

교육 과정 소개서.

파이썬을 활용한 데이터 전처리 Level UP



강의정보

- 강의장 : 온라인 | 데스크탑, 노트북, 모바일 등
- 기간 : 평생 소장
- 담당 : 패스트캠퍼스 온라인팀
- 문의 : 02- 501-9396

학습목표

- 데이터를 필요에 맞게 가공하는 데이터 핸들링을 배웁니다.
- 데이터의 의미를 파악하는 탐색적 전처리를 배웁니다.
- 결과값의 정확도를 높이고 머신러닝의 성능을 향상시킬 수 있는 고급 전처리 스킬을 배웁니다.

한줄요약

- 데이터 분석에 꼭 필요한 필수 전처리부터, 데이터의 품질 및 머신러닝의 성능 향상을 위한 고급 스킬까지 완전 정복

강의특징

나만의
속도로

낮이나 새벽이나
내가 원하는 시간대에 나의 스케줄대로 수강

원하는 곳
어디서나

시간을 쪼개 먼 거리를 오가며
오프라인 강의장을 찾을 필요 없이 어디서든 수강

무제한
복습

무엇이든 반복적으로 학습해야
내것이 되기에 이해가 안가는 구간 몇번이고 재생



Curriculum

01. 시작에 앞서: 데이터 전처리는 왜 중요할까?

- 01. 데이터 전처리의 중요성 및 개요
- 02. 데이터 전처리를 잘하는 방법

02. 데이터를 어떻게 표현할 것인가: 파이썬 주요 데이터 구조

- 01. 리스트와 튜플
- 02. 사전
- 03. 반복문과 comprehension
- 04. Numpy의 데이터 구조
- 05. Pandas의 데이터 구조

03. 재료 준비하기: 데이터 불러오기

- 01. txt, csv, tsv 데이터 불러오기
- 02. excel 데이터 불러오기

04. 한 눈에 데이터 보기: 데이터 통합 및 집계

- 01. merge를 이용한 데이터 통합
- 02. concat을 이용한 데이터 통합
- 03. 기초 통계 함수를 활용한 데이터 집계
- 04. pivot을 이용한 데이터 집계
- 05. groupby를 이용한 데이터 집계

05. 보고 싶은 데이터 찾아보기: 데이터 정렬과 인덱싱

- 01. 데이터 정렬
- 02. 인덱서를 활용한 인덱싱
- 03. 마스킹 검색
- 04. 문자열 검색



Curriculum

06. 데이터 이쁘게 보기: 데이터 시각화

01. matplotlib 기초
02. 꺾은선 그래프와 산점도 그리기
03. 막대 차트 그리기
04. 파이 차트 그리기
05. 박스 플롯 그리기

07. 직접 해보기

01. 주문서 정리하기
02. 뉴스 기사 요약하여 정리하기

08. 시작에 앞서: 탐색적 데이터 분석이란?

01. 정의 및 중요성
02. 사례 소개

09. 변수가 어떻게 생겼나: 기초 통계 분석

01. 기초 통계 분석을 해야 하는 이유
02. 대표 통계량
03. 산포 통계량
04. 분포 통계량
05. 머신러닝에서의 기초 통계 분석

10. 둘 사이에는 무슨 관계가 있을까: 가설 검정과 변수 간 관계 분석

01. 영가설과 대립가설, p-value
02. 독립 표본 t 검정
03. 쌍체 표본 t 검정
04. 일원분산분석
05. 상관분석
06. 머신러닝에서의 가설 검정



Curriculum

11. 비슷한 애들 모여라: 군집화

- 01. 군집화 기본 개념
- 02. 계층적 군집화
- 03. k-평균 군집화

12. 어디서 많이 봤던 패턴이다 싶을 때: 빈발 패턴 탐색

- 01. 연관 규칙 탐색
- 02. 빈발 시퀀스 탐색
- 03. 빈발 시계열 패턴 탐색
- 04. 머신러닝에서의 빈발 패턴 탐색

13. 직접 해봐야 내것이 된다

- 01. A/B 테스트: 홈페이지 화면을 어떻게 구성할 것인가?
- 02. 고객 세분화를 통한 마케팅 전략 수립
- 03. 이탈 고객의 고객 여정 탐색하기

Chapter 14 부터 배경지식 즉, 머신러닝에 대한 기초 이해가 없으면 수강에 어려움이 있을 수 있습니다.

14. 이걸 꼭 알아야 해: 지도학습 모델의 핵심 개념

- 01. 지도학습 개요
- 02. 모델 개발 프로세스
- 03. 주요 모델의 구조 및 특성

15. 이럴땐 이걸 쓰고, 저럴땐 저걸 쓰고: 지도학습 모델 & 파라미터 선택

- 01. 그리드 서치
- 02. 기준 (1) 변수 타입
- 03. 기준 (2) 데이터 크기
- 04. 복잡도 파라미터 튜닝 방법



Curriculum

16. 흩어진 데이터 다 모여라: 데이터 파편화 문제

- 01. 문제 정의
- 02. 유형 (1) 파일 자체가 나뉘어 저장된 경우
- 03. 유형 (2) 명시적인 키 변수가 있는 경우
- 04. 유형 (3) 포맷이 다른 키 변수가 있는 경우
- 05. 유형 (4) 거리 기반 병합이 필요한 경우
- 06. 유형 (5) 데이터 요약이 포함되는 경우

17. 왜 여기엔 값이 없을까: 결측치 문제

- 01. 문제 정의
- 02. 해결 방법 (1) 삭제
- 03. 해결 방법 (2) 대표/근처 값으로 대체
- 04. 해결 방법 (3) 결측치 예측 모델 활용

18. 문자보다는 숫자: 범주형 변수 문제

- 01. 문제 정의 및 해결 방법
- 02. 관련 문법 및 실습

19. 이상적인 분포를 만들순 없을까: 변수 분포 문제

- 01. 특징과 라벨 간 약한 관계 및 비선형 관계
- 02. 이상치 제거 (1) IQR 규칙 활용
- 03. 이상치 제거 (2) 밀도 기반 군집화 활용
- 04. 특징 간 상관성 제거
- 05. 변수 치우침 제거
- 06. 스케일링



Curriculum

20. 편향된 모델은 쓸모없어: 클래스 불균형 문제

01. 문제 정의 및 탐색 방법
02. 재샘플링: 오버샘플링과 언더샘플링
03. 비용 민감 모델

21. 많다고 좋은게 아니다: 차원의 저주 문제

01. 문제 정의
02. 특징 선택

22. 배운걸 적용해 보자: 실전 모델 개발 프로세스

01. 실전 그리드 서치
02. 최종 실습 예제

23. 진짜 문제를 해결해 보자

01. 상점 신용카드 매출 예측
02. 아파트 실거래가 예측
03. IEEE - CIS Fraud Detection
04. Bosch Production Line Performance