

빅데이터 전처리

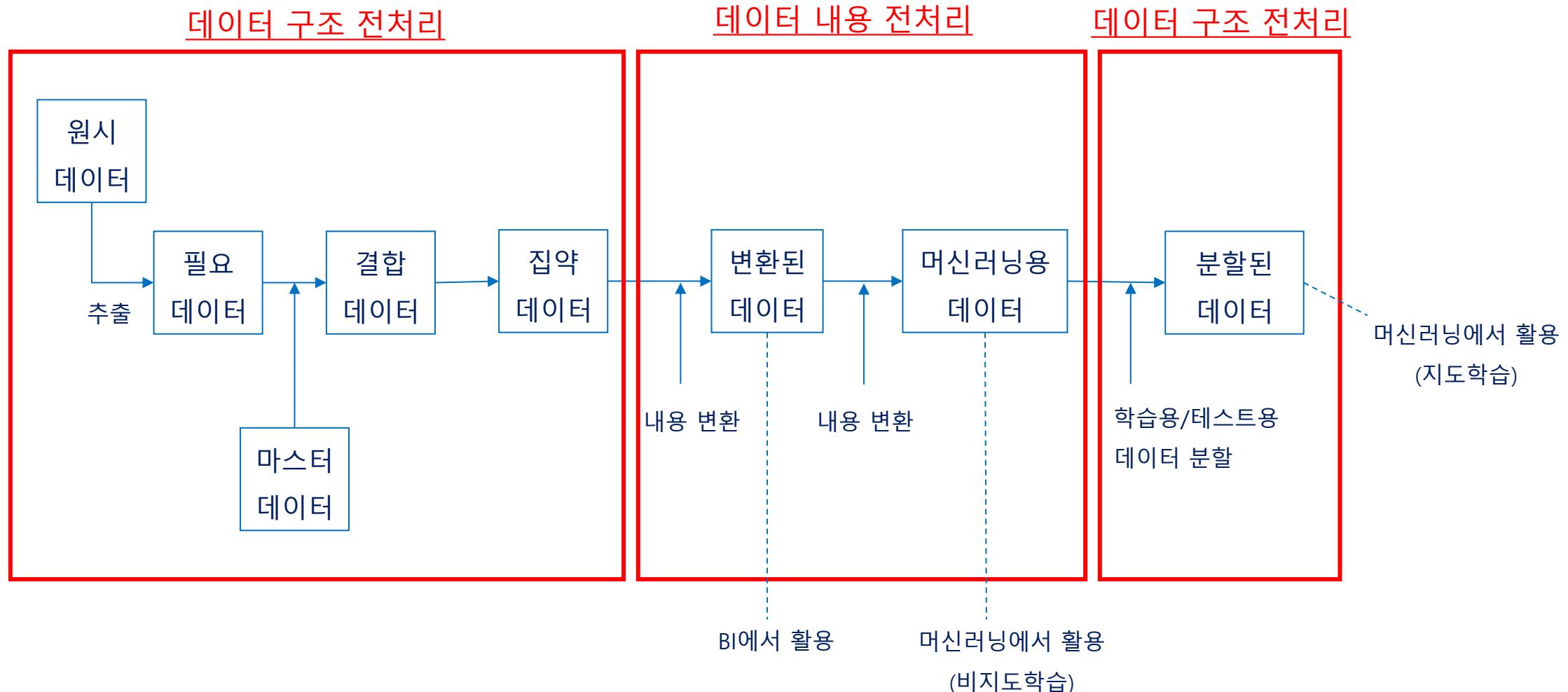
1회차 – 전처리 개요

전처리의 역할

- 데이터 분석에서 전처리란 데이터 분석을 준비하기 위한 처리를 의미한다.
- 데이터 전처리는 다음 세 종류로 나눌 수 있다.
 - 지표, 표, 그래프 작성용 전처리 : 지표의 계산, 표나 그래프로 쉽게 변환할 수 있는 데이터를 준비하기 위한 전처리 과정
 - 지도 학습용 전처리 : 학습 데이터, 테스트 데이터, 적용 데이터 세 종류를 준비하는 것을 목표로 한다.
 - 비지도 학습용 전처리 : 비지도학습 모델이 이용할 설명 변수를 가진 데이터를 준비하는 전처리 과정이다.

전처리의 흐름

- ▶ 전처리에서 학습할 내용의 흐름을 정리하면 아래 그림과 같이 표시할 수 있다.



본 학습에서 사용할 데이터셋

➤ 본 학습에서 다음 데이터 셋을 활용한다.

- 호텔 예약 레코드(reserve_tb) : 호텔 예약 정보를 모아둔 데이터 세트
- 호텔 마스터(hotel_tb) : 호텔의 기본 정보를 모아둔 데이터 세트
- 고객 마스터(customer_tb) : 고객의 기본 정보를 모아둔 데이터 세트
- 공장 제조 레코드(production_id) : 공장 제조물 내용과 제조 결과를 묶은 데이터 세트
- 월별지표 레코드(monthly_index_tb) : 월별 소매점 매출 등의 지표를 묶어둔 데이터 세트

실습 환경 구축

Step1> 브라우저를 열고 본인 계정으로 github에 로그인 한다.

Step2> 브라우저를 열고 주소창에 아래 주소를 복사하여 붙여넣기 한다.

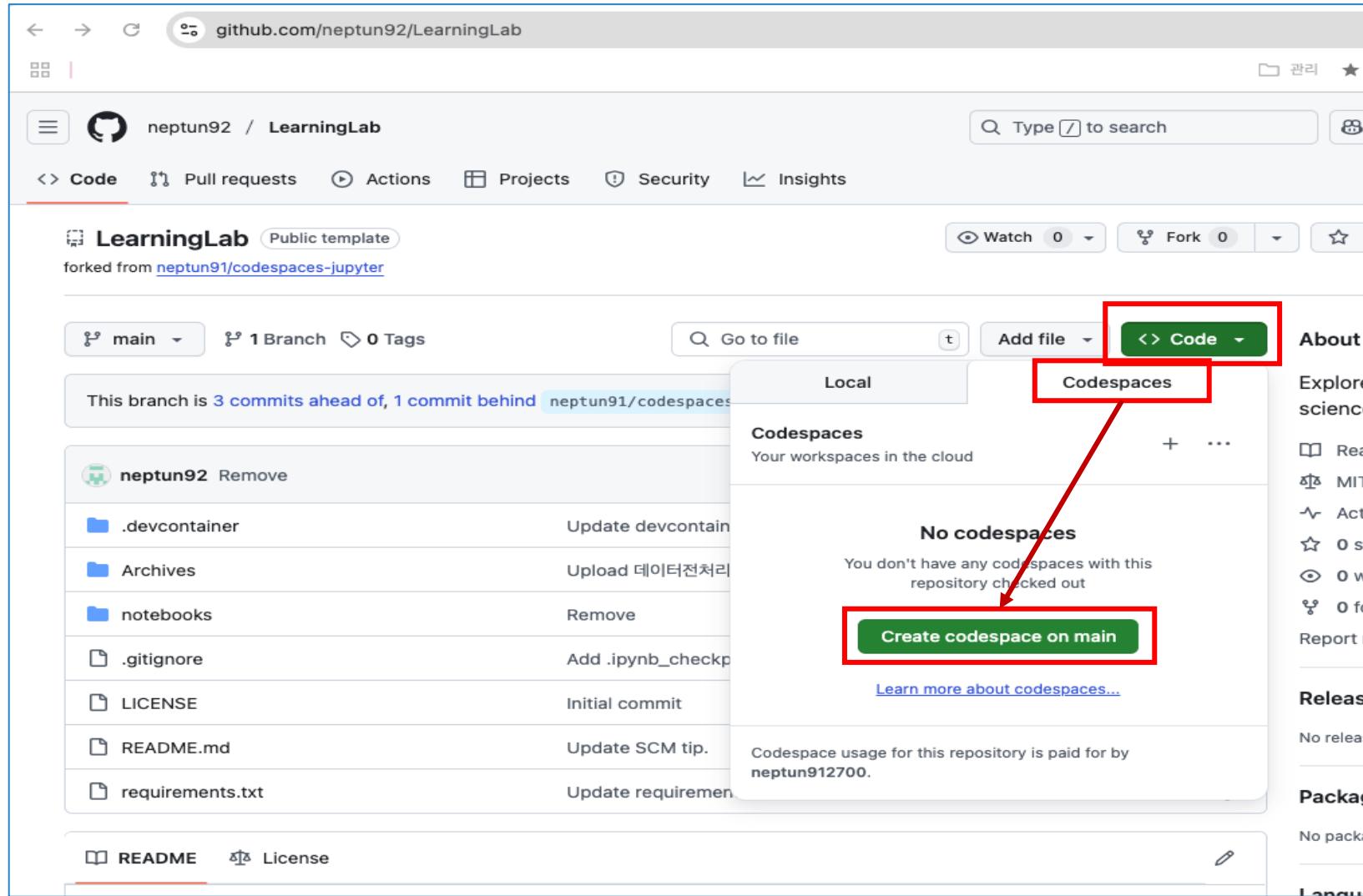
- <https://github.com/neptun92/Preprocessing-excersize>

The screenshot shows a GitHub repository page for 'neptun92 / LearningLab'. The repository is a public template forked from 'neptun91/codespaces-jupyter'. It has 1 branch and 0 tags. The main branch is 3 commits ahead of 'neptun91/codespaces-jupyter:main'. The commit history lists several changes made by 'neptun92' over time, including removing files like '.devcontainer', 'Archives', 'notebooks', '.gitignore', 'LICENSE', 'README.md', and 'requirements.txt', and updating others like 'devcontainer.json' and 'SCM tip'. The commits range from 4 minutes ago to 11 months ago.

Commit	Message	Time
Remove	1dfaee2 · 4 minutes ago	82 Commits
.devcontainer	Update devcontainer.json	3 years ago
Archives	Upload 데이터전처리대전 소스	last week
notebooks	Remove	4 minutes ago
.gitignore	Add .ipynb_checkpoints/ to .gitignore	3 years ago
LICENSE	Initial commit	3 years ago
README.md	Update SCM tip.	3 years ago
requirements.txt	Update requirements.txt	11 months ago

실습 환경 구축

Step3> '<> code' 버튼을 누른 후 'codespaces' 탭에서 'Create codespace on main'을 클릭한다.



실습 환경 구축

Step3> 실습 환경 구축 화면

