

이름	조영민	학과	경제학부
학년	4학년	학번	2013121219
		핸드폰 번호	010-4052-0514
컴퓨터언어	Python, R		
수강 과목	통계학입문, 통계방법론, 미분적분학, 선형대수학		
지원동기			
<p>데이터 분석 기술과 인사이트를 얻고 싶습니다. 교수님, 전문가, 그리고 학우들과의 교류를 통해 지식 뿐만 아니라 실제분석의 응용도 기대합니다.</p>			
프로젝트 경험			
<p>Uber China Competitor Tracking team leader  스마트스터디 앱사업팀 데이터분석가  연세대학교 데이터 분석 경진대회 장려상</p>			

이름	전예슬	학과	응용통계학 대학원
학년	1학기	학번	20193111256
		핸드폰 번호	010-3657-9677
컴퓨터언어	R		
수강 과목	통계학입문, 통계방법론, 선형대수학, 수리통계학 1, 수리통계학 2		
지원동기			
<p>세미나를 통해, 최근 기술 및 트렌드를 파악하고 교류를 통해 다양한 관점에서 데이터를 이해하고 분석하는 것을 배우고자 지원하게 되었습니다.</p>			
프로젝트 경험			
<ul style="list-style-type: none"> <li>- 빅콘테스트 2014 경진대회 본선진출</li> <li>- 신한카드 빅데이터 센터 인턴을 통한 대학생 카드 데이터 프로젝트 참여</li> <li>- 국민건강 알람서비스 진료건수 기반 예측 모델링 프로젝트 참여</li> <li>- 현재 다음소프트 재직 중 <ul style="list-style-type: none"> <li>* 공공기관 및 사기업 다수 프로젝트 참여</li> <li>예측 모델링 및 위험도 모델링</li> </ul> </li> </ul>			

이름	홍찬희	학과	응용통계학과
학년	3학년	학번	2013115022
		핸드폰 번호	010-7708-3650
컴퓨터언어	Python, R, SQL		
수강과목	통계학입문, 통계방법론, 미분적분학, 선형대수학, 수리통계학 1, 수리통계학 2		
지원동기			
<p>박태영 교수님께 소개받고 지원하게 되었습니다. 학회를 통해 학부생들, 교수님들과 대학원생들과 교류하며 데이터 사이언스에 대해 좀 더 심층적으로 알 수 있는 기회가 될 수 있을 것이라 생각하여 지원하게 되었습니다.</p>			
프로젝트 경험			
<p>교환학생 시기 머신러닝 관련 수업에서 프로젝트 과제를 통해 데이터 분석 프로젝트를 경험했습니다. 그때 저는 와인 데이터 세트로 와인의 평점을 회귀분석하고 랜덤 포레스트 기법을 통해 와인의 종류를 분류하는 프로젝트를 진행했습니다.</p>			

이름	구한모	학과	경제학과
학년	4학년	학번	2014121136
		핸드폰 번호	010-5141-6069
컴퓨터언어	Python, Java, C		
수강과목	통계학입문, 통계방법론, 선형대수학, 인공지능, Machine Learning(Georgia Institute of Technology 교환학생)		
지원동기			
<p>안녕하세요, 연세대학교 경제학과 14 학번 구한모라고 합니다. 경제학과 컴퓨터과학을 복수전공하면서 Data-driven IT 서비스 기획자가 꿈인 저는 서비스를 통한 혁신과 개선을 더욱 잘 이룰 수 있도록 데이터 사이언스랩에 지원하게 되었습니다. 경제학에서 계량 경제학을 들으며 데이터를 실제로 분석하고 이에 대한 의미를 도출하는 것에 관심을 가지기 시작하였으며 전 학기에 교환학생으로 수강중이던 미국 조지아텍에서 Machine Learning 강의를 수강하면서 이러한 데이터 분석을 통계적 방법론 뿐만 아니라 여러 모델을 이용한 코딩으로서 그 결과를 도출하고 그 의미를 분석하는 데에 대해 많은 호기심이 생겼습니다. 특히 통계학적인 방법론과 컴퓨터과학적인 방법론이 그 출발점은 사뭇 다르나 그 과정과 결과에 있어서 상통하는 부분이 많음을 느끼고 이에대해 조금 더 체계적으로 배우고 경험을 쌓고 싶다는 생각이 들었습니다. 또한 머신러닝을 이용하여 데이터 분석을 진행할 때에도 같은 데이터를 이용하여 같은 목표를 위한 분석을 진행하더라도 서로 다른 모델을 이용한다면 그 결과가 달라질 수 있고 또 그 효율성 측면에서도 차이가 나는 것을 보면서 해당 분야에 대한 호기심을 느끼고 경험에 대한 필요성을 더욱 절실히 느끼게 되었습니다. IT 서비스와 그로 인해 생긴 새로운 연결을 통해 쌓이는 대량의 데이터를 처리하는 것과 분석하는 것은 모든 산업에 있어서 더 나은 상품을 제작하고 더 좋은 서비스를 맞춤으로 제공하는 데에 있어서 필수불가결합니다. IT 서비스 관련 전략을 계획하는 데에 있어 데이터 능력은 시장을 읽고 개선점의 출발 뿐 아니라 새로운 시장에 대한 인사이트를 도출할 수 있는 큰 무기가 될 것입니다. 이렇게 성립된 전략은 직관이 개입될 수밖에 없는 인간의 의사결정을 도와 더욱 효율적이고 혹은 그를 넘어선 창조적인 사회적 혁신을 이룰 수 있다고 믿습니다.</p>			
프로젝트 경험			

2018 년 봄학기에는 연세대학교에서 경제학 전공인 계량경제학 수업을 들으면서 전국 지자체별 쓰레기 배출량과 그 요인을 주제로 팀 프로젝트 분석을 진행한 적 있습니다. 해당 프로젝트에서 전국 지자체별 쓰레기 배출량을 종속변수로 놓고 각 요인의 지역별 변수를 독립변수로 하여 Stata 를 통한 선형회귀분석을 진행하였습니다. 이러한 실제 분석을 위해 단순한 MLS 선형회귀에 대한 개념 뿐만 아니라 Heteroskedasticity, multi-collinearity, auto-correlation 에 대한 검증 방식과 이에 대한 보완 방법 등을 배우고 이를 실제 팀 프로젝트에 적용하기도 하였습니다. 또한 같은 학기에 컴퓨터과학 전공인 인공지능 수업에서는 파이썬을 이용한 기본적인 supervised-learning, reinforcement learning 등을 과제로 직접 코드를 짜고 제출하였으며 해당 강의 후반부에는 Convolutional Neural Network 이나 Recursive Neural Network 등의 개념들을 통해 실제 딥러닝의 개념들을 배울 수 있었고, 간단하게 MNIST 문제를 텐서플로우 코드로 작성하여 과제로써 제출하였습니다. 그리고 2018 년 가을학기에는 교환 수학중이던 Georgia Institute of Technology 에서 수강한 Machine Learning 과목에서는 감독학습, 비감독학습과 강화학습 등의 머신러닝의 여러 모델들과 개념들, 그리고 그 사이의 차이점을 직접 데이터를 이용하여 분석하고 비교함으로써 자세하게 알 수 있었습니다. 또한 같은 학기 수강하였던 Machine Learning for trading 수업에서는 여러 강의와 과제를 통하여 머신러닝 방법론을 이용한 주식 가격 예측시스템을 배웠고 이를 위한 파이썬 과제들을 통하여 Numpy 와 Pandas, Scikit 등의 패키지들에 대한 경험 역시 쌓을 수 있었습니다.

이름	임지현	학과	응용통계학 대학원			
학년	2학기	학번	2018321087			
		핸드폰 번호	010-5472-2455			
컴퓨터언어	Python, R, SAS					
수강과목	통계학입문, 미분적분학, 선형대수학, 수리통계학 1, 수리통계학 2					
지원동기						
<p>데이터 사이언스 전문가가 되고 싶습니다. 빅데이터 분석은 통계를 배우는 학생으로 필요하다고 느껴 연세대학교 대학원 지원 후, 2018 년 6 월부터 8 월까지 경희대학교에서 빅데이터 분석 프로그램을 수료했습니다. 프로그램에서 프로젝트를 진행하였고, 최우수상을 수여 받았습니다. 그 이후로 빅데이터 분석에 흥미를 느끼던 중, 좋은 기회로 세미나를 진행하고 학술적 교류와 경진대회 기회를 얻을 수 있는 학회를 알게 되어 지원합니다. 빅데이터 분석에 대한 관심과 흥미가 있는 학생들과 함께하면 더 큰 배움을 얻을 수 있을 거라 확신합니다.</p>						
프로젝트 경험						
<p>경희대학교에서 개최하는 빅데이터 분석 프로그램에서 프로젝트 경험을 했습니다. 내일배움카드 사용자를 중심으로 한 취업 예측 모델 개발 프로젝트였습니다. 내일 배움 카드는 취업을 목적으로 직업훈련이 필요한 구직자에게 고용노동부에서 국비를 지원하여 실시하는 훈련제도인데, 취업 훈련비를 받아 취미생활을 하는 등 악용하는 사례가 잇따랐고, 그에 따라 내일배움카드의 효율성을 높이기 위한 취업 여부 및 취업 준비 기간 예측모델을 제안했습니다. 정부를 통해 내일 배움 카드 관련 데이터를 받았습니다. 결측치와 이상치를 처리하고, 교육 프로그램명에 맞는 통계청 산업대분류 산업별 변수를 입력했습니다. 또한 공공데이터를 활용해 코스피 지수, 코스닥 지수를 추가하였습니다. 그 후, 상관관계를 분석하여 데이터를 정하고, LASSO 를 활용하여 사용할 칼럼을 정했습니다. 취업 여부 예측모델은 xgboost 를 사용했고, 취업 준비 기간 모델은 랜덤포레스트를 사용했습니다.</p> <p>빅콘테스트에서 블레이드앤소울 사용자 이탈 예측 모델을 생성하는 경험을 했습니다. 인 게임 활동량, 파티 구성 관계, 결제 금액, 문파 구성 관계, 유저간 거래 이력에 해당하는 14 만 개의 데이터를 받았습니다. 2117 개의 변수를 생성했습니다. 예를 들어, 유저간 거래 이력에 해당하는 데이터는 판매자 계정 아이디 기준, 구매자 계정 아이디 기준, 거래 활동 시간 기준을 정하고, 그에 맞는 변수를 추가하였습니다. 판매자 계정 아이디 기준은 아이템 판매 개수와 판매 물품 종류와 판매 개수, 구매자 계정 아이디 기준은 아이템 구입 개수와 구매 물품 종류와 구매 개수, 거래 활동 시간 기준은 아침, 점심, 저녁, 새벽으로 분류하고 월-목요일인지, 금-일요일인지 나누었고, 구매 물품 종류와 구매 개수에 해당하는 변수를 추가하였습니다. 이렇게 생성된 변수들을 pca 차원 축소를 통해 568 개의 변수로 변경 해주었고, xgboost 기법을 활용하여 모형의 정확도가 73.05%인 예측 모형을 만들었습니다.</p>						

이름	김미연	학과	경제학과
학년	2학년	학번	2018121184
		핸드폰 번호	010-4957-5456
컴퓨터언어	Python, R		
수강과목	통계학입문, 선형대수학		
지원동기			
<p>바로 직전학기인 1 학년 2 학기에, 김재경 교수님의 컴퓨팅적 사고와 sw 프로그래밍을 수강했습니다. 프로그래밍에 대해 아는 것이 전혀 없었지만, 이 수업에서 파이썬을 처음으로 배우고 프로그래밍에 흥미가 생겼습니다. 문제 상황을 파악하고 문제 해결을 위한 알고리즘을 구상하여 직접 제 손으로 프로그램을 짜서 문제를 해결하는 것이 저에게 큰 성취감을 주었기 때문입니다. 대학에서 배운 파이썬을 흘러가는 한 학기 교양수업의 지식정도로 두기에는 너무 아쉬웠고, 그래서 방학동안 파이썬을 심화해서 배우거나 씨언어, 자바, 자바스크립트, html 등 다른 프로그래밍 언어를 배워보고 싶었습니다. 그러나 지금껏 앱 개발이나 웹 개발에 관심을 갖고 있던 않았던 터라 무엇을 목표로 프로그래밍 언어를 공부할 지 고민했습니다. 다음 소프트 부회장님이 티비에 나와서 특정 연예인에 대한 여론을 언어 클라우드로 보여주는 것을 보고, 저도 제가 궁금한 분야를 검색하거나 과제를 위한 자료조사를 할 때 인터넷의 방대한 데이터를 언어 클라우드로 표현하면 좋겠다고 생각해왔었고, 빅데이터 분야에 대한 막연한 흥미를 갖고 있었습니다. 이번 기회에 검색을 해보니 파이썬으로 크롤링 코드를 짜서 통계적 처리를 통해 표현할 수 있음을 알았고, 데이터 분석을 공부하면 자연스레 도구로서 파이썬을 사용하게 되고 두 공부의 시너지를 얻을 수 있겠다고 생각했습니다. 아직 진로에 대해 고민해보는 시기이기에 데이터와 관련된 공부, 실습을 하면서 관련된 진로로 구체화 해보고 싶어 지원하게 되었습니다. 2 학기 이상 활동 가능하고 아직 아는게 많이 없기 때문에 붙게되면 따로 열심히 공부하면서 학회활동을 하겠습니다.</p>			
프로젝트 경험			
<p>통계에대한 경험은 통계학 입문 수업에서 배운 내용이 전부입니다. 만약 학회원이 된다면 따로 독학을 하면서라도 성실히 활동하겠습니다!!</p>			

이름	민형규	학과	응용통계학과
학년	3학년	학번	2014122007
		핸드폰 번호	010-9946-1878
컴퓨터언어	R		
수강과목	통계학입문, 통계방법론, 미분적분학, 선형대수학, 수리통계학 1, 수리통계학 2, 회귀분석, 금융 리스크 관리 실무		
지원동기			
<p>3 학년이 되어 취업을 위한 스킬을 쌓고자 기회를 알아보던 중 데이터 사이언스 학회가 설립된다는 소식을 들어서 지원하게 되었습니다. 데이터 사이언스에 관심을 가지게 된 것은 지난 학기의 시사 통계학 수업을 통해서 LG 디스플레이와 SKT 데이터 연구소에서 근무하시는 선배님들의 이야기를 통해서 였습니다. SKT 연구소에서 일하시는 선배님은 첨단기술이라 저희와는 연이 없을 것 같았던 인공지능 'NUGU'의 개발에 참여하고 계신 분이었고 LG 디스플레이에서 근무하고 계시는 선배님은 공정의 신호를 포착하여 공정오류를 예측하는 일을 하시는 분이었습니다. 이렇듯 현장에서 데이터를 이용하여 실제 가치를 창출하시는 선배님들의 이야기를 듣고 큰 감명을 받았습니다. 따라서 저도 선배님들처럼 현장에서 전공지식을 살리고자 데이터 사이언스 공부를 하고자 합니다.</p> <p>그렇기에 18-2 학기에 들은 전공수업에 더 열의를 가지고 임하게 되었습니다. 수리 통계학 2 수업에서는 교수님께 적극적으로 질문을 드려 보다 포괄적인 지식을 갖고자 하였습니다. 회귀분석 수업에서는 수업시간에 배운 내용 중 궁금한 내용을 해외 사이트 서핑을 하면서 탐구하였습니다. 이러한 노력이 빛을 발해 두 수업에서 좋은 성적을 받게 되었습니다. 데이터 사이언스 학회에서도 많은 탐구를 통해 데이터에 대한 심도깊은 이해를 하고 많은 것을 배우고자 합니다.</p>			
프로젝트 경험			
<p>데이터 분석 프로젝트의 경험은 많지는 않습니다. 1 학년 2 학기 김철웅 교수님의 통계방법론 강의와 2 학년의 컴퓨터 자료처리를 통해 R 을 사용한 과제를 다수 한 경험은 있습니다. 특히 통계방법론 강의에서는 당시 R 을 따로 배우지 않은 상황에서 R 을 독학으로 학습했고, 다른 친구들과 함께 모여 의논하면서 과제를 해나갔습니다. 또한 컴퓨터 활용능력 1 급 자격증이 있어 엑셀과 엑세스를 이용한 데이터 처리 기술을 배웠습니다. 이 때의 경험을 살려 수리통계학이나 회귀분석 강의에서 애매한 것이나 궁금한 것이 있을 때에는 엑셀이나 R 을 활용하여 결과를 내보거나 그래프를 그려보거나 하였습니다. 현재는 독학으로 인터넷에 무료로 풀려있는 파이썬 강의를 듣고 있습니다.</p> <p>차후 학회에 들어가게 된다면 기업의 재무제표를 활용하여 여러 가지를 예측하는 활동을 해보고 싶습니다. 2 년간 사회복무요원을 하면서 회계 공부를 하였는데 이때의 지식을 살려 보다 회계를 이용한 더 가치있는 무언가를 만들어 보고자 합니다. 또한 금융 리스크 실무 강의의 내용들을 기업의 재무제표를 활용해 실습해보고 현장에서 어떤 감각으로 신용평가가 이뤄지는 지를 체험해보고 싶습니다. 또한 회귀분석 강의에서 배운 회귀분석 스킬들을 직접 활용하고 매트릭스를 이용한 자료처리를 경험하며 더 나아가 딥 러닝과 빅 데이터에 관한 내용들을 배워 보고자 합니다.</p>			



이름	이승은	학과	응용통계학과
학년	3학년	학번	2017122011
		핸드폰 번호	010-5191-9770
컴퓨터언어	Python, R, C++		
수강과목	통계학입문, 통계방법론, 미분적분학, 선형대수학, 수리통계학 1, 회귀분석		
지원동기			
<p>고 3 때 응용통계학과를 지원했을 때부터 저는 데이터 분석을 바라보고 지원하였습니다. 학부 과정에서 배우는 것들은 모두 이의 초석이 되는 것이라 열심히 배우고 있지만, 데이터 분석을 꿈꾸고 전공을 선택했기 때문에 조금이라도 빨리 통계와 코딩에 친숙해지고 싶어 컴과 부전공을 바라보고 저번학기 부터 컴과 전공을 듣기 시작했습니다.</p> <p>그러나 컴과 전공 수업에서는 제가 바라던 데이터들을 다루는 비중은 없었고 컴퓨터 언어를 배우는 고 문제를 푸는 것에 그쳤습니다. 물론 컴퓨터를 사용해서 데이터를 분석하기 위해서는 컴퓨터 언어와 친숙해지는 것도 필요하지만 조금 더 제 목적에 맞게 실용적으로 이 언어를 사용하고 싶습니다.</p> <p>‘모든 기초를 쌓고 데이터를 분석하자’ 라는 마인드로 기반만 닦아 왔는데, 시간이 지날수록 실전을 경험해보고 부딪히면서 배우는 것이 더 크다는 생각과, 너무 늦은 것은 아닐까 하는 걱정이 생겼습니다. 이에, 예정되어 있던 교환학생보다 1 년의 학회 활동으로 얻는 지식과 경험들이 지금의 저에게는 더 중요한 것이 되었습니다.</p> <p>공통의 관심사를 가진 사람들과 함께 데이터 분석 프로젝트를 진행하면서 함께 성장하고 싶습니다. 물론 제가 아직 데이터 분석 경험이 없기 때문에 초반에는 힘들 수 있지만, 그 만큼 열심히 따라잡을 각오가 되어 있습니다. 경험이 없는 만큼 열심히 배우고 노력하겠습니다.</p>			
프로젝트 경험			
없습니다.			

이름	차수만	학과	응용통계학과			
학년	4학년	학번	2015122025			
		핸드폰 번호	010-5911-1939			
컴퓨터언어	Python, R					
수강과목	통계학입문, 통계방법론, 미분적분학, 선형대수학, 수리통계학 1, 수리통계학 2					
지원동기						
<p>2018 년 여름방학 때 빅콘테스트와 빅데이터 경진대회를 시작으로 실제 데이터 분석에 처음 발을 들였습니다. 상크모라는 상경대 기독교 동아리 회원들과 함께 팀을 꾸려 대략 3~4 개월동안 데이터 분석 기법에 대해 배우면서 이 분야에 대해 큰 흥미를 갖게 되었습니다. 또한 이 분야에 먼저 발을 들여 치열하게 고민하고 노력하는 사람들과 경쟁해봄으로써 체계적인 공부의 필요성을 체감하였습니다. 지난 3~4 개월의 노력이 헛되지 않고 이 분야로 나아가는 데에 있어 더 큰 원동력이 되었지만 한가지 아쉬웠던 점은 기존에 함께 모여 꾸준히 공부해왔던 것이 아닌 점이었습니다. 따라서 데이터사이언스랩 학회에 참여하여 꾸준한 공부와 노력을 통해 데이터 분석에 정진하고자 지원하였습니다.</p>						
프로젝트 경험						
<p>위에서 언급했듯이 2018 빅콘테스트와 교내 빅데이터 경진대회에 참가한 경험이 있습니다. 빅콘테스트에서는 Analysis 분야 중 챔피언 리그에 참가하였습니다. 여기서 NC soft 의 '블레이드앤소울' 게임이탈자 분석을 시행하였습니다. 훈련 기간이 8 주간 주어지고 8 주를 기준으로 다음 12 주 간의 게임 활동에서 이탈을 언제 하는지를 분석하는 것이었습니다. 먼저 데이터 구조가 각 변수값들이 표준화되어 있는 점을 감안해 각 변수들의 관측값의 상관성 계산과 feature engineering 을 위해 관측값들의 최저값을 1 로 통일하는 작업을 시행했습니다. 그뒤 유튜브와 게임 커뮤니티 조사를 바탕으로 해당 게임을 이탈하는 이유에 대한 도메인 연구를 실시하였으며 이를 통해 몇가지 지수들을 만들었습니다. 그뒤 주어진 데이터들을 병합한 후 한 아이디어에 대해 한 행으로 데이터가 나열되도록 전처리하였습니다. 그후 모델을 xgboost 로 선정하여 학습을 시켰고 파라미터를 최적화하였습니다. 이렇게 분석을 하여 제출하였지만 지수의 논리적 타당성, 모델 파라미터의 불안정성, 시간관리에 대해 문제점을 발견하였고 더 안정적 분석을 위한 받침돌이 되었습니다.</p> <p>교내 빅데이터 경진대회에서는 마크로밀 엠브레인에서 제공된 설문 데이터와 앱사용 데이터를 바탕으로 마크로밀 엠브레인 사용자들에 대한 분석을 실시하였습니다. 빅콘테스트에서 주어진 수량화된 데이터와 달리 설문 데이터는 명목형 변수와 텍스트로 이루어져 있어 데이터를 전처리하는 과정이 까다로웠고 분석 기법 또한 까다로웠습니다. ROCK 라는 클러스터링 기법을 사용하려 했지만 데이터의 방대함과 모델에 대한 이해 부족으로 실패하였습니다. 이에 눈을 돌려 사용자들의 생활정보를 담고 있는 앱사용 데이터를 기준으로 K-means 클러스터링을 시행하였고, Nonparametric MANOVA(비모수 다중분산분석)으로 클러스터링의 유효성을 검증하였습니다. 이에 그치지 않고 앱 중에서 어떤 앱이 클러스터를 나누는데 큰 기준을 차지하는지를 의사결정나무를 사용하여 분석하였습니다. 개인적으로 아쉬웠던 점은 많은 텍스트와 설문 데이터가 있었지만 이를 곧장 바로 분석하지 못했다는 점입니다. 이번에 학회에 참여하여 텍스트 분석 기법과 분석을 하기 위한 기초를 더욱 단단히 다지고 싶습니다.</p>						

이름	이재성	학과	기계공학과
학년	3학년	학번	2014145110
		핸드폰 번호	010-6663-8739
컴퓨터언어	Python, C, C++, C#		
수강과목	통계학입문, 미분적분학, 선형대수학		
지원동기			
<p>현재 응용통계학과 부전공을 하고 있는 이유가 제 본과인 기계공학과로 취업을 할 것이 아니기 때문입니다. Data specialist 로서 직장을 가지며 커리어를 쌓기에 이번 기회가 좋다고 생각되어 지원하게 되었습니다.</p>			
프로젝트 경험			
<p>설계 및 생산공학에서 떨어지는 낙하 물체 설계 모양에 따른 anom, anova 분석을 해본 경험이 있습니다. 비록 많은 경험을 가지고 있지는 않지만 배우기 위해 지원하였습니다.</p>			

이름	홍주혁	학과	응용통계학과
학년	3학년	학번	2013122049
		핸드폰 번호	010-7270-1080
컴퓨터언어	Python, R		
수강과목	통계학입문, 통계방법론, 미분적분학, 선형대수학, 수리통계학 1		
지원동기			
<p>What's next? 어떤 정보를 접할 때, 언제나 생각하게 되는 문구입니다. 응용통계학과 전공생으로 늘 데이터를 다루지만 무엇을 위해 이 강의를 듣고, 이런 공부를 하고 있는지 혼자서 많이 고민을 했었습니다. 완벽한 답은 아니지만, 그 목적에 대한 답을 생각하는데 있어서 선배들과의 대화, 동기들과의 토론, 사람들과의 상호작용, 협동이 많은 도움을 주었습니다. 혼자서 공부하는 데이터 사이언스가 아닌 다른 사람들과 도우며 공부하는 데이터 사이언스를 하고 싶어 이 학회에 지원하게 되었습니다.</p>			
프로젝트 경험			
<p>2018 년 하반기에 SK telecom 의 ICT 기술원에서 인턴으로 있었던 덕분에 실무적인 분석 프로젝트에 참여할 수가 있었습니다. 토픽은 Cloud Server Data 의 Clustering 을 통한 Anomaly Detection 입니다. 클라우드 서버를 운영하는 시스템에서는 엄청나게 많은 Log data 들과 여러 종류의 Metric data 들이 나오게 됩니다. 이 데이터들을 일일이 확인할 수가 없기 때문에 운영 개선과 오류 탐지에 도움이 되는 정보들을 얻는데에 문제가 있었는데, 이를 몇 가지의 clustering 알고리즘으로 해결하는 프로젝트에 있었습니다. Log data 는 time stamp 가 찍힌 text data 였는데, 이런 raw text 들을 변환하여 일정 포맷으로 만들고, EDA 를 통해 불필요한 정보들을 제거하는 작업을 해준 후 나누어진 포맷에 text 별 거리계산 알고리즘을 만들어 군집화하는 작업을 진행했습니다. 이를 통해 (예시로) 60 만개의 로그 데이터를 10 초안에 10 개 정도의 log type 으로 분류를 하게 되었고, 이를 통해 type 별 발생 빈도를 통한 anomaly detection 이 가능하게 되었습니다. 실제 실무를 통해 분석을 해 보니, 문제 이해부터 목표 설정, preprocessing, 모델 설정 등 데이터 분석의 한 사이클을 다 경험해 본 것 같습니다.</p>			

이름	양준성	학과	응용통계학과
학년	3학년	학번	2015122009
		핸드폰 번호	010-5307-9392
컴퓨터언어	R		
수강과목	통계학입문, 통계방법론, 미분적분학, 선형대수학, 수리통계학 1		
지원동기			
<p>군 전역 후 진로에 대해 고민하고 있습니다. 전역을 한 지 한 달도 채 안된 저는 지금 무엇을 해야 하는지 몰라 미래에 대해 막연한 걱정만 하고 있었습니다. 그러던 중 모집글을 톡방에서 보게 되었고, 입대 전 컴퓨터자료처리 강의에 굉장히 흥미를 느꼈던 것을 떠올려 지원하게 되었습니다. 비록 지금은 강의를 들은지 오래되어 기억이 거의 나지 않지만, 열심히 해서 다시 그 때의 흥미를 느껴보고 지금 느끼고 있는 막연한 걱정을 떨쳐버리고 싶습니다.</p>			
프로젝트 경험			
<p>경험 없습니다.</p>			

이름	이승한	학과	(본)경영학/(복수)응용통계학			
학년	3학년	학번	2015123127			
		핸드폰 번호	010-8768-8472			
컴퓨터언어	Python, R, SQL, ( SQL 은 기초적인 수준으로 알고 있습니다 )					
수강과목	통계학입문, 통계방법론, 컴퓨터와 자료처리					
지원동기						
<p>저는 경영학과에 재학 중이고 응용통계학과를 복수전공 할 예정인 3 학년 학생입니다. 경영학과지만, 통계 수업에 더 관심이 많고, 데이터를 분석하는데 흥미가 있어서 이쪽 분야로 더 깊게 파고들고자 통계 복수 전공과 더불어 이 학회에 지원하게 되었습니다. 이러한 관심사를 바탕으로 올해 초부터 Python 과 R 독학을 시작으로, 빅데이터 분석 동아리에 들어가서 이 두 언어를 활용하여 머신러닝과 텍스트 마이닝을 학습해왔습니다. 하지만 동아리에서 하던 활동들은, 돌아가면서 각자 공부해온 부분들을 코딩과 개념을 학습하고 발표하는 위주로 이루어져서 실제로 복잡한 데이터를 심층적으로 분석하는 실질적인 프로젝트 활동은 많지 않았습니다. 그래서 배워온 것들을 실질적으로 적용할 수 있는 기회가 없어서 늘 아쉬움을 느끼고 있었습니다. 하지만 이 학회는 그러한 활동들도 할 수 있으면서도, 동아리보다도 더 강도 높은 학습까지 할 수 있어서 저에게 더 적합한 단체라고 생각했습니다. 그래서 이 학회에 들어가게 되면, 지금까지 배워온 데이터 분석 기법들을 더 심화적으로 학습하고 싶고, 프로젝트에도 참여하면서 이러한 것들을 활용할 기회를 가지고 싶습니다. 뿐만 아니라, 교수님들과 대학원생 분들과의 학술적인 교류를 통해서 학부생으로서 쉽게 접할 수 없는 것들을 배우고 싶습니다. 현업에서는 이러한 것들이 실제로 어떻게 활용되고 있는지에 대해서도 더 많이 알게 됨으로써 앞으로 이 분야로 나아가기 위해 어떠한 방향으로 공부를 해야할 지에 대해 알아 갈 수 있는 좋은 기회라고 생각했습니다.</p>						
프로젝트 경험						
<p>제가 활동했던 빅데이터 분석 동아리에서 아직 '프로젝트'라고 할 수 있을 만한 큰 활동을 한 경험은 없습니다. 대신 매주 새로 배운 모델들을 활용하자는 차원에서, 매주 다양한 모델들을 이용하여 여러 작은 문제들을 해결한 경험은 있습니다. 인터넷에 있는 데이터들을 가져와서 해당 데이터를 잘 분석할 수 있는 가장 적절한 모델을 찾고 그를 활용하여 데이터를 분석하는 방식으로 진행해왔습니다.</p> <p>해당 동아리에서는 크게 '머신 러닝'과 '텍스트 마이닝' 둘을 통해 학습 및 분석을 했습니다. 머신러닝에서 해결한 여러 작은 문제들은,</p> <p>(1) K-fold CV 를 적용한 부스팅 모델을 만들어서 "인디언들의 당뇨병 여부" 결과 예측하기</p> <p>(2) 미국 51 개 주를 대상으로 특정 시기에 발생한 범죄들에 대해 군집화 모델을 통해 비슷한 주들끼리 묶거나</p> <p>서울시에서 발생한 교통사고 사망자/사상자 관련 데이터를 회귀분석 모델을 이용하여 분석하기</p> <p>(3) 여러 자동차들을 그것의 외형의 특징들(차의 폭,너비,길이의 비율 등)을 사용하여 트리 모델을 활용하여 분류하기</p> <p>등이 있었습니다.</p> <p>텍스트 마이닝의 경우,</p> <p>(1) Amazon 사의 고객 평가를 끌어와서 전처리하고 Polarity 함수를 사용한 분석</p>						

(2) 트럼프와 문재인 대통령의 연설문으로 각각 영어/한글의 wordcloud 생성 등이 있었습니다.

위의 머신러닝 및 텍스트 마이닝 활동들은 동아리에서 전부 R 을 활용해서 진행해왔고, 동아리 내 소모임에서는 Python 을 통해 위와 비슷한 활동들을 진행해 왔습니다. 현재 하고 있는 동아리의 (의무적) 커리큘럼이 2 월 첫째 주면 마무리 되고, 그것이 끝난 뒤에는 (선택적) 프로젝트 활동이 있습니다. 하지만 저는 프로젝트를 하게 될 경우, 동아리보다는 더 전문적이고, 교수님들과 대학원생 분들까지도 함께 계신 학회에서 활동을 하면 더 좋을 것 같다고 생각했습니다. 아직까지는 큰 규모의 프로젝트를 한 경험은 없기 때문에, 부족하다면 혼자서라도 따로 더 열심히 공부해와서 이번 기회를 통해서 이 학회에서 꼭 프로젝트를 진행해보고 싶습니다.

이름	김민채	학과	철학과/사회학과
학년	4학년	학번	2014117022
		핸드폰 번호	010-5171-7042
컴퓨터언어	Python, SQL, Spss		
수강과목	사회통계학		
지원동기			
<p>사회통계학 수업 수강 이후 통계를 좀 더 심도있게 배워보고 싶었는데 학회에 참여하는 것이 좋은 기회인 것 같아서 지원하게 되었습니다.</p> <p>사회 통계학 이전에 데이터 사이언스를 처음 접한 것은 작년 여름 방학이었습니다. 데이터진흥원의 빅데이터 전문가 과정에 한달 정도 참여 python 과 sql 등 데이터 처리 과정에 대해 짧게나마 배운 경험이 있습니다. 그 이후, 지난 학기에 사회통계학 수업을 수강하며 기본적인 통계에 대해 배웠고, 수업 이후에도 계속 통계와 통계 프로그램을 배워보고 싶다는 생각을 하게 되었습니다.</p> <p>사회 통계학 시간에 배운 spss 보다 좀 더 높은 수준의 통계프로그램을 다루고 익히는데 학회가 좋은 기회가 될 것이라고 생각합니다. 적지 않은 나이인 만큼 제가 도울 수 있는 부분이 있다면 열심히 돕고, 참여하여 배우고 싶습니다. 감사합니다.</p>			
프로젝트 경험			
<p>데이터프로그램 분석 프로젝트 경험은 없지만, 사회 통계학 시간에 소논문을 작성한 경험이 있습니다. 한국 청소년의 학교폭력 경험 유형에 따른 주관적 행복감에 대한 연구로 평일 수면시간의 조절효과를 밝혔습니다. SPSS 를 이용하였고, 2018 년 방정환 데이터를 사용하여 직접 코딩, 분석하였습니다. 아직까지는 프로젝트 경험이 없지만, 학회에서 기회가 주어진다면 분석 프로젝트에 꼭 참여하고 싶습니다.</p>			



이름	오승미	학과	응용통계학과
학년	4학년	학번	2015122051
		핸드폰 번호	010-4566-1056
컴퓨터언어	Python, R, SQL		
수강과목	통계학입문, 통계방법론, 미분적분학, 선형대수학, 수리통계학 1, 수리통계학 2, 회귀분석, 확률과정 등		
지원동기			
<p>저와 같은 방향성을 가진 사람들을 만나 다양한 프로젝트를 하면서 많이 배우고 싶어서 이렇게 지원하게 되었습니다.</p> <p>작년 한 해동안 미국으로 교환학생을 갔는데, 그곳에서 배우면서 프로젝트의 중요성을 깨달았습니다. 한국에서 수업을 들을 때는 프로그래밍을 잘 알지도 못했고, 특히나 파이썬은 학과수업에서 접해볼 기회가 없었습니다. 그런데 미국에 가보니 현재 통계학의 트렌드에 맞추어 각종 데이터사이언스와 관련한 수업을 활발히 진행하고 있었고, 이를 통해 프로젝트로 발자취를 남겨 놓는 것이 취업에 얼마나 중요한지 깨달을 수 있었습니다. 일년동안 수업을 통해 다른 많은 데이터 사이언스 쪽을 희망하는 친구들을 만나서 프로젝트를 진행하며 많이 배울 수 있었습니다. 비단 파이썬,git 과 같은 기술적인 측면 뿐만 아니라 제가 생각만하고 있던 커리어를 구체화하고 실현시키는 친구들을 보면서 자극도 많이 받았습니다.</p> <p>프로젝트의 중요성을 몸소 체감했지만, 현실적으로 저의 능력치의 한계로 수업외 프로젝트를 혼자서 진행하기엔 무리가 있다는 판단이 들었습니다. 이곳 데이터 사이언스 랩에서 같은 꿈을 꾸는 사람들을 만나 같이 고민하고 배워가면서 성장하고 싶습니다.</p>			
프로젝트 경험			
<p>작년 한 해 동안 파이썬을 다루는 수업을 꾸준히 수강하고, 당연히 이를 사용한 프로젝트를 진행했습니다. 가장 힘들었지만 기억에 남는 프로젝트는 stackoverflow 사이트에서 답변할 만한 질문을 추천해주는 시스템을 만드는 것이었습니다. 학부생이다보니 결과적으로 봤을 때는 많이 미흡하지만 프로젝트 주제를 선정하는 과정에서 부터 결과물을 만들어 사이트에 배포하고 하는 과정에서 성취감을 느낄 수 있었습니다. 교수님이 한 학기의 절반을 EDA 과정에 힘을 줘야한다고 강조하셨는데, 프로젝트가 끝나고 나서야 이를 깨달을 수 있었습니다. 수업시간에 주어지는, 가이드라인이 있는 데이터셋을 분석하는 것은 큰 어려움이 없었지만 막상 방대한 양의 데이터, 그 사이트의 설립부터 지금까지의 누적된 질문과 답변 데이터들, 를 마주하니 어디서부터 손을 대야할지 눈앞이 캄캄했습니다. 양이 너무 많아서 뭐하나 손대보려하면 프로그램이 돌아가는데 시간을 많이 허비하다보니 너무 막막했습니다. 다행히 교수님의 도움과 조원들의 고민끝에 여러가지 그래프를 뽑아내면서, 질문의 만족도나 답변속도에 영향을 미치는 변수들이 어떤 것인지 조금씩 알아갔습니다. 하지만 매번 '이정도면 됐어' 라고 생각했어도, 교수님은 면담 때마다 이런 방향, 저런 방향에서의 접근 방법을 알려주셨고 감히 저희의 힘으로는 단시간내에 그많은 인사이트를 도출해내는게 거의 불가능에 가깝다는 것을 깨달았습니다. 또한 제가 발견한 것들을 다른 사람에게 효과적으로 전달하기 위해서 피피티는 어떻게 구성되어야하고 그래프를 낭비하지 않으려면 어떻게 해야하는지, 심지어 색상을 선정하는 방식도 교수님께 많이 배웠습니다. 10 주간의 시간을 통해 느낀건, 제가 생각해왔던 이상속의 데이터 분석보다 훨씬 치열하고, 많은 시간을 들인다고 해서 답이 나오진</p>			

않았습니다. 하지만 이를 통해 더 다양한 데이터 분석 프로젝트에 도전해보고 싶다는 열정이 생겼습니다.

이름	강태규	학과	응용통계학과 대학원
학년	3학기	학번	
		핸드폰 번호	010-9487-3589
컴퓨터언어			
수강과목	통계학입문, 통계방법론, 미분적분학, 선형대수학, 수리통계학 1, 수리통계학 2		
지원동기			
<p>실제 데이터를 분석하는 능력 및 경험을 얻고 그에 필요한 이론들이 어떠한 것들이 있는지 알아 다른 학회원들과 함께 공부하고 싶어서 지원합니다.</p>			
프로젝트 경험			
<p>대내 및 대외적인 프로젝트에 대한 경험은 전무합니다.</p>			

이름	권다희	학과	응용통계학과
학년	3학년	학번	2016122006
		핸드폰 번호	010-6287-7674
컴퓨터언어	Python, R		
수강과목	통계학입문, 통계방법론, 미분적분학, 선형대수학, 수리통계학 1, 확률과정, 실험계획법, 금융리스크관리 및 통계, 회귀분석		
지원동기			
<p>통계학이라는 학문이 좋았던 가장 큰 이유는 여러 학문 중 가장 실용적이고 많은 사회현상을 설명할 수 있는 기반이라 생각했기 때문입니다. 그중에서도 ‘통계를 기반으로 한 데이터 분석’이 요즘 통계학의 가장 핵심이라 생각했기에 이와 관련된 공부를 꾸준히 해왔습니다.</p> <p>데이터 분석의 툴을 처음 접한 것은 ‘컴퓨터와 자료처리’라는 과목을 수강했을 때입니다. 이 수업을 통해 R 프로그램에 대한 기본 지식을 쌓았고, 코딩을 통한 데이터 분석에 관심을 두게 되었습니다. 이를 계기로 R 에 대해 더 다양한 기법들을 배울 수 있는 과목들을 수강하였고, 데이터 처리 및 머신 러닝의 여러 기법을 배웠습니다. 또한, 지난 학기 빅데이터 경진대회에 참가하면서 실제 기업의 데이터를 통해 지금까지 배운 기법들을 적용해보는 시간을 가졌습니다. 현재는 R 프로그램뿐만 아니라 파이썬으로도 다양한 기법들을 공부하며 데이터 사이언티스트로서 필요한 공부들을 하고 있습니다.</p> <p>이런 저의 노력을 바탕으로 데이터 분석에 대해 더 깊게 공부할 여러 기회를 찾아보던 중, 데이터 사이언스 학회를 설립한다는 공고를 보게 되었습니다. 평소 공부하고 싶었던 ‘통계학을 기반으로 한 데이터 분석’을 관심 있는 사람들과 함께 전문적으로 공부하면 훨씬 좋을 것 같다고 생각하여 데이터 사이언스 랩에 지원하게 되었습니다. 학회원으로서 항상 열심히 공부하면서 데이터 분석에 관해 많은 공부를 해나가고 싶습니다.</p>			
프로젝트 경험			
<p>지난 학기 상경대학에서 주관한 빅데이터 분석 경진대회에 참가하였습니다. 저희 팀은 리서치 회사의 패널 데이터를 받아 분석하였습니다. 주제는 약 2 만 명 패널들의 개인정보 및 카드 결제 정보를 바탕으로 직업에 대해 분석을 하는 것이었습니다. 생활 패턴 및 개인정보와 직업의 연관성을 찾고 설문조사를 통해 그 모델을 직접 적용해보는 시간도 가졌습니다. 또한, 이 정보들을 정리하여 임시 웹사이트를 구축해 정보제공용 사이트를 만들기도 하였습니다.</p> <p>사실, 여러 수업에서 데이터 분석과 관련된 프로젝트를 해보았지만, 실제 회사 데이터로 프로젝트에 참가한 일은 처음이었기에 프로젝트를 시작할 땐 조금 막막했습니다. 특히, 팀의 대표자로서 이 대회를 준비하면서 어려웠던 점 중 하나는 팀원들 간 알고 있는 기법과 그 정도가 각자 다르다는 점이었습니다. 머신 러닝 기법에 대해서 잘 알지 못하는 친구도 있었고, 같은 기법을 다른 방법의 코드로 풀어내는 친구도 있었습니다. 이런 점들을 해결하고자, 초반 모임에서 각자 자신 있는 부분에 관해 설명해주는 시간을 가졌고, 부족한 부분은 개인적으로 공부해오기로 하였습니다. 그렇게 한 후, 정리한 여러 데이터 가공 방법 및 분석 기법들을 바탕으로 데이터 분석을 진행하였습니다. 처음부터 각자 잘하는 부분을 분담하여 프로젝트를 더 빨리 진행할 수도 있었지만, 팀원들과 함께 만들어내는 프로젝트이기에 그 결과물에 대해 모든 팀원이 잘 알고 있는 것이 중요하다고 생각하였습니다. 실제로, 이렇게 공부하며</p>			

프로젝트를 진행하였기에 데이터 분석 실력뿐만 아니라 협동심도 더 많이 배울 수 있었고, 예상보다 좋은 결과도 얻을 수 있었던 것 같습니다. 앞으로 데이터 사이언스 학회에서도 다양한 프로젝트를 진행하며 열심히 배우고, 데이터 사이언티스트로서 중요한 역량을 쌓고 싶습니다.

이름	최병준	학과	응용통계학과
학년	2학년	학번	2018122021
		핸드폰 번호	010-8311-6280
컴퓨터언어	Python, R		
수강과목	통계학입문, 통계방법론, 미분적분학, 선형대수학, SW 프로그래밍		
지원동기			
<p>고등학교 때부터 제 꿈은 데이터 분석가였습니다. 데이터들을 적절한 통계 분석 방법에 따라 분석하고 결과물을 통해 유의미한 결론을 도출하는 일이 하고 싶었습니다. 송도에서 앞으로 어떤 일을 구체적으로 하고 싶은지 알아보다 몇 가지 깨달은 점이 있습니다. 앞으로는 AI 와 블록체인 기술이 자동화, 공유경제 사회를 만들 것입니다. 저는 이 학회에 들어와 블록체인 기술과 데이터 분석간의 연관성에 대해 좀 더 연구해보고 싶습니다. 또한 막연하게만 생각하는 데이터 전문가가 되기 위해서 어떤 준비를 해야 하는지 구체적으로 알아가는 기회가 되었으면 좋겠습니다.</p>			
프로젝트 경험			
<p>고등학교 시절, 비슷한 전공을 희망하는 학생들끼리 모여 진로 탐색을 정기적으로 하는 프로그램이 있었습니다. 그 시간에 몇몇 친구들과 통계학개론 책을 준비하여 공부해보았는데, 책에 나온 예제 문제들을 바탕으로 실제 실험을 준비하고 실행하고 결과를 해석하는 일을 해본 경험이 있습니다. 간단하게는 동전 던지기를 통해 중심 극한 정리 보이기, 학생들의 등교 시간 조사를 통해 등교시간의 95% 신뢰구간 구하기 등이 있었습니다. SPSS 라는 분석 프로그램을 사용한 실험도 있었는데 인근 지역의 커피숍 매장의 상권과 청결도와 매출과의 선형 관계가 있는지 알아보는 실험을 계획하기도 했습니다. 당시 매장의 청결도와 상권을 수치화하는 과정에서 어려움을 겪은 기억이 있으나 기준을 정해 문제를 해결하기 위해 노력했습니다. 대학교에 입학하여 통계학 입문과 통계방법론 수업을 들으며 R을 이용하여 주어진 데이터를 t-검정, F-검정, 카이제곱 검정, 분산 분석 등 적절한 분석 방법에 따라 분석해보았습니다.</p>			

이름	송민정	학과	계량위험관리			
학년	2학년	학번	2017195085			
		핸드폰 번호	010-2359-3185			
컴퓨터언어	R					
수강과목	Research Design and Quantitative Methods					
지원동기						
<p>대학교 입학 후 네 학기를 보내면서 강점과 관심사를 찾기 위해 폭넓은 경험을 해왔습니다. 이제는 저만의 특화된 능력을 길러야 하는 시기이며 실무에 활용할 수 있는 기술을 배우고 싶습니다. 다양한 경험 중에서 가장 큰 흥미를 느꼈던 것은 데이터 분석이었습니다. 국제대 필수교양 수업 ‘Research Design and Quantitative Methods’에서 R 을 처음 배웠는데, 분산된 데이터에서 유의미한 정보를 끌어내는 것이 흥미로웠습니다. 더 깊이 있게 배우고자 독학을 시도했지만, 혼자 공부하는 것에는 한계가 있었습니다. 학회에서 회원들과 함께 공부한다면 팀워크에서 비롯된 시너지로 더 빠르고 크게 발전할 수 있으리라 생각하여 DS.lab 에 지원하였습니다. 또한, 응용통계학과 교수님과 대학원생분들의 조언을 들으며 졸업 후 진로에 대하여 뜻깊은 도움을 받을 수 있는 것도 DS.lab 의 큰 혜택이라 생각합니다. UIC 는 전공을 2 학년 이후에 선택하는데, 따라서 전공 내 네트워크가 미약하여 진로에 대한 조언을 구하기가 쉽지 않습니다. DS.lab 을 통해 데이터 사이언스라는 같은 관심 분야 내에서 다양한 진로를 준비하거나 혹은 이미 현직자 자리에 계신 선배님들께 조언을 얻으며 비전을 넓히고 싶습니다. 데이터사이언스에 대한 경험이나 지식이 많지 않습니다만 열정과 도전정신을 바탕으로 DS.lab 과 함께 데이터사이언스 전문가로 성장하고 싶습니다.</p>						
프로젝트 경험						
<p>저는 앞서 언급한 ‘Research Design and Quantitative Methods’ 수업에서 R 을 활용하여 주어진 데이터셋을 분석하고 그에 따라 기업의 향후 전략을 제안하는 프로젝트를 진행했습니다. 구체적으로, Prediction and Classification of Readers’ Response to a Newspaper Promotion 이라는 제목의 프로젝트로, 신문 고지서에 기재 된 아동 신문에 대한 프로모션 승낙 여부를 나이, 성별, 소득 등으로 분석하였습니다. 먼저, 구독자의 전반적인 특징을 파악하기 위해 3D Data Visualization 을 통해 데이터를 시각화했고, 그 결과 소득이 40-50,000 이상인 40 대 남성이 프로모션을 승낙하는 빈도가 높을 것이라는 가설을 수립했습니다. 다음은 그룹별 프로모션 승낙 비율을 구체적으로 파악하기 위해 decision tree 를 산출하여 네 개의 생성규칙을 구했습니다. 각각의 규칙에 따른 rule accuracy 와 rule coverage 를 구하여 각 규칙이 예측과 분류 중 어느 것에 도움이 될지 평가했습니다. 예를 들어, 44 세 이하 여성일 때 아동 신문 프로모션에 승낙하는 경우에 rule accuracy 는 1.0 이고 rule coverage 는 약 0.67 이었습니다. 따라서 이 규칙은 향후 고객 행동 예측에 유용하리라 판단했습니다. 이후 confusion matrix 로 한 개의 데이터가 잘못 카운트됨을 발견했고 오류율은 6%로 다소 낮음을 확인했습니다. 결론으로, 앞의 과정을 바탕으로 가설을 검증했고, 기업의 차후 프로모션 계획에 대하여 제안을 하며 프로젝트를 마무리 지었습니다. 이 프로젝트를 통해 데이터 분석의 전반적인 과정을 경험했고, 참신한 아이디어를 얻기 위해서 이미 가공된 정보를 활용하기보다는</p>						

데이터셋에서 직접 정보를 끌어내는 것이 중요하다는 것도 느꼈습니다. 앞으로 R 을 통해 더욱 방대한 데이터를 분석하고 R 이외의 다양한 데이터 분석 기법을 배우고 싶습니다.



이름	임채확	학과	응용통계학과
학년	4학년	학번	2014122050
		핸드폰 번호	010-3061-6943
컴퓨터언어	R		
수강과목	통계학입문, 통계방법론, 미분적분학, 선형대수학, 수리통계학 1, 수리통계학 2, 확률과정, 회귀분석		
지원동기			
<p>지금 이 데이터사이언스 학회에 지원하게 된 동기는 데이터를 분석하는 역량을 키우고 싶어서입니다. 통계학과 학생으로서 로스쿨이나 cpa 같은 다른 예외적인 분야로 가지 않은 이상 어떠한 진로를 가더라도 데이터를 분석하는 역량을 키우는 것이 중요하다고 생각합니다. 제가 금융수학 쪽에 관심이 많은데 이 분야에서도 결국에 쌓여진 데이터를 분석하는 것이 정말 중요하다는 것을 깨닫게 되었습니다. 뿐만 아니라 시간이 지날수록 사회전반에서 데이터의 중요성이 커지고 있는 만큼 이것을 분석하는 능력이 통계학과로서 갖춰야할 것이라는 생각을 갖게 되었습니다. 학부 3년 동안 학교를 다니면서 전공과 관련된 이론 수업은 어느정도 들었던 것 같습니다. 물론 들어보고 싶은 것도 많고, 남은 과목도 많지만 졸업에 필수적인 과목은 많이 들어놓은 상태입니다. 그렇지만 이론위주의 수업을 듣다보니 프로그램 관련 과목을 거의 듣지 못했습니다. 1 학년 때 들었던 김철웅 교수님 통방이나 2-1 에 들었던 컴자처가 사실 마지막으로 프로그램을 쓰는 강의였습니다. 이론위주의 강의를 듣다보니 많이 편중이 된 것 같습니다. 또한 생각보다 학과의 커리큘럼이 파이썬과 같은 다른 프로그램 수업은 잘 열리지 않다보니 관련 프로그램을 많이 다루어 보지 못했습니다. 그런데 마침 이번학기에 통계학을 베이스로 하는 데이터사이언스 학회가 창립되었다고 해서, 지금까지 공부한 이론들을 적용하여 실질적으로 분석하고 배울 수 있는 좋은 기회라고 생각해서 지원을 하게 되었습니다.</p>			
프로젝트 경험			
<p>이 부분에 대해서는 제가 코딩이나 프로그램 실력이 없다보니 안타깝게도 관련 경험이라고 할 것이 없습니다. 처음부터 새로 열심히 배우겠다는 마음으로 지원을 했습니다. 들어간다면 열심히 해서 데이터 분석 프로젝트에 참가할 수 있을 정도의 실력을 쌓겠습니다. 그렇게 해서 학회의 다른 구성원들에게도 도움을 주고 서로 배울 수 있게 되고 싶습니다.</p>			

이름	김아영	학과	국어국문학과
학년	2학년	학번	17
		핸드폰 번호	010-6428-3120
컴퓨터언어	Python		
수강과목	통계학입문		
지원동기			
<p>한스 로슬링 교수의 TED 강의를 보고 저는 데이터 분석에 매료되었습니다. 두루뭉실한 이상적 미래를 그리는 것이 아니라 축적된 데이터를 기반으로 그 데이터를 분석해 미래를 예측하는 모습에 통계학과 데이터 분석을 꼭 배워야겠다고 마음먹었습니다. 인권과 문화는 모든 사람들이 누려야 할 것이지만 이것만으로 그것을 이룰 순 없습니다. 이를 위해선 경제 성장과 기술 발전이 꼭 필요합니다. 그리고 삶의 많은 부분이 인터넷에서 이루어지고 있는 시대에 경제와 기술 발전의 중심은 데이터라고 생각합니다. 그리고 이 데이터를 분석하고 분석된 데이터를 통해 미래를 예측하는 데이터 과학을 배우는 것은 누구보다 앞서나가고, 경제 성장과 기술 발전에 도움이 될 능력을 갖추는 일이라 생각합니다. 이러한 이유로 저는 데이터과학을 비롯한 이에 기본이 되는 통계적 지식을 습득하고 싶습니다. 통계학과 대학원 진학을 목표로 학부에서 학회의 경험을 쌓고 싶어 지원하게 되었습니다.</p>			
프로젝트 경험			
<p>데이터를 분석해 본 프로젝트는 없습니다. 그렇지만 학회 활동에 있어 부족함이 없도록 학회 활동 이외에도 따로 열심히 준비하고 공부하겠습니다. 응용통계학과 전과를 계획하고 있어 관련한 학부 수업도 많이 들을 예정입니다.</p>			

이름	김재익	학과	응용통계학과
학년	3학년	학번	2017122037
		핸드폰 번호	010-7343-9667
컴퓨터언어	Python, R		
수강과목	통계학입문, 통계방법론, 미분적분학, 선형대수학, 수리통계학 1		
지원동기			
<p>응용통계학과 재학생이며, 이쪽 수업을 들으며 데이터 분석 및 통계에 대해 아직 막연하지만 흥미를 갖게 되었습니다. 전공에 대한 적성도 어느 정도 맞다고 느끼고 있으며 현재까지 들은 수업들도 즐겁게 들었습니다. 이처럼 막연한 흥미를 가지고 무엇을 해야 이쪽 분야에 좀 더 깊게 진입할 수 있을까 고민하던 중이었습니다. 주변 동기들을 보면 와이빅타, 보아즈, 이에스씨 등등 수많은 데이터관련 학회 및 동아리를 통해 자신들의 커리어를 쌓아가는 모습을 보며 살짝 걱정도 되던 차였습니다. 수리통계학 1 을 지난학기에 수강하며 조교님을 뵈었는데, 마침 조교님께서 홍보하시는 글을 보고 이쪽 분야에 대해 더욱 탐구해보고자 지원하게 되었습니다.</p>			
프로젝트 경험			
<p>없습니다.</p>			

이름	강다연	학과	(본)경영학/(복수)응용통계학과			
학년	4학년	학번	2016123270			
		핸드폰 번호	010-4526-7739			
컴퓨터언어	Python, R					
수강과목	통계학입문, 통계방법론, 미분적분학, 선형대수학, 수리통계학 1					
지원동기						
<p>경영학과 학생으로서 아마존과 같이 빠르게 성장하는 기업을 보며 데이터를 비즈니스에 활용하는 것의 중요성을 느꼈습니다. 인터넷 의존도가 높아지고 사물인터넷이 발달함에 따라 데이터는 기하급수적으로 증가하고 있고, 그 데이터는 곧 현재 고객만이 아니라 잠재적 고객에 대한 모든 것을 파악할 수 있는 귀중한 자산이라는 생각이 들었습니다. 빅데이터는 특히나 제 관심 분야인 마케팅에 있어 고객의 니즈 파악, 수요 예측, 가격/위치 선정 등의 일련의 과정에서 중대한 인사이트를 제공한다고 생각합니다. 그렇기에 직관보다는 데이터 분석에 기반한 의사 결정의 필요성을 실감했고, 준비된 마케터가 되기 위해 경영학과 내에서 빅데이터 관련 수업을 수강하고 데이터 사이언스의 기반이 되는 응용 통계학을 복수 전공하고 있습니다. 그러나 경영학과이기에 데이터 사이언스에 관심있는 학우들과 교류하기 힘들다는 점과 실제 프로젝트에 참여할 기회가 많지 않다는 점이 항상 아쉬웠습니다. 응용 통계학과에 데이터 사이언스 관련 학회가 있으면 꼭 참여하고 싶다고 생각을 하곤 했습니다. 데이터 사이언스 랩의 창립 멤버로 활동하는 것은 그런 저에게 대학 생활에 있어 가장 소중한 경험일 것입니다. 학회원으로 데이터 분석에 대한 학술적인 지식을 습득하고 기업 프로젝트 및 경진 대회에 참가하며 실무 경험을 쌓아 학문과 실전 능력 모두 키우고 싶습니다.</p>						
프로젝트 경험						
<p>미국 교환학생 중 'Data Mining for Business Analytics' 수업을 들으며 검증용 데이터를 나누는 bootstrap, k-fold cross validation 등의 방법과 변수선택 방법, 회귀분석, KNN, Random Forest, Clustering, Logistic Regression 과 같은 모델링을 배우며 데이터 마이닝의 기초를 길렀습니다. 이를 기반으로 유튜브의 장르, 구독자 수, 태그 등의 독립 변수로 YouTube 조회수를 예측하는 프로젝트를 수행했습니다. 회귀 분석과 Random Forest 를 사용하며 다양한 모델을 시험해 가장 유의미한 결과를 보이는 모델을 선별했습니다.</p> <p>'마케팅 애널리틱스' 수업에서는 panel regression, factor analyses, web crawling 과 text mining 등을 학습하며 다양한 프로젝트를 수행했습니다. 그 중 2018 년 1 년간 빌보드 weekly top100 안에 든 곡을 웹사이트에서 직접 크롤링하여 전체 데이터 중 2018 년에 가장 인기있었던 10 곡을 선별하는 프로젝트, 음식점에 관한 conjoint data 를 직접 수집해 메뉴, 가격에 따른 상대적 민감도와 시장 점유율을 예측한 프로젝트가 기억에 남습니다. 또한 Logit 모형을 사용해 미국의 자치주 별 인종 분포, 성별, 생업, 소득 등의 변수가 트럼프를 지지할 확률에 어떤 영향을 미치는지 분석을 하기도 했습니다. 기말 프로젝트로는 약 2 년간 상영된 75 개의 한국 영화 상영 날짜 별 관객수 데이터를 사용해 관객 수를 예측하는 모델을 만들었습니다. 배급사, 요일, 월, 장르, 등급, 상영 일 전까지의 평점을 독립 변수로 회귀 분석을 했는데, 더미 변수로 변환하는 데에 있어 베이스라인을 바꾸는 것, 변수에 로그 또는</p>						

제곱을 취하는 방법, interaction term 을 확인하는 방법 등을 사용하며 모델의 설명력을 높일 수 있었습니다.

이름	김민준	학과	응용통계학과			
학년	2학년	학번	2017122029			
		핸드폰 번호	010-5288-5090			
컴퓨터언어	Python, R, SAS 와 파이썬은 기초적인 정도입니다.					
수강과목	통계학입문, 통계방법론, 미분적분학, 선형대수학, 수리통계학 1, 컴퓨터자료처리, 회귀분석					
지원동기						
<p>안녕하세요. 응용통계학과 17 학번 김민준입니다. 저는 전공 과목을 듣던 중 아쉬움이 생겼습니다. 강의에서는 주로 이론적인 내용을 다루었습니다. 그러다 보니 이런 내용이 실제로 어떻게 쓰일 수 있는지 알고 싶었습니다. 데이터를 가지고 분석하여 결과를 얻는 과정을 직접 경험해보고 싶었습니다. 그럴 능력을 키우고자 방학 중에 혼자 R 과 SAS 를 공부하고 자격증을 준비하기도 했습니다. 하지만 혼자서 하니 무엇을 어떻게, 얼마나 공부해야할지 잘 알지 못했습니다. 또한 도중에 느슨해져서 잘 진행되지 않았습니다. 그러던 중 이번 학회 지원 공지를 보고 정말 좋은 기회라고 생각했습니다. 제가 부족함을 느끼고 아쉬워했던 부분을 채워줄 수 있을것 같았습니다. 데이터를 가지고 직접 분석을 해보면서 이론적인 부분이 어떻게 활용되는지 배우고 싶습니다. 또한 같은 주제를 가지고 다른 학회원분들과 함께 공부하며 서로에게 도움이 되고 같이 성장하는 즐거움도 느껴보고 싶습니다. 저는 향후 대학원 진학을 희망하고 있습니다. 먼저 대학원에 진학하신 선배님들을 보면서 배우고 교류하며 막연하게 그리고 있는 대학원 생활에 관한 정보도 구체적으로 얻고 싶습니다. 저는 이런 이유로 학회에 지원하게 되었습니다. 꾸준히 관심을 가졌던 분야이고, 제 전공이니만큼 들어가게 된다면 적극적으로 활동하며 열심히 배우겠습니다.</p>						
프로젝트 경험						
<p>아직 데이터 분석 프로젝트 경험은 없습니다. 방학중에 파이썬과 R 을 공부하며 필요한 능력을 키우고 있습니다. 꾸준히 공부하여 향후 학회에서 활동할때 부족하지 않도록 준비하겠습니다. 데이터사이언스랩에 들어가면 프로젝트 경험을 쌓을 수 있을 것이라고 기대하고 있습니다.</p>						

이름	박건희	학과	응용통계학과
학년	3학년	학번	13학번
		핸드폰 번호	010-8511-1362
컴퓨터언어	R		
수강과목	통계학입문, 통계방법론, 미분적분학, 선형대수학, 수리통계학 1, 수리통계학 2		
지원동기			
<p>지난 학기 통계학과 세미나가 열렸습니다. 그 당시 실리콘 벨리에서 근무하시고 계시는 통계학 박사님이 오셨습니다. 현재 통계학자들이 데이터 사이언티스트라고도 불리는 현재에 R 뿐만 아니라 여러 프로그램을 통해서 어떤 과정을 통해서 데이터 분석을 하는 지에 관한 대략적인 설명을 해주셨습니다. 학부 수준에서는 이미 정돈되어 분석하기에 용이한 자료들을 이용하여 데이터 결과를 낼 수 있지만 실무에서는 자료를 크롤링 하는 것에서 부터 알맞게 선택하고 빼내는 작업들이 필요함을 역설해주셨습니다. 당장에 통계학과에 이론들은 많이 쓰이고 분석하는데 도움을 주고 있지만 자료를 모으는 것과 정제하는 과정 역시도 앞으로 통계학도로서 매우 중요함을 느끼게 되었습니다. 파이썬과 같은 프로그램 언어 역시도 다른 분야를 전공하신 동료 직원들이 쓰고 있음을 말해주셨습니다. 학회의 방향이 데이터 분석과 그 프로젝트를 행하는 것임을 확인하였습니다. 친구들이 기상청 빅데이터 공모전 등 여러 프로젝트를 하면서 많은 도움을 받았다는 얘기를 많이 들었습니다. 이 학회를 통해서 저도 많이 배우고 프로젝트를 참여하면서 단순히 통계 이론에만 머무르는 것이 아니라 , 여러 컴퓨터 언어와 프로그램을 다루고 자료 수집부터 결론까지 해나갈 수 있는 사람이 될 수 있기를 희망하고 있습니다. 잘부탁드립니다.</p>			
프로젝트 경험			
<p>죄송합니다. 데이터 분석 프로젝트 경험은 전무합니다.</p>			

이름	김경률	학과	언더우드 경제학과			
학년	3학년	학번	2014190335			
		핸드폰 번호	010-8894-3138			
컴퓨터언어	Python					
수강과목	통계학입문					
지원동기						
<p>데이터사이언스랩에 지원하게 된 가장 큰 동기는 함께 성장할 수 있는 사람들을 만나기 위해서입니다. 데이터 사이언스를 알게 된 후 여러 경로를 통해 공부를 하고 있지만, 체계적인 환경에서 같은 목표를 가지고 있는 이들과 교류하고 협력하는 것이 얼마나 큰 시너지 효과를 내고, 어려운 순간에도 동기부여가 될 수 있는 지 여러 연합동아리 활동에서 이미 체감했습니다. 데이터사이언스 전반에 대한 지식 교류와 실습을 통해 해당 분야에 대한 전문성을 키우고 싶어 학회에 지원하게 되었습니다.</p> <p>데이터사이언스랩에서 가장 하고 싶은 활동은 실제로 데이터를 활용하는 것입니다. 제가 앞으로 하고 싶은 일은 오랜 해외 생활에서 얻은 경험과 데이터 분석을 결합하여 새로운 가치를 창출할 수 있는 일입니다. 세계화가 되었다고 하지만, 사업을 위한 외국 소비자 이해에 있어서는 단순한 인식이나 막연한 인사이트만을 활용하는 경우가 아직 많다고 생각합니다. 기존의 다양한 데이터를 가공하고 객관화하여 이를 적재적소에 활용할 수 있는 빅데이터 시대의 새로운 인재로 성장하고 싶습니다. 예전 활동에서의 귀중한 경험과 데이터 사이언스에 대한 열정을 토대로 데이터사이언스랩에서 역시 성실히 활동하고 발전하는 모습 보여드리겠습니다.</p>						
프로젝트 경험						
<p>공식적인 프로젝트는 아니었지만, 캐글의 타이타닉 생존자 예측과 자전거 공유 시스템 대여량 예측 과제를 수행한 적 있습니다. 두 과제에서 모두 파이썬의 Seaborn, Pandas 패키지와 사이킷런 라이브러리를 사용하였습니다. 자전거 대여량 예측 과제 같은 경우, 전 처리 작업, Seaborn 시각화 분석, 머신러닝 하이퍼패러미터 튜닝과 마지막으로 Random Forest 를 통해 값을 예측하였습니다. 그 과정에서 어려웠지만 가장 재미있었던 부분은 통찰력을 가지고 시각화된 데이터를 분석하는 작업이었습니다. 데이터 값 중 월에 따른 대여량을 확인하기 위해 시각화를 진행하였는데 같은 겨울이지만 1월보다 12월의 데이터가 높았습니다. 처음에는 이에 대한 의문을 가지지 않았으나, 캐글의 점수를 개선하기 위한 고민을 하다 년과 월의 데이터를 합쳐보니 대여량의 차이가 꾸준히 성장해온 회사의 11년 1월과 12년 12월의 차이였음을 알게 되었습니다. 즉 자전거를 대여하는 고객 입장에서 12월이라고 자전거를 더 많이 빌려야 할 이유가 없었습니다. 알고리즘이 과적합되는 것을 피하기 위해 연도와 달을 같이 묶어 분석하니 정확도가 개선되었습니다.</p> <p>또한 자전거 대여량 같은 경우, 시각화를 통해 분석해보니 하루 자전거 대여량이 1-20 대인 비중이 매우 높았지만 대여량이 천대로 분포에서 많이 벗어난 값도 있었습니다. 이러한 분포가 데이터를 왜곡되게 한다고 생각했고 방법을 찾다 로그 변환이라는 개념을 알게 되어 적용하였더니 보다 대칭적인 패턴을 가지게 되었고, 극도로 띄는 값들이 사라지는 것을 발견할 수 있었습니다.</p>						



같은 데이터를 가지고도 사람들마다 다른 예측치를 얻어내는 것을 보면서, 데이터를 가공하는 것은 단순히 툴을 활용하는 것뿐만 아니라 개인의 인사이트, 통계적 지식이 많은 비중을 차지한다는 것을 한다는 것을 알게 되었습니다. 데이터사이언스랩에서 다양한 경험을 가진 분들과 활동하며 툴 활용 능력뿐만 아니라 새로운 관점으로 데이터를 보고 분석하는 눈도 기르고 싶습니다.

이름	이종빈	학과	경제학과
학년	4학년	학번	13
		핸드폰 번호	010-8775-2650
컴퓨터언어	Python, R, SAS, STATA		
수강과목	통계학입문, 미분적분학, 선형대수학, 계량경제학, 금융계량경제학		
지원동기			
<p>작년 8 월 UNIST 빅데이터 경진대회에 참가하였을 때, 스스로 부족함을 많이 느꼈습니다. 학부를 다니며 다양한 모형들과 통계적 분석방법을 배웠습니다만, 실제 대회에서는 주어진 데이터 전처리조차 쉽지 않았습니다. 또 수업에서 다루는 모형들에 대한 이론적 지식은 많지만 실제로 그것을 적용해 본적이 드물어 사용해보지 못한 것이 많이 아쉬웠습니다. 게다가 빅데이터 분석방법에는 수업에서 다루지 않았던 '랜덤포레스트', '신경망' 등 처음들어보는 용어들이 많아 따로 공부하고싶다는 생각을 많이 했습니다. 그래서 학회에 들어가 이러한 모형들도 공부하고 실제 데이터 분석경험을 늘려보고 싶습니다. 또한 지금까지 데이터 분석한 것은 아무래도 경제학적 이론 검증을 위해 실시한 것들이 많은데, 조금 더 다양한 방면에 데이터 분석을 활용하는 경험을 해보고 싶습니다.</p>			
프로젝트 경험			
<p>1. 계량경제학 팀프로젝트: 계량경제학 수업 팀프로젝트에서 SAS 를 이용해 대학교수들 연봉의 결정요인을 다양한 변수 데이터를 구해 회귀분석으로 찾아 보았습니다. 특히 노동경제학의 가장 중요한 이론인 임금 = 한계생산성이 실제 사회에서도 만족하는지를 알아보기 위해 SCI 급 논문의 연간 편찬 횟수를 설명변수로 선정하였는데, 유의미한 결과가 도출되었습니다.</p> <p>2. 한국경제정책분석 팀프로젝트: 수업 팀프로젝트에서 STATA 를 이용해 MLB 의 사치세(Competitive balance tax)가 리그의 전력평준화에 긍정적인 효과를 가져왔는가를 분석하기 위해, 메이저리그 20 년간의 승률 패널데이터를 이용해 RD 모형과 DD 모형을 사용해 분석해봤습니다.</p> <p>3. UNIST 빅데이터 경진대회 참가:작년 8 월에 개최된 UNIST 빅데이터 경진대회에 참가하였습니다. 울산항만공사의 액체화물 물동량 예측이 과제였고, 이를 위해 VAR 모형을 활용해 예측해 보았습니다.</p> <p>4. 그 외 다수의 수업에서 최저임금의 고용효과, 최저임주연령 변화에 따른 청년 사망률 변화, 소비패널데이터를 활용한 DTI LTV 변화에 따른 소비변화 분석 등 소소한 데이터 분석 경험이 있습니다.</p>			

이름	안정욱	학과	응용통계학과
학년	2학년	학번	2016122057
		핸드폰 번호	010-9431-1585
컴퓨터언어	R		
수강과목	통계학입문, 통계방법론, 미분적분학, 선형대수학		
지원동기			
<p>아직 전공 수업을 많이 들어보지는 못했지만 통계 공부가 적성에 맞는 것 같아 재밌었습니다. 다만 학교에서 배우는 내용들을 실무에서는 어떻게 활용하는지가 늘 궁금했습니다. 최근 떠오르는 데이터사이언스 분야가 정확하게 어떤 일을 하는지, 또한 제 적성에 맞는지 알아보고 싶어서 데이터사이언스랩에 지원하게 되었습니다.</p>			
프로젝트 경험			
<p>없습니다.</p>			

이름	황진혁	학과	경제학과
학년	3학년	학번	2013121203
		핸드폰 번호	010-6892-7898
컴퓨터언어	없음		
수강과목	통계학입문, 미분적분학, 선형대수학, 수리통계학 1		
지원동기			
<p>우연히 계리직에 관심을 가지게 되어 관련 공부하면서 자연스레 통계학에대한 관심이생겼습니다. 그 뒤 공부를 계속하며 실무적인 영역에 까지 생각에 미쳤을때 데이터적인 인사이트가 필요함을 느꼈습니다. 학회를 통해 데이터를 다룰 기회를 얻고, 여러 선배들과 학우분들과 교류를 하고자 학회에 지원합니다. 커리어와 직접적으로 영향이 있고 의지가 강한만큼 부족하지만 열심히, 폐안끼치며 활동해보도록 하겠습니다. 감사합니다.</p>			
프로젝트 경험			
<p>없습니다.</p>			

이름	이현탁	학과	응용통계학과
학년	3학년	학번	2015122044
		핸드폰 번호	010-6776-0905
컴퓨터언어	R		
수강과목	통계학입문, 통계방법론, 미분적분학, 선형대수학, 수리통계학 1		
지원동기			
<p>저는 응용통계학과 15 학번 이현탁입니다. 그리고, 저는 통계학도가 아닙니다.</p> <p>저는 재수 끝에 겨우 연세대학교에 들어올 수 있었습니다. 사실 딱히 가고 싶은 과가 있었던 것도 아니고, 공부에 큰 뜻이 있었던 것도 아니었습니다. 단지 고 3 때의 결과에 웬지 분해서, 납득할 수 없어서 재수를 했던 것 뿐입니다. 재수에 성공해서 연세대학교 응용통계학과에 원서를 쓰게 된 것도 그냥 재수학원 선생님이 점수에 맞춰서 추천해주신 대로였습니다.</p> <p>그래서 저는 공부를 하지 않았습니다. 하고 싶지도 않고 머리에 들어오는 것 같지도 않았습니다. 특히 전공 과목 학점은 2 학년을 마친 지금 a 권에 해당하는 점수는 단 하나도 없습니다. 저는 통계학도가 아니었습니다.</p> <p>그렇게 흘러가는 대로 살다 어느새 군을 제대하고 복학하게 되었습니다. 그런데 말입니다, 예전에는 똑같이 흘러가는 대로 살던 친구들이 다들 통계학도가 되어버렸습니다. 배신감을 느꼈다는 것이 아니라, 오히려 저 자신에게 실망감을 느끼게 되었습니다. 언제까지 원하지 않았다는 이유로 더욱 원hat 않는 삶으로 향해 갈 것인지 제 자신에게 묻고 있는 그 때는, 저만 통계학도가 아니었습니다.</p> <p>그리고 지금은, 동기 모두가 통계학도가 아닌 데이터 사이언스 학도로 이름이 바뀌게 되었습니다. 응용통계학과가 격변하는 이 때, 학과는 그저 이름만 바뀌게 되는 것일지도 모르겠습니다만, 저는 이런 기회를 핑계로라도 변하지 않으면 안 되겠다고 느꼈습니다. 통계학도 되기에 실패한 제가, 데이터 사이언스 학도가 되기 위해서 생각해낸 것은 계속 제 전공에, 전공을 공부할 수 밖에 없는 환경에 노출시키는 것이었습니다. 학회에 지원한 것은 겨우 이런 이유에서 입니다. 누군가는 자신의 통계학적 기량을 물리기 위해, 누구는 경험을 쌓고 실무로 나아가기 위해 지원하는 곳에 저는 저 자신을 찾으러 문 앞에 섰습니다. 그러나 저는, 제 근본으로 찾아가는 기회인 만큼, 활동에 진심을 다 할 것입니다. 저는 데이터 사이언스 학도이고 싶습니다. 그리고 데이터 사이언스 학회원이고 싶습니다.</p>			
프로젝트 경험			
통계학 입문 과제로 분석했던 것 외 전무합니다.			

이름	김자령	학과	경제학과
학년	4학년	학번	2015121012
		핸드폰 번호	010-2799-2196
컴퓨터언어	R		
수강과목	통계학입문, 미분적분학, 선형대수학, 계량경제학		
지원동기			
<p>2018 년, 네덜란드로 교환학생을 갔던 경험에서 저는 데이터 이용, 분석에 대한 유용성과 필요성을 크게 느낄 수 있었습니다. 국제경제학 수업에서 저는 북한과 우리나라의 사이에서의 경제적 이론을 세워 유의미한 결과를 도출해 내는 과제를 하게 되었습니다. 그 때, 저는 학교에서 이론으로 배우기만 했던 경제학적 이론이 실제 사례에서 데이터를 통해 구체화 되는 것을 미약하게나마 체험할 수 있게 되었습니다. 하지만 프로젝트를 하면서 저는 이론적 지식은 충분하지만 데이터 처리 부분에서 많은 어려움을 느꼈고, 이 때문에 사회적 실제 현상의 데이터 처리를 조금 더 효율적으로 처리 할 수 있는 방법과 분석능력 그리고 그 데이터 분석결과를 이론과 실제에 적용하는 능력을 기르고 싶었기에 이 학회에 지원하게 되었습니다.</p>			
프로젝트 경험			
<p>데이터 분석 프로젝트를 대외적으로 한 경험은 없습니다. 하지만 학교 수업에서 과제의 일원으로 데이터 분석을 해 본 경험은 있었습니다. 예로 계량경제학 수업에서 미연준 금리와 국민총생산과의 관계를 분석하기 위해 회귀분석을 진행해 보았습니다. 또 네덜란드 교환학생 시절 북한의 상황과 우리나라 금융시장 사이의 관계, 그리고 그 둘 사이의 상호작용을 보기위해 약간의 데이터 분석을 진행해 본 경험이 있습니다. 두 과제를 통해서 저는 데이터 분석의 필요성과 폭 넓은 적용 범위에 대해 크게 느낄 수 있었습니다.</p>			

이름	정해용	학과	응용통계학 대학원
학년	1학기	학번	2019311266
		핸드폰 번호	010-4100-0210
컴퓨터언어	R, SQL, SAS		
수강과목	통계학입문, 통계방법론, 미분적분학, 선형대수학, 수리통계학 1, 수리통계학 2, 탐색적 자료분석, 범주형 자료분석 등		
지원동기			
<p>저는 데이터 분석을 통해 언제나 더 가치있는 결과를 찾아낼 수 있다고 확신합니다. 하지만 좋은 데이터라도 올바른 분석 과정이 없다면 오히려 정보가 왜곡되는 결과를 초래할 수 있습니다. 따라서 다양한 데이터에 대한 분석 능력과 지식을 갖추는 것은 제게는 그 무엇보다 중요한 일입니다. 새로 대학원에 입학하며 데이터 사이언스에 대한 공부를 시작하려던 제게 DS.lab 은 너무나 매력적인 선택이었습니다. 정기 세미나를 통한 데이터 사이언스 스킬에 대한 스터디는 물론이고, 현업에서 근무하는 전문가분들과 대화할 수 있다는 것도 DS.lab 에서만 얻을 수 있는 기회라고 생각합니다.</p> <p>하지만 무엇보다 기업과의 협력을 통해, 학교 수업에서는 구하기 힘든 다양한 raw data 들을 접하고 이를 처리 및 분석할 수 있는 기회를 얻을 수 있다는 것이 제게는 가장 큰 매력으로 다가왔습니다. 기업들이 진행하는 프로젝트에 참여하면서 현업에서는 데이터 사이언스가 어떻게 활용되는지, 어떤 러닝 스킬이 주로 쓰이며 분석 결과에 따라 기업의 전략에 어떻게 활용되는지 느껴보고 싶습니다.</p> <p>마지막으로 많은 학회원들과 함께 활동하며 서로에게 도움을 주고 받으며 성장하는 기회를 얻고 싶습니다. 수많은 분석 방법들이 존재하는 요즘, 서로가 가진 능력을 공유한다면 더욱 빠르게 성장할 수 있는 계기가 될 것이라 확신합니다.</p>			
프로젝트 경험			
<p>저는 교내 학회에서 학기말 Final project 로 진행했던 House price 예측 분석 경험을 말씀드리겠습니다. 주어진 Data 는 미국의 집 가격과 관련된 자료로 총 81 개의 변수와 1,310 개의 값을 가지고 있습니다. 집의 위치, 방의 개수, 지하실의 유무 등 다양한 범주형, 수치형 변수가 혼재되어 있고 변수 내에는 NA 나 . 등 결측치들이 포함되어 있습니다. Project 의 최종 목적은 이 변수들을 활용하여 집의 가격에 대해 최대한 높은 예측력을 갖는 모델을 만드는 것이었습니다.</p> <p>저희는 먼저 데이터를 정제하는 작업을 거쳤습니다. 데이터 정제에 있어 가장 집중한 두 가지는 첫 째, 결측치 제거와 둘 째, 데이터의 형식을 성격에 맞게 가공하는 것이었습니다. 결측치의 경우 자료가 "비어 있음"을 뜻하는 NA 와 "0"임을 뜻하는 NA 를 구분했습니다. 또한 연도와 같이 수치형이지만 범주형의 성격을 갖는 변수나, 범주형이지만 연속형 성격을 갖는 변수들을 각각의 성격에 맞는 자료들로 가공했습니다.</p> <p>산점도나 Box plot 등을 그려보면서 변수들의 상태를 파악한 뒤 모델에 활용할 변수들을 고민했습니다. 저희 조는 모든 변수가 포함된 Full Model 로 ANOVA 분석을 진행하고, 그 결과 산출된 변수 중요도에 따라 변수를 선택했습니다. 마지막으로 집 값을 적은 오차로 맞춰내야 하는 과제의 목적에 맞게 예측력이 좋은 랜덤 포레스트 기법으로 모델을 구축하여 예측 집</p>			

값을 구했습니다.

또한 저는 금융사에서 금융 데이터를 분석한 경험이 있습니다. 카드사 신용 부문에서 근무하며 회원들의 한도운영 현황을 모니터링 하였습니다. 전체 회원들의 연체율, 한도 상향 및 하향 전략 적용 현황, 회원들의 한도 조정 인입 채널 경로 등 다양한 데이터들을 통해 롯데카드의 리스크를 분석하는 업무를 담당 했습니다. 또한 우량 회원에 대한 금융 서비스 한도 상향 전략 프로젝트에 머신러닝 앙상블 기법을 활용하여 타겟 고객군을 선별하는 과제도 수행하였습니다.

이러한 분석 경험을 바탕으로 DS.lab 에서 성장할 수 있는 기회를 얻는 학회원이 되고 싶습니다.



이름	최석영	학과	응용통계학과
학년	4학년	학번	2015122037
		핸드폰 번호	010-4591-4223
컴퓨터언어	Python, R, SAS		
수강과목	통계학입문, 통계방법론, 미분적분학, 선형대수학, 수리통계학 1, 수리통계학 2, 시계열분석, 표본조사론, 확률과정		
지원동기			
<p>데이터 사이언스라는 포괄적인 분야에는 교환학생을 다녀오면서 관심을 가지게 되었습니다. 미국에서 시계열분석 수업을 들으면서 실제로 월별로 업데이트 되는 시계열 데이터를 활용해 다음 달의 월 데이터를 예측해보는 프로젝트를 여러 번 했었는데, 이런 주어진 과제 해결을 목표로 하는 프로젝트들이 재밌다는 것을 느끼고 데이터 사이언티스트라는 직업에 관심을 가지게 되었던 것입니다. 또한, 작년에 제약회사에서 인턴을 하면서 여러가지 데이터 분석 과제를 맡았었는데, 그 중에 예측 모델을 세워보는 과제들이 가장 재밌었지만 동시에 제가 가진 경험과 지식의 한계를 느끼게 되었습니다. 실제 데이터를 다룰 때는 학부 지식만으로는 유의한 예측 모델을 세우기가 어려우며, 통계 지식과 함께 경험도 중요하다는 것을 느꼈기 때문입니다. 그래서 데이터 사이언스 관련 지식을 공유하고, 데이터 분석 경진대회를 같이 참여할 수 있는 장을 만들 수 있는 본 학회에 지원하게 되었습니다. 특히, 국내에서는 아직 데이터 사이언티스트라는 직업이 데이터 분석가, 데이터 엔지니어들과의 경계가 모호하고 분류가 제대로 되지 않은 것 같은데, 이런 학회에서 실제로 해당 업무에 종사하고 계신 분들은 정확히 어떤 업무를 맡으시는지 공유 받을 수 있다는 점에 더욱 매력을 느꼈습니다.</p>			
프로젝트 경험			
<p>작년에 5 개월 간 제약회사 바이엘코리아에서 마케팅 부서 통계 직무 담당 인턴을 하는 중, 크게 두 가지 데이터 분석 프로젝트를 받았었습니다.</p> <p>첫 번째 과제는 2 년치 월별 처방량 데이터를 이용해 다음 달 약의 마켓 쉐어를 예측하는 것이었습니다. 시계열 데이터가 주어졌기 때문에 시계열 분석으로 각 경쟁제품의 처방량 데이터를 예측하고 그 결과값으로 마켓 쉐어를 예측하고자 했습니다. 그 과정에서 첫 번째로는 월별 처방량 데이터만으로 ARIMA 모델을 사용했었는데, 유의한 모델링을 도출하지 못했습니다. 그 다음으로는 경쟁사별 프로모션 비용을 설명변수로 더해 VECM 모델을 사용했는데, 이 모델 역시 유의한 모델링이 쉽게 나오지 않았습니다. 마켓 쉐어 자체를 예측하는 방법으로 시계열 자료라는 특징을 무시하고 로짓 분석을 쓸 수 있었지만, 로짓 분석에 대해서는 제대로 공부해본 적이 없었고 설상가상 다른 업무로 바빠지면서 더 이상 진행하는 것은 중단하게 되었습니다.</p> <p>두 번째로 받은 과제는 주별로 제공되는 약의 처방량 데이터로 컨설팅 업체에서는 그 다음 달 15 일에 제공되는 월별 처방량을 미리 추정해보는 것이었습니다. 제일 큰 문제점은 처방량이 일별로는 제공되지 않는다는 것이었습니다. 주차별로 제공된 데이터로는 월별 데이터 중에서도 21 일치 또는 28 일치의 데이터만 정확히 구할 수 있고, 나머지는 일별 데이터가 제공되거나 비슷하게 추정되어야 월별 데이터를 예측할 수 있었습니다. 일별 처방량을 추정할 수 있는 방법을 찾아봤지만 처방량 관련 일별 데이터가 저에게는 제공되지 않았기에, 결국 토요일과</p>			

일요일에는 처방이 거의 되지 않는다는 사실을 이용해 주차 데이터를 5로 나눠 나온 평균값으로 일별 처방량을 추정한 후, 21일치 또는 28일치를 제외한 날들의 일별 추정 처방량을 구해서 더했습니다. 결과적으로는 비슷하게 예측이 가능하다는 것을 확인 후 이 방법으로 월 처방량을 컨설팅 업체에서 제공하는 것보다 1~2주 앞서 예측하게 되었습니다.

이름	이동석	학과	응용통계학과
학년	3학년	학번	2015122040
		핸드폰 번호	010-8884-8461
컴퓨터언어	Python, R		
수강과목	통계학입문, 통계방법론, 미분적분학, 선형대수학, 수리통계학 1		
지원동기			
<p>통계학과에 재학중임에도 통계학을 실제로 응용해볼 수 있는 기회가 적었고, 또 과목 과목 하나로 들을때는 종합적인 통계 활용의 기회가 적었다. 그래서 그런지 주변에서는 데이터사이언스가 그렇게 유망하다며 부럽다고 하지만 막상 재학 2 년이 지난 지금에도 내가 무엇을 할 수 있는지, 또 무엇을 하게 될 것인지 감을 잡지 못하고 있다. 이런 막연한 불안감 때문인지 나름 실전에서 스스로 통계적 분석을 할 수 있을 실력을 키우고자 응용통계학 전공에 다레 컴퓨터과학과 수업, 그리고 의사결정과정을 다루는 경제학과, 경영학과 수업을 수강할 계획으루짜고, 저번학기부터 조금씩 실행해나가고 있었다. 하지만 혼자 막연하게 수업만 골라듣는 것으로는 바다에 표류하고 있는 느낌을 없잖아 느끼던 와중에, 과의 지원으로 데이터사이언스 학회가 생긴다는걸 접하고, 과 선후배와의 교류 그리고 데이터사이언스에 대해 방향성 있고 체계적이며, 동질적인 생각을 갖고 공부할 수 있는 기회라 생각해 지원합니다.</p>			
프로젝트 경험			
<p>이제 3 학년으로 올라가는 응용통계학과 학생으로 데이터 분석 프로젝트의 기회가 없던것이 사실입니다. 응용통계학 전공 수업에서 컴퓨터자료처리수업에서 간단한 r 명령어 처리를 했던 것이 첫 통계툴의 사용이었으며, 컴퓨터과학과 전공 기초인 컴퓨터프로그래밍 수업에서 그나마 분석이라할 만한(국회의회에서 의원들의 투표내역을 기반으로 의원들간의 정치적 거리를 구하는, 그리고 텍스트파일에서 특정 단어를 추출해 처리하는) 과제 두어개를 제출한 것이 전부였습니다. 작년 여름 연세대학교 상경대 주최로 열린 빅데이터경진대회에 참여하려했으나 아쉽게 예선에서 떨어져 실제 데이터 분석에 참여할 기회를 갖지 못했습니다. 다만 이 빅데이터 대회 참여 준비를 통해 현대 통계학이 데이터사이언스로, 왜 빅데이터가 뜨고 있는지, 그리고 대체 빅데이터가 무엇인지 어렵풋하게나마 알게되었습니다. 그리고 그냥 아무생각 없이 학과 수업 채우다 보면 뭔가 있겠지 하는 미련함을 떨치고 내가 데이터사이언스를 위해 알아야할 지식들이 무엇인가 고려하고 수강 계획을 짜는 그런 변곡점이 되었습니다. 데이터 분석 경험이 없다시피 한 것이 사실입니다, 하지만 부족함을 메꿀 수 있을정도의 노력을 보일 자신이 있어 지원합니다.</p>			

이름	박상우	학과	응용통계학과
학년	4학년	학번	2013122053
		핸드폰 번호	010-8317-7907
컴퓨터언어	SAS,JMP / R 과 Python 은 이번학기 수업수강예정입니다.		
수강과목	통계학입문, 통계방법론, 미분적분학, 선형대수학, 수리통계학 1, 수리통계학 2		
지원동기			
<p>제가 처음 통계학과를 입학하면서 상상했던 통계학이란 학문의 모습은 데이터를 분석하여 그 기저를 파악하고 다양한 분야에 적용하는 것이었습니다. 하지만 제가 5학기 동안 배웠던 전공 수업은 대부분 기초를 다지고 이론적인 부분을 배우는 데에 초점을 맞춰진 수업들 뿐이었습니다. 물론 저희 학과에도 데이터 분석 수업이 있지만 아직까지 그러한 수업들을 들을 기회가 없어 실제 데이터를 분석하고 실무적으로 적용하는 부분에 대한 배움에 목 말라 있었습니다. 그런데 저번학기 미국에 있는 대학교에서 교환학생 생활을 하게 되었고 그곳에서는 많은 수업들이 실제 데이터분석에 초점을 맞춰져 있었습니다. 그로 인해 데이터 분석에 대한 관심과 애정이 더욱 커져갔습니다. 그 예로 그 대학에서는 고전 문학들 속의 단어들을 분석하여 작가 개개인 별로 자주 쓰는 단어나 조사 등을 도출해내고 그것을 통해 여러 작자미상의 고전 작품들의 실제 작가를 파악해내는 프로젝트가 활발히 진행되고 있었습니다. 그래서 그에 관한 여러 모임이 있었고 그러한 모임에 참가하면서 데이터 사이언스에 대해 폭 넓고 깊이 있는 지식을 쌓고자 하는 열망이 생겼습니다. 그에 따라 이번 학기에는 데이터 마이닝과 데이터 사이언스 실전 등 다양한 데이터 사이언스 수업을 듣고자 합니다. 거기에 더해 더욱 더 전문적인 지식과 다양한 현장경험을 얻고자 데이터 사이언스랩 학회에 지원하게 되었습니다.</p>			
프로젝트 경험			
<p>저는 아직 외부에서 진행하는 데이터 분석 프로젝트에는 참가한 적이 없습니다. 하지만 저번학기 다변량 통계학 수업과 표본조사론 수업을 들으면서 여러 예제를 통해 데이터 분석을 접해보게 되었습니다. 그 과정에서 다양한 통계프로그램을 활용하여 매주 데이터를 분석하는 연습을 하게 되었습니다. 구체적으로 다변량 통계학 수업에서는 JMP 를 활용하여 Principal components analysis, Factor analysis, Cluster analysis, Discriminant analysis, Multidimensional Scaling 과 Correspondence analysis 를 배웠습니다. 표본조사론 수업에서는 SAS 를 활용하여 simple random sampling, stratified sampling, regression, ratio estimation, domain estimation, poststratification, cluster sampling 과 systematic sampling 을 배웠습니다.</p>			

이름	이상효	학과	(본)심리학/(복수)응용통계학과
학년	3학년	학번	2014119045
		핸드폰 번호	010-4543-3853
컴퓨터언어	Python, R		
수강과목	통계학입문, 통계방법론, 미분적분학, 선형대수학, 수리통계학 1, 컴퓨터자료처리		
지원동기			
<p>연세 데이터사이언스랩은 저 스스로의 데이터분석 역량 신장과, 더 중요하게는 팀으로서 함께 데이터를 분석하고 처리하는 능력을 함양하는 데에 꼭 필요하다고 생각합니다.</p> <p>그동안 홀로 응용통계학과를 복수전공하면서 학기 중에는 이론수업을 열심히 수강해 왔습니다. 물론 아직까지 배운 내용 자체가 많이 부족하지만, 새로운 내용을 배울 때마다 이를 조금이라도 데이터분석에 활용해보고 싶었습니다. 그러나 혼자서는 이를 어떻게 할지 몰랐습니다. 막연하게 통계학입문서적을 뒤적이며 R 에서 예시코드를 끼적여 보던 게 다였습니다.</p> <p>지난 여름방학 때에는 R 을 이용하여 수강신청 마일리지내역을 나름대로 분석해보고 이를 활용하여 수강신청에 성공해 보기도 했습니다. 그러나 작년 말 한 데이터분석학회의 컨퍼런스에 참여하며 제가 매우 비효율적으로 데이터분석을 독학(?)하고 있음을 깨달았습니다. 가령, 저는 마우스로 일일이 자료를 복사하고 붙여 넣고 R 로 불러오는 등의 번거로운 과정을 거쳐 데이터를 수집해 왔으나, 컨퍼런스에서 이러한 일련의 작업을 '크롤링'이라고 하며, 이는 코드 몇 줄로 간단히 할 수 있음을 알게 되었습니다.</p> <p>물론 오늘날은 검색을 통해 이러한 문제를 비롯한 대부분의 문제를 혼자서 해결할 수 있습니다. 스스로 고민의 시간을 가지며 데이터를 분석하여 예측하고, 이것이 들어맞을 때의 희열도 큼니다. 그러나 저는 미흡하게나마 제가 알고 있는 것과 다른 사람이 알고 있는 것을 공유하며 같은 문제라도 더욱 다양한 방법으로, 효율적으로 해결하고 싶습니다.</p>			
프로젝트 경험			
<p>저는 아직 다른 사람과 함께 데이터를 분석하는 경험을 해 본 적이 없습니다. 이 점은 저의 큰 지원동기입니다. 그러나 혼자서 서툴게나마 데이터를 살펴보며 느낀 것도 많았습니다.</p> <p>우선 지원동기에서도 언급했듯 저는 수강신청 과정에서 R 을 이용하여 초보적인 수준의 데이터분석을 해 보았습니다. 쌓여가는 마일리지데이터들을 보며 뭐라도 해 보고 싶었습니다. 이 과정에서 단순히 데이터를 정리하는 과정만으로도 이전까지는 생각하지 못 했던 새로운 의사결정 기준을 세울 수 있어 신기했습니다.</p> <p>예를 들어, 기존에는 수강신청 매뉴얼을 아무리 참고하더라도 주어진 수치들을 어떻게 이용할지 막막했고, 단순히 정원 대비 얼마나 많은 수요가 꾸준히 있었는지 정도만을 고려했습니다. 그러나 같은 학년 내에서의 누적 credit 의 양 비교가 중요하며, 따라서 같은 학년이라면 1 학기보다 2 학기를 보낼 때 더 경쟁력을 갖출 수 있다는 점을 비로소 알게 되었습니다. 이를 통해 같은 자료를 다양하게 정리하는 것만으로도 훨씬 많은 정보를 얻어낼 수 있다는 점과 더불어, 자료공개방식을 다양하게 할 필요성을 느꼈습니다.</p> <p>또한, 아직 시작단계이지만, 이번 방학에는 우리학교의 교환학생 정보 공유 커뮤니티인 교정공의 댓글자료를 다루며 파이썬을 익히고 있습니다. 교정공은 현재 우리학교의 유일한</p>			

교환학생 관련 사이트이지만, 그 활성도가 점점 떨어져 그동안 축적된 데이터를 바탕으로 일종의 새로운 사이트를 구축하고 싶었습니다.

가령 현재 가장 수요가 높은 자료인 각 학교들에 대한 합격자들의 총점들을 기록한 댓글들만이라도 누구나 쉽게 찾아볼 수 있도록 정리하고 싶었습니다. 이 과정에서 많은 사람로부터 입력된 자료는 다루기가 어려움을 느끼고 있고, 이를 해결하기 위해 유사한 자료(텍스트)를 판단하여 제시하는 프로그램이나 연관된 검색어를 추천해 주는 프로그램을 이용해보고자 하고 있습니다. 학회에서 함께 이를 프로젝트로 다룬다면 더욱 좋은 성과가 있을 것이라고 생각합니다!

이름	이명진	학과	대기과학과, 산업공학과
학년	3학년	학번	2016136004
		핸드폰 번호	010-8870-2373
컴퓨터언어	Python		
수강과목	미분적분학, 자료분석(산공전공): 회귀분석, 분산분석, 실험계획법 관련 수업/기상통계분석(대기과학전공): 가설검정 관련 수업		
지원동기			
<p>산업공학과 전공을 들으면서 데이터를 가공하고, 분석하고, 새로운 가치를 부여하는 작업들에 크게 매료되어 그 이후로도 또 다른 프로젝트에 대한 갈증이 있었습니다. 하지만 스스로 이 갈증을 해결하기에는 데이터 분석 툴(Tool)을 다루는 실력 면에서 많이 부족할뿐더러, 실제 문제를 접할 기회조차 많지 않다는 생각이 들었습니다. 그러던 중 우연히 이 학회에 대한 안내를 접하게 되었고, 이 학회라면 제 갈증을 시원하게 해소해줄 수 있을 것으로 생각하여 이렇게 지원하게 되었습니다. 보다 실무에 가까운 데이터 분석을 많이 접해보고, 그 과정에서 교수님들과 선배님들로부터 조언을 얻음으로써 데이터 사이언티스트로서의 초석을 다지고 싶습니다.</p>			
프로젝트 경험			
<p>지난 학기에 열린 산업공학과 전공수업, ‘경영과학설계’를 들으면서 데이터 분석실습을 했습니다. 당시 프로젝트의 주제는 주어진 건강검진데이터(진료내역, 의약품처방, 건강검진정보)를 이용하여 문제점을 찾아내고, 해결방안을 도출하는 것이었습니다. 저희 조는 ‘농촌 마을의 의료사각지대 주민들을 위한 의료서비스 제공’을 주제로 잡았었습니다.</p> <p>우선 저희 조는 기존의 건강검진데이터들을 합치고 가공하는 과정에서 연령대별로 중요한 질병을 알게 되었습니다. 그다음, 공공데이터 포털로부터 전국의 병원들과 농촌 마을들의 위치를 서로 떨어진 정도를 계산함으로써 의료사각지대(주변에 병원이 없는)에 해당하는 농촌 마을을 찾아냈습니다. 그리고 의료사각지대에 해당하는 농촌 마을 주민들의 연령대별 분포 데이터와 앞서 알게 된 연령대별 주요 질병정보를 결합함으로써 농촌 마을 별로 중요한 질병을 알게 되었습니다. 이로써 의료사각지대에 해당하는 농촌 마을의 의료서비스 수요를 파악한 후, 이에 대한 방안으로 의료서비스를 제공하는 버스를 만들어 국가에서 주기적으로 운행하도록 하는 프로젝트를 제시했었습니다.</p> <p>데이터와 데이터를 연결하여 새로운 데이터를 만들어내는 과정, 그리고 어떤 데이터와 결합하느냐에 따라 전혀 다른 문제들이 도출되는 과정들을 지켜보면서 데이터 사이언스에 대한 흥미가 생겼고, 더욱더 잘하고 싶다는 욕심이 생겼습니다.</p>			

이름	박세희	학과	응용통계학 대학원
학년	2학기	학번	2018321090
		핸드폰 번호	010-7185-2886
컴퓨터언어	Python, R		
수강과목	통계학입문, 통계방법론, 미분적분학, 선형대수학, 수리통계학 1, 수리통계학 2, 선형통계이론, 데이터마이닝, 머신러닝과응용 2		
지원동기			
<p>이번 학기에 수강한 머신러닝과 응용 2 라는 수업에서 배운 딥러닝은 여태껏 알고 있던 통계적 데이터마이닝 기법들과는 달랐습니다.</p> <p>학부시절 경제학과 통계학을 복수전공한 본인은 실제 데이터에 통계적 기법을 활용할 수는 없는지 궁금했었습니다. 그러나 경제 변수들은 서로서로 연관성이 크기 때문에 회귀분석을 이용한 학문인 계량경제학 역시 그래서 다중공선성 등을 중요한 문제로 다뤘습니다. 딥러닝은 여러 레이어가 있어 설명변수들간의 관계를 직접 지정하지 않아도 복잡한 관계를 반영할 수 있다는 점, 심지어는 그 설명변수들이 무엇인지 전부 알 필요도 없다는 점이 매력적이었습니다. 이는 꼭 경제모델 뿐 아니라 다른 학문, 또는 기업 모델에서도 마찬가지로 적용될 수 있다고 생각합니다.</p> <p>또한, CNN, RNN, unsupervised learning 등은 회귀분석, 랜덤포레스트 등으로는 아예 할 수 없었던 새로운 종류의 과제를 해결한다는 점에서 흥미로웠습니다. 다른 학회와 수업 과제 등을 통해 여러 번의 데이터 분석을 해봤지만, 대부분의 경우, 0 또는 1, 혹은 수치와 같은 형태의 종속변수만을 예측할 수 있었습니다. 그에 비해 neural network 는 다수의 설명변수를 포함한 모델을 만들 수 있기에 데이터를 쪼개 컴퓨터가 인식할 수 있는 형태로 입력할 수만 있다면 데이터형식에 제한이 적습니다.</p> <p>그래서 더 공부해보고 싶다고 생각했습니다. 데이터사이언스랩의 일원이 된다면 특히 RNN 과 Natural Language Processing 등의 원리에 대해 같이 공부하고, 현재 관련해서 개발되고 있는 기법들에 대해 알아보고 싶습니다. 아직까지는 지식이 부족해 구체적으로 제시하지는 못하겠지만 이를 이용해 예시 프로젝트로 그치는 것이 아니라 현실에 도움이 될 만한 프로젝트도 함께 진행해보고 싶습니다.</p> <p>제가 이 학회의 일원이 된다면 최선을 다해 매주 미팅과 프로젝트 아이디어 등을 준비하고 다른 학회원들이 기피하는 업무 등을 적극적으로 맡아 기여할 것을 약속드립니다.</p>			
프로젝트 경험			



저는 ESC 학회에서 두 학기 동안 활동했습니다. 첫 학기는 회귀분석, 두 번째 학기는 Introduction to Statistical Learning 이라는 책에 나오는 데이터 마이닝 기법들 (로지스틱 회귀분석부터 랜덤포레스트까지) 에 대해 일주일에 한 번씩 학회원들과 함께 공부하고 학기말에 그 기법을 이용한 조별 프로젝트를 진행했습니다. 첫 학기에는 소개팅 데이터를 가지고 소개팅 성공 여부를 예측해봤습니다. 두번째 학기에는 영화 평점 데이터를 가지고 다양한 기법들을 이용해 어떤 모델이 가장 높은 정확도를 보이는지 분석해봤습니다. 학회에서의 데이터분석 활동을 통해 배운 점은 모델을 적용하는 것은 코드 몇 줄이면 빠르게 할 수 있다는 것이었습니다. 오히려 흔히 생각하는 것과는 달리 전후의 과정들이 데이터 분석에서는 못지않게 중요하다고 느꼈습니다. 데이터를 원하는 형태로 가공하고 missing value 를 처리하는 것, 풍부한 EDA 를 통해 데이터를 파악하고 알맞은 가설을 세우는 것, 채택한 모델의 결과를 해석하는 것, 그리고 비전문가도 이해할 수 있게 결과를 시각화하는 법 등이 모두 중요하다고 생각했습니다. 저의 이런 경험이 데이터사이언스랩에서 활동하는데 역시 도움이 되지 않을까 생각합니다.

아직 데이터사이언스랩의 활동방향을 구체적으로는 모르겠으나 필요하다면 데이터를 취득하는 과정 역시 학회원들과 알아보고 싶은 마음이 있습니다.

또한, 저는 머신러닝과 응용 2 라는 수업 최종 조별 과제에서 bidirectional LSTM 을 활용한 RNN 기법을 이용해 영화평점 데이터 (앞서 언급한 데이터와 전혀 다른 데이터임. 텍스트로 이루어진 데이터)를 분석해보기도 했습니다. 아예 처음 배워본 내용이라 아주 솔직히 말씀드리면 RNN 구조를 처음부터 끝까지 직접 짜지는 못했습니다. 그러나 coursera deep learning 강의 등을 주도적으로 병행하며 neural network 의 구조와 원리를 완벽히 익히고 이를 바탕으로 수업 중에 교수님이 제공해주신 코드 등을 활용해 분석을 진행했습니다. 데이터사이언스랩에서 비슷한 프로젝트를 할 수 있는 기회가 주어진다면 조금 더 많은 부분을 직접 짜서 프로젝트 의도에 맞는, 더 높은 정확도를 보이는 모델을 저희가 직접 만들 수 있도록 노력하겠습니다. 또한, 처음이라 가장 간단한 과제를 실습해봤는데 다른 조들의 발표 (영상에서 비 제거하기 등의 과제를 시행함)에서 자극받았던 점을 토대로 저희만의 창의적이고 실용적인 프로젝트 주제를 다뤄보도록 하겠습니다.

이름	조성호	학과	정보산업공학과			
학년	2학년	학번	2014147037			
		핸드폰 번호	010-5199-1176			
컴퓨터언어	Python, R					
수강과목	통계학입문, 통계방법론, 미분적분학, 선형대수학					
지원동기						
<p>1. 이번 2 학기 전공 강의에서 처음으로 통계 수업과 머신러닝 패키지를 다루는 수업을 들었습니다. 이전까지는 무엇을 공부하고 목표로 해야 하는지 몰랐습니다. 모델을 세우고 의사 결정을 내리는 것을 배우면서 관심이 생겼습니다. 그러면서 진로를 데이터 사이언스 쪽으로 생각을 했습니다.</p> <p>데이터 사이언스 혹은 빅데이터라는 말을 들었을 때는 저는 어떤 컴퓨터 프로그래밍의 영역을 떠올렸었습니다. 하지만 수업 과제를 하면서 단순히 코딩 하고 시뮬레이션 돌려가면서 파라미터 세팅하는 것이 정답이 아니라는 것을 깨달았습니다. 또 데이터 사이언스 관련해 찾아보면서 통계학에 기반해 데이터 전처리와 모델링을 강조하는 것을 보면서 그것을 확인했습니다.</p> <p>그래서 응용통계를 복수전공하면서 데이터 분석에 대한 공부를 하고자 합니다. 또 실무자들이 강조했던 것이 협업이었습니다. 다양한 직무와 백그라운드를 가진 사람들과 함께 해나가야 하는 일이라고 말이었습니다. 데이터 사이언스 랩에서 관련 공부와 함께 프로젝트를 진행하는 경험을 쌓아가고 싶습니다.</p>						
프로젝트 경험						
<p>2 학기 전공 수업 과제로 나왔던 프로젝트였습니다. 두 종류의 데이터를 갖고 조별로 python 의 패키지를 사용해보는 과제였습니다.</p> <p>첫번째는 kaggle 의 wine 데이터를 변형한 데이터를 갖고 분류 모델을 학습시켜 보았습니다. 먼저 수업 시간의 배웠던 decision tree 모델을 갖고 데이터 시각화부터 전처리, 학습, 성능개선 등을 고민해보고 최종적으로 학습한 tree 모델을 시각화해 보았습니다. 전처리와 변수선택, 파라미터를 조절해가며 성능을 살펴봤을 때 주어진 데이터에서 트리 모델은 test set 에 대한 성능이 낮게 나와 과적합되어 있다 판단을 했습니다. 또 학습한 트리 모델을 시각화 해봤을 때 매우 복잡하여 데이터에 대한 어떤 직관을 얻기는 힘들었습니다. 그리하여 보다 트리 모델보다 일반화될 수 있는 모델을 찾았고 logistic regression 과 SVM 그리고 Random Forest 모델을 사용했습니다.</p> <p>두번째는 반도체 공정 시운전에서 얻어진 데이터에 예측 모델을 적용해보는 문제였습니다. 처음 문제를 받아봤을 때는 첫번째 문제보다 더 쉽게 생각했지만 그것은 잘못이었습니다. 데이터를 관찰해보니 운전 횟수에 따라 경향을 갖는 시계열 데이터였습니다. 또 데이터의 개수에 비해서 데이터들의 노이즈가 너무 커서 어떤 모델을 적용해도 성능이 아주 나쁘게 나왔습니다. 처음 모델의 성능을 봤을 때는 좌절의 순간이었지만 조원끼리 시계열 자료에 대한 책과 자료도 찾아보고 계속 고민해 나갔습니다 그렇게 하여 얻어진 최종 모델의 성능 역시 좋지는 않았습니다. 하지만 공부를 하고 또 실제 데이터를 다루는 것이 얼마나 어려운지를 알</p>						

수 있었던 귀중한 경험이었다고 생각합니다.

프로젝트에서 python 의 pandas 와 sklearn, matplotlib 등의 패키지를 사용했습니다.

이름	김단아	학과	응용통계학 대학원
학년	3학기	학번	18
		핸드폰 번호	010-5173-6600
컴퓨터언어	Python, R, SQL		
수강과목	통계학입문, 통계방법론, 미분적분학, 선형대수학, 수리통계학 1, 수리통계학 2, 학부 - 데이터마이닝, 실험계획법, 다변량통계분석, 해석학(청강) 외. 대학원 - 수리통계학, 선형통계모형, 응용공간통계학, 다변량분석 1, 위험성 및 생존분석 1, 비모수적합수추정론, 데이터마이닝이론		
지원동기			
<p>안녕하세요. 응용통계학과 대학원 석사 3 학기생 김단아입니다. 4 차산업 시대와 데이터의 혁명 속 데이터와 통계의 중요성은 더욱 커지고 있습니다. 저는 학교 수업에서는 배우지 못했던 데이터 분석 실전 경험을 쌓아 데이터 사이언스 전문가로 거듭나고 싶어 데이터사이언스랩에 지원하게 되었습니다.</p> <p>데이터분석 역량을 함양하기 위해 응용통계학과 소속 학회인 ESC 에서 활동하며 심화된 통계적 이론 공부와 R, Python 을 이용한 다양한 데이터분석을 했습니다. 분석 과정에서 한계를 최소화하기 위해 관심 변수를 새로이 정의하는 등 유연한 사고를 키울 수 있었고, 데이터분석은 통계적 기본이 필요하다는 것을 느낄 수 있었습니다.</p> <p>현재는 통계적 전문성을 함양하기 위해 대학원에 진학하여 박상언 교수님 밑에서 연구를 준비하고 논문 지도를 받는 중에 있습니다. 통계를 공부하는 학우들과 함께 고민하고 공부하고, 저 또한 초기멤버로서 학회의 발전에 기여를 하며 데이터 사이언스 전문가로의 도약을 준비하고 싶습니다.</p>			
프로젝트 경험			
<p>IoT 수요자·공급자 분석 프로젝트(2016) / R 의 Decision Tree 를 이용하여 분석</p> <p>집값 예측 프로젝트(2017) / R 의 Regression 과 Random Forest 를 이용하여 예측 / 학회 내 프로젝트 1 등</p> <p>스피드 데이팅 실험 데이터 프로젝트(2017) / R 의 사용하여 분석 / 학회 내 프로젝트 1 등</p> <p>연세대·고려대 통계학회 연합 세션 프로젝트(2017) / 다변량분석을 사용한 이미지 압축 논문과 Python 을 이용한 실습 발표</p> <p>시장조사론 수강(2016) / SPSS 와 R 을 이용한 현대약품 미에로화이바 소비자 분석 및 브랜드 개선안 / A+학점 취득</p> <p>데이터마이닝 수강(2017) / SAS E-miner 를 이용한 지리정보와 날씨정보를 이용한 서울 콜택시 수요분석 프로젝트</p> <p>다변량분석(1) 수강(2018) / SAS 를 이용한 국민건강영양조사 청소년의 우울증 위험군 조사 / A+학점 취득</p>			

이름	전상후	학과	응용통계학과
학년	2학년	학번	2018122037
		핸드폰 번호	010-6858-0716
컴퓨터언어	R		
수강과목	통계학입문, 통계방법론, 미분적분학, 선형대수학		
지원동기			
<p>1학년 때 과 전공 기초 공부를 하면서 뭔가 허전함을 느꼈습니다. 기초 공부가 재미가 없지는 않았지만, 아무래도 기초를 쌓아가는 과정이다 보니 실제 통계를 다뤄볼 수 있는 기회가 많이 부족했습니다. 특히 2 학기 때 국제대 친구의 설문조사를 도와주면서 이를 많이 느꼈습니다. 통계자료 수집 과정부터 분석까지 제가 입문 과정에서 배운 것으로는 제대로 된 통계적 결과를 도출하기 어려웠고, 저는 심화 과정 학습의 필요성을 느낌과 동시에 실생활 경험의 중요성을 깨달았습니다. 앞에 언급했던 경험을 통해 실제 통계자료 조사를 하며 직접 부딪히면서 알아가는 것이 얼마나 값진 것인지 알게 되었습니다. 그 과정에서 케글이라는 데이터 사이언스 관련 사이트에 대해 알게 되었습니다. 케글 코리아 페이스북을 팔로우하면서 소식을 받아 보며 데이터 사이언스 대회 등에 대해서도 조금씩 알아가던 찰나, 데이터 사이언스 랩에 대해 알게 되었습니다. 경진대회 참가, 프로젝트 등의 랩 활동 내용에 대해 알게 되면서 흥미를 느껴 지원하게 되었습니다.</p>			
프로젝트 경험			
<p>아직은 공식적으로 프로젝트를 경험해 본 적은 없습니다.</p>			

이름	황주헌	학과	물리학과
학년	3학년	학번	201732023
		핸드폰 번호	010-9391-5127
컴퓨터언어	Python, R		
수강과목	통계학입문		
지원동기			
<p>저는 2018 년에 sw 프로그래밍과 전자계산(1)수업을 수강하면서 파이썬을 배우게 되었습니다. 파이썬을 배우다 보니 이 컴퓨터 언어가 인공지능이나 머신러닝에 주로 사용된다는 것을 알게 되었고 자연스럽게 제 관심이 이쪽으로 기울게 되었습니다. 관심이 생긴 후 여름방학 때 친구들과 함께 데이터 사이언스 대회에 등록하여 큰 연구 성과를 내지는 않았지만 R 언어에 대해서 배우고 통계학에 대해서도 배우는 시간을 가지면서 데이터 사이언스에 대해서 더 공부하고 싶다는 생각을 하게 되었습니다. 그렇게 대학원도 진학하고자 희망하고 있는데 데이터 분석을 하는 것은 응용통계학과가 제가 원하는 목표에 가장 적합할 것 같다는 생각을 하게 되었습니다. 그러던 중 데이터사이언스 학회를 만든다는 것을 알게 되어서 이 학회에서 활동을 하게 된다면 저에게 큰 밑바탕이 될 수 있을 것 같아서 이렇게 지원을 하게 되었습니다.</p>			
프로젝트 경험			
<p>네이버에서 주관하였던 데이터 사이언스 대회에 등록하여서 제공된 데이터들 중 하나를 골라 주제를 정하여 결과를 도출하는 프로젝트를 하였었습니다. 비록 기초가 부족하여 기본 강의들을 수강하는 데에 시간을 대부분 할애하여 제대로 된 보고서를 작성하지는 못하였으나 그 경험을 통하여 R 언어를 어느 정도 익히게 되었습니다. 그와 함께 통계학도 공부를 같이 하여서 R 언어를 통해 분석할 때, 어떤 모형을 써서 상관관계를 찾아갈 수 있는지에 대해서 알게 되었습니다.</p>			