automated ML, Designer (which is highlighted with a red box), Assets, Datasets, Experiments, Pipelines, Models, Endpoints, Manage, Compute, Datastores, and Data Labeling. The main area has a title 'Welcome to the studio!' and four cards: 'Create new' (with a plus icon), 'Notebooks' (with a code icon), 'Automated ML' (with a gear icon), and 'Designer' (with a cluster icon). The 'Designer' card is also highlighted with a red box. Below these cards is a section titled 'My recent resources' with a table of 'Runs'.

Run number	Experiment	Updated time	Status
35	ASO_AUTOML	Feb 19, 2020 10:55 AM	Completed
4	Designer_test_v1	Feb 19, 2020 10:53 AM	Completed
1	Designer_test_v1	Feb 19, 2020 10:35 AM	Failed
36	ASO_AUTOML	Feb 19, 2020 10:34 AM	Completed
1	ASO_AUTOML	Feb 19, 2020 10:30 AM	Canceled

- Azure ML 시작 화면
- 1. Azure ML의 Home 화면에서 왼쪽 메뉴의 Designer 혹은 중앙 부분 Designer 파트의 Start Now 버튼 클릭

03 | Azure ML Designer 홈화면

ASO_MachineLearning > Designer

Designer

1 pipeline + 2

Easy-to-use prebuilt modules ⓘ

Sample 1: Regression - Automobile Price Prediction... ⓘ

Sample 2: Regression - Automobile Price Prediction... ⓘ

Sample 3: Binary Classification with Feature Selection - Inc... ⓘ

Show more samples ⓘ

Pipelines

Pipeline drafts Pipeline runs

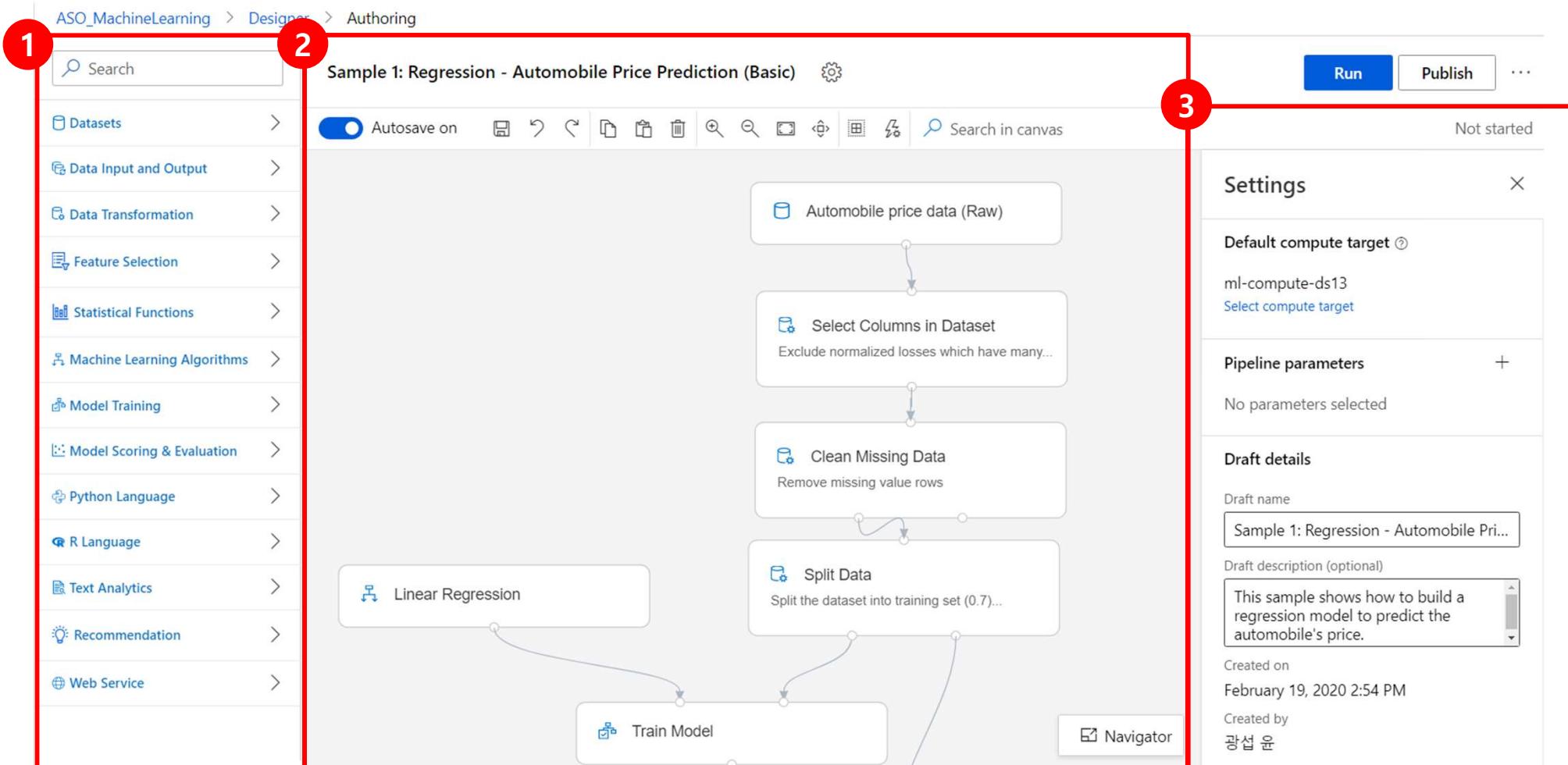
Refresh Delete Search to filter items...

Name	Pipeline type	Updated on ↓	Created by
R_script 활용	N/A	Feb 19, 2020 11:08 AM	광섭 윤

Select a pipeline from list to preview

- Azure ML Designer 홈 화면
- +버튼을 클릭하여 새로운 Pipeline 생성
- Regression, Text Classification, Python Script 활용, SVD 활용 Recommendation 등 다양한 Designer Pipeline 예제 제공
- Show more samples를 통해 전체 예제 확인

04 | Azure ML Designer 화면



1. Module 선택 메뉴

2. Module 실행 Canvas

3. Module 설정 창

05 | 분석 개요

분석 주제

- 영화 리뷰 긍/부정 감성 분류

분석 목적

- 머신러닝을 활용한 영화 리뷰 긍/부정 분류 모델 생성
- 비정형 텍스트 데이터를 활용한 머신 러닝 기반의 정형 분석 수행

데이터 건수

- 네이버 영화 리뷰 긍정 부정 데이터(150,000개)
- 부정:75,173건, 긍정:74,827건

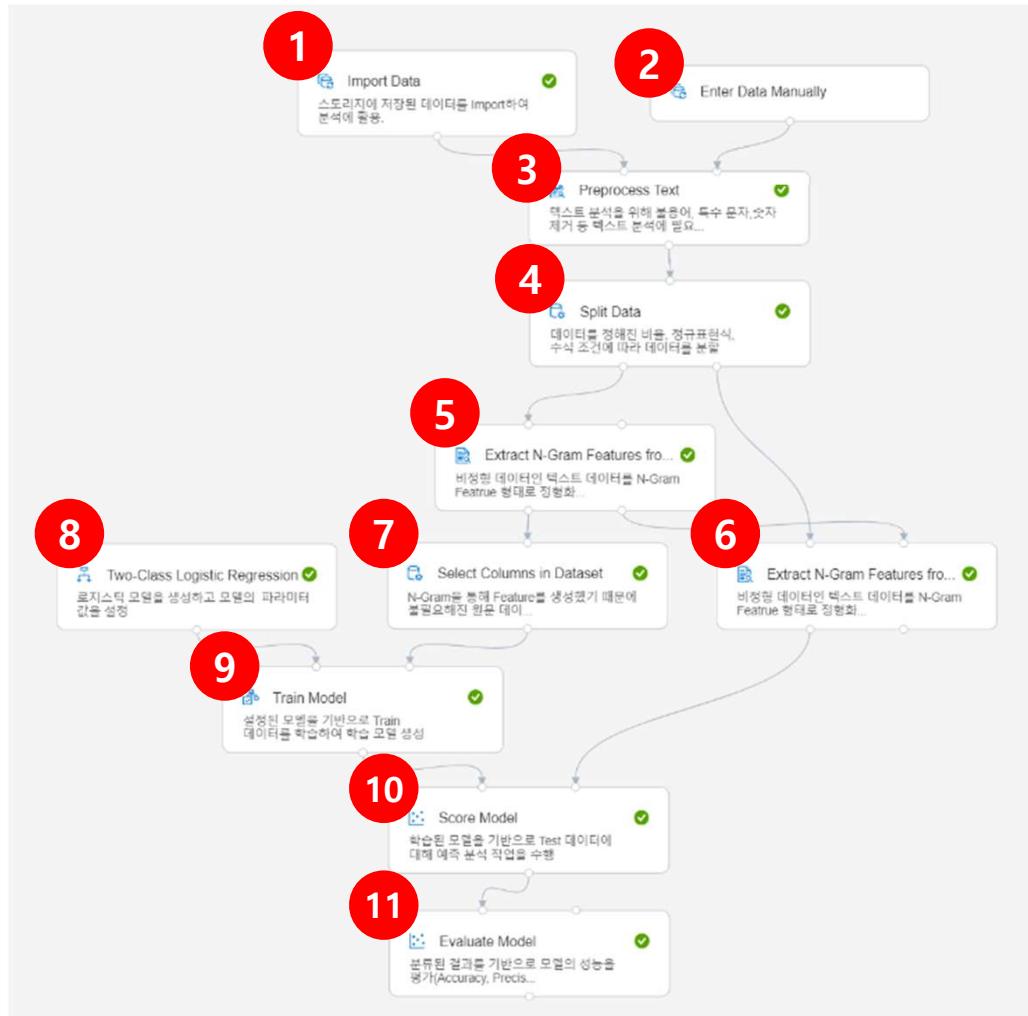
데이터 특징

- 140자 이하의 짧은 영화 리뷰
- 긍정(10점 만점 기준 평점 9~10점), 부정(10점 만점 기준 평점 1~4점)
- 5 ~ 8점의 평점의 중립 감성 문서는 제외

활용 알고리즘

- Logistic Regression을 활용한 이진 분류

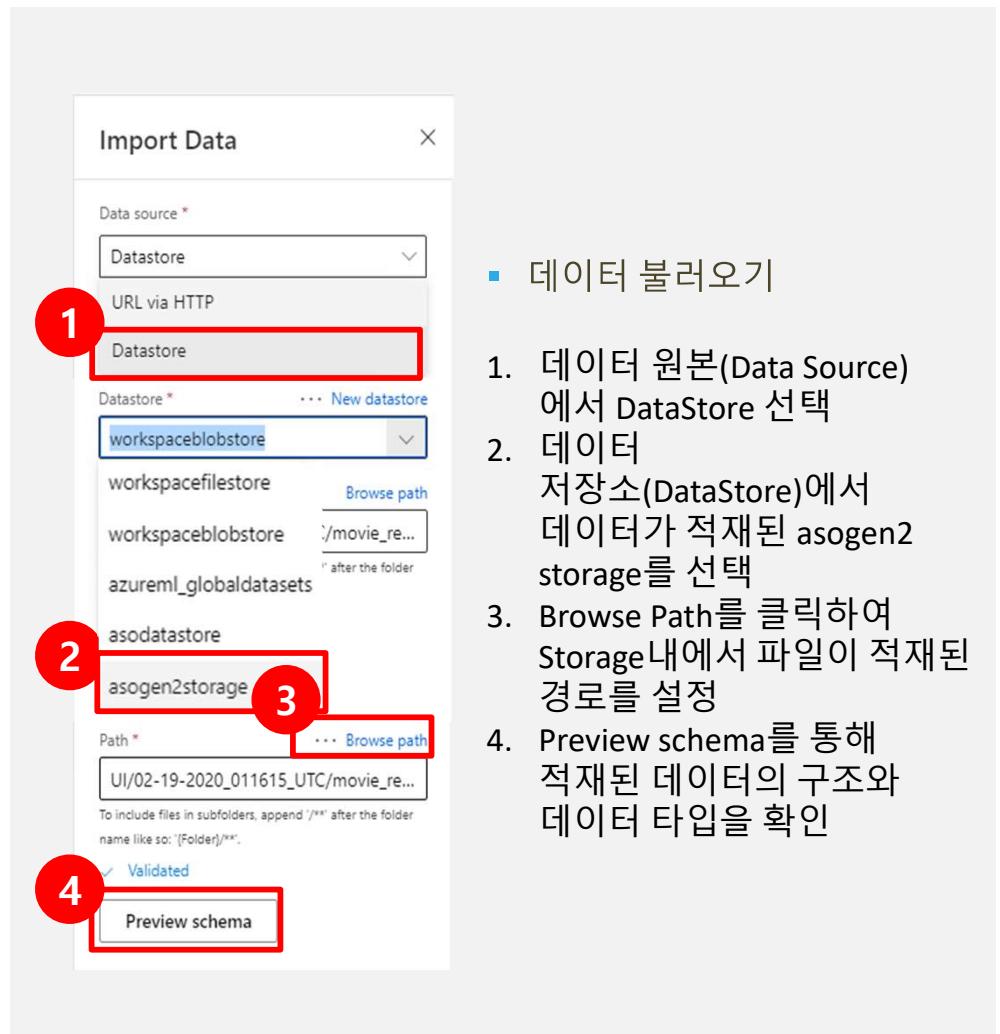
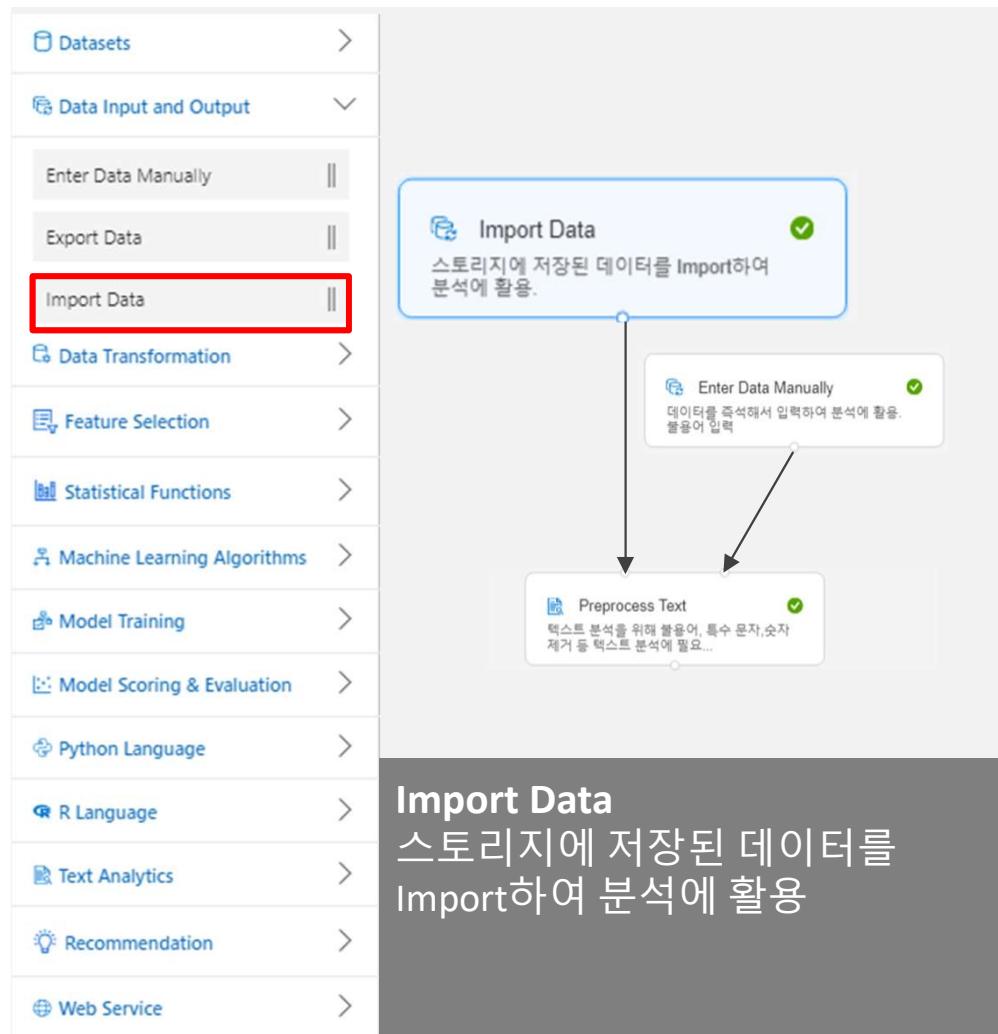
06 | 전체 프로세스



■ 분석 프로세스

1. 데이터 불러오기
2. 데이터 직접 넣기
3. Text 데이터 전처리
4. Train/Test 데이터 분할
5. Train 데이터 N-gram 추출
6. Test 데이터 N-gram 추출
7. 컬럼 선택
8. 모델 생성
9. 모델 학습
10. Test 데이터 예측
11. 모델 평가

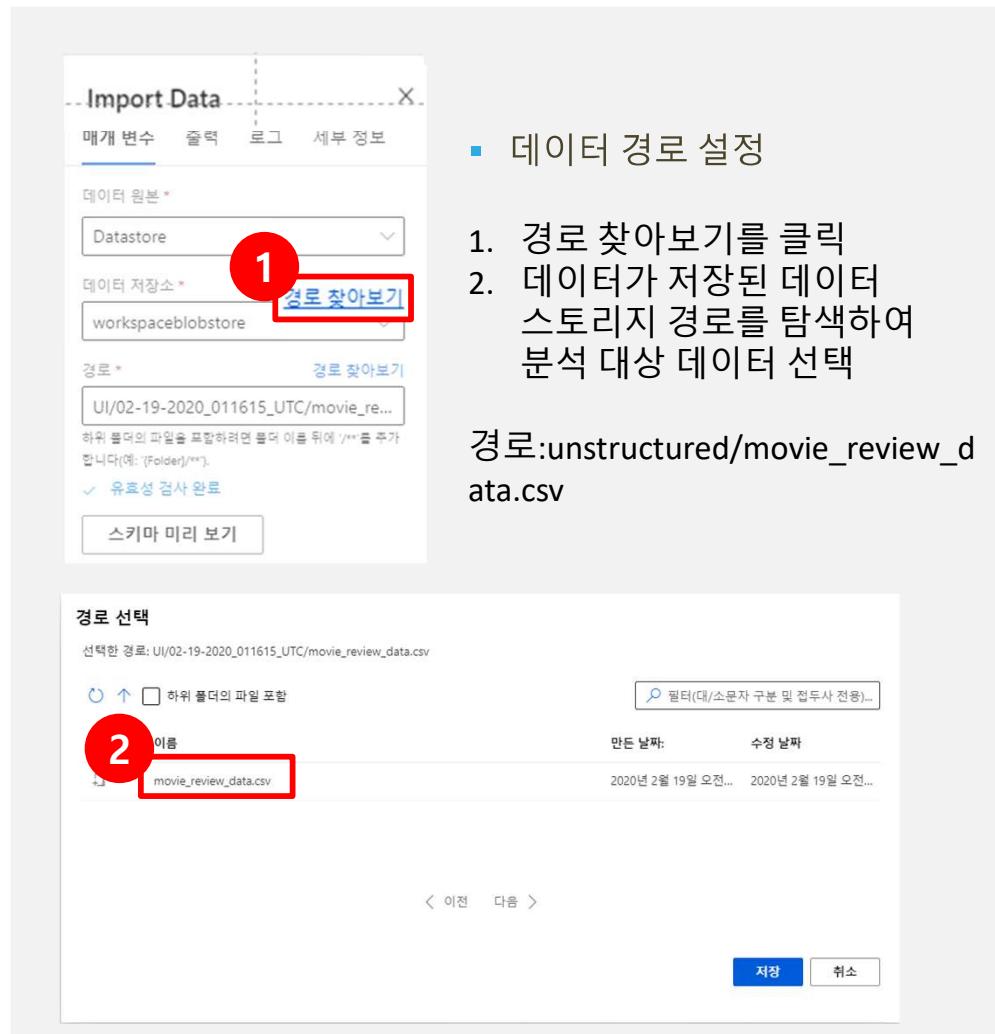
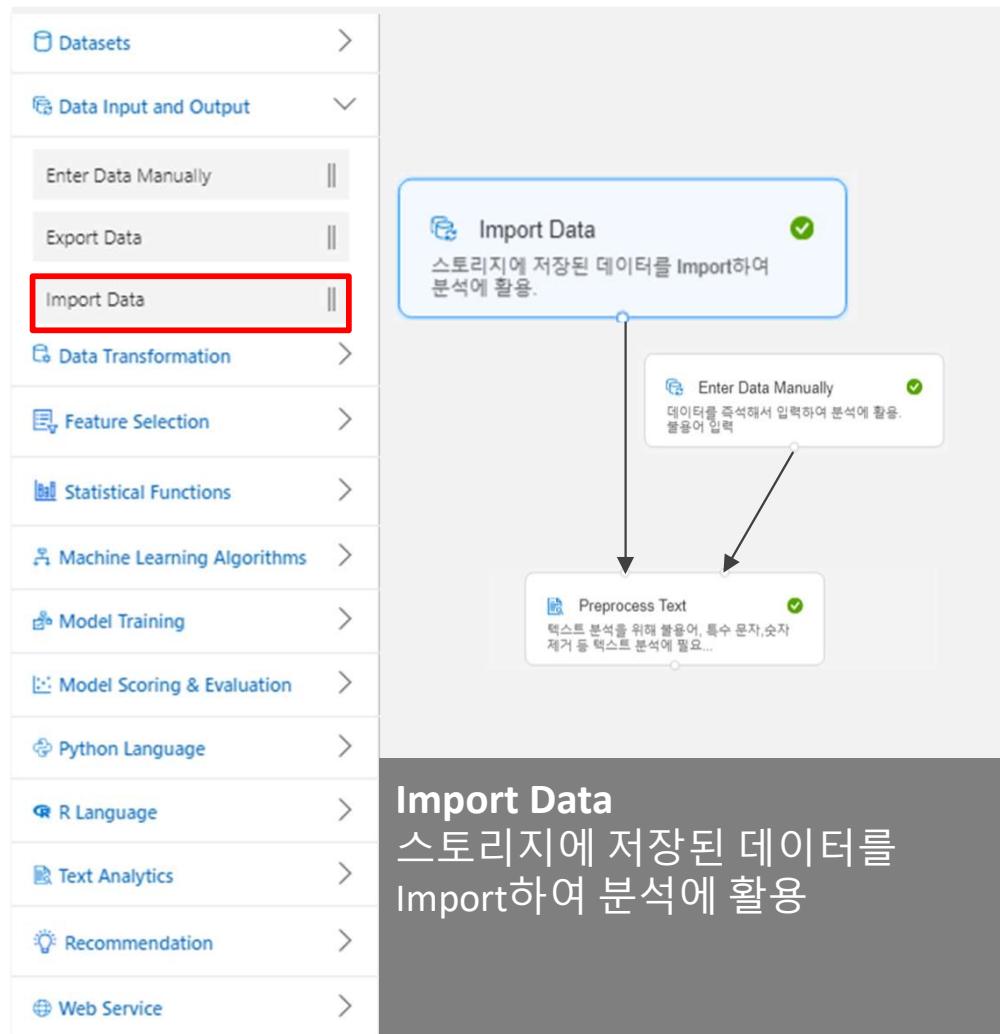
07 | 데이터 불러오기



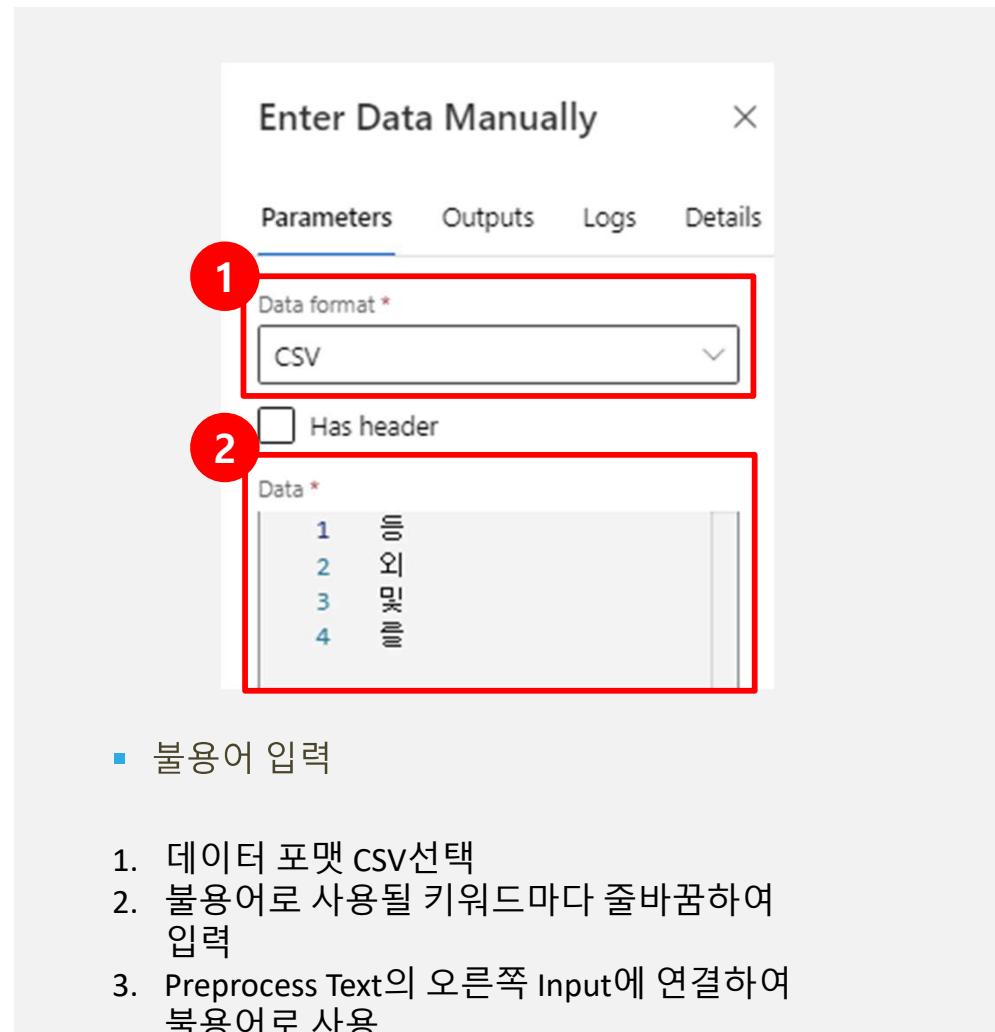
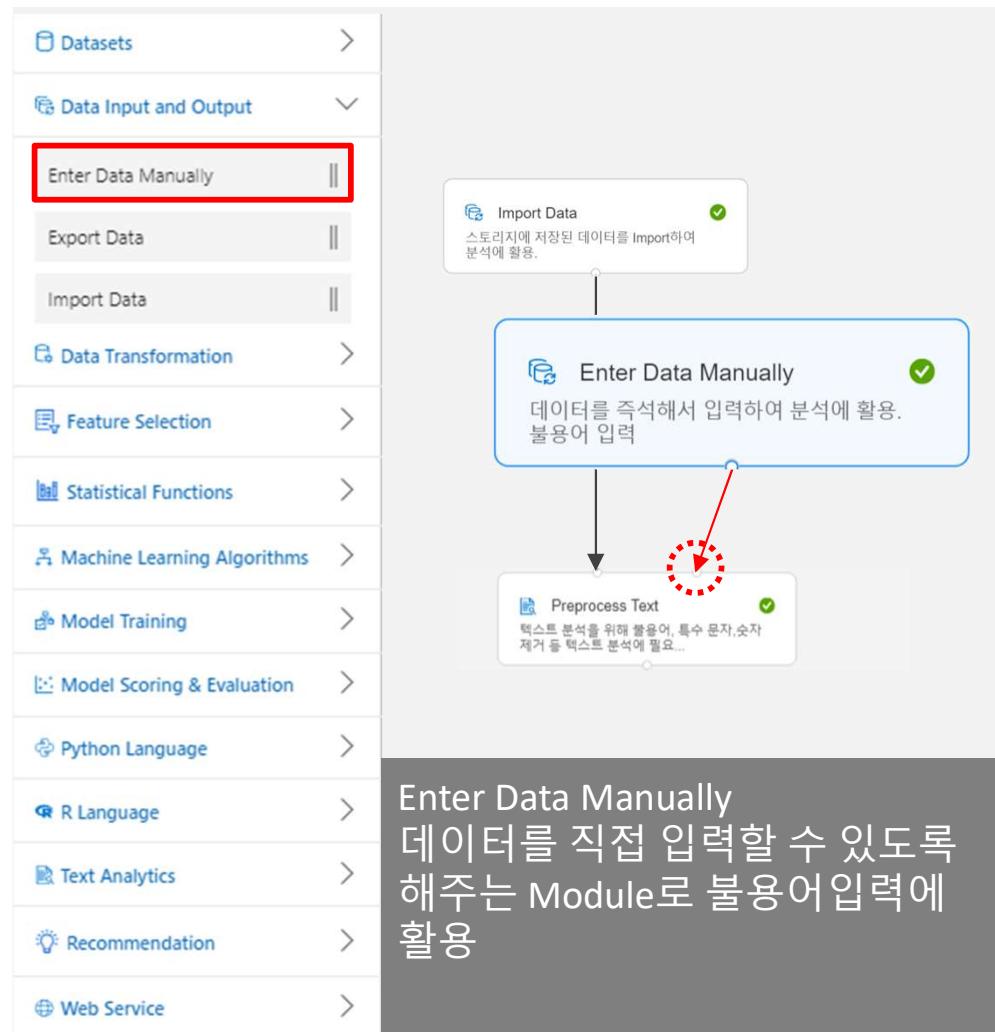
■ 데이터 불러오기

1. 데이터 원본(Data Source)에서 DataStore 선택
2. 데이터 저장소(DataStore)에서 데이터가 적재된 asogen2 storage를 선택
3. Browse Path를 클릭하여 Storage내에서 파일이 적재된 경로를 설정
4. Preview schema를 통해 적재된 데이터의 구조와 데이터 탑입을 확인

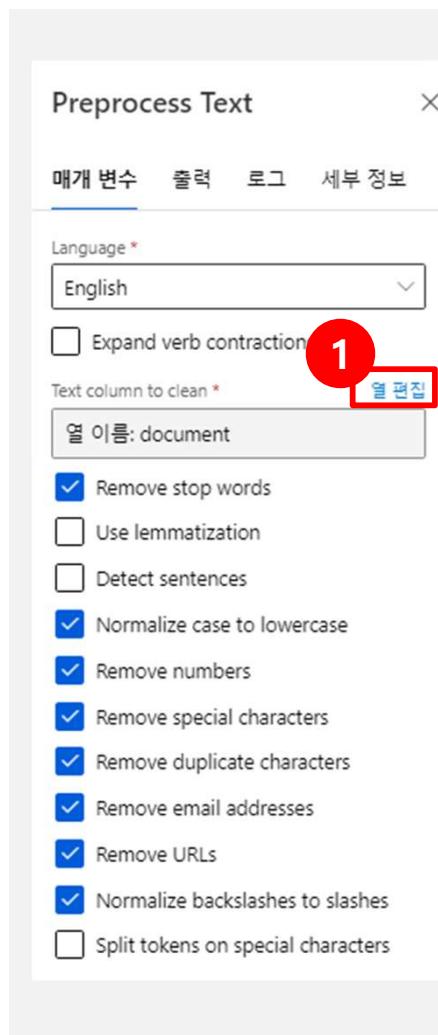
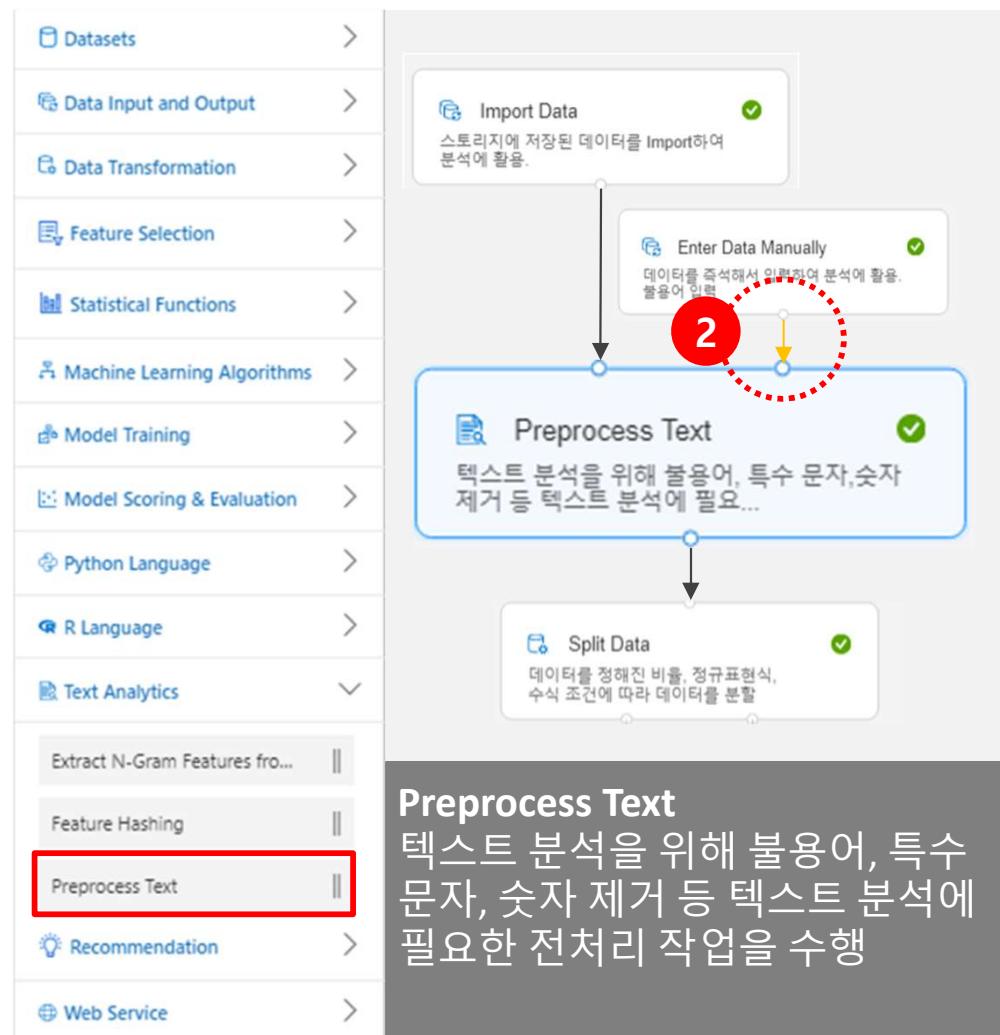
08 | 데이터 불러오기



09 | 데이터 직접 넣기(불용어 입력)

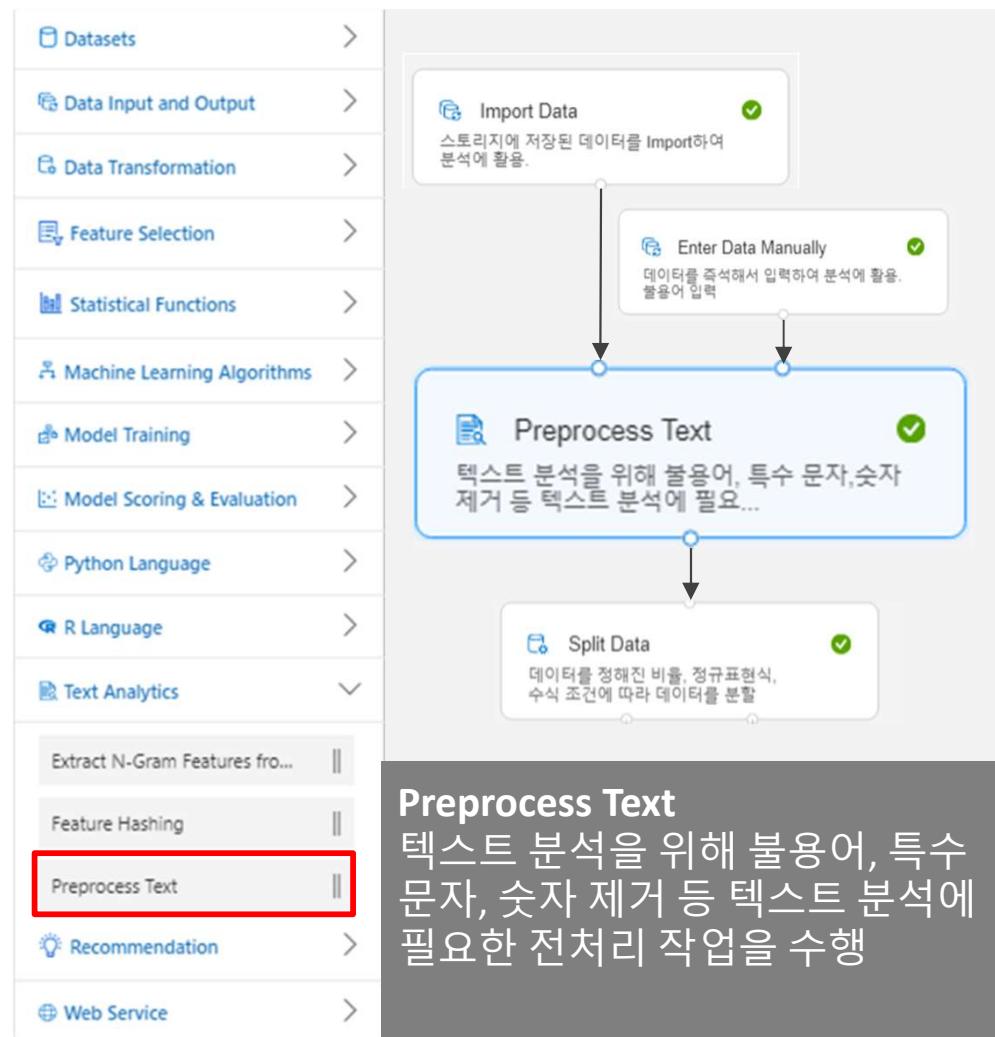


10 | Text 데이터 전처리 1



- Text 데이터 전처리
- 1. 전처리 컬럼 선택
- 2. Remove Stop Words
 - 두 번째 입력 포트를 통해 입력된 불용어 사전을 활용하여 불용어 삭제
 - 영어의 경우, 기본적인 불용어 삭제가 가능하며, 그 외 언어는 사전을 직접 입력 필요
- 3. Normalize Case to lowercase
 - 모든 영어를 소문자로 변환
- 4. Remove numbers
 - 숫자를 제거
 - 문자와 붙어 있는 경우, 삭제 불가

11 | Text 데이터 전처리 2



Preprocess Text

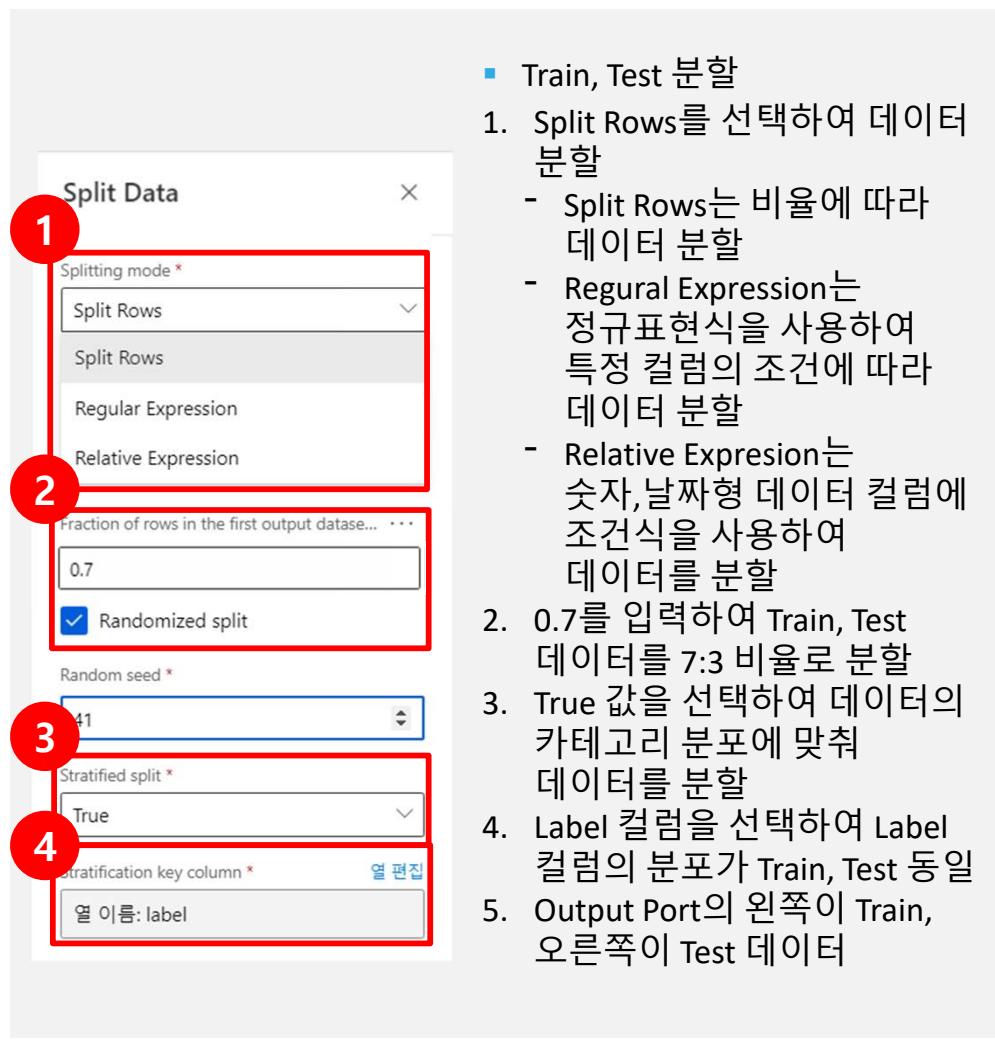
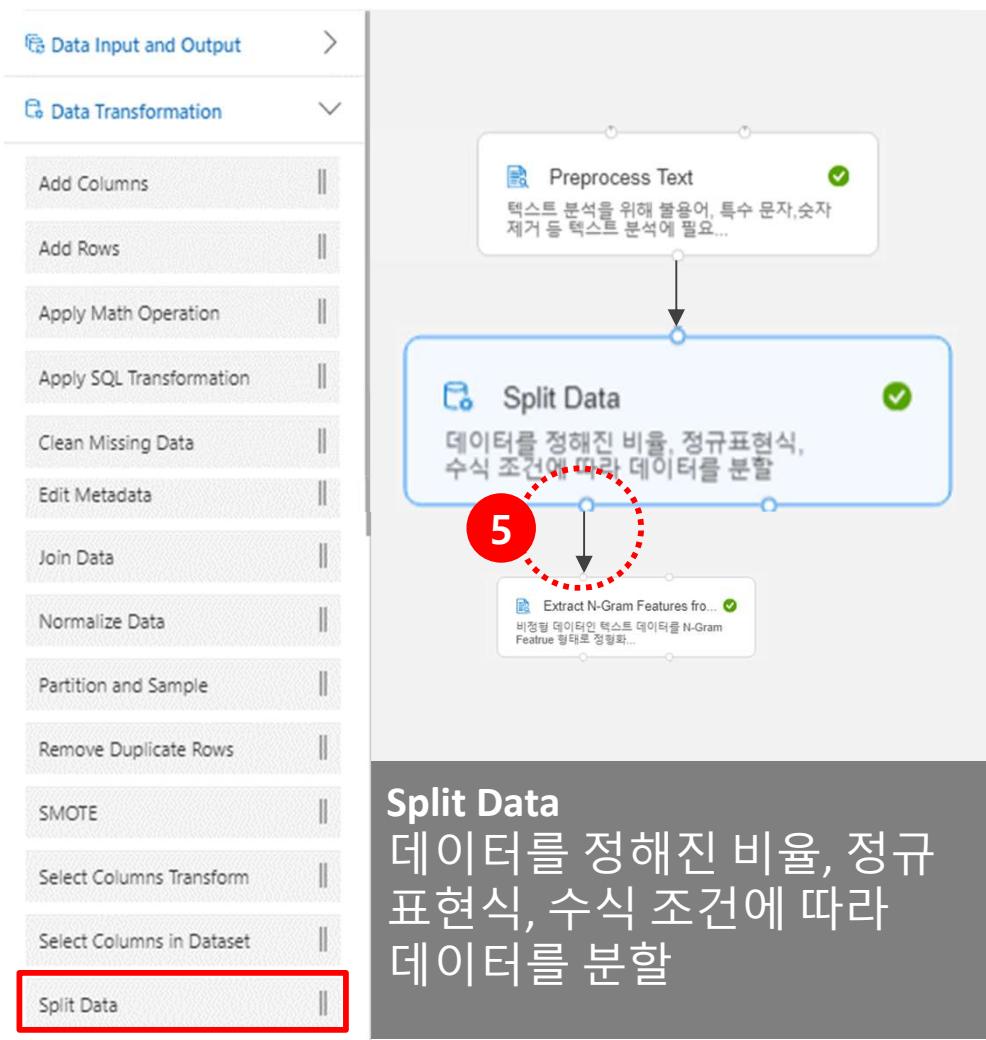
Preprocess Text settings:

- Language ***: English
- Text column to clean ***: 열 이름: document
- Operations** (checkboxes):
 - Remove stop words
 - Use lemmatization
 - Detect sentences
 - Normalize case to lowercase
 - Remove numbers
 - Remove special characters
 - Remove duplicate characters
 - Remove email addresses
 - Remove URLs
 - Normalize backslashes to slashes
 - Split tokens on special characters

Text 데이터 전처리

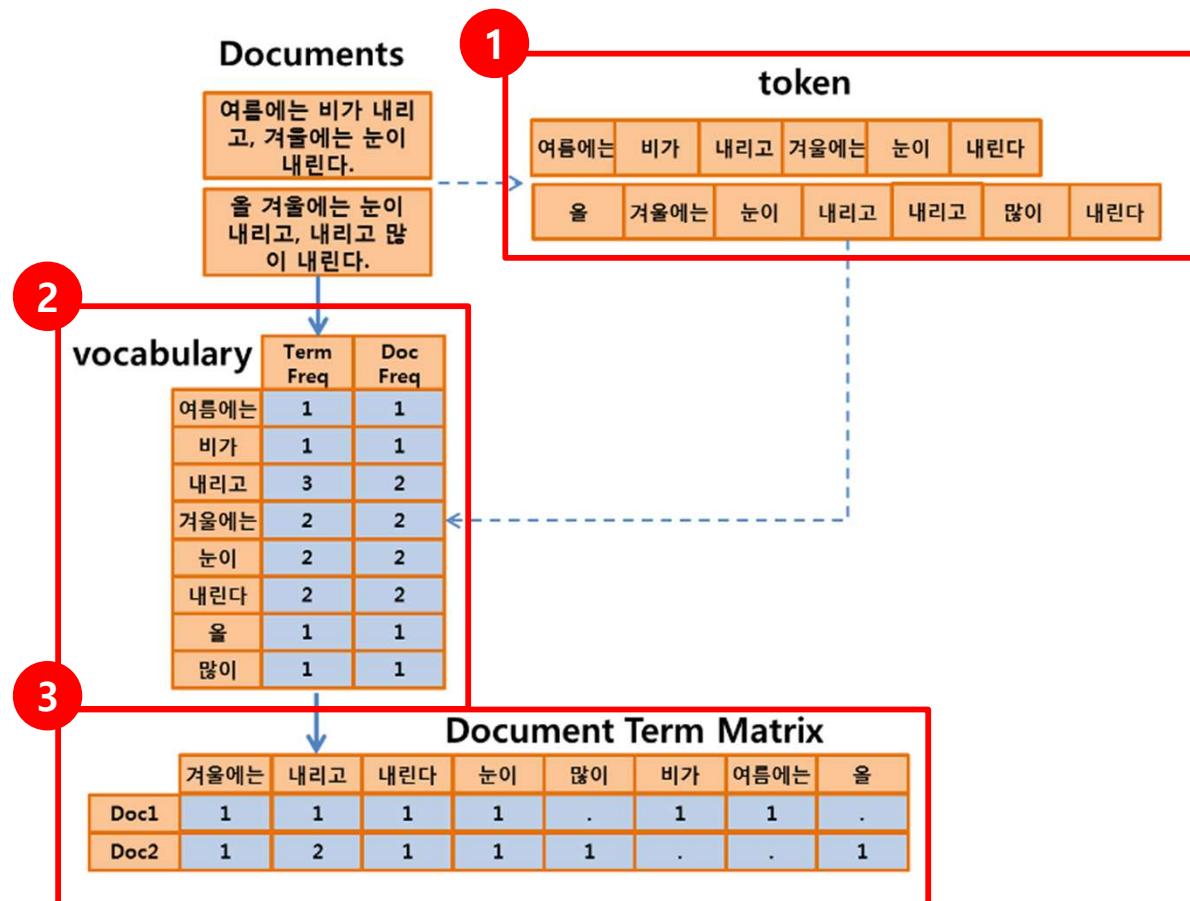
5. Remove special characters
 - 특수 문자 제거
6. Remove duplicate characters
 - 동일한 문자가 연속해서 사용된 경우 중복 제거
 - "aaaaa" ->"aa"
7. Remove email addresses
 - 문자열@문자열. 형태의 이메일 주소 제거
8. Remove URLs
 - http, ftp, www와 같은 url 주소 삭제
9. Normalize backslash to slash
 - \ (백슬래시)를 / (슬래시)로 통일

12 | Train/Test 데이터 분할



13 | Text 데이터 정형화

■ Document Term Matrix를 활용한 Text 데이터 정형화



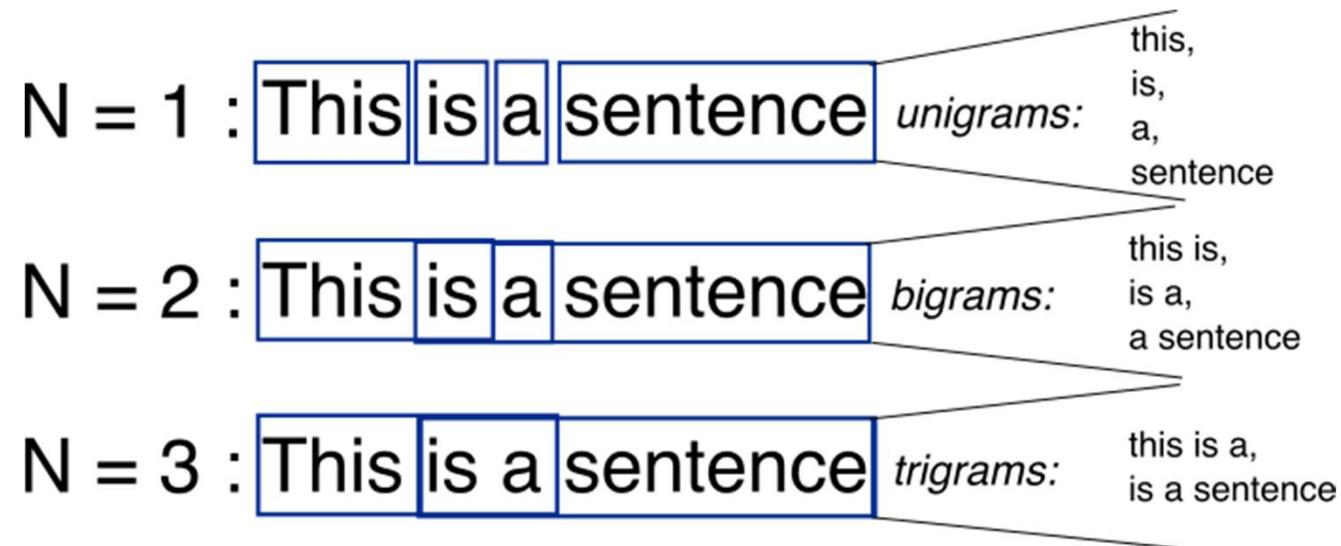
- DTM을 활용한 비정형 정형화

텍스트로 구성된 문서를 DTM을 통해 정형화하여 ML알고리즘을 사용하여 분류 작업 수행

1. 전체 문서에 대해 형태소, N-gram, White Space 기준으로 Token을 생성
2. 생성된 Token을 기준으로 출현 빈도 혹은 TF-IDF 값을 Value로 사용하여 테이블 형태로 표시
3. 2번 과정에서 생성된 Token을 특정 기준에 따라 필터링하여 문서별로 Token의 가중치(빈도, 출현 여부, TF-IDF)를 통해 테이블(DTM) 형태로 생성
4. Token을 하나의 Feature로 사용하여 DTM을 통해 비정형 데이터를 정형 분석 방법론을 통해 분석 가능

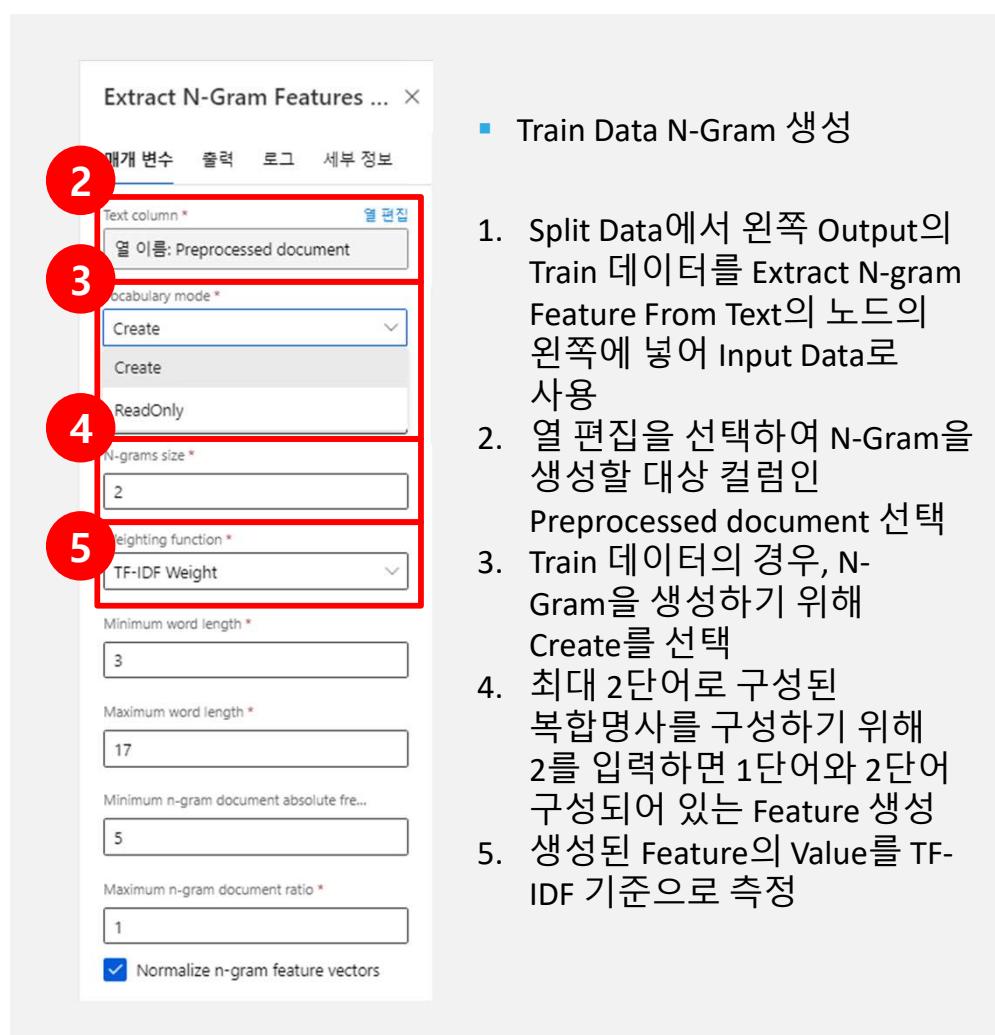
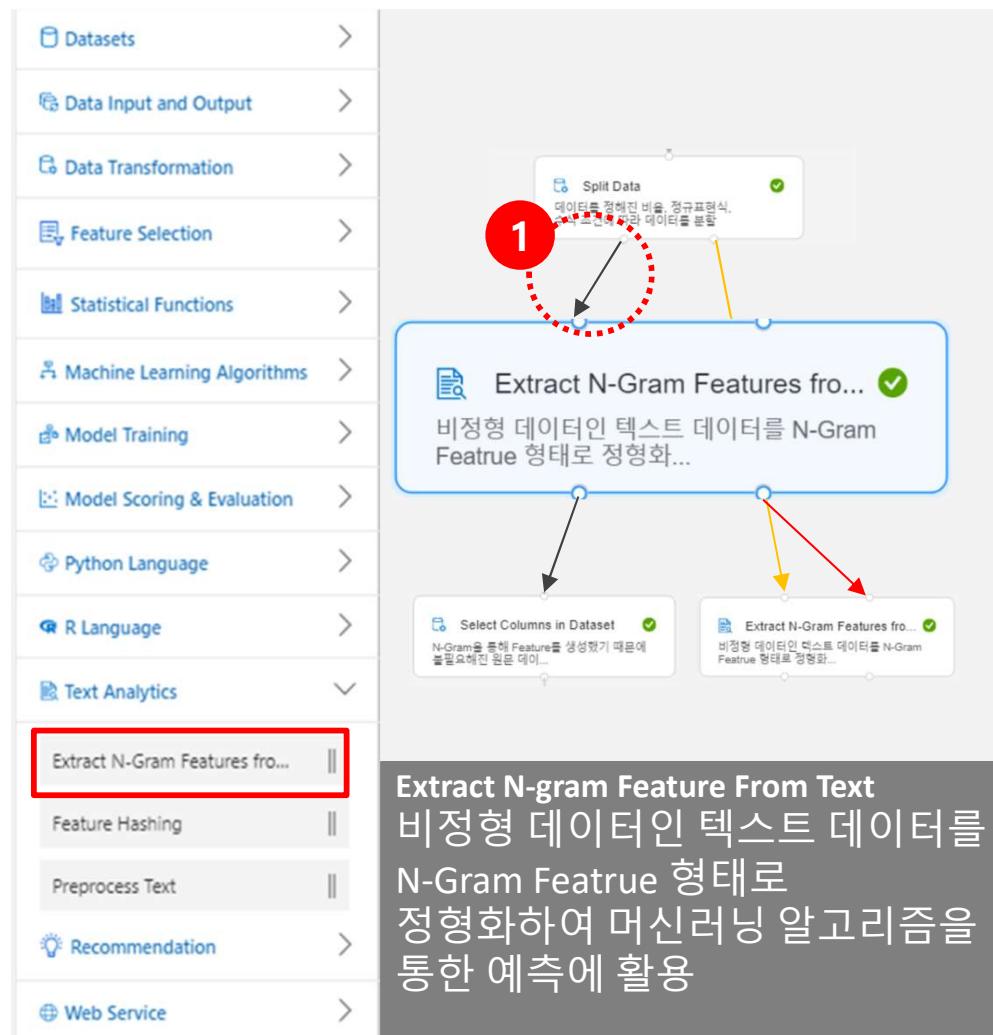
14 | Train/Train 데이터 N-gram Feature 추출

■ N-Gram 알고리즘



- 문서 내에 N-gram은 문장에서 원하는 단위(글자수 혹은 단어수)로 Term을 만들어 추출하여 반환하는 알고리즘
- Uni-gram: 1단어 혹은 1글자로 이루어진 Term 추출
- Bi-gram: 2단어 혹은 2글자로 이루어진 Term 추출
- Tri-gram: 3단어 혹은 3글자로 이루어진 Term 추출
- N-gram을 통해 추출된 단어를 분석 Feature로 활용

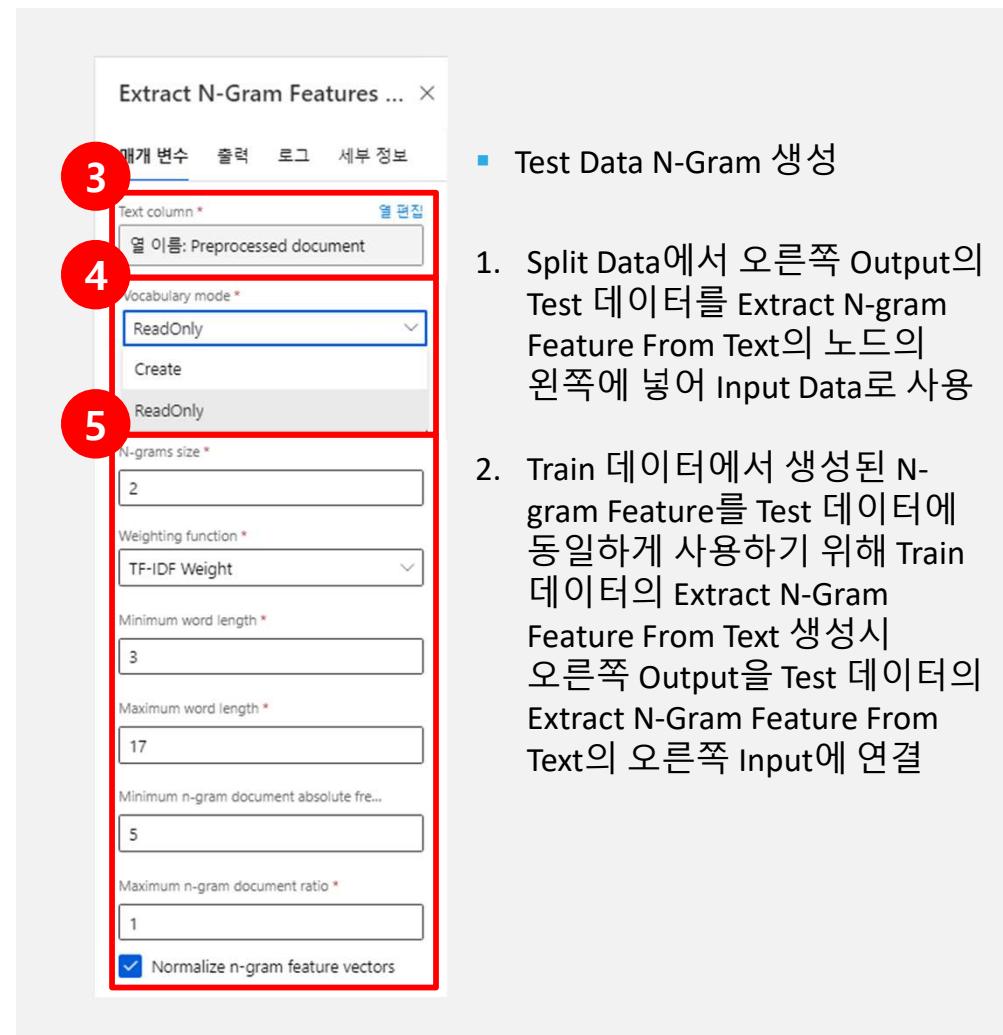
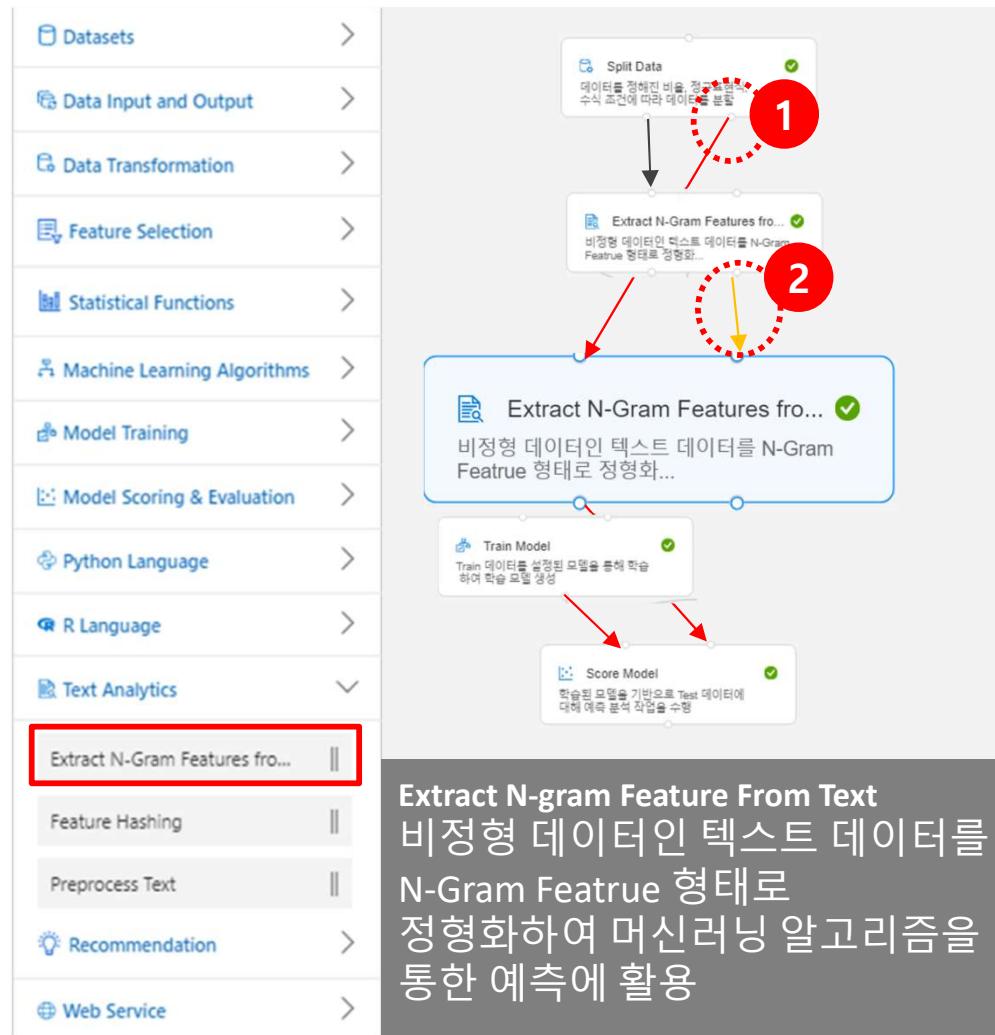
15 | Train 데이터 N-gram 추출



▪ Train Data N-Gram 생성

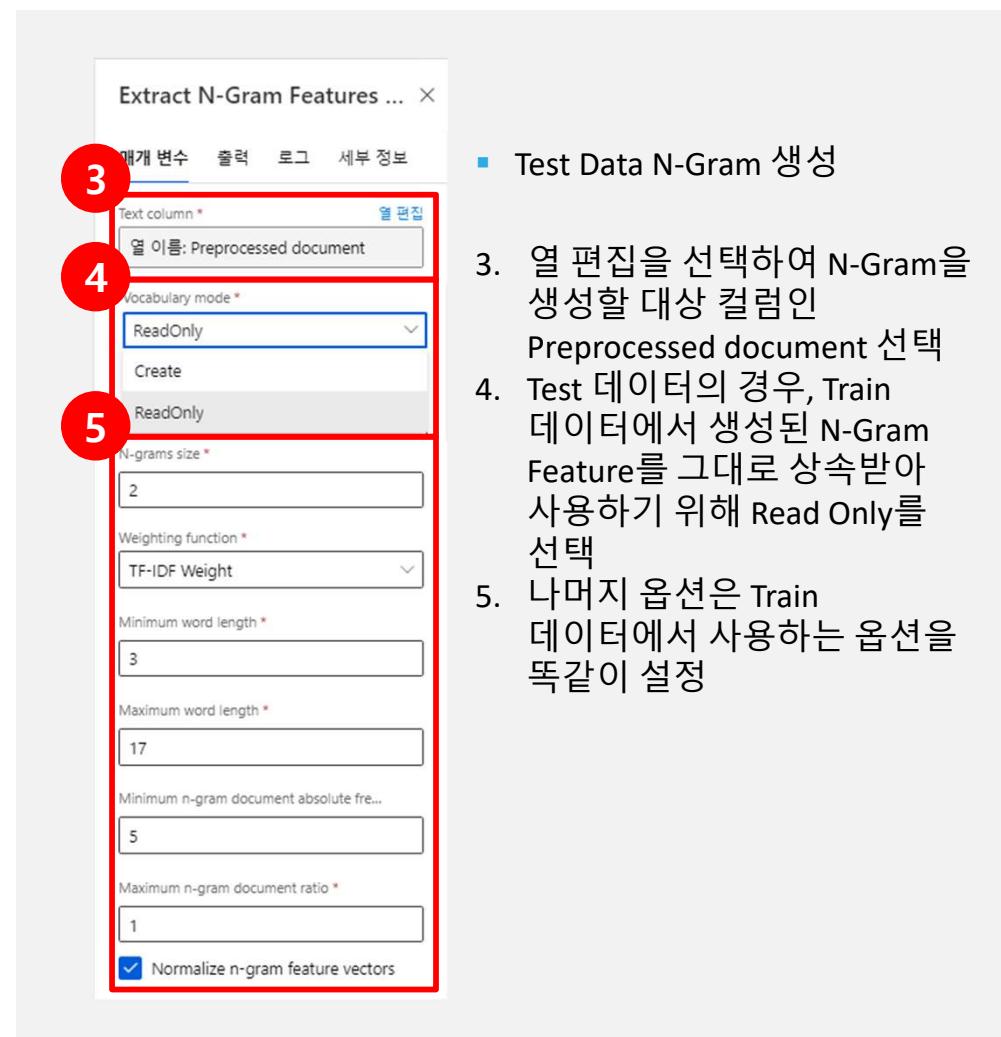
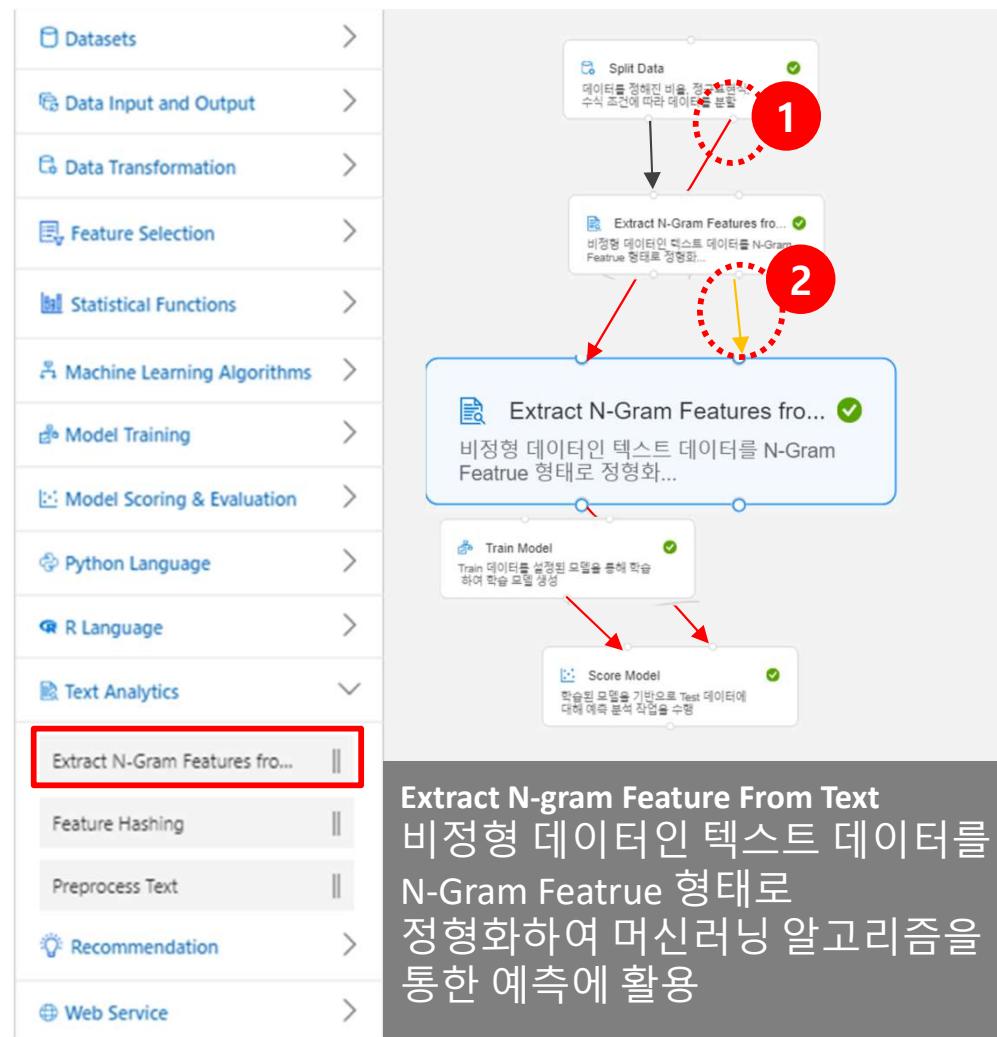
1. Split Data에서 왼쪽 Output의 Train 데이터를 Extract N-gram Feature From Text의 노드의 왼쪽에 넣어 Input Data로 사용
2. 열 편집을 선택하여 N-Gram을 생성할 대상 컬럼인 Preprocessed document 선택
3. Train 데이터의 경우, N-Gram을 생성하기 위해 Create를 선택
4. 최대 2단어로 구성된 복합명사를 구성하기 위해 2를 입력하면 1단어와 2단어 구성되어 있는 Feature 생성
5. 생성된 Feature의 Value를 TF-IDF 기준으로 측정

16 | Test 데이터 N-gram 추출 1



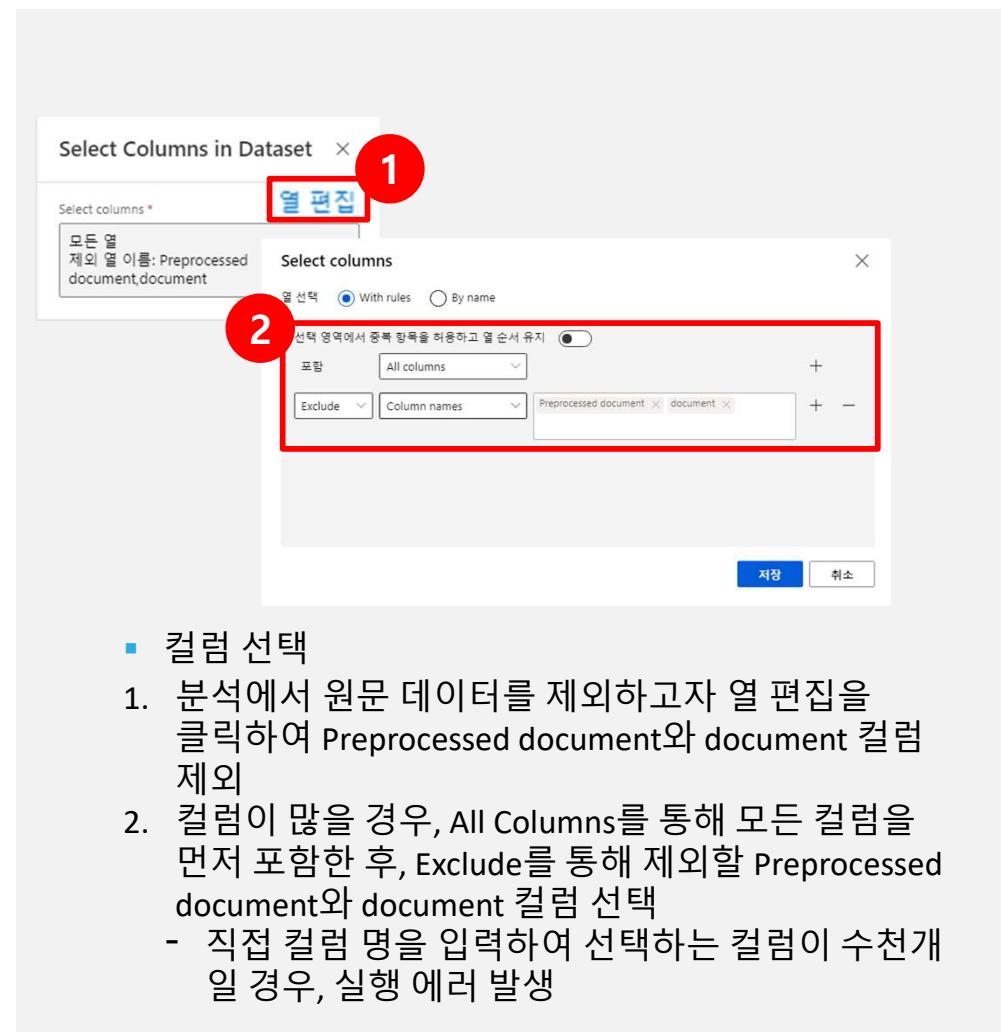
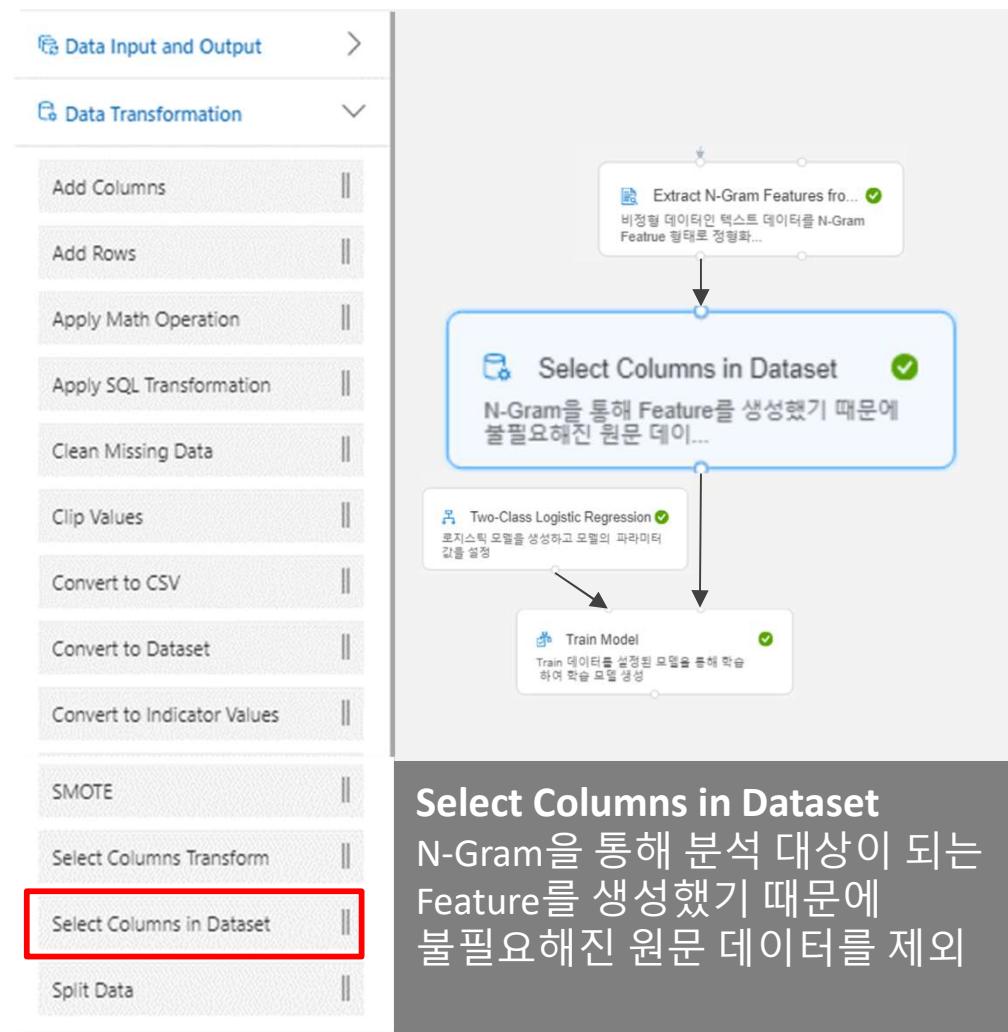
- **Test Data N-Gram 생성**
- 1. Split Data에서 오른쪽 Output의 Test 데이터를 Extract N-gram Feature From Text의 노드의 왼쪽에 넣어 Input Data로 사용
- 2. Train 데이터에서 생성된 N-gram Feature를 Test 데이터에 동일하게 사용하기 위해 Train 데이터의 Extract N-Gram Feature From Text 생성시 오른쪽 Output을 Test 데이터의 Extract N-Gram Feature From Text의 오른쪽 Input에 연결

17 | Test 데이터 N-gram 추출 2



- **Test Data N-Gram 생성**
- 3. 열 편집을 선택하여 N-Gram을 생성할 대상 컬럼인 Preprocessed document 선택
- 4. Test 데이터의 경우, Train 데이터에서 생성된 N-Gram Feature를 그대로 상속받아 사용하기 위해 Read Only를 선택
- 5. 나머지 옵션은 Train 데이터에서 사용하는 옵션을 똑같이 설정

18 | 컬럼 선택



■ 컬럼 선택

- 분석에서 원문 데이터를 제외하고자 열 편집을 클릭하여 Preprocessed document와 document 컬럼 제외
- 컬럼이 많을 경우, All Columns를 통해 모든 컬럼을 먼저 포함한 후, Exclude를 통해 제외할 Preprocessed document와 document 컬럼 선택
 - 직접 컬럼 명을 입력하여 선택하는 컬럼이 수천개 일 경우, 실행 에러 발생

19 | 모델 생성

Machine Learning Algorithms

Regression

- Boosted Decision Tree Regre...
- Decision Forest Regression
- Linear Regression

Classification

- MultiClass Boosted Decision...
- Multiclass Decision Forest
- Multiclass Logistic Regression
- Multiclass Neural Network
- One-vs-All Multiclass
- Two-Class Averaged Percept...
- Two-Class Boosted Decision ...
- Two-Class Decision Forest
- Two-Class Logistic Regression
- Two-Class Neural Network**
- Two-Class Support Vector M...

Two-Class Logistic Regression
로지스틱 모델을 생성하고
모델의 파라미터 값을 설정

Select Columns in Dataset

N-Gram을 통해 Feature를 생성했기 때문에
불필요해진 원문 데이터는 제거되었습니다.

Two-Class Logistic Regression

로지스틱 모델을 생성하고 모델의 파라미터
값을 설정

Train Model

Train 데이터를 설정된 모델을 통해 학습
하여 학습 모델 생성

Two-Class Logistic Regre... ×

매개 변수 출력 로그 세부 정보

1 Create trainer mode *

ParameterRange

SingleParameter

ParameterRange

2 optimization tolerance *

0.00001; 0.00000001

3 L2 regularization weight *

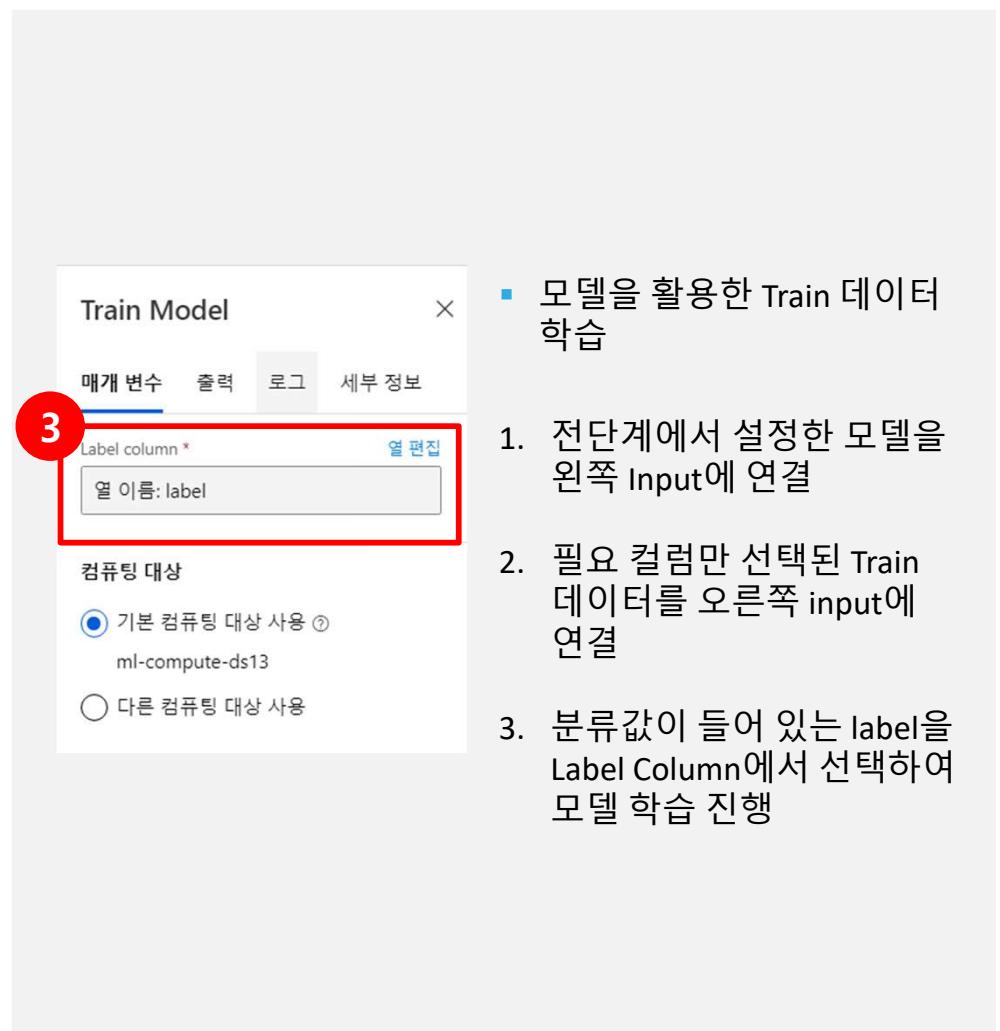
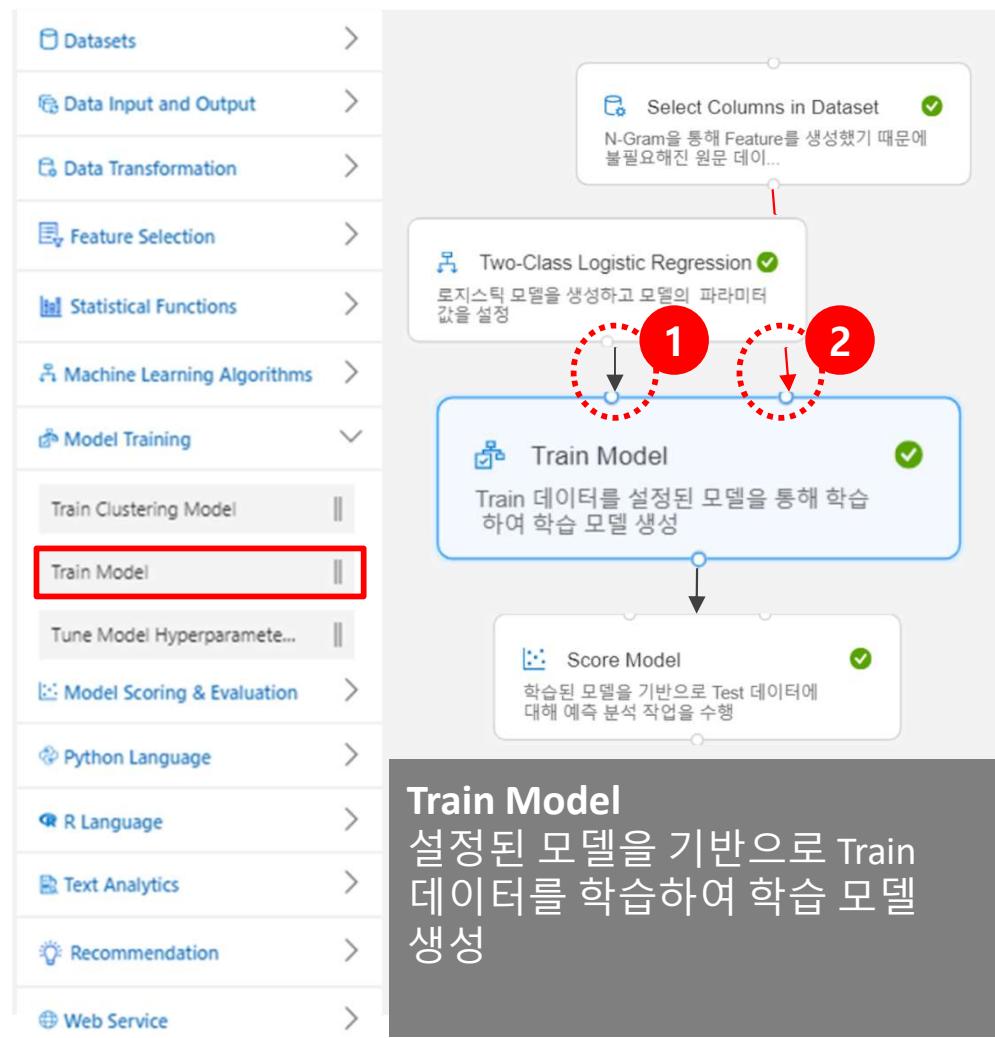
0.01; 0.1; 1.0

Random number seed

41

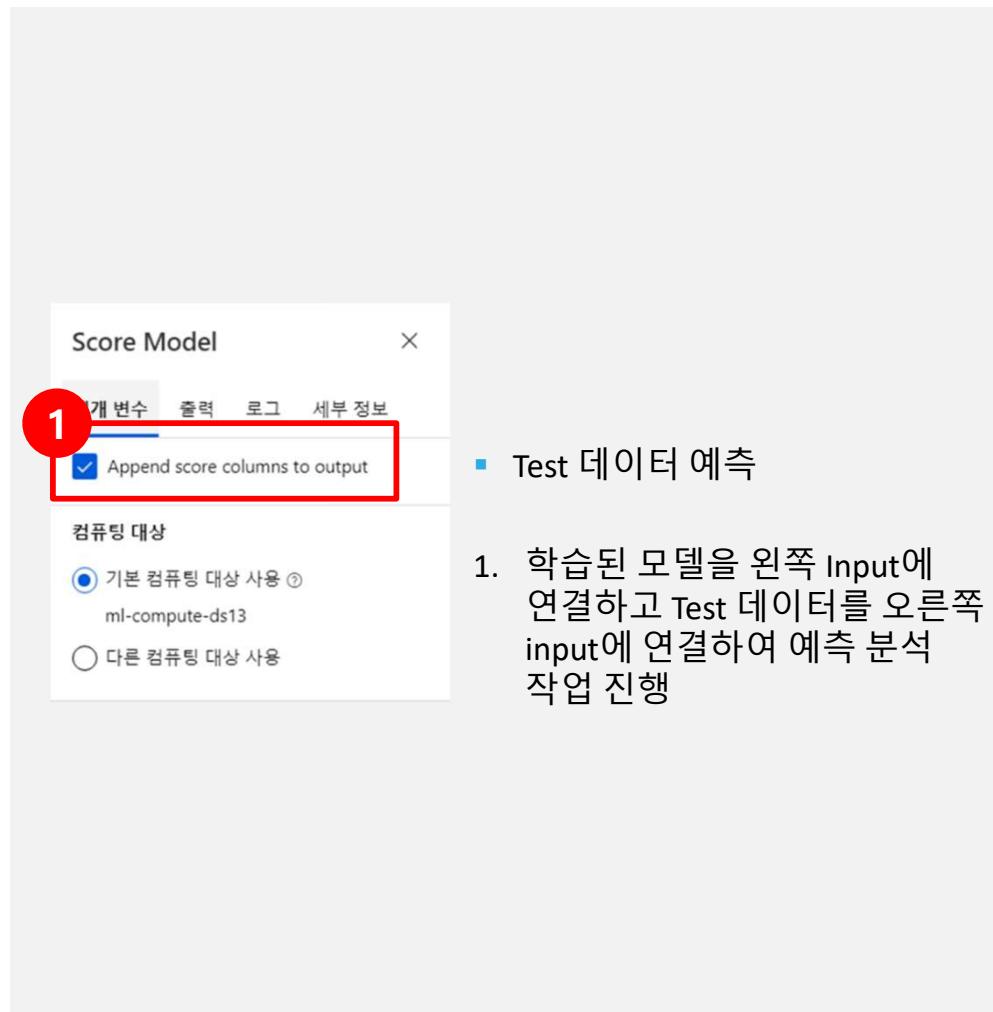
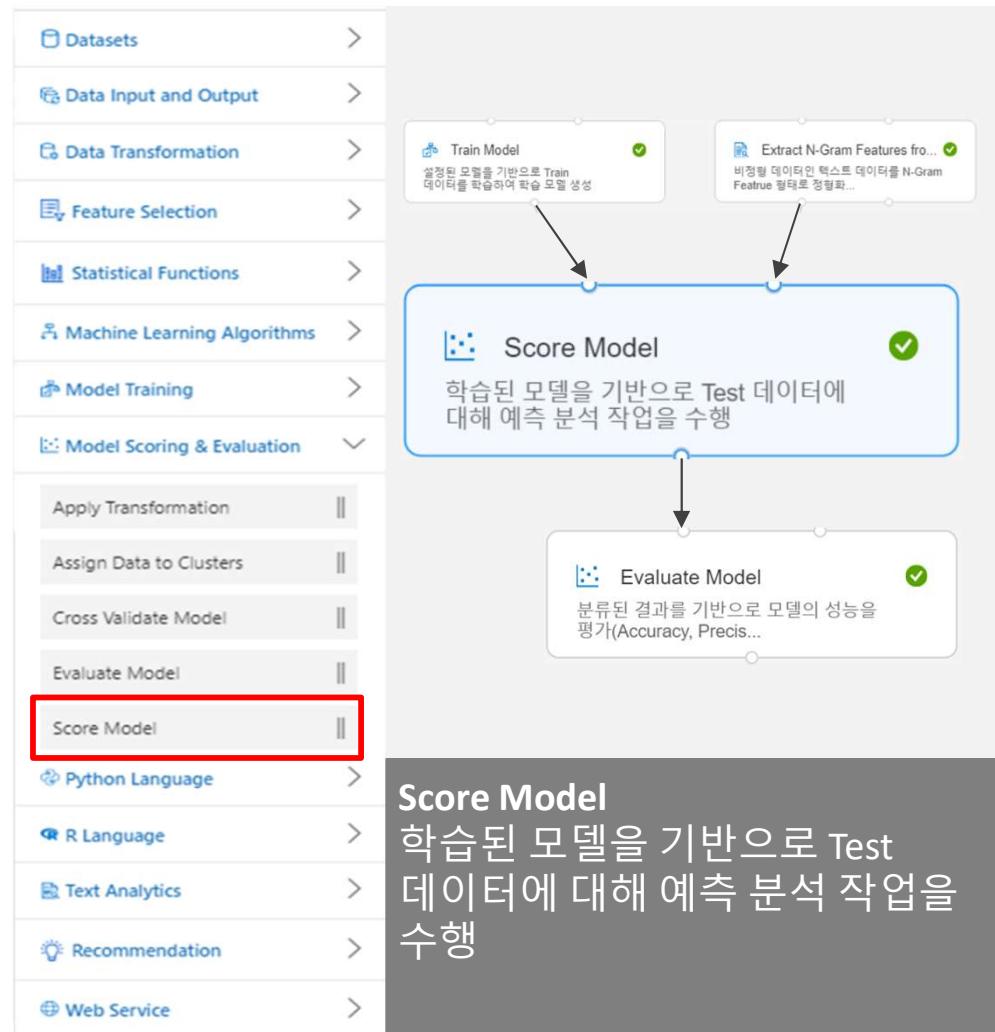
- 학습 모델 설정
- 1. 학습 모델을 생성할 때,
ParameterRange를 선택하여
HyperParameter 값을 단일
값이 아닌 여러 개의 값을
이용하여 학습을 진행하여
값을 비교
- 2. 모델을 최적화할 때, 이 값에
따라 모델이 학습을
반복하여도 일정 커버리지
이상 개선되지 않을 때 학습을
멈추도록 하는 커버리지 설정
- 3. 과적합을 막기 위한 규제
값으로 이 값이 높을 수록
일반화된 모델로 학습

20 | 모델 학습



- 모델을 활용한 Train 데이터 학습
- 1. 전단계에서 설정한 모델을 왼쪽 Input에 연결
- 2. 필요 컬럼만 선택된 Train 데이터를 오른쪽 input에 연결
- 3. 분류값이 들어 있는 label을 Label Column에서 선택하여 모델 학습 진행

21 | Test 데이터 예측



- Test 데이터 예측

1. 학습된 모델을 왼쪽 Input에 연결하고 Test 데이터를 오른쪽 input에 연결하여 예측 분석 작업 진행

22 | 모델 평가 1

The screenshot shows the Azure Machine Learning Studio interface. On the left, there is a sidebar with various modules listed under 'Model Scoring & Evaluation'. The 'Evaluate Model' module is highlighted with a red box. The main workspace contains two 'Evaluate Model' modules. One is a standard module, and the other is a larger, expanded view. This expanded view has a dark gray background and displays the text: 'Evaluate Model Test데이터에 대한 예측 결과를 기반으로 학습된 모델의 성능을 평가하고 시각화'. Above these modules is a 'Score Model' module with a green checkmark icon.

This screenshot shows the same workspace as the previous one, but the 'Evaluate Model' module is now fully expanded. A context menu is open over the expanded module, with the '시각화 Evaluation results' option highlighted by a red box. A green checkmark icon is visible in the top right corner of the expanded module area.

- 모델 성능 평가
 1. Module에서 시각화 Evaluation result를 클릭하여 시각화된 평가 결과 확인

23 | 모델 평가 2

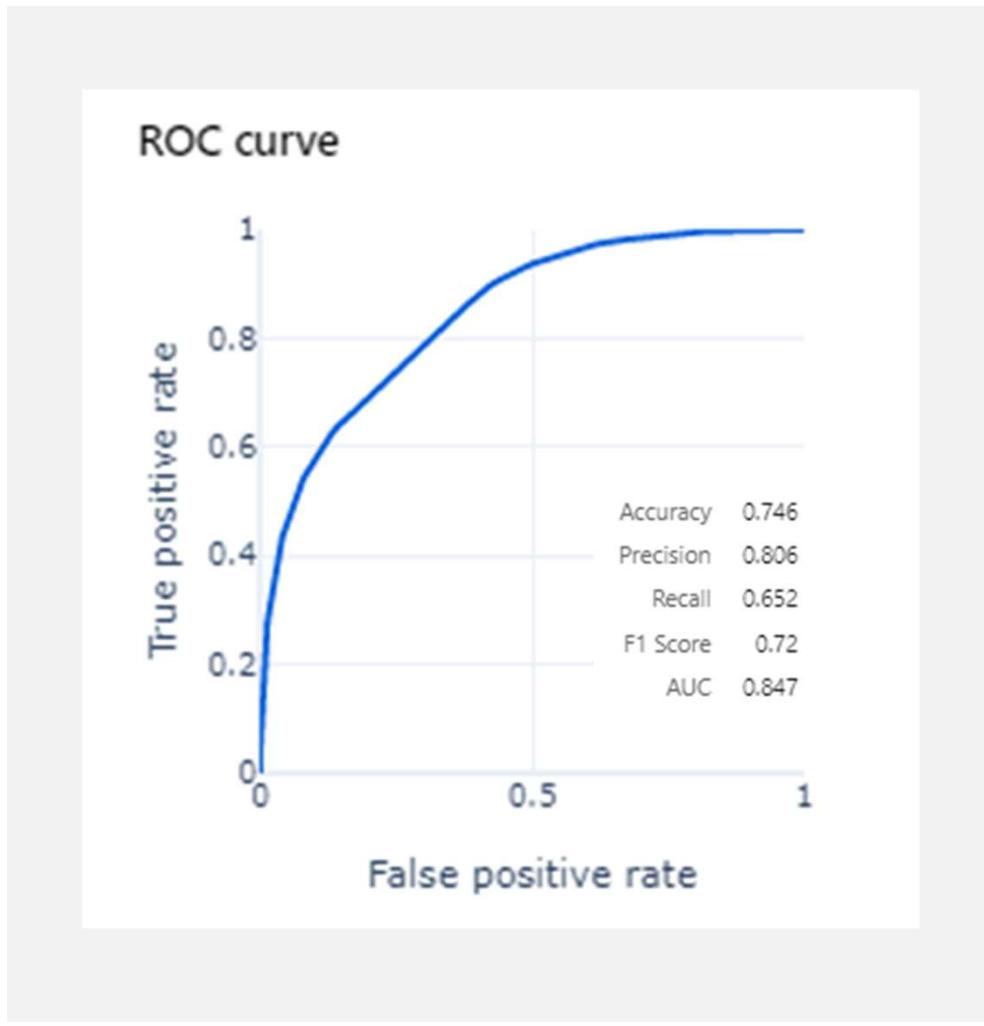
The screenshot shows the Azure Machine Learning Studio interface. On the left, there's a sidebar with various steps: Datasets, Data Input and Output, Data Transformation, Feature Selection, Statistical Functions, Machine Learning Algorithms, Model Training, Model Scoring & Evaluation, Apply Transformation, Assign Data to Clusters, Cross Validate Model, Evaluate Model (highlighted with a red box), Score Model, Python Language, R Language, Text Analytics, Recommendation, and Web Service. The main area displays two cards: 'Score Model' and 'Evaluate Model'. The 'Evaluate Model' card has Korean text: '분류된 결과를 기반으로 모델의 성능을 평가(Accuracy, Precis...)'.

Evaluate Model

Test데이터에 대한 예측 결과를 기반으로 학습된 모델의 성능을 평가하고 시각화



24 | 모델 평가 3 (ROC Curve)



ROC Curve

- Roc Curve란 모델의 성능을 비교하기 위한 지표
- 정상을 정상으로 비정상을 비정상으로 잘 분류했는지 한번에 보는 지표
- Roc Curve 아래의 범위인 Auc(Area Under Curve) 값이 클 수록 좋은 성능을 가진 모델

Auc 평가 기준

0.90-1 = excellent

0.80-0.90 = good

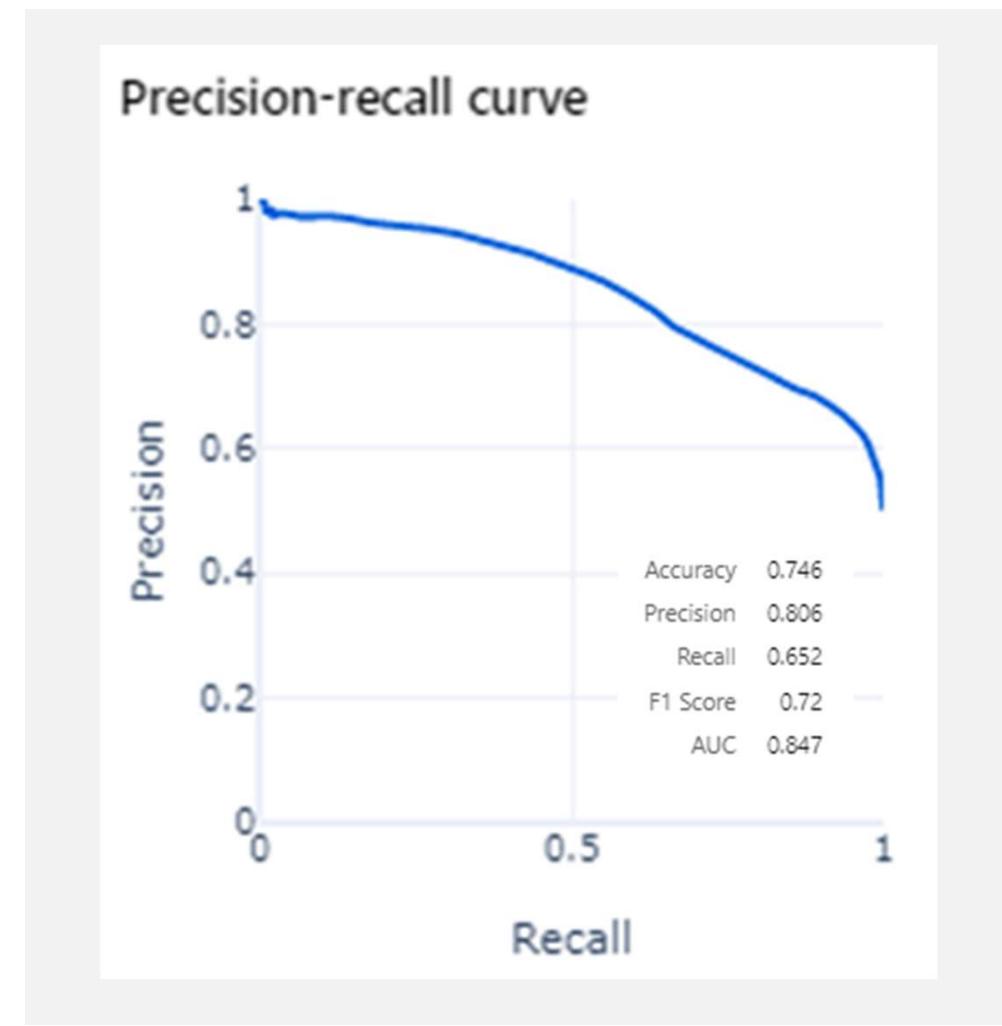
0.70-0.80 = fair

0.60-0.70 = poor

0.50-0.60 = fail

해당 모델의 경우, AUC 값이 0.847임으로 분류 모델로서 우수한 성능을 가짐

25 | 모델 평가 4 (Precision-Recall Curve)



■ Precision-Recall Curve

- True라고 분류한 것 중에서 실제 True인 것의 비율을 나타내는 Precision
- 실제 True인 것 중에서 모델이 True라고 분류한 비율인 Recall
- 전체 데이터를 True라고 예측하면 Precision이 극단적으로 높아지지만 Recall값은 감소
- 일반적으로 Precision과 Recall은 Trade-Off 관계이며 모델 평가에 있어 상호 보완적
- Precision과 Recall 모두 좋은 모델이 좋은 모델

해당 모델의 경우, Precision과 Recall과의 반비례 관계가 상당히 완만하게 감소하고 있어 Precision과 Recall 모두 좋은 것으로 볼 수 있음