**[DScover 2021-2 syllabus]**

* **Machine Learning** 
  + 1주차 : 강의 소개 / 머신러닝 개념 / 실습환경 준비 (섹션0, 섹션1, 섹션2)

머신러닝 개념

분류 문제 vs 회귀 문제

지도학습, 비지도학습, 강화학습

트레이닝 데이터, 테스트 데이터

검증용 데이터, 오버피팅

Google Colab 소개

* + 2주차 : numpy, pandas, matplotlib, seaborn 복습 (섹션3)

머신러닝과 데이터 분석을 위한 도구 소개

Numpy, 배열과 행렬을 다루는 도구

Numpy 실습

Pandas, 행과 열을 가진 테이블을 다루는 데이터 분석 도구

Pandas 실습

고급 pandas 함수들 – concat, groupby, query

고급 pandas 함수 실습

그래프를 그리는 plotting 도구 – matplotlib, seaborn

Matplotlib, seaborn 실습

* + 3주차 : scikit-learn 입문 (섹션4)

선형회귀 소개

Regression 알고리즘의 성능평가 지표 – MSE, RMSE, MAE

Scikit-learn 소개

Linear Regression으로 키에 대한 몸무게 예측해보기

* + 4주차 : Kaggle 입문 (섹션5, 섹션6)

Kaggle 및 Kaggle competition 소개

결정 트리 소개

Titanic 사고 데이터 소개

범주형 컬럼 & 수치형 컬럼

범주형 컬럼 다루기 (LabelEncoder)

* + 5주차 : Linear Regression 알고리즘1 (섹션7)

K-fold Cross Validation

Feature Engineering – Feature Selection

상관 분석(Correlation Analysis) & regplot()

Regression 알고리즘으로 보스턴 부동상 가격 예측해보기 (EDA & Feature Selection)

* + 6주차 : Linear Regression 알고리즘2 (섹션7)

Feature Engineering – Feature Normalization

Feature Engineering – Feature Generation

Ridge & Lasso & ElasticNet Regression

보스턴 부동산 가격 예측 성능 향상 (Feature Generation & Advanced Estimator)

* + 7주차 : Random Forest (섹션8)

Random Forest 소개

DataFrame의 .replace() 함수

범주형 데이터 처리하기 – One-hot Encoding

Rain in Australia 데이터셋 소개

Random Forest Classifier로 내일 비가 올지 안 올지 예측

이상치(Outlier)제거 기법 – IQR(Inter Quantile Range)

Outlier 제거를 통해 Random Forest Classifier 성능 향상

* + 8주차 : XGBoost (섹션9)

XGBoost 소개

Stroke Prediction 데이터셋 소개

XGBoost를 이용해서 뇌졸증(Stroke) 발생유무 예측

* + 9주차 : 현업 머신러닝 엔지니어 설명 (섹션10)

현업 머신러닝 엔지니어 소개

회사의 머신러닝 팀 구성원

머신러닝 팀에서 업무를 분배해서 일하는 법

현업 엔지니어의 하루 – 데이터 엔지니어

현업 엔지니어의 하루 – 데이터 사이언티스트

현업 엔지니어의 하루 – 머신러닝 엔지니어

머신러닝 업무에서 어려웠던 점

머신러닝 업무의 장/단점

* + 10주차 : 실무자가 알려주는 DS/ML 실전 팁 (섹션11)

실제로 회사에서 진행하는 머신러닝 업무 프로세스

가장 중요한 것은 데이터

Kattle/데이콘과 다르게, 실제 업무에서 중요하게 여기는 것은?

머신러닝 엔지니어, 일잘러로 거듭나보자

더 똑똑하게 일 잘하는, 머신러닝 코드 만드는 방법

가장 몸 값이 비싼 Role, 역략이 무엇일까?

* **Deep Learning** 
  + 1주차 : Lecture Orientation / Tensor 개념 (1강, 2강, 3강)

Constant and Variable Tensors

Making Tensors with APIs

Tensor Operations

* + 2주차 : Model 이해 및 생성 (4강, 5강, 6강)

Gradient Tape in Tensorflow

Model Implementation

Binary Classification

* + 3주차 : Model 생성 심화 및 평가 (7강, 8강, 9강)

Various Models in the Wild

Reset\_states of Metrics

Validation, Test, and tf.function Decorator

* + 4주차 : Dataset 생성 및 preprocessing(10강, 11강, 12강, 13강)

Dataset Load1

Dataset Load2

Dataset Split

Dataset Preprocessing

* + 5주차 : losses/metrics와 Classification 심화 (14강, 15강, 16강)

Losses and Metrics

Multiclass Classification

MNIST Classification

* + 6주차 : 학습모델 오류 확인 (17강, 18강, 19강)

Model Build and Summary

Model Configuration and Params

Gradient Vanishing Problem

* + 7주차 : CNN (20강, 21강)

Correlation in Deep Learning

Conv and Pooling Layer

* + 8주차 : LeNet (22강, 23강)

LeNet Implementation with Sequential APIs

LeNet Imp with Model Sub classing

* + 9주차 : utility funtions setting (24강, 25강, 26강, 27강)

Utility Functions – learning env setting1

Utility Functions – learning env setting2

Utility Functions – 3 More Utils

Utility Functions – cp Utils

* + 10주차 : 정리 및 Serial Training(28강, 29강)

Training LeNet5 Revisited

Serial Training

* **Data Visualization** 
  + 1주차 : Orientation / 태블로 이해 (1강, 2강, 3강, 4강)

디지털로의 전환

태블로 이해 및 설치 (학생용 1년 무료 버전 설치 가능)

태블로 기본 컨셉 이해하기\_1

태블로 기본 컨셉 이해하기\_2

* + 2주차 : 태블로 기초1 (5강, 6강, 7강)

태블로 부분 전체 분석

태브로 상관 관계 분석

태블로 이중 축 vs 결합된 축

* + 3주차 : 태블로 기초2 (8강, 9강, 10강)

태블로 그룹 vs 집합

태블로 테이블 구성하기\_개층과 총계

태블로 맵 만들기

* + 4주차 : 태블로 중급 탐험1 (11강, 12강, 13강, 14강)

대시보드 만들기

스토리 만들기

퀵 테이블 계산하기\_1

퀵 테이블 계산하기\_2

* + 5주차 : 태블로 중급 탐험2 (15강, 16강, 17강)

계산된 필드 만들기

매개 변수 만들기\_1

매개 변수 만들기\_2

* + 6주차 : 태블로 고급 맛보기1 (18강, 19강, 20강)

태블로 작동 순서

대시보드 액션 적용하기\_1

대시보드 액션 적용하기\_2

* + 7주차 : 태블로 고급 맛보기2 (21강, 22강, 23강)

효율적인 대시보드 만들기\_1

효율적인 대시보드 만들기\_2

효율적인 대시보드 만들기\_3

* + 8주차 : 태블로로 포트폴리오 만들기1 (24강, 25강)

주제에 적합한 시각화 방법\_1

주제에 적합한 시각화 방법\_2

* + 9주차 : 태블로로 포트폴리오 만들기2 (26강, 27강, 28강, 29강)

데이터 시각화를 더 돋보이게 만드는 요소\_1

데이터 시각화를 더 돋보이게 만드는 요소\_2

우수한 대시보드 구성하기

데이터 구하기

* + 10주차 : 태블로로 포트폴리오 만들기3 (30강, 31강, 32강)

분야별 대시보드 만들기\_1

분야별 대시보드 만들기\_2

분야별 대시보드 만들기\_3

* **Natural Language Processing** 
  + 1주차 : 강좌 및 자연어 처리 소개, 웹/이미지 등에서 글 추출1
  1. 강좌 소개
  2. 실습에 필요한 환경설정

2-1. 웹페이지 탐색 및 크롤링 계획

2-2. 웹에서 글 추출하기 – BeautifulSoup 사용법

2-3. 웹에서 글 추출하기 – BeautifulSoup 사용한 크롤링

2-4. 웹에서 글 추출하기 – Selenium 소개 및 사용법

2-5. 웹에서 글 추출하기 – Selenium을 활용한 크롤링

* + 2주차 : 웹/이미지 등에서 글 추출2

2-6. 이미지에서 글 추출하기 – pyTesseract 설치 및 기본 사용법

2-7. 이미지에서 글 추출하기 – OCR 인식률을 높이는 방법(1)

2-8. 이미지에서 글 추출하기 – OCR 인식률을 높이는 방법(2)

2-9. PDF에서 글 추출하기 – 라이브러리 소개 및 사용법(1)

2-10. PDF에서 글 추출하기 – 라이브러리 소개 및 사용법(2)

* + 3주차 : 데이터 정제1

3-1. 정규표현식 – 기본 사용법

3-2. 정규표현식 – 불필요한 문자 및 문자열 제거하기(1)

3-3. 정규표현식 – 불필요한 문자 및 문자열 제거하기(2)

3-4. 정규표현식 – 문장, 문단 분리하기

* + 4주차 : 데이터 정제2

3-5. 맞춤법 교정하기 – 라이브러리 소개 및 사용법

3-6. 맞춤법 교정하기 – 직접 수정하기(1)

3-7. 맞춤법 교정하기 – 직접 수정하기(2)

3-8. 자동화된 한글 텍스트 교정기 만들기(1)

3-9. 자동화된 한글 텍스트 교정기 만들기(2)

3-10. 유니코드(Unicode)

* + 5주차 : 데이터 분석1

4-1. 형태적 분석 – 형태소 분석의 정의와 의의

4-2. 형태적 분석 – Konlpy 소개 및 사용법(1)

4-3. 형태적 분석 – Konlpy 소개 및 사용법(2)

4-4. 형태적 분석 – 형태소 분석의 활용(1)

4-5. 형태적 분석 – 형태소 분석의 활용(2)

* + 6주차 : 데이터 분석2

4-6. 빈도 분석 – BoW(Bag of Words)

4-7. 빈도 분석 – Co-occurrence Matrix

4-8, 빈도 분석 -BPE(Byte Pair Encoding)

4-9. 빈도 분석- 문서 통계정보 나타내기(1)

4-10. 빈도 분석 – 문서 통계정보 나타내기(2)

* + 7주차 : 데이터 벡터화1

5-1. 표현 학습과 단어 표현

5-2. 원-한 인코딩(One-hot Encoding)

5-3. 카운트 백터(Count Vector)

5-4. TF-IDF(Term Frequency-Inverse Document Frequency)

5-5. Word2Vec(Distributed Representations of Words)

* + 8주차 : 데이터 벡터화2

5-6. GloVe (Global Vectors for Word Representation)

5-7. Gensim 라이브러리 활용한 벡터화(1)

5-8. Gensim 라이브러리 활용한 벡터화(2)

5-9. 시퀀스 벡터(Sequence Vectors) (1)

5-10. 시퀀스 벡터(Sequence Vectors) (2)

* + 9주차 : 모델 적용하기 (토픽 모델링)

6-1. 토픽 모델링 과제 소개

6-2. 잠재 의미 분석(LSA) (1)

6-3. 잠재 의미 분석(LSA) (2)

6-4. 잠재 디리클레 할당(LDA) (1)

6-5. 잠재 디리클레 할당(LDA) (2)

6-6. 잠재 디리클레 할당(LDA) (3)

* + 10주차 : 모델 적용하기 (감정 분류)

6-7. 감정분류 과제 및 모델 소개

6-8. 네이버 영화리뷰 데이터를 활용한 감정 분류(1)

6-9. 네이버 영화리뷰 데이터를 활용한 감정 분류(2)

6-10. 네이버 영화리뷰 데이터를 활용한 감정 분류(3)