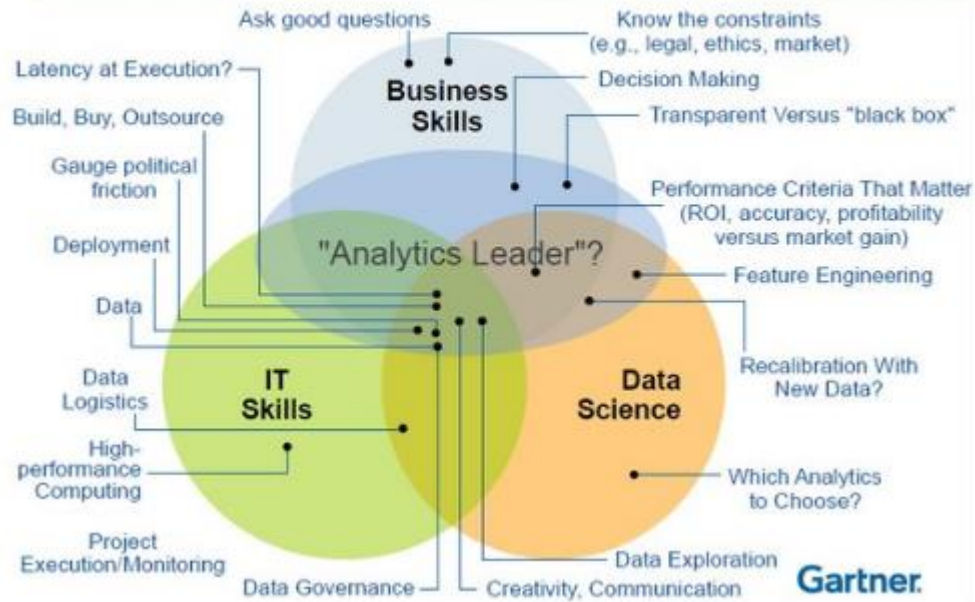


스터디 간략한 소개

Driving the Success of Data Science Solutions: Skills, Roles and Responsibilities ...



<https://cinema4dr12.tistory.com/1016>

GOAL

- 데이터분석의 끈은 잡고 있어보자.
- 데분/데싸 공부한 기록도 남기기

소프트웨어 개발자의 관점에서

- 데이터에 대한 관점을 키우기
- 데이터를 활용도는 점점 높아질 것
 - 추천시스템....

스터디 개요

- **학습하기**

난이도 의논 후 유다시티 커리큘럼 선택하여 일주일에 4-5시간 정도 분량 듣기

- **모여서 공유하기**

일주일에 1-2번 정도 모여서 말은 부분에 대해서 요약, 브리핑 시간 갖기

일주일 동안 학습한 것 리마인드용

학습 내용 중 모르는 것/ 다시 짚고 싶은 부분 공유

- **포폴/프로젝트**

1-2주일에 한번 학습내용 토대로 프로젝트 진행

- 주간 프로젝트를 진행하게 된다면

최대한 간략한 것으로 진행

유다시티에서 제공하고 있는 퀴즈나 미니 프로젝트 진행

- 장기 프로젝트로 진행 시

배운 것을 응용해 볼 수 있는 케글 또는 공모전

Udacity?

Lesson 6: Datasets and Questions

What Is A POI

What is a person of interest?

- indicted
- settled without admitting guilt

2. What Is A POI

3. Accuracy vs. Training Set Size

4. Downloading Enron Data

5. Types of Data Quiz 1

6. Types of Data Quiz 2

7. Types of Data Quiz 3

8. Types of Data Quiz 4

9. Types of Data Quiz 5

- Youtube 기반의 강의

Udacity?

The screenshot displays the Udacity learning interface. On the left is a dark sidebar with a navigation menu. The top of the sidebar shows 'Lesson 6: Datasets and Questions'. Below this are sections for '검색' (Search), '자료' (Data), and 'CONCEPTS'. The 'CONCEPTS' section is expanded, showing a list of quizzes and projects. '5. Types of Data Quiz 1' is highlighted with a green checkmark. Other items include '6. Types of Data Quiz 2' through '10. Types of Data Quiz 6', and projects like '11. Enron Dataset Mini-Project Vid...', '12. Datasets and Questions Mini-...', and '13. Size of the Enron Dataset'. The main content area has a dark header with 'Types of Data Quiz 1'. Below the header is a large white box containing a handwritten-style quiz question. The title 'Types of Data' is underlined in blue. The question asks 'What type of data is salary info?'. To the left of the question are five options in red: 'numerical', 'categorical', 'time series', 'text', and 'other', each followed by a radio button. At the bottom of the main area are three buttons: '인트로 보기' (View Intro), '정답 보기' (View Answer), and '답 제출하기' (Submit Answer).

Lesson 6: Datasets and Questions

Types of Data Quiz 1

Types of Data

Quiz

What type of data is salary info?

numerical

categorical

time series

text

other

인트로 보기

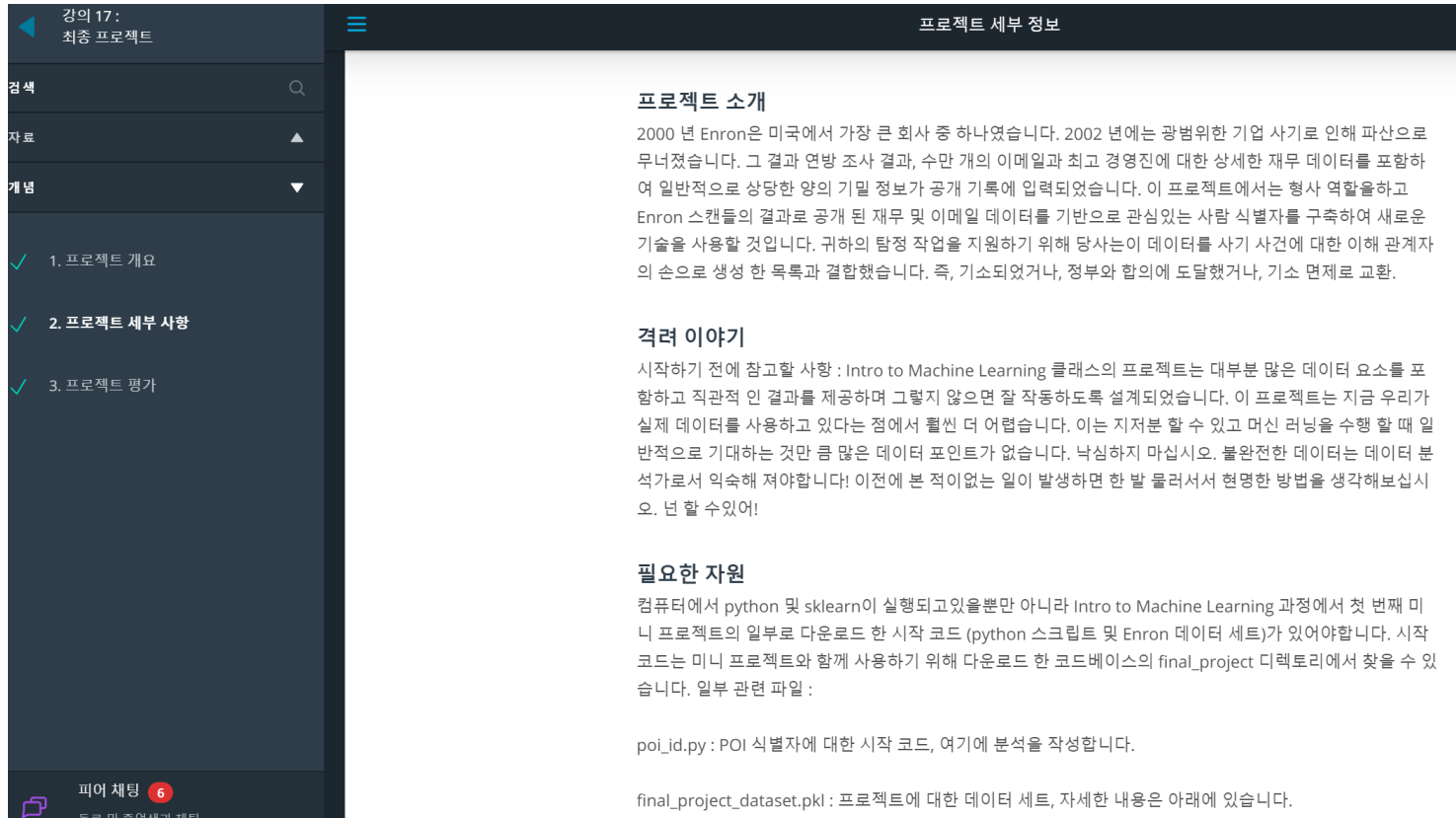
정답 보기

답 제출하기

- 강의 내용 중간 중간 확인하는 퀴즈, 또는 미니 코딩이 있음
- 계속해서 강의만 듣는 것 보다 효율적이고 쉽게 학습 가능

Udacity?

Projects



강의 17: 최종 프로젝트

프로젝트 세부 정보

프로젝트 소개

2000 년 Enron은 미국에서 가장 큰 회사 중 하나였습니다. 2002 년에는 광범위한 기업 사기로 인해 파산으로 무너졌습니다. 그 결과 연방 조사 결과, 수만 개의 이메일과 최고 경영진에 대한 상세한 재무 데이터를 포함하여 일반적으로 상당한 양의 기밀 정보가 공개 기록에 입력되었습니다. 이 프로젝트에서는 형사 역활을하고 Enron 스캔들의 결과로 공개 된 재무 및 이메일 데이터를 기반으로 관심있는 사람 식별자를 구축하여 새로운 기술을 사용할 것입니다. 귀하의 탐정 작업을 지원하기 위해 당사는이 데이터를 사기 사건에 대한 이해 관계자의 손으로 생성 한 목록과 결합했습니다. 즉, 기소되었거나, 정부와 합의에 도달했거나, 기소 면제로 교환.

격려 이야기

시작하기 전에 참고할 사항 : Intro to Machine Learning 클래스의 프로젝트는 대부분 많은 데이터 요소를 포함하고 직관적 인 결과를 제공하며 그렇지 않으면 잘 작동하도록 설계되었습니다. 이 프로젝트는 지금 우리가 실제 데이터를 사용하고 있다는 점에서 훨씬 더 어렵습니다. 이는 지저분 할 수 있고 머신 러닝을 수행 할 때 일반적으로 기대하는 것만 큼 많은 데이터 포인트가 없습니다. 낙심하지 마십시오. 불완전한 데이터는 데이터 분석가로서 익숙해 져야합니다! 이전에 본 적이없는 일이 발생하면 한 발 물러서서 현명한 방법을 생각해 보십시오. 오, 년 할 수있어!

필요한 자원

컴퓨터에서 python 및 sklearn이 실행되고있을뿐만 아니라 Intro to Machine Learning 과정에서 첫 번째 미니 프로젝트의 일부로 다운로드 한 시작 코드 (python 스크립트 및 Enron 데이터 세트)가 있어야합니다. 시작 코드는 미니 프로젝트와 함께 사용하기 위해 다운로드 한 코드베이스의 final_project 디렉토리에서 찾을 수 있습니다. 일부 관련 파일 :

poi_id.py : POI 식별자에 대한 시작 코드, 여기에 분석을 작성합니다.

final_project_dataset.pkl : 프로젝트에 대한 데이터 세트, 자세한 내용은 아래에 있습니다.

참고

https://github.com/hyunohenn/Communicate-Data-Findings/blob/main/explanatory_communicate_data_findings.ipynb

프로젝트

<https://anaconda.org/wangxbl/identify-point-of-interest-id-for-enron-dataset-using-machine-learning/notebook>

깃헙

<https://github.com/llwang8/Udacity-intro-machine-learning>

예상(?) 커리큘럼

- 2월
(3주~4주)

- Intro to Data Analysis
- [Link](#)
- 내용: numpy, pandas
- 유다시티 주장 학습소요시간: 약 15시간

예상(?) 커리큘럼

- 3월 ~ 5월

(약 12주 60시간)

- Introduction to Machine Learning Course
- [Link](#)
- 내용: Machine Learning, Supervised Learning, Unsupervised Learning 나이브베이지스, SVM, Decision Tree, 회귀, 클러스터링 등
- 유다시티 주장 학습소요시간: 약 62시간

예상(?) 커리큘럼

- 난이도에 따라 조절

- PyTorch / TensorFlow lite 코스
- 하둡 (대용량 분산처리)
- 데이터사이언스 과정(데이터 수집, 정제부터 맵리듀스까지)