

TEXT MINING for PRACTICE

Python을 활용한 비정형 데이터 분석 - 강의소개

연세대학교 | 서중원

강사소개 - 서중원

주요이력

- (現) IT 교육/컨설팅 업체 코드도둑 대표
- (前) 핀테크 스타트업 (주) 코스폴 개발팀장
- (前) NASA & Yonsei 큐브위성 소프트웨어 엔지니어

학력

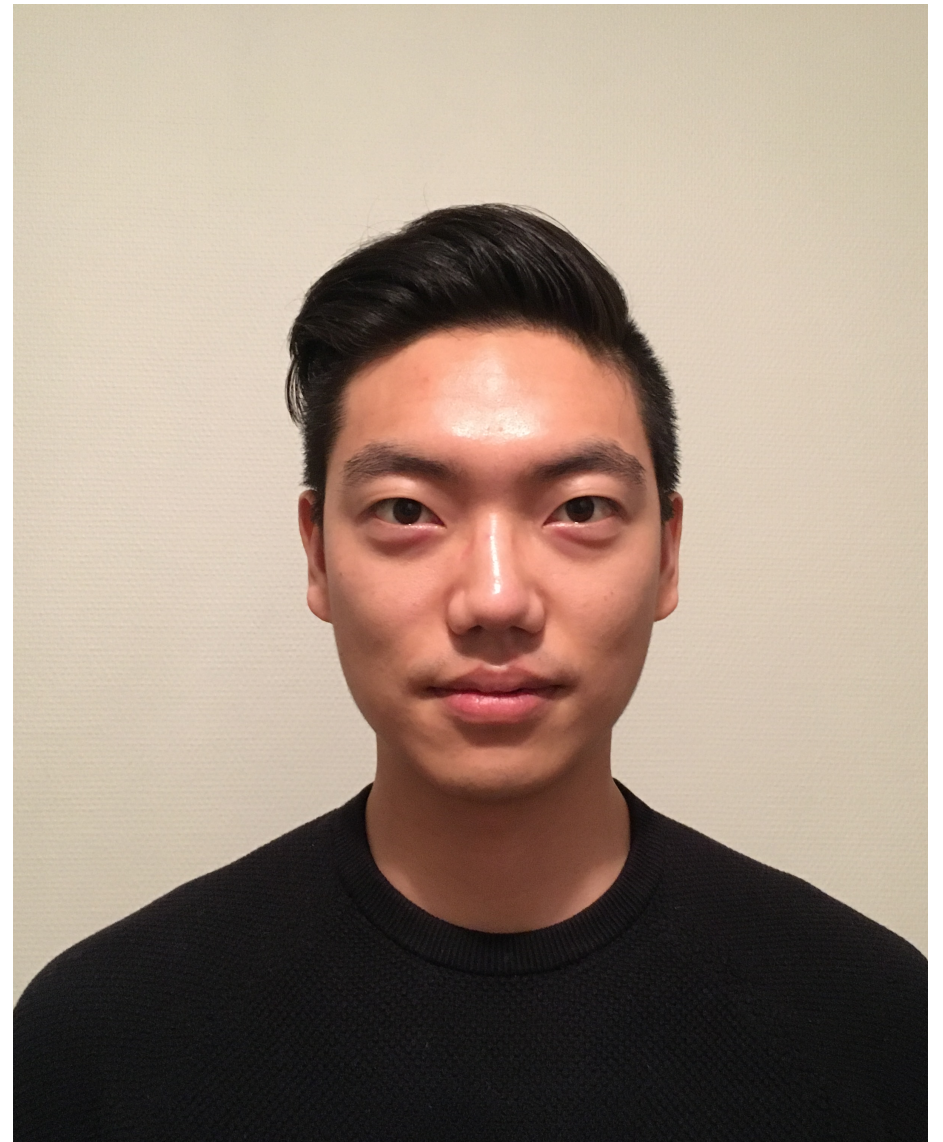
- (석사) University of Stavanger, Computer Science
- (학사) 연세대학교, 컴퓨터과학과
- 대전과학고등학교

전문분야

- 음성처리, 검색엔진
- 데이터 마이닝, 머신러닝, 빅데이터
- 웹/앱 프로그래밍

Contact

- Email : js@codethief.biz



강의계획 (강의 1시간, 실습 2시간)

- 1주차
 - 비정형 텍스트 데이터 분석 기본개념
 - 실습환경 구축
- 2주차
 - 텍스트 데이터 실무 활용사례
 - 텍스트 데이터 핸들링
- 3주차
 - 텍스트 데이터 유형 및 수집
 - BeautifulSoup4 & Selenium
- 4주차
 - 웹크롤링 실전활용법 및 적용사례 소개
 - 한국어/영어 텍스트 데이터 전처리
- 5주차
 - 형태소분석과 개체명인식
 - KoNLPy & NLTK
- 6주차
 - 단어빈도분석
 - TF-IDF
- 7주차
 - 동시출현분석(연관어분석)
 - 단어 네트워크 분석
- 8주차
 - 단어/문서 군집화
 - 토픽모델링
- 9주차
 - 텍스트 감성분석과 활용
 - 문서요약과 키워드 추출
- 10주차
 - 단어 임베딩: Word2Vec
 - 임베딩 기반 연관어분석
- 11주차
 - 비정형 데이터를 활용한 예측 모델링
- 12주차
 - 비정형 데이터와 머신러닝
 - 비정형 데이터 분석 실무 적용사례 소개

강의자료

- **Github**

- 모든 실습자료 및 실습에 관한 정보들은 Github에 업데이트될 예정입니다.

- **Google Colaboratory & Python3**

- 개발환경은 Google Colaboratory을 기반으로 Python3를 사용합니다.

- **Kahoot**

- 매 강의 마지막 10분 동안, 해당 주차 강의 내용을 기반으로 한 Quiz가 있습니다.

- **Kaggle**

- 몇 과제들은 Kaggle을 이용해 제출합니다.

E.O.D