

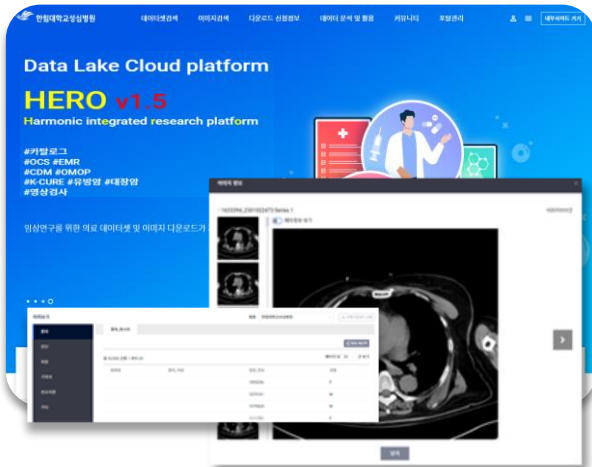
2024 한림대의료원 Clinical data 활용 경진대회 용 HERO Guide book

- INDEX -

1. HERO & 분석솔루션
2. 사용법
3. 문의처

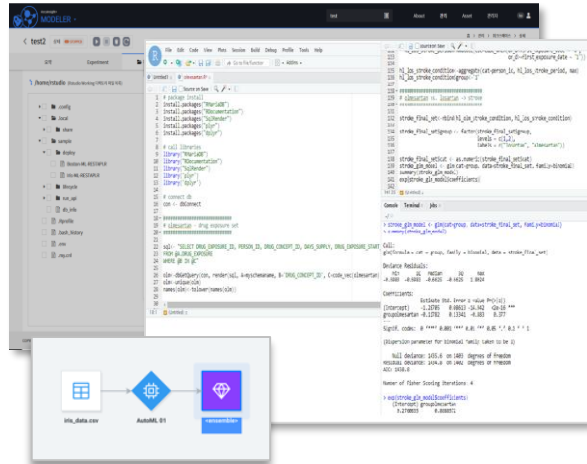
HERO & 분석솔루션

HERO



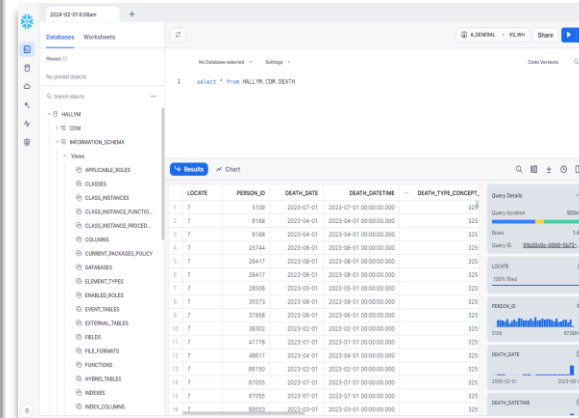
- 한림대학교의료원 차세대 연구 분석 포털 (2022.03.02 v1.0 오픈)
- **H**armonic int**E**grated **R**esearch plat**F**orm
- **D**ata **L**ake : OCS/EMR, CDM, K-CURE, 임상특화 DB, PACS 등의 Data 탑재
- Hybrid Cloud 환경 (Private & Public)
- 개인정보보호 및 비식별화 적용
- 경진대회 전용 데이터 셋 개방

AccuInsight+3.5



- SK의 클라우드 기반 다기능 분석솔루션
- 워크플로우를 통한 객체형 Task 코딩
- HDFS, Local file, DB 소스 연결 가능
- AutoML 기능으로 초심자 범용 AI 가능
- Sandbox를 통해 R, Python 코딩 환경
- 다기관 연계 분석 가능

Snowflake



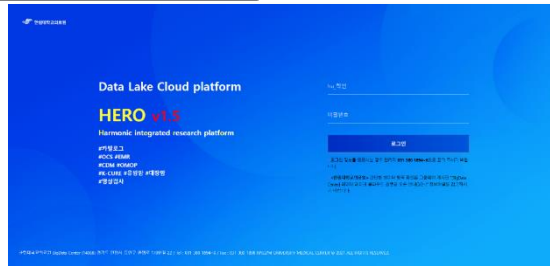
- AWS Public SaaS 설치 방식이며, Public의 RDBMS 개념
- 컴퓨팅 리소스 관리 용이
- DB/Schema/Table/Column Role 정의 가능 (부분 공유, 읽기/쓰기 등)
- On demand 형식 유연한 과금 정책
- Data를 안전하게 로드 및 사용자 간 공유
- 금융/의료/교육/경영 등 7200여개 Reference
- 현재 가장 Hot한 솔루션 중 하나

DataRobot



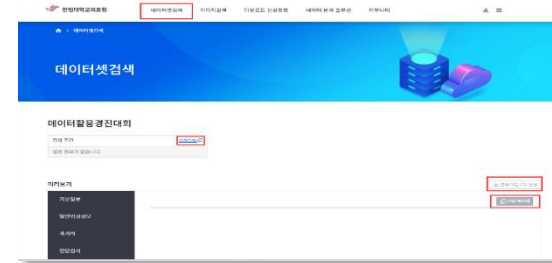
- AWS Public 설치
- AutoML 위한 모든 자동화 (데이터 준비/구축/배치/모니터링)
- Start 클릭의 손쉽고 편리한 UI
- 최적화 순서대로 수 십여 개 AI 알고리즘 구현
- 선택 알고리즘의 결과 리포트 출력
- 다양한 업무 적용 및 효율 기대

1. 접속하기



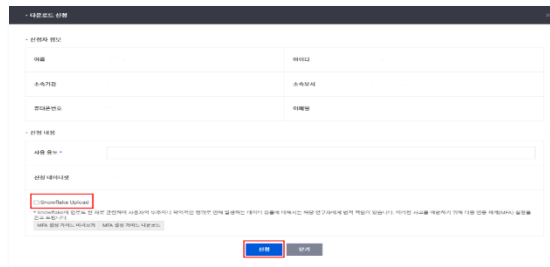
- www.hero.re.kr로 접속
- ID: hu_학번/ 초기 PW: 생년월일 6자리(yymmdd)
- 초기 PW 변경 후 사용 시작
- Chrome 또는 Edge 브라우저 이용

2. 데이터셋 검색하기



- 상단메뉴 {데이터셋검색} 클릭
- [검색설정] 클릭하여 기간 [검색]
- <미리보기> 확인 및 [전체 다운로드 신청] 버튼 선택

3. 데이터셋 다운로드 신청



- [사용 용도] 입력
- [snowflake 체크박스]를 체크 시 snowflake로 전송 가능 옵션(선택 요망)
- 체크 하지 않을 시 .csv파일로만 다운로드

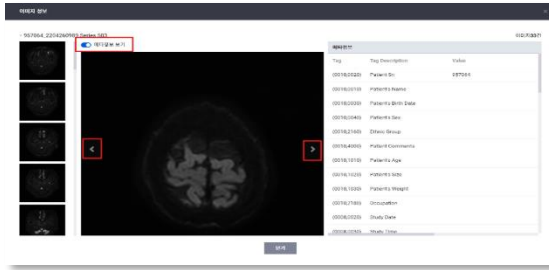
4. 이미지 검색하기



- 상단메뉴 {이미지검색} 클릭
- 좌측 이미지 목록 [폴더 트리] 클릭 및 중앙의 진단명/환자번호/기간조건을 입력하여 [검색]
- [돋보기 버튼] 클릭 시 상세 이미지 조회

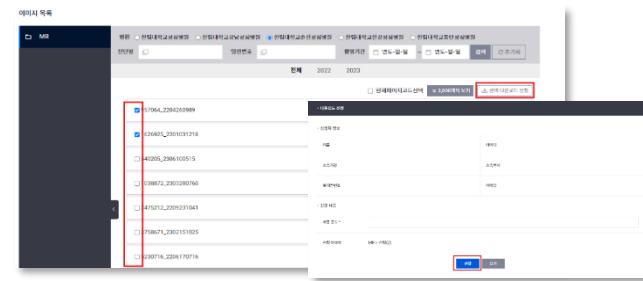
HERO 기본 사용법(계속)

5. 이미지 상세조회



- [메타정보 보기] 버튼을 클릭하여 메타 정보를 열기 또는 닫기
- Series 좌측 목록 또는 [화살표 버튼] 클릭하여 전/후 이미지로 이동

6. 이미지 다운로드 신청



- 다운로드 받을 [이미지 선택]한 후 [선택 다운로드 신청]버튼 클릭
- [사용 용도]를 입력한 후 [신청]

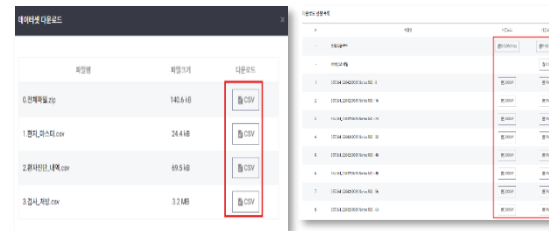
7. 다운로드 신청 확인



- 상단메뉴 [다운로드 신청정보] 클릭
- 데이터셋 또는 이미지 메뉴에서 신청된 목록 확인
- 승인된 목록만 다운로드 가능

8. 다운로드

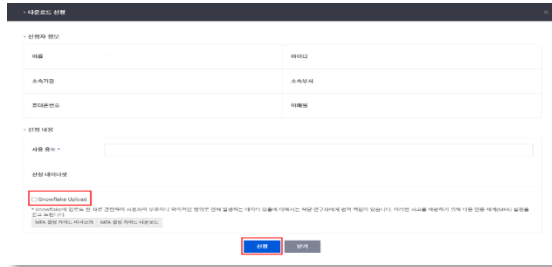
가산포인트



- 데이터셋은 csv 다운
- 이미지는 dcm 또는 png로 선택하여 다운 가능

HERO to Snowflake Pipeline 방법

1. 데이터셋 다운로드 신청



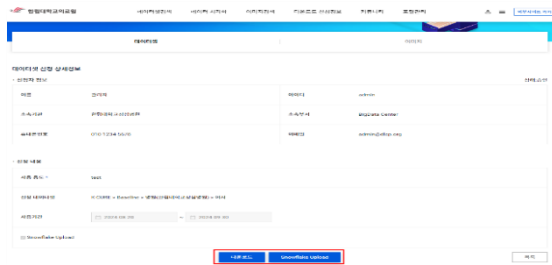
- 데이터셋 검색 후 다운로드 신청에서 [snowflake] 체크박스를 선택

2. 다운로드 승인



- 상단메뉴 [다운로드 신청정보]
- 데이터셋 신청 목록 중 승인 목록 선택

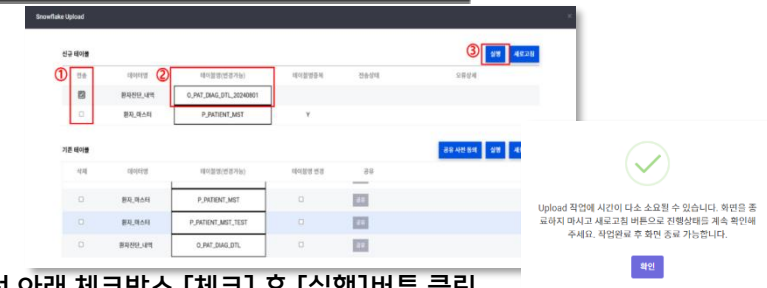
3. Snowflake Upload



- [Snowflake Upload] 버튼 선택

4. 데이터 업로드

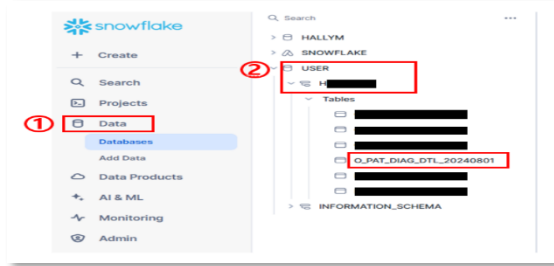
가산포인트



- 전송 컬럼 아래 체크박스 [체크] 후 [실행]버튼 클릭
- 전송 실행 완료 후 화면 종료 가능
- 테이블(신규,기존) 명 변경 가능 중복에 Y체크 시 필수 변경
- 테이블명은 영문 대문자,숫자,언더바(_)만 사용 가능, 첫 글자 영문 대문자

HERO to Snowflake Pipeline 방법

5. 데이터 확인하기



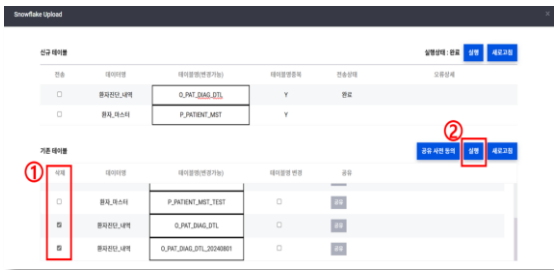
- 상단 메뉴 {데이터 분석 솔루션} > Snowflake 접속 후
- Data>Databases>USER>schema (H+hu_학번)>입력한 테이블 명

6. 기존 테이블 명 변경



- 기존 테이블 목록에서 변경할 테이블명 변경 체크 박스 선택 후
- 테이블 명 변경 후 실행(여러 테이블 동시 선택 가능)
- 테이블 명은 영문 대문자,숫자,언더바(_)만 사용 가능, 첫 글자 영문 대문자

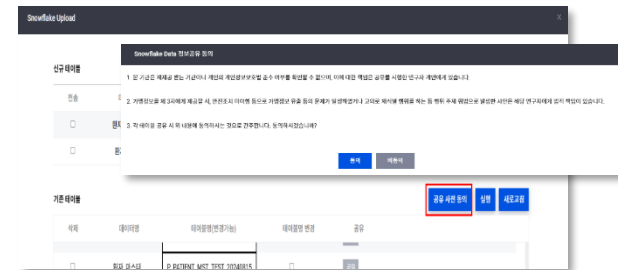
7. 기존 테이블 삭제



- 기존 테이블 목록에서 [삭제] 체크 박스 선택 후 [실행] (여러 테이블 동시 선택-실행 가능)

8-1. 테이블 공유

가산포인트

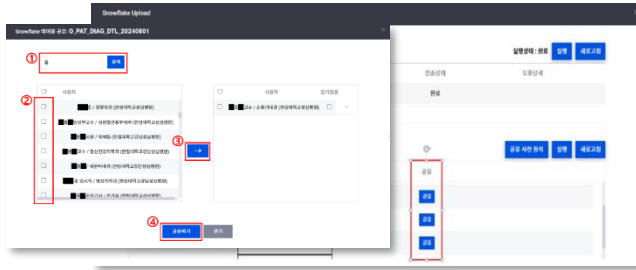


- 자신의 snowflake 테이블을 다른 사용자가 접근 할 수 있도록 하는 기능
- [공유 사전 동의] 버튼 선택 후 공유 버튼 활성화

HERO to Snowflake Pipeline 방법

8-2. 테이블 공유

가산포인트



- 공유 사전 동의 후 [공유] 버튼 클릭(공유는 개별 테이블 별로 가능)
- 사용자 성명으로 [검색] 후 해당 사용자를 [체크] 후 우측 화살표로 이동

8-3. 테이블 공유

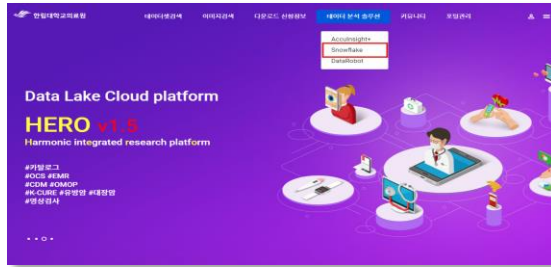
가산포인트



- [읽기전용] 체크박스 선택 시 select(읽기) 권한만 부여
- 체크 하지 않으면 읽기, 쓰기, 삭제(select, insert, update, delete) 권한이 부여
- 공유 권한을 회수하려면 오른쪽 사용자 목록에서 [X] 버튼 클릭 후 [공유하기] 버튼 클릭하여 반영

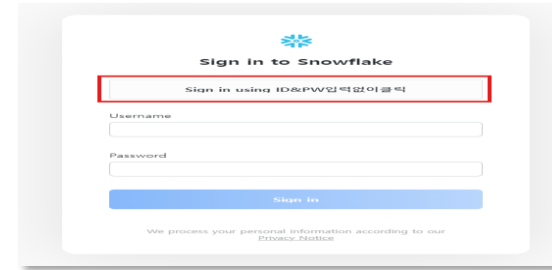
Snowflake 기본 사용법

1. 접속하기



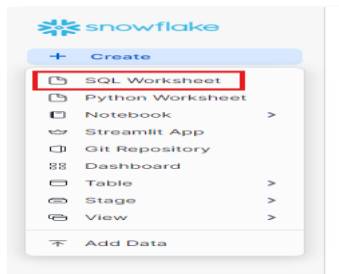
- www.hero.re.kr로 접속 & 로그인 후
- 상단메뉴 {데이터 분석 솔루션} > Snowflake

2. SSO 로그인



- [Sign In Using ID&PW입력없이클릭] Click

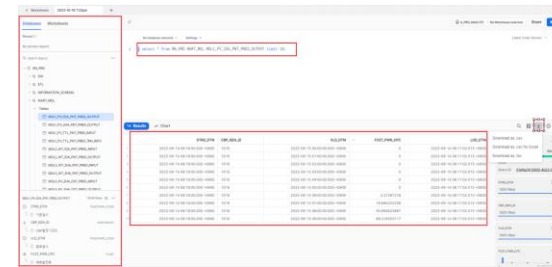
3. SQL Worksheet 생성



- 왼쪽 상단 +Create
- SQL Worksheet 선택

4. 데이터 조회

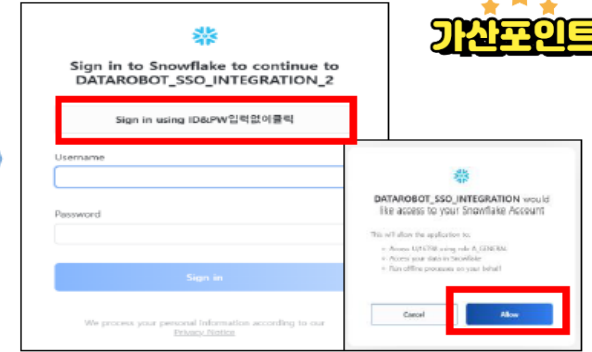
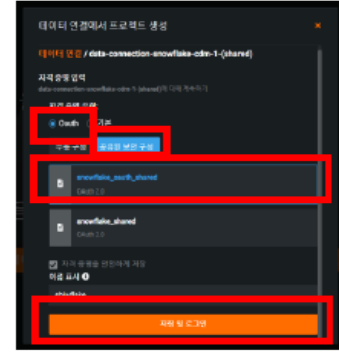
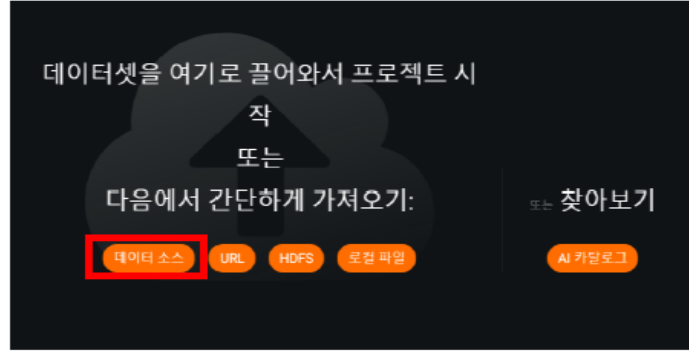
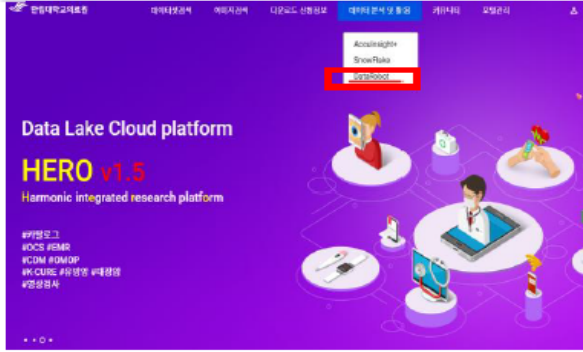
가산포인트



- 왼쪽 메뉴에 데이터베이스, 스키마, 테이블 구조 확인 가능
- SQL로 데이터 핸들링
- 결과 다운로드 가능

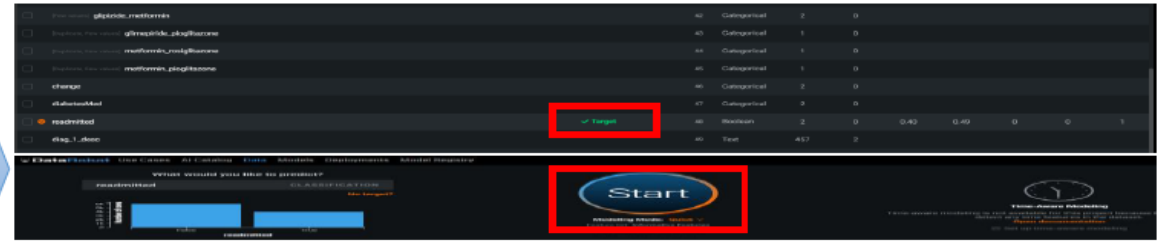
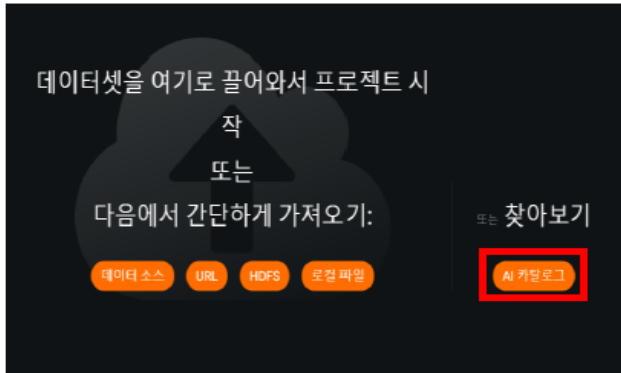
DataRobot 기본 사용법

01 DataRobot 로그인 및 계정 연동



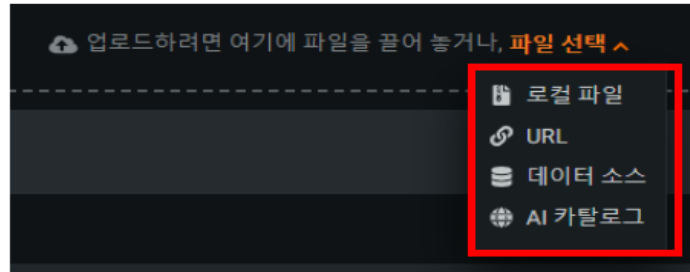
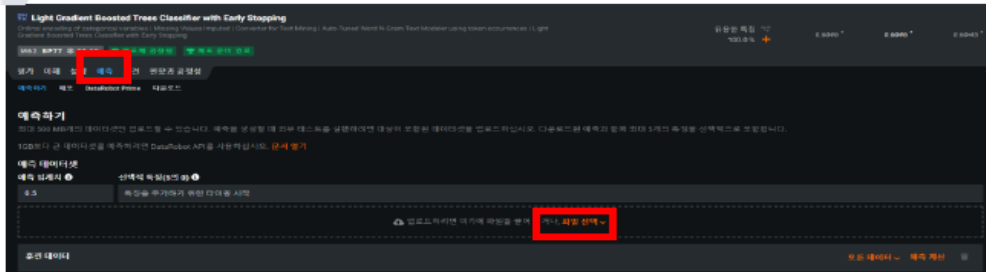
가산포인트

02 데이터 연결 및 모델을 학습



- Autopilot(오토파일럿) 모드는 가장 컴퓨팅 리소스를 철저히 최적화하여 계산하는 모드입니다
- Quick(신속) 모드는 Autopilot 모드의 간략한 버전입니다
- Manual(수동) 모드를 사용하면 리포지토리에서 실행할 개별 청사진을 선택할 수 있습니다. 어떤 종류의 모델링 접근 방식을 사용할지 이미 알고 있는 경우에 유용합니다.
- Comprehensive(종합) 모드를 사용하면 모델의 정확도를 높이기 위해 최대 샘플 크기를 사용하여 리포지토리의 모든 청사진을 실행할 수 있습니다

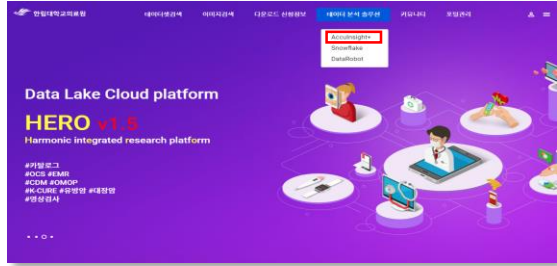
03 학습된 모델로 결과 예측



- 로컬파일은 개인이 가지고있는 데이터를 예측할때 사용합니다.
- URL은 URL 형태의 데이터를 예측할 수 있습니다.
- 데이터 소스는 Snowflake에 생성된 데이터를 사용하여 예측할때 사용합니다.
- AI 카탈로그는 모든사용자에게 공유된 데이터를 예측할때 사용합니다.

AccuInsight+ 기본 사용법

1. 접속하기



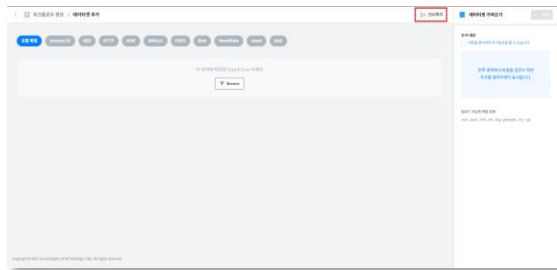
- www.hero.re.kr로 접속 & 로그인 후
- 상단메뉴 {데이터 분석 솔루션} > AccuInsight+

2. 워크플로우 생성



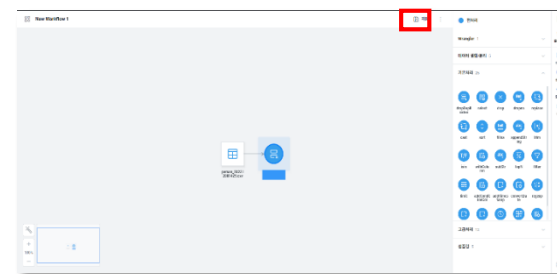
- 상단 {프로젝트} 메뉴 or 중앙화면 [프로젝트 생성] 후
- [워크플로우] 생성 후 [다음]

3. 데이터셋 가져오기



- 워크플로우의 [데이터 셋 추가] (건너뛰기 가능)
- [로컬 파일] Browse로 csv 로드 후 우측 상단 [추가] → 로드 완료

4. 전처리

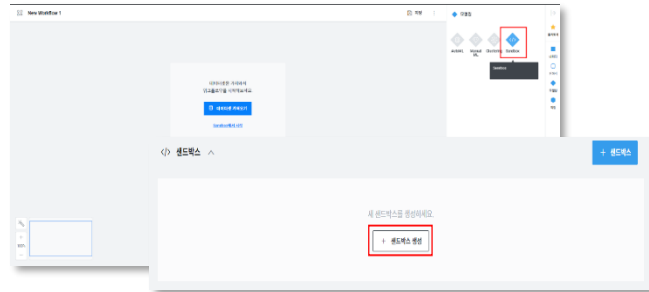


- 우측 패널의 전처리 항목 설정 후 [저장]

AccuInsight+ 기본 사용법(계속)

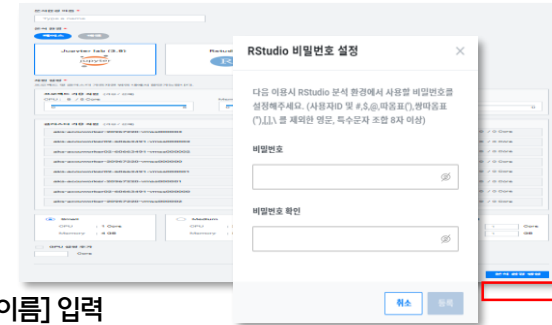
5-1. 모델링

가산포인트



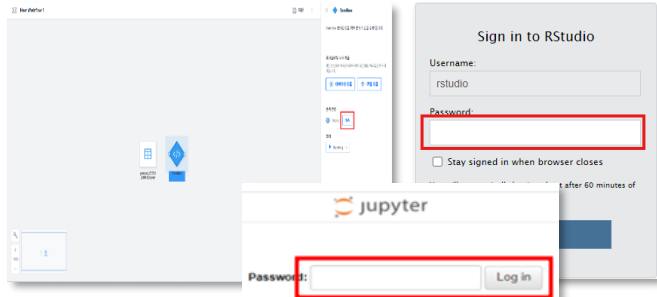
- 전처리 후 모델링에서 Sandbox 선택
- 또는 프로젝트에서 [샌드박스] 생성 가능

5-2. Sandbox 생성



- [분석환경 이름] 입력
- 분석환경(Jupyter lab 또는 Rstudio) 선택 후 [분석 환경 생성] 클릭
- 비밀번호 설정

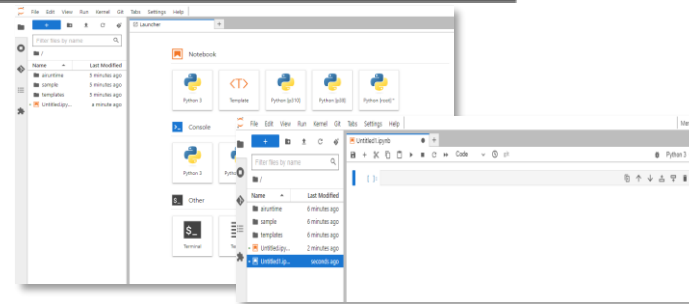
6. Sandbox 접속



- 본인 샌드박스 선택 후 [접속] 클릭
- 본인 설정한 비밀번호 [입력]

7. 분석 환경 활용

가산포인트



- 노트(코딩) 환경 선택
- 클라우드 환경 코딩

문의처 (한림대학교의료원 BigData Center)

- **경진대회** **장수연** 매니저 031-380-1805 suyeon@hallym.or.kr
- **HERO** 기능 **김윤희** 매니저 031-380-1947 yoonhee@hallym.or.kr
- **Snowflake** **임성빈** 매니저 031-380-1895 sbim@hallym.or.kr
- **DataRobot** **도웅** 매니저 031-380-1945 wdoh@hallym.or.kr
- **Acculnsight+** **김윤희** 매니저 031-380-1947 yoonhee@hallym.or.kr