

바이오화학 데이터과학 재직자 교육 운영 계획(안)

1 추진배경 및 기대효과

□ 추진목적

- 4차산업혁명에 대비하여 바이오 및 화학 관련 데이터 분석 과정들을 통해 데이터에 대한 통찰력, 전공데이터 분석능력, 분야간 융합능력이 높은 미래지향형 창의 융합 인재 양성
- 대학이 보유하고 있는 우수인력을 활용하여 지역기업의 근로자들의 직무능력 향상을 통한 인적자원 개발 및 인력양성 등 기업의 경쟁력 확보

□ 기대효과

- 가족회사 및 지역기업이 가진 기술적 애로사항, 경영문제 등에 대해 대학의 우수한 전문인력을 활용하여 산업체 맞춤형 전문인력 양성
- 대학의 우수한 인적·물적자원 공유를 통하여 지역산업 경쟁력 확보 및 활성화 도모
- 대덕연구단지 등 지리적 접근을 활용하여 특화분야(바이오메디컬 IGC)와 연계한 재직자 교육 프로그램 운영 등 지속가능한 산학협력 사례 및 성과 창출

2 사업개요

- 모집기간 : 2023.08.21. ~ 09.03.
- 교육기간 : 2023.09.04. ~ 12.31. / 2학기 (15주)
- 교육방식 : 온라인 강의
- 교육비 : 1인당 600,000원
- 교육대상 : 본교에 학적이 없는 기업인, 정부출연연구원 등
- 주요내용
 - 통계 및 데이터전처리, 회귀/분류모델의 종류 및 생성/평가(앙상블, 딥러닝)
 - Rapidminer 프로그램을 통한 데이터과학 모델 개발 및 Jamovi 프로그램을 통한 통계 실습
- 접수방법 : 이메일 접수 (산업성장센터 박정후 연구원 jh1389@hnu.kr)

3 추진계획

○ 세부 교육계획

- 수료기준 : 전체 교육기간 중 60% 이상 수강 시 수료증(산학협력단장명) 발급

주차	수업 내용	수업 방법
1주차	강의소개 및 데이터 과학 서론	이론
	Rapidminer 기본구성, 가시화	실습
2주차	기초통계(1) - 데이터 종류 확률, 정규분포, 중심값	이론
	Rapidminer 기본속성, 변환, titanic data 분석	실습
3주차	기초통계(2) - 분산, 정규화, 특이값, 공분산	이론
	Rapidminer 표준화, 단변량 계산, 상관행렬, 공분산 행렬	실습
4주차	분산분석, t-test, f-test, chi-test	이론
	Rapidminer, jamovi ANOVA 분석, t-test, chi-test	실습
5주차	회귀(1): 선형회귀 데이터 변환	이론
	Rapidminer 가중치, 속성선택, 변환, 다중선형회귀식	실습
6주차	회귀(2) : 잔차분석, 로지스틱회귀식	이론
	Rapidminer 신뢰구간, 표준잔차, 로지스틱회귀식	실습
7주차	분류(1) : decision, k-NN	이론
	Rapidminer 의사결정나무, k-NN 모델최적화	실습
8주차	분류(2) : Naive Bayes, SVM	이론
	Rapidminer Naive Bayes, SVM 모델최적화	실습
9주차	모델평가 : 성능값, 과적합, 검증법	이론
	Rapidminer 성능, threshold 결정, 교차검증, bootstrap	실습
10주차	데이터전처리 : 결측치, 잡음, 특이값, 데이터통합/축소	이론
	Rapidminer 결측치 처리, 전진선택	실습
11주차	특징공학: 특징선택, 특징생성 주성분 분석	이론
	Rapidminer 유전자알고리즘, 주성분 분석	실습
12주차	군집분석 : 계층, 중심, 밀도, 모델기반	이론
	Rapidminer k-means, top-down 계통분석, DBSCAN	실습
13주차	복합모델 : bagging, boosting, stacking	이론
	Rapidminer vote 모델, RF, GBT 모델실습	실습
14주차	인공신경망 : perceptron, MLP, 최적화 방법	이론
	Rapidminer 인공신경망 최적화	실습
15주차	딥러닝 : Drop-out, weight, CNN, RNN	이론
	Rapidminer Deep NN 최적화	실습