

[Data Science] 데이터 사이언스 개념 - 1.데이터 과학이란? — 나무늘보의 개발 블로그

노트북: 첫 번째 노트북

만든 날짜: 2021-01-13 오후 5:45

URL: <https://continuous-development.tistory.com/210?category=833358>

Data Science

[Data Science] 데이터 사이언스 개념 - 1.데이터 과학이란?

2021. 1. 8. 00:56 수정 삭제 공개



01.데이터 과학이란?

데이터 과학자 - 컴퓨팅 기술을 활용해 데이터 수집 및 처리, 통계학이나 머신러닝으로 분석, 의사결정과 상품 개발까지 이어지는 일련의 흐름으

효과적으로 처리하는 기능을 가진 사람

업무 계열 - 비즈니스를 중시

IT 계열 - 구현을 중시

데이터 과학

- 통계학 , 머신러닝

- 컴퓨터 과학

- 응용력

02.빅데이터란?

빅데이터란 최근 얻을 수 있게 된 데이터의 총칭으로서 크게 4가지 측면에서 기존 데이터와 다른 특징을 보인다.

빅데이터의 5V

- 크기

- 속도

- 다양성

- 정확성

- 가치

03.인공지능과의 관계

인공지능과 데이터 과학은 머신러닝을 공통항으로 가진 관계

자동화할 수 있는 부분은 제대로 자동화하고 전체의 흐름을 제어하는 것이 데이터 과학자의 업무

04.데이터 과학자의 역할

무엇을 할 것인지 명확히하고(가설수립)

어떤 데이터를 모을 것인지 생각

필요한 이론과 요소기술을 조합해 프로그램으로 구현

피드백된 정보를 바탕으로 개선하는 일종의 PDCA(plan - do - check-action) 주기를 반복

주기를 반복할 때 중요한 것은 자신이 하는 작업이 최종 목적과 일치하는지 강하게 의식하는 것

05.데이터 수집법(1) - 오픈 데이터

오픈데이터란? 누구나 널리 활용할 수 있도록 데이터를 만들어 공개한 것
종류

각국 정부가 제공하는 사회 전반에 관한 통계 데이터

연구자가 머신러닝 알고리즘의 벤치마크용으로 공개하는 데이터

위키피디아 등 공동 협력으로 모은 정보를 데이터 베이스화 한 것

파나마 문서 등 특정단체가 공개한 데이터

데이터 과학 블로그가 공개한 데이터,

영화데이터 베이스

케글데이터

DeepAnalystics 등

등등이 있다.

06.데이터 수집법(2) - Web API

API - 애플리케이션 프로그래밍 인터페이스(Application Programming Interface)의 줄임말

서비스 제공자가 소프트웨어 일부나 보유중인 데이터를 다른 사용자도 효과적으로 활용할 수 있도록 공개하는 서비스

종류

국회도서관

라쿠텐

구루나비

리크루트

Google

Facebook

New York Times

Associated Press

Foursquare

Yelp

API는 프로그래밍에서 이용한다. 응답으로 돌아오는 데이터 형식은 서비스에 따라 다르지만 보통 RSS, XML, JSON등의 특수한 형식으로 받는다.

07.데이터 수집법(3) - 웹 스크레이핑

웹 스크레이핑 - 웹사이트에 있는 정보를 수집하는 행위

수집 가능한지 확인하는 방법 -> 홈페이지 주소/robots.txt 로 확인

허가된 것에 대해 * 가 붙어있고 상대 서버에 지나친 부하가 걸리지 않게 프로그래밍을 해야 한다.

본 내용은 그림으로 배우는 DataScience 데이터 과학을 참고한 내용입니다

'Data Science' 카테고리의 다른 글

[Data Science] 데이터 사이언스 개념 - 4.회귀 모델

[Data Science] 데이터 사이언스 개념 - 3.과적합과 모델 선택

[Data Science] 데이터 사이언스 개념 - 2.머신러닝의 기본

[Data Science] 데이터 사이언스 개념 - 1.데이터 과학이란?

Data Science

데이터 사이언스

데이터 사이언스 개념

데이터 사이언스 기초

데이터 사이언스 기초개념



나아무늘보

혼자 끄적끄적하는 블로그 입니다.