와!
데이터 사이언스!
아시는구나!

Hyunseung Jeon

School of CSE, Kyungpook Nat’l Univ.

hyunseung.jeon@knu.ac.kr

발표자 소개

- 전현승
- 컴퓨터학부 심화컴퓨터전공 3학년 2학기
- 2019. 03. ~ 2020. 06. Gori 회장

- ~~6개월만에 스터디를 하게 되다니...~~

스터디 소개

- 그냥 가볍게 들으시면 될 것 같습니당
- 저도 가볍게 준비해 왔습니당
- 회장님 한 번쯤은 도와줘야 할 것 같아서 가벼운 주제로...

- 알고리즘이라는 주제에서 잠깐 벗어나서, “데이터 사이언스” 이야기
- 그런데 갑자기 왜?

– 0x00

이번 대경권 ML, 다들 하셨나요?

대경권 대학생 프로그래밍 경진대회

- 매년 “대경권 대학생 프로그래밍 경진대회” (**대경권**) 이라는 이름으로 PS 대회가 열렸었음
- 2018년을 첫 해로 1회가 개최됐었고,
  2019년에는 두 번이나 열렸었고,
  올해는 5월에 한 번, 11월(지금) 한 번
- 경북대, 계명대, 한동대 등 대경권 대학교 재학생들만 참가할 수 있는 대회
- (조금만 노력한다면) 수상실적 올리기에 꿀 같은 대회
  - 그래도 ICPC 본선 가서 상 타는 것보단 쉽잖아요...?

대경권 대학생 AI 프로그래밍 경진대회

참가 방법

대상 지역선도대학육성사업(경북대 컨소시엄) 참여 대학생

※ 참여대학 : 경북대, 계명대, 한동대, 경운대, 동국대(경주캠퍼스)

접수 대회 홈페이지 https://programmers.co.kr/competitions/581/dg-univ-2020

일 정

온라인 접수 2020. 10. 12.(월) ~ 11. 4.(수)

참가자 선발 2020. 11. 5.(목)

※ 대학별로 참가자 확인 후 공지 예정

온라인 경진대회 2020. 11. 7.(토) ~ 11. 14.(토)

시 상

상명	시상금(원)	인원(명)	총 시상금(원)	비고
대상	1,000,000	1	1,000,000	경북대학교 총장상
최우수상	750,000	4	3,000,000	경북대학교 총장상
우수상	500,000	6	3,000,000	SW교육센터 센터장상
장려상	300,000	10	3,000,000	SW교육센터 센터장상
합 계		21	10,000,000	

기타 사항

- ○ 운영방식: Data set 제공시 D/S, 머신러닝을 활용한 문제해결 코드와 데이터 제출 형태
- ○ tool : 웹페이지 제공
- ○ language: 자유형식

대경권 대학생 “인공지능” 프로그래밍 경진대회

- 근데 갑자기 “인공지능” 프로그래밍 경진대회?
- ML? 머신러닝?
- 아이고 이게 무슨 일이야!

- 대회의 양상이 크게 달라진 것을 알 수 있다
- 이전의 알고리즘 문제풀이(PS) 대회와는 달리,
  이번 대경권은 **머신러닝 컴페티션** 형태로 진행됨
- 이젠 더 이상 PS로는 먹고살 수 없는 시대가 온 걸까요...

대회 방식을 간단히 소개하자면...

- 이번 대회는 **“중고차 가격 예측하기”**

- 대회 참가자에게는 중고차에 대한 데이터셋이 주어진다
- 데이터셋에는 여러 중고차의 데이터들이 담겨져 있고,
  데이터 한 행은 모델명, 연식, 연료, 구동 방식 등 여러 특성Feature으로 이루어져 있음

대회 방식을 간단히 소개하자면... (cont'd)

- 주어진 훈련 데이터셋$_{\text{training set}}$을 이용해 모델을 학습시켜서,
  평가 데이터셋$_{\text{test set}}$의 데이터를 추측해라!
- 모델명이랑 연식, 연료, 구동 방식 등 이런 특성들을 줄 테니까,
  중고차 가격이 얼마일지 맞춰봐라!

주어지는 데이터들과 설명

- 주어지는 파일은 train.csv와 test.csv
- “제출은 이런 식으로 하세요” 하는 submission.csv
- 그리고 각 Feature들에 대한 설명
  - 모델명은 어떤 형식으로 주어지고...
  - 연식에는 빈 데이터도 있으며...

train.csv

input > train.csv

```
no,모델명,연월,연식,연료,주행거리,인승,최대출력(마력),기통,최대토크(kgm),구동방식,자동수동,국산/수입,신차가(만원),가격(만원)
0,기아 더 뉴 K7 2.4 GDI 프레스티지 스페셜,12/12(13년형),2013.0,가솔린,4만km,,201.0,4.0,25.5,FF,,국산,3141.0,1870.0
1,현대 YF쏘나타 2.0 Y20 LPi 프리미어,10/03,2010.0,LPG,9만km,,157.0,,20.0,FF,,국산,,700.0
2,현대 그랜져HG 220 디젤 프리미엄,14/09(15년형),2015.0,디젤,3만km,,202.0,4.0,45.0,FF,,국산,3389.0,2990.0
3,쌍용 뉴카이런 2.0 LV5 2WD 고급형,07/06(08년형),2008.0,디젤,16만km,,151.0,4.0,33.8,FR,,국산,2499.0,420.0
4,현대 뉴스타렉스 점보 밴 TCI 3인승 GX 윈도우밴 일반형,04/03,2004.0,디젤,16만km,3.0,103.0,,24.0,FR,수동,국산,1445.0,450.0
5,현대 그랜져TG Q270 LPI 장애인용,06/01,2006.0,LPG,23만km,,165.0,,25.0,FF,,국산,24120.0,550.0
6,기아 오피러스 프리미엄 GH270 스페셜 럭셔리,11/09,2011.0,가솔린,5만km,,195.0,6.0,25.6,FF,,국산,3927.0,1390.0
7,기아 K7 VG350 노블레스 프리미엄,10/02,2010.0,가솔린,10만km,,290.0,6.0,34.5,FF,,국산,4130.0,1530.0
8,현대 제네시스 쿠페 200 터보 P,08/11(09년형),2009.0,가솔린,14만km,,210.0,4.0,30.5,FR,,국산,2641.0,490.0
9,GM대우 뉴 다마스 코치 5인승 슈퍼,13/03,2013.0,LPG,2만km,5.0,43.0,4.0,6.7,FF,수동,국산,930.0,630.0
10,현대 엑센트 1.6 VGT 모던,14/11(15년형),2015.0,디젤,14만km,,136.0,4.0,30.6,FF,,국산,1796.0,600.0
11,기아 올 뉴 카니발 하이리무진 2.2 디젤 9人 프레스티지,15/04(16년형),2016.0,디젤,1만km,,202.0,4.0,45.0,FF,,국산,4985.0,5200.0
12,현대 그랜져IG 3.0 LPi 익스클루시브,17/01,2017.0,LPG,9만km,,235.0,6.0,28.6,FF,,국산,3295.0,1890.0
13,르노삼성 뉴 SM3 LE20,10/11(11년형),2011.0,가솔린,9만km,,141.0,4.0,19.8,FF,,국산,1860.0,550.0
14,현대 더 뉴 아반떼 1.6 GDi 텐 밀리언 리미티드,15/04,2015.0,가솔린,6만km,,140.0,4.0,16.9,FF,,국산,2005.0,1850.0
15,기아 봉고 프론티어 1톤 초장축 더블캡,02/06,2002.0,디젤,7만km,5.0,,,,FR,수동,국산,,650.0
16,현대 아반떼AD 1.6 T-GDi 스포츠 M/T,16/06(17년형),2017.0,가솔린,6만km,5.0,204.0,4.0,26.9,FF,수동,국산,1965.0,1450.0
17,기아 더 뉴 스포티지R 2.0 디젤 2WD 에이스,15/07(16년형),2016.0,디젤,6만km,,184.0,4.0,41.0,FF,,국산,2470.0,1690.0
18,기아 스포티지R 2.0 디젤 2WD TLX 프리미엄,11/03(12년형),2012.0,디젤,16만km,,184.0,4.0,40.0,FF,,국산,2615.0,750.0
```

test.csv

input > test.csv

```
no,모델명,연월,연식,연료,주행거리,인승,최대출력(마력),기통,최대토크(kgm),구동방식,자동수동,국산/수입
11769,제네시스 GV80 3.0 디젤 AWD 5인승,20/04,2020.0,디젤,8천km,,278.0,,60.0,AWD,,국산
11770,쉐보레 어메이징 뉴 크루즈 2.0 디젤 LT 디럭스팩,15/05,2015.0,디젤,9만km,,163.0,4.0,36.7,FF,,국산
11771,현대 그랜져HG 240 럭셔리,12/07,2012.0,가솔린,11만km,,201.0,4.0,25.5,FF,,국산
11772,기아 더 뉴 모닝 밴 ,15/04(16년형),2016.0,가솔린,2만km,,78.0,,9.6,FF,,국산
11773,쉐보레 더 넥스트 스파크 1.0 LT 플러스,16/07(17년형),2017.0,가솔린,1만km,,,4.0,,FF,,국산
11774,현대 그랜져TG Q270 럭셔리,06/07,2006.0,가솔린,18만km,,195.0,6.0,25.6,FF,,국산
11775,쉐보레 스파크 LPGi 밴,14/09,2014.0,LPG,5만km,,,,,,,국산
11776,현대 싼타페DM 2.0 VGT 2WD 익스클루시브,13/05,2013.0,디젤,2만km,,184.0,4.0,41.0,FF,,국산
11777,현대 더 럭셔리 그랜져 Q270 프리미어 스마트팩,09/10(10년형),2010.0,가솔린,12만km,,195.0,6.0,25.5,FF,,국산
11778,기아 로체 LEX20 고급형,07/03,2007.0,가솔린,21만km,,144.0,4.0,19.1,FF,,국산
11779,GM대우 라세티 프리미어 SX 일반형,09/06,2009.0,가솔린,6만km,,114.0,4.0,15.5,FF,,국산
11780,현대 쏘나타 더 브릴리언트 2.0 LPi 모던,13/11(14년형),2014.0,LPG,6만km,,157.0,,20.0,FF,,국산
11781,기아 더 뉴 K3 1.6 GDi 노블레스,17/01,2017.0,가솔린,7만km,,132.0,4.0,16.3,FF,,국산
11782,현대 그랜져HG 240 모던,13/01,2013.0,가솔린,9만km,,201.0,4.0,25.5,FF,,국산
11783,쉐보레 올란도 2.0 LPI LTZ 프리미엄,14/07,2014.0,LPG,10만km,,140.0,4.0,18.8,FF,,국산
11784,쌍용 뉴 체어맨W CW600 4트로닉 프레스티지,13/12(14년형),2014.0,가솔린,4만km,,225.0,6.0,30.2,AWD,,국산
11785,기아 더 뉴 스포티지R 2.0 디젤 4WD 트렌디,13/09(14년형),2014.0,디젤,9만km,,184.0,4.0,41.0,4WD,,국산
11786,현대 투싼ix 디젤 LX20 2WD 럭셔리,10/11(11년형),2011.0,디젤,14만km,,184.0,4.0,40.0,FF,,국산
11787,기아 더 뉴 모닝 바이퓨얼 럭셔리,16/07,2016.0,가솔린/LPG겸용,4만km,,78.0,3.0,9.6,FF,,국산
11788,현대 LF쏘나타 2.0 CVVL 프리미엄,14/06(15년형),2015.0,가솔린,8만km,,168.0,4.0,20.5,FF,,국산
```

submission.csv

submission.csv

```
no,가격(만원)
11769,0
11770,0
11771,0
11772,0
11773,0
11774,0
11775,0
11776,0
11777,0
11778,0
11779,0
11780,0
11781,0
11782,0
11783,0
11784,0
11785,0
11786,0
```

대회 방식을 간단히 소개하자면... (cont'd)

- train.csv에는 여러 Feature들과 함께 중고차 가격(정답)이 존재함

모델명	연식	연료	...	중고차 가격
			...	

대회 방식을 간단히 소개하자면... (cont'd)

- 하지만 test.csv에는 정답은 없고 Feature들만 존재!

모델명	연식	연료	...

대회 방식을 간단히 소개하자면... (cont'd)

- 따라서 test.csv의 중고차 가격들을 직접 예측해서 제출하면

no.	예상 중고차 가격

대회 방식을 간단히 소개하자면... (cont'd)

- 요렇게 제출한 csv에 대한 공개 점수들이 바로바로 나온다!

제출 목록	공개 점수	최종 점수 사용 데이터
v1.2.1.xgb_lgb_cv.csv 2020-11-11 01:24:53	66.3103891174790 3	○ 선택
v1.2.1.xgbcv.csv 2020-11-11 01:14:32	66.3276611114419 7	● 선택
v1.3.rfr.csv 2020-11-11 00:53:19	52.877202632515 52	○ 선택
v1.2.rfr.csv 2020-11-11 00:19:26	63.901317658689 81	○ 선택
ensemble.csv 2020-11-10 21:00:58	50.027760295132 2	○ 선택

대회 방식을 간단히 소개하자면... (cont'd)

- 대회 기간은 일주일 정도로 넉넉하게 주어지고, 기간동안 최대한 높은 점수를 따내면 됨
- 대회 동안에는 제출한 데이터의 60% 정도로만 채점한 '공개 점수'가 점수판에 표시되고,
- 대회가 끝나면 전체 데이터로 다시 채점한 '최종 점수'가 발표됨

- 어? 그러면 막 숫자만 조금씩 바꿔서 엄청 많이 제출하면 100점 받을 수 있지 않나?
- 그럴 줄 알고 하루에 최대 5회만 제출할 수 있지롱!
  - 이는 다른 머신러닝 컴페티션에서도 보통 적용하는 사항

저는 잘 했을까요?

공개 리더보드 파이널 리더보드

최종 결과는 전체 데이터를 기준으로 재 산출 하므로 공개 리더보드의 순위와는 달라질 수 있습니다.

	개발자	총점	제출 횟수	마지막 제출
1	[illegible]	75.26	9	3일 전
2	[illegible]	73.847	25	3일 전
3	[illegible]	73.623	29	3일 전
4	[illegible]	73.434	24	7일 전
5	[illegible]	72.987	8	5일 전
6	[illegible]	72.5	22	3일 전
7	전*승 — do******@knu.ac.kr	72.262	27	3일 전
8	[illegible]	72.164	14	3일 전
9	[illegible]	71.626	25	3일 전

방금 보니까 최우수상 탔네요 아싸! > ▽ <

순위	학교명	성명	상명
1	경북대학교	[illegible]	대상
2	경북대학교	[illegible]	최우수상
3	경북대학교	[illegible]	최우수상
4	경북대학교	[illegible]	최우수상
5	경북대학교	[illegible]	최우수상
6	경북대학교	[illegible]	최우수상
7	경북대학교	전현승	최우수상
8	경북대학교	[illegible]	우수상
9	경북대학교	[illegible]	우수상
10	경북대학교	[illegible]	우수상
11	경북대학교	[illegible]	우수상
12	한동대학교	[illegible]	우수상
13	경북대학교	[illegible]	우수상

어떻게 했냐면요...

- 일단 이 스터디(세미나?)에서는 머신러닝 기법을 이야기하고 싶지는 않아서...
- 나중에 다른 세션이 마련된다면, 또는 고리에서 머신러닝 스터디를 열게 된다면
- 경험담을 자세히 풀어보도록 할게요

- 아니면 참가하셨던 분들 중 궁금하신 분들은 개인적으로 물어보셔도 됩니당 ><

– 0x01

데이터 사이언스란 대체 뭘까요...

Data Science

- 다양한 데이터로부터 지식과 인사이트를 추출하는데
  과학적 방법론, 프로세스, 알고리즘, 시스템을 동원하는 융합분야
- 데이터가 주어졌을 때, 데이터 간의 관계를 파악하거나,
  파악된 관계를 사용해서
  우리가 원하는 새로운 데이터를 만들어내는 과정
- 타이타닉에 탑승했던 사람들의 특성과 생존 여부가 주어졌을 때,
  이를 예측하는 것을 생각해 보면...

Data Science (cont'd)

- 저는 개인적으로 이런 거라고 생각해요
- **"주어진 Feature들을 가지고 결과물을 얼마나 잘 뽑아내느냐?"**

$$f(\text{모델명},\ \text{연식},\ \text{연료},\ \ldots) = \text{중고차 가격}$$

- 이러한 함수를 얼마나 잘 모델링하느냐 아닐까요?

어떤 것들이 필요할까?

- **수리통계학을 비롯한 수학**
  - 정말로 '데이터 사이언스'를 하려면 수학은 기본
  - 선형대수, 함수론, 미적분, 최적화 등등

어떤 것들이 필요할까?

- **코딩 능력**
  - 손으로 하는 게 아닌, 코드로 짜서 결과를 내는 것인 만큼
  - **Python** (또는 R) 정도는 다룰 줄 알아야 한다

  - 행렬이랑 숫자 다루는 **numpy**
  - 데이터 다루는 **pandas**
  - 그래프 그리는 **matplotlib**
  - 정도는 알아야 한다고 생각해요
  - 하면서 배워도 되긴 함

  - 알고리즘 했으면 이 정도는 금방 배울 것 같아요

어떤 것들이 필요할까? (cont'd)

- **정말 '분석력', '인사이트'**
  - **EDA(Exploratory Data Analysis, 탐색적 데이터 분석)**
  - 본격적으로 머신러닝을 돌리기 전에, 데이터들 간의 관계를 먼저 분석해야 한다!
  - 두 데이터 간의 상관관계표 등등... 다양한 방법을 동원해서
    데이터에서 얻을 수 있는 것들을 최대한 짜내기
  - 솔직히 저는 경험이랑 직관이 가장 중요하다고 생각해요

  - *'타이타닉 생존자를 예측하는 데에 성별 데이터가 중요할까? 어 그런데 여성 승객이 생존 확률이 74%로 더 높네?'*
  - *'호실(Cabin)은 데이터 양도 적고, 중복도 많네? 일부 손님들이 호실을 공유한 건가?'*

어떤 것들이 필요할까? (cont'd)

- **머신러닝 알고리즘**
  - 데이터를 정제하고 분류했으면, 이제 머신러닝이나 다른 알고리즘을 통해 결과를 내 볼 차례!
  - 기본적인 Linear Regression부터
  - 심화 기법인 Random Forest 같은 것들을 거쳐
  - 최근에 자주 쓰이는 Ensemble 기법인 XGBoost 등등...
  - 이 과정에서 아주 핫한 '**딥러닝**' 도 써볼 수 있겠죠?

– 0x02

추천하는 공부 방법과 Material

Kaggle!

- 데이터 사이언스 계의 ‘백준’
- https://www.kaggle.com/

Kaggle! (cont'd)

- 공부용으로 시작하기에 적합한 가장 대표적인 예제 컴페티션들이 많음
- 타이타닉 생존자 예측하기, 집값 예측하기
- MNIST 데이터 '손글씨 숫자 인식하기' 등등...

Titanic: Machine Learning from Disaster

Start here! Predict survival on the Titanic and get familiar with ML basics

Knowledge • Getting Started • Ongoing • 17036 Teams

House Prices: Advanced Regression Techniques

Predict sales prices and practice feature engineering, RFs, and gradient boosting

Knowledge • Getting Started • Ongoing • 4381 Teams

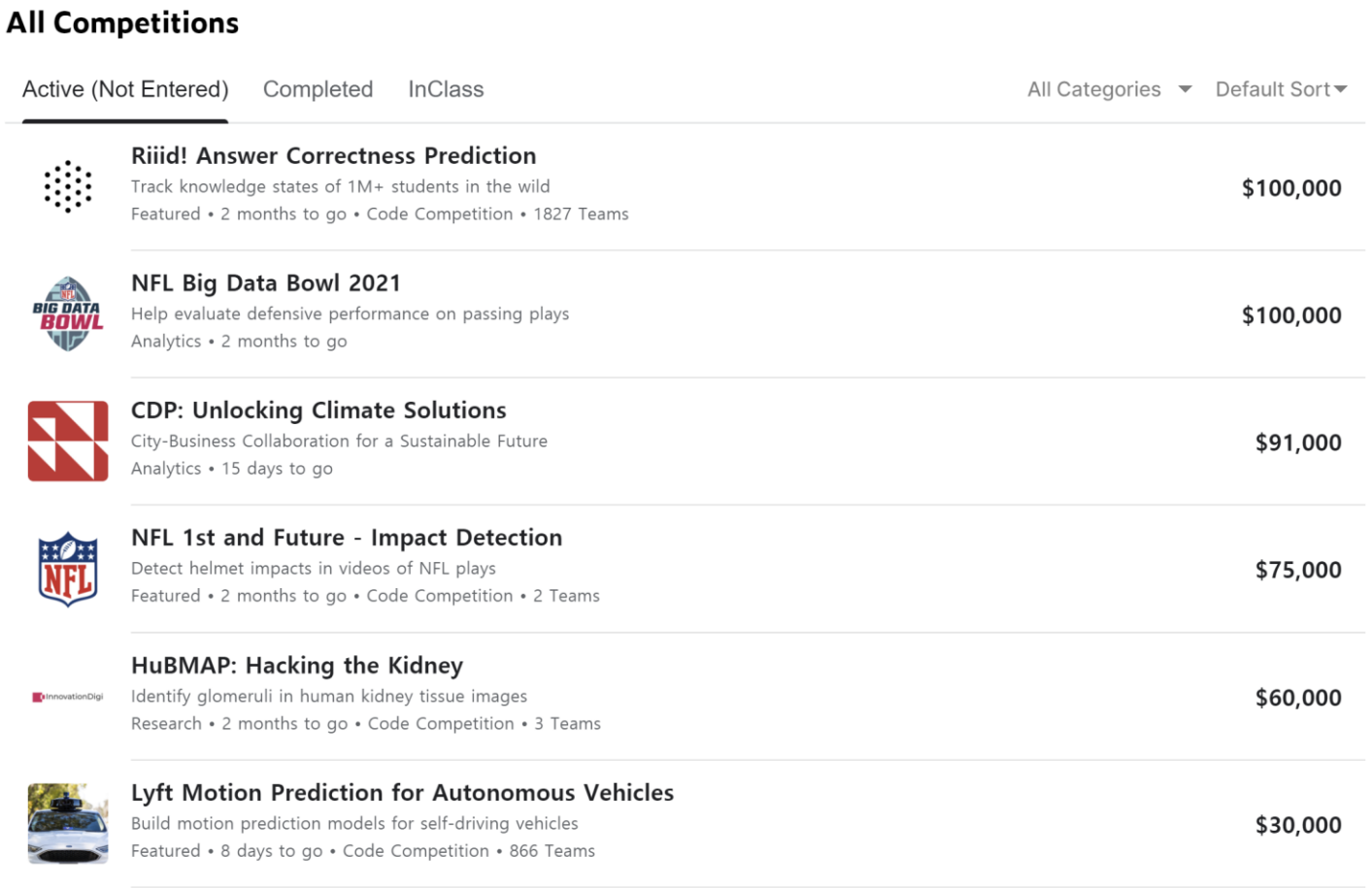
Digit Recognizer

Learn computer vision fundamentals with the famous MNIST data

Knowledge • Getting Started • Ongoing • 2214 Teams

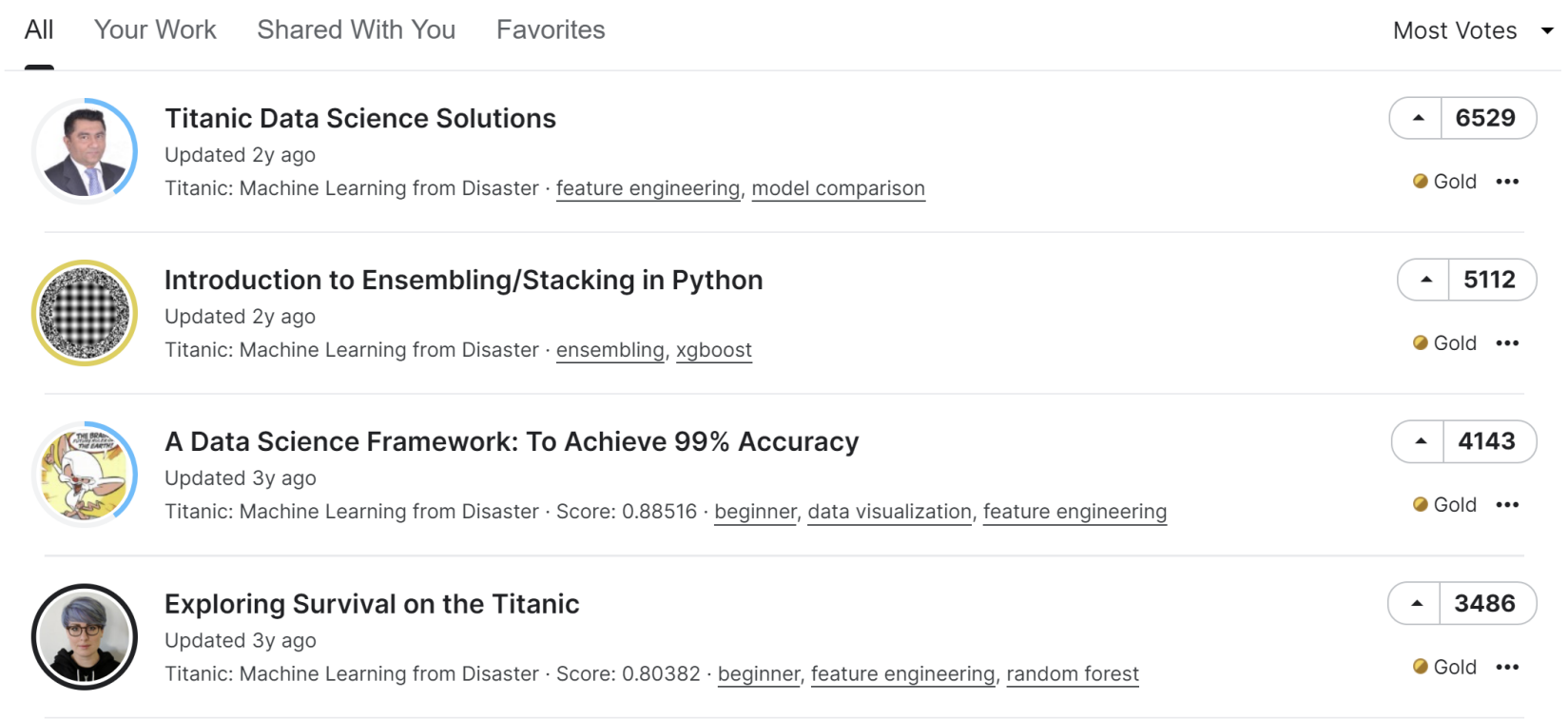
Kaggle! (cont'd)

- 돈 주는 컴페티션들도 많아요 (이런거 잘해서 돈 벌고싶다)



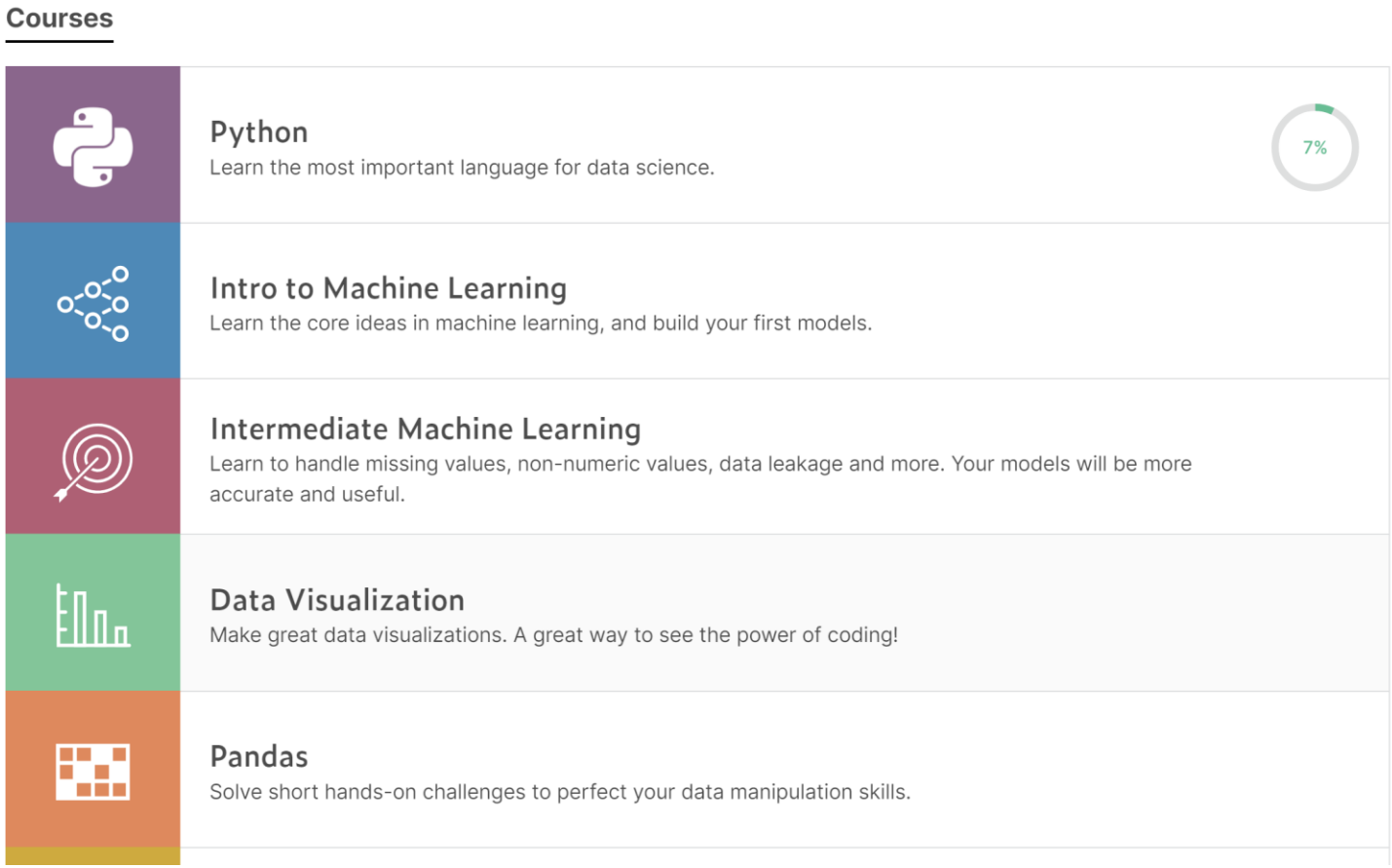
Kaggle! (cont'd)

- 각 컴페티션마다 고수들이 제출한 코드와 결과물을 튜토리얼 형태로 올려두기도 함
- 이를 '커널'이라고 합니다



Kaggle! (cont'd)

- 튜토리얼도 제공해요! 영어가 된다면 들어 보심이...



DACON

- 캐글의 국내 버전이라고 생각하셔도...?
- https://dacon.io/

DACON (cont'd)

- 역시 돈 걸고 여러 기업이나 학교에서 주최하는 데이터 사이언스 컴페티션들이 있구요

45개 대회 개최 | 30,183 팀 참여 | 3억 6150만원 상금

Y&Z세대 투자자 프로파일링 시각화 경진대회

금융 | NH투자증권 | 시각화 | 투표 및 심사평가 | 중복참가 불가, 대학 재학생만 참가 가능

NH투자증권

참가신청중
D-43 · 142팀
총 5,000만원(League1,2 통합)

AI야 , 진짜 뉴스를 찾아줘! AI 경진대회

금융 | NH투자증권 | 텍스트 분류 | Accuracy + Time | 중복 참가 불가, 대학 재학생만 참가 가능

NH투자증권

참가신청중
D-43 · 216팀
총 5,000만원(League1,2 통합)

한국어 문서 추출요약 AI 경진대회

Bflysoft | 추출 요약 | ROUGE-N

Bflysoft 고려대학교 DSBA

참가신청중
D-21 · 261팀
1,000만원

한국어 문서 생성요약 AI 경진대회

Bflysoft | 생성 요약 | ROUGE-N

Bflysoft 고려대학교 DSBA

참가신청중
D-21 · 165팀
1,000만원

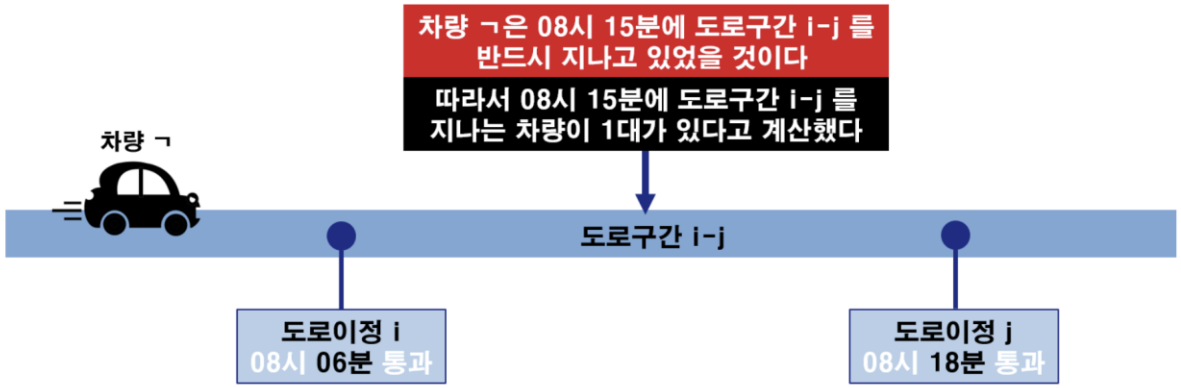
DACON (cont'd)

- 이렇게 컴페티션 별로 코드와 설명, 테크닉 공유까지 이루어짐

DACON (cont'd)

- 설명이 정말 자세한 커널들이 많았어요

밀도 교통량 구하기



- **밀도 교통량은 도로구간에 차량이 얼마나 빽빽하게 들어서있는지를 나타내는 지표로, 15분 간격으로 밀도 교통량 값을 계산**하였다. 예를 들어 2월 4일 오후 2시 45분 A 도로구간의 밀도 교통량 값이 높다면, 2월 4일 오후 2시 31분부터 오후 2시 45분 사이 A 도로구간이 정체되었다는 의미이다.
- 밀도 교통량을 계산한 방법은 다음과 같다. 하루치 DSRC 데이터를 이용해 모든 차량의 경로를 딕셔너리로 만든 뒤, 이전 도로이정을 지나간 시각과 이후 도로이정을 지나간 시각을 토대로 도로구간에 머문 차량의 대수를 구하였다. 예를 들어, i 도로이정 다음에 j 도로이정이 나올 때, 차량 ㄱ이 i 도로이정을 지났다고 찍힌 시각이 오전 8시 6분이고 j 도로이정을 지났다고 찍힌 시각이 오전 8시 18분이라면, 차량 ㄱ은 오전 8시 15분에 반드시 i-j 도로구간을 지났다는 점에 착안하여 도로구간에 머문 차량의 대수를 계산하였다.
- **그런데 동일한 대수의 차가 있더라도 차종에 따라 밀도 교통량이 달라질 것이다**. 예를 들어 길이가 같은 a 도로구간과 b 도로구간에 동일하게 10대의 차량이 있다고 가정하자. 이때 a 도로구간에는 버스 10대가 있고 b 도로구간에는 승용차 10대가 있다면, 같은 10대이더라도 a 도로구간에서 차가 더 밀릴 가능성이 있다. 따라서 차종별로 서로 다른 계수값을 적용하여 밀도 교통량을 구했다.

DACON (cont'd)

- 아직 돈 걸고 하는 컴페티션이 부담스럽다면 교육용 컴페티션들도...

책보다는 인강을 추천

- 처음 공부하는 분야라면 책보다 동영상 강의가 훨씬 효율이 좋습니다

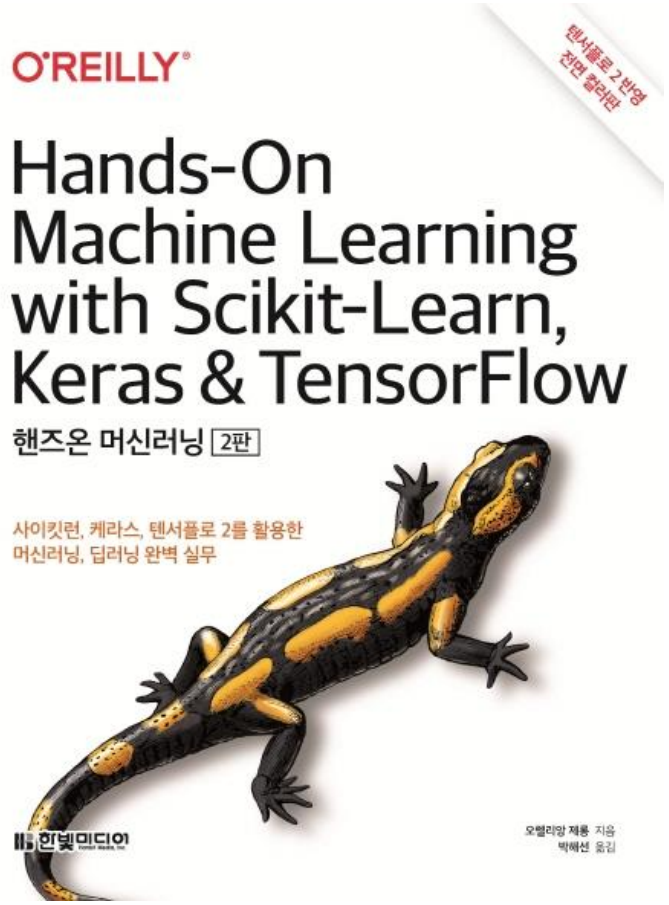
- **인프런**이나 **패스트캠퍼스** 같은 IT 강의 사이트에서 자신에게 맞는 강의 찾아보기
- 제가 들었던 건 이거 (홍보 아님)
  - **처음하는 파이썬 데이터 분석 [전처리, pandas, 시각화까지 전과정 기본 기술 쉽게 익히기]**
  - https://inf.run/mudN

책보다는 인강을 추천 (cont'd)

- 유튜브에도 무료로 풀리는 좋은 자료들이 많습니다

- **수비니움의 캐글 따라하기**
  - https://www.youtube.com/playlist?list=PLxpjN8IqlPc4IzZnA3F513qxuE4DkqpoZ
- **수비니움의 머신러닝 튜토리얼**
  - https://www.youtube.com/playlist?list=PLxpjN8IqlPc7RvrVglO7T3F49KGTc7y4A
- **Stanford CS229: Machine Learning | Autumn 2018**
  - https://www.youtube.com/playlist?list=PLoROMvodv4rMiGQp3WXShtMGgzqpfVfbU

그래도 책 한권쯤 갖고 있으면 좋죠



제가 갖고 있는 책은 요렇게 두 권

기타 좋은 블로그 및 사이트들

- **수비니움의 캐글 따라하기**
  - https://subinium.github.io/kaggle-tutorial/
  - 타이타닉 생존자 예측, 주택 가격 분류 등 기초 컴페티션 5가지
  - 한국어로 친절히 설명까지 되어 있는 커널
  - 자신이 노리는 수준에 따라 난이도별로 커널 세분화까지 해 주심
- **데이터 사이언스 스쿨**
  - https://datascienceschool.net/intro.html
  - 파이썬, 수학, 머신러닝 세 가지에 대해 심도 있게 배울 수 있음
  - 다소 어려울 수 있지만 내용이 알차다
  - 확실하게 공부하고 싶으신 분들에게 추천

기타 좋은 블로그 및 사이트들 (cont'd)

- **Machine Learning 강의노트**
  - https://wikidocs.net/book/587
  - Andrew Ng 교수님의 Coursera 강의 내용을 개인적으로 정리하신 것
  - 이외에도 wikidocs에 좋은 출판물들이 많으니 찾아보심을 추천

– 0x03

Summary

그래서 결론은?

- 알고리즘 PS 공부하다가 질리면, 또는 회의감이 느껴진다면
- 한 번쯤은 공부해볼만한 주제가 아닐까? 싶습니다
- 물론 막막하다는 느낌을 받을 수는 있겠지만, 알고리즘도 그랬잖아요!

- 그리고 PS와는 달리, 이건 정말로 “현업에서 바로 써먹을 수 있는 지식들”
  - 이라고 생각해요...
- 최근 빅데이터/AI 관련 공모전, 해커톤 등도 많이 열리는 추세고...
- 데이터 사이언스 하실래요?

끝!