

KUBIG CONTEST

2022 겨울 분반 스터디 9주차

분반 스터디

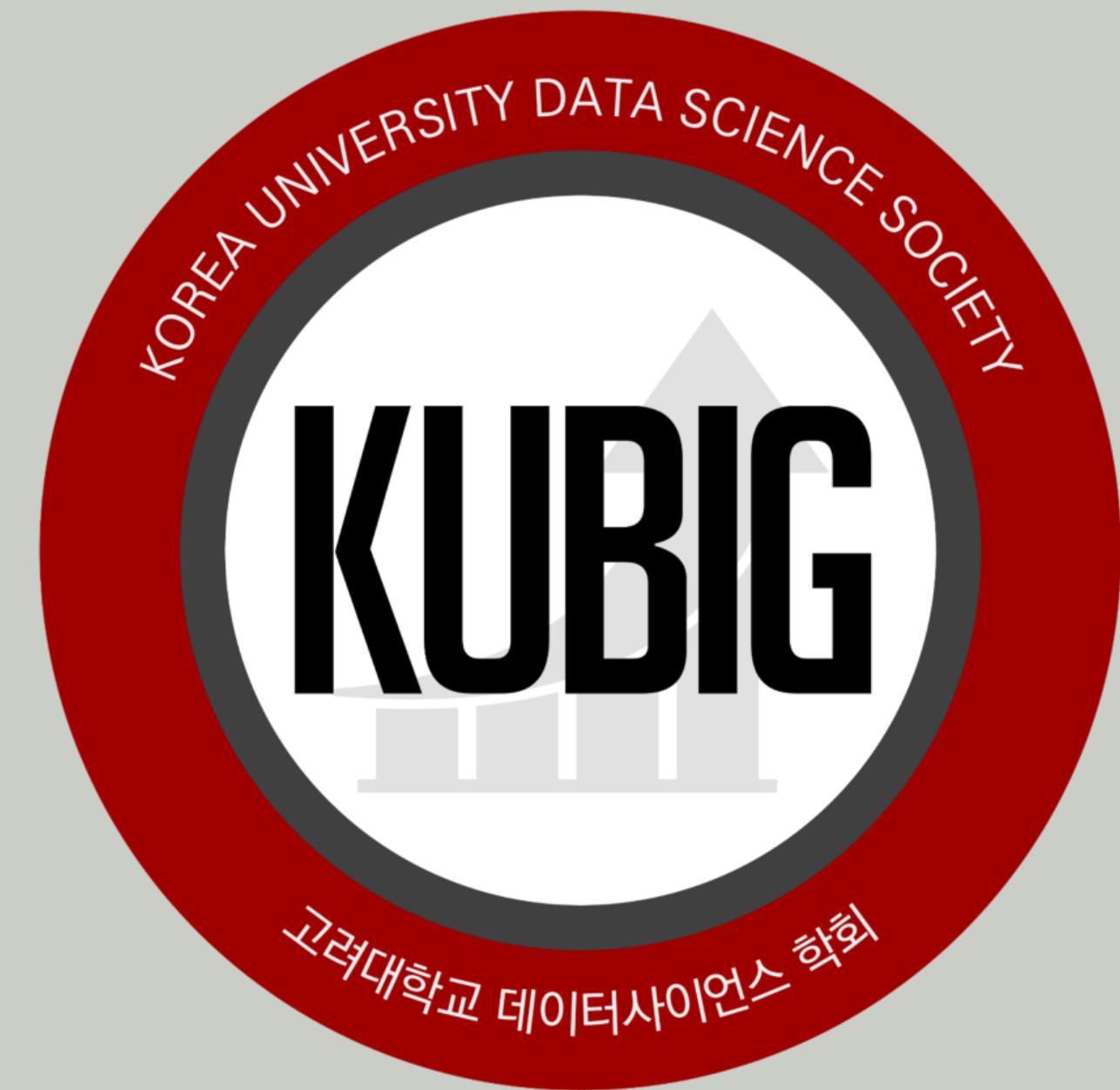
ML / DL / NLP / CV



CONTENTS

- 1부: 머신러닝 4팀 발표
- 2부: 딥러닝 / NLP / CV 총 4팀 발표
- 공지사항: 기획 1팀 / 재무팀

01	ML 팀 발표	- 분반장 분반 소개 - 병원 개/폐업 분류 예측
02	DL 팀 발표	- 분반장 분반 소개 - NLP: 한국어 문장 관계 분류 - CV: hair segmentation
03	NLP 팀 발표	- 분반장 분반 소개 - 뉴스 토픽 분류
04	CV 팀 발표	- 분반장 분반 소개 - 사용자 인터페이스 기반 이미지 조작
05	기획 1팀 공지	- 심화 스터디 진행방식 - 장기 프로젝트 진행방식
06	재무팀 공지	- 오프라인 활동 지원금 - 분반장 수고비



ML 분반소개

분반장: 13기 박주영, 13기 박무성

- 기계학습 입문자를 대상으로 머신러닝 기초이론과 모델 학습
- 핵심 모델에 대한 직관적인 이해와 수리적 원리에 집중
- 코드 리딩, 개인 과제, 그룹 프로젝트 등 다양한 방식으로 매주 코딩 과제를 수행

방학 세션 목표

1. 머신러닝 전반적인 흐름에 대한 이해: How to 'Machine' Learning, Data preprocessing
2. 통계적 머신러닝을 위한 수리적 접근: MLE & Loss Function, Regularization, Kernel, KKT
3. 모델에 대한 수리적, 직관적 이해: Logistic Regression, LDA/QDA, SVM, Ensemble
4. 기초 데이터 분석 실습: 코드 리딩, Classification/Regression 실습, 그룹 프로젝트

방학 세션 커리큘럼

1주차

OT

- MLE
- Likelihood
- Optimization
- Loss Function

2주차

End to End Machine Learning

- Data preprocessing
- Learning Process
- Model Evaluation
- Prediction

3주차

Classification

- Logistic Regression
- Likelihood & Loss Function
- KNN
- LDA, QDA
- Naïve Bayes

4주차

Regression

- Likelihood & Loss Function
- Stein's Paradox
- Regularization: Ridge, Lasso

5주차

설 연휴 휴식

6주차

SVM / Decision Tree

- Lagrange Multiplier
- KKT
- Support Vector Machine
- Decision Tree

7주차

Ensemble1

- Ensemble 학습의 종류 및 직관적 이해
- Random Forest
- Adaptive Boosting
- Gradient Boosting Machine
- Extreme Gradient Boosting Machine

8주차

Ensemble2

- Light Gradient Boosting Machine
- Categorical Boosting Machine
- Auto ML
- 그룹 프로젝트 발표

대회 소개



커뮤니티 대회 교육 랭킹 더보기

병원 개/폐업 분류 예측 경진대회

금융 | 병원 재무 데이터와 AI로 개업/폐업 예측 분석 | 분류 | Accuracy

₩ 상금 : \$3,500 + 40,000ZPR

⌚ 2018.09.15 ~ 2018.10.13 23:59 [+ Google Calendar](#)

👤 761명 마감

병원 개/폐업 분류 예측 경진대회

최근 2개년의 재무정보와 병원 정보를 사용하여 병원의 폐업 여부를 예측하는 '정형데이터' 대회

참가 인원

- 1팀: 15기 염윤석, 15기 고태영, 15기 반민정, 15기 최경석
- 2팀: 14기 박상준, 15기 김제성, 15기 신윤, 15기 정영희
- 3팀: 15기 조우영, 15기 김지호, 15기 우명진, 15기 장수혁
- 4팀: 15기 공도웅, 15기 박지우, 15기 이제윤, 15기 최민경

대회 특징

liquidLiabilities1 – 유동부채, 2017(회계년도)년 데이터를 의미함
 shortLoan1 – 단기차입금, 2017(회계년도)년 데이터를 의미함
 NCLiabilities1 – 비유동부채, 2017(회계년도)년 데이터를 의미함
 longLoan1 – 장기차입금, 2017(회계년도)년 데이터를 의미함
 netAsset1 – 순자산총계, 2017(회계년도)년 데이터를 의미함

ctax2 – 법인세비용, 2016(회계년도)년 데이터를 의미함
 profit2 – 당기순이익, 2016(회계년도)년 데이터를 의미함
 liquidAsset2 – 유동자산, 2016(회계년도)년 데이터를 의미함
 quickAsset2 – 당좌자산, 2016(회계년도)년 데이터를 의미함
 receivableS2 - 미수금(단기), 2016(회계년도)년 데이터를 의미함

- 도메인 지식이 필요한 재무제표 데이터
- 58개의 다양한 feature
- 까다로운 결측치 처리
- 예측해야하는 Test set에 결측치 포함

발표 시작하겠습니다 :D

최종 성능

Accuracy

1팀: 0.8889

2팀: 0.9048

3팀: 0.8730

4팀: 0.8889

대회 결과



우승한 2팀 축하드립니다!!

다른 팀들도 모두 고생 많으셨어요~