

통합구현 포트폴리오

A7팀: 김용현(팀장), 조아진, 김현주, 김승현

B7팀: 임성구(팀장), 신정훈, 신유라, 양정연

C7팀: 정성수(팀장), 이해동, 조성운

팀 별로 다음 과제를 수행하고 결과와 R code, python code를 포함한 과제보고서를 제출하시오.

(빅데이터 분석기법 비교)

1) 본 과정에서 이미 학습한 그리고 본 교과목에서 학습하는 통계기반 데이터분석 기법 및 머신러닝 기반 데이터분석 기법(선형회귀, 회귀, 시계열, 로지스틱회귀, 의사결정나무, 인공신경망, SVM, 앙상블(부스팅, 배깅), 랜덤포레스트, xgboost) 대상으로

(1) R code 내 training model

(2) caret package 내 training model

(3) scikit-learn package 내 training model

을 수립하고 실행하고 비교하시오.

(패키지내 기능 설명)

2) 기타 본 과정에서 이미 학습한 내용 중 caret package 와 scikit-learn package내 기능과 비교 가능한 주제(예, 전처리, 데이터 분할 등) 3개를 선정하여 설명하시오.

- 데이터는 통계/머신러닝 기반 데이터분석 기법 적용 code 실행시 사용 데이터 또는 팀별 자유선택
- Option의 default값으로으로 실행 및 비교 후 R code로 학습한 training model과 같은 결과가 산출되도록 caret package와 scikit-learn package 내 training model의 option을 조정(parameter tuning 실시)

■ Caret Package

<https://topepo.github.io/caret/>


■ Scikit-learn Package

https://scikit-learn.org/stable/user_guide.html

제출기한: 2023년 02월 09일(목) 09:30

제출처: 카페 내 '과제제출' 게시판, LMS(보고서 only)

제출물: 보고서, R codes, python codes, 데이터(데이터가 큰 경우 URL만 제공)

		평가기준표		훈련과정	빅데이터&AI(머신러닝,딥러닝) 인공지능 B_혼합
				훈련기간	2022-08-25 ~ 2022-02-23
훈련교사	손 형 진 (인)	평가항목	통합구현	성 명	
구분	평가세부내용			교수자 평가	
				배점	평가 점수
평가문항 1	팀 보고서 제출여부 (기한 내 제출: 20점/ 당일 제출: 13점/ 후일 제출: 6점/ 미제출: 0점)			20	
	팀 보고서 항목별 codes 포함 여부 (전 항목 포함: 20점, 전 항목의 2/3 이상 포함: 13점, 전 항목의 2/3 미만 포함: 6점, 미제출: 0점)			20	
	팀 보고서 내용 충실도(비교설명 포함) 여부 (충분한 비교/설명: 20점, 간단한 비교/설명: 13점, 미흡한 비교/설명: 6점, 설명 없음: 0점)			20	
	본인의 과제 수행 여부 (전과정 참여: 20점, 1/2 이상 참여: 13점, 1/2 미만 참여: 6점, 미참여: 0점)			20	
평가문항 2	팀 보고서 제출여부 (기한 내 제출: 5점/ 당일 제출: 3점/ 후일 제출: 1점/ 미제출: 0점)			5	
	팀 보고서 항목별 codes 포함 여부 (전 항목 포함: 5점, 전 항목의 2/3 이상 포함: 3점, 전 항목의 2/3 미만 포함: 1점, 미제출: 0점)			5	
	팀 보고서 내용 충실도(비교설명 포함) 여부 (충분한 비교/설명: 5점, 간단한 비교/설명: 3점, 미흡한 비교/설명: 1점, 설명 없음: 0점)			5	
	본인의 과제 수행 여부 (전과정 참여: 5점, 1/2 이상 참여: 3점, 1/2 미만 참여: 1점, 미참여: 0점)			5	
종합평가				훈련생 평가점수	