경상국립대학교
Gyeongsang National University

UNLV
Summer Research Program

2023 빅데이터 혁신융합대학 사업단
하계 해외 연구프로그램 결과보고서

경상국립대학교
Gyeongsang National University

UNLV
Summer Research Program

2023 하계 해외 연구프로그램 결과보고서

빅데이터 혁신융합대학 사업단

인사말

안녕하십니까? 반갑습니다.

빅데이터 혁신융합대학사업단은 빅데이터 분야의 융합성, 수월성, 다양성을 지닌 인재를 양성하기 위해 교육부 지원으로 2021년부터 지금까지 7개 대학이 컨소시엄을 구성하여 사업을 수행하고 있으며, 이번 빅데이터 해외 연구프로그램도 학생들의 데이터 분석 능력과 글로벌 마인드 함양을 위해 비교과활동 지원의 일환으로 기획되었습니다.

이번 프로그램은 경상국립대학교가 기획하고 빅데이터 혁신융합대학 사업에 함께 참여 중인 전북대학교와 경기과학기술대학교 등 3개 대학 26명(경상국립대학교 10명, 전북대학교 10명, 경기과학기술대학교 6명)의 학생이 참여한 가운데 지난 6월 26일부터 7월 21일까지 4주간 미국의 네바다주립대학교(UNLV)에서 진행되었습니다. 참여 학생들은 빅데이터 및 인공지능(AI) 분야의 최신 기술 습득 및 프로젝트 수행 등을 통하여 글로벌 인재로서의 소양을 쌓기 위해 인생 최고로 뜨거운 여름을 라스베이거스에서 보냈습니다.

해외 연구프로그램에 참여한 학생들은 영어 구사 능력뿐만 아니라 데이터 분석과 머신러닝 기술, 그리고 팀원들과의 협업으로 문제 해결 능력을 키울 수 있었습니다. 또한 다양한 문화와 사람들을 접함으로써 사고의 폭이 넓어졌으며, 이번 연구프로그램이 더욱 열심히 노력하여 부족한 부분을 개선하고 발전시킬 것임을 다짐하는 귀중한 시간이 되었으며 재학생들에게 프로그램 참여를 적극적으로 추천할 것이라고 합니다.

귀국 후 가진 최종보고회에서 이번 해외 프로그램을 통해 쌓은 다양한 경험과 새로운 지식을 발판 삼아 우리 학생들이 한층 더 성장할 것임을 확인할 수 있어 매우 기뻤으며, 앞으로 해외 프로그램 외에도 빅데이터 창업동아리, 경진대회 등 다양한 프로그램 운영을 통해 빅데이터 인력 육성을 위해 노력해 나아가겠습니다.

이 결과보고서에는 연구프로그램에 임했던 학생들이 수행했던 Big Data & AI 프로젝트, Field Trip, American Buddy Program, 그리고 Industry Visit 등에 대한 생생한 경험담이 담겨져 있습니다. 앞으로 해외 연구프로그램에 참여할 후배들에게 유익한 도움을 줄 수 있는 뜻깊은 자료가 되리라 생각됩니다.

처음 시행하는 프로그램으로써 기획부터 보고서 발간까지 단 하나도 쉬운 것이 없었음에도 끝까지 믿고 따라와 준 26명의 학생과 물심양면으로 이번 프로그램을 적극적으로 지원해주신 서울대학교 및 컨소시엄 참여 대학 관계자 여러분의 격려와 성원에 깊은 감사의 말씀을 드립니다.

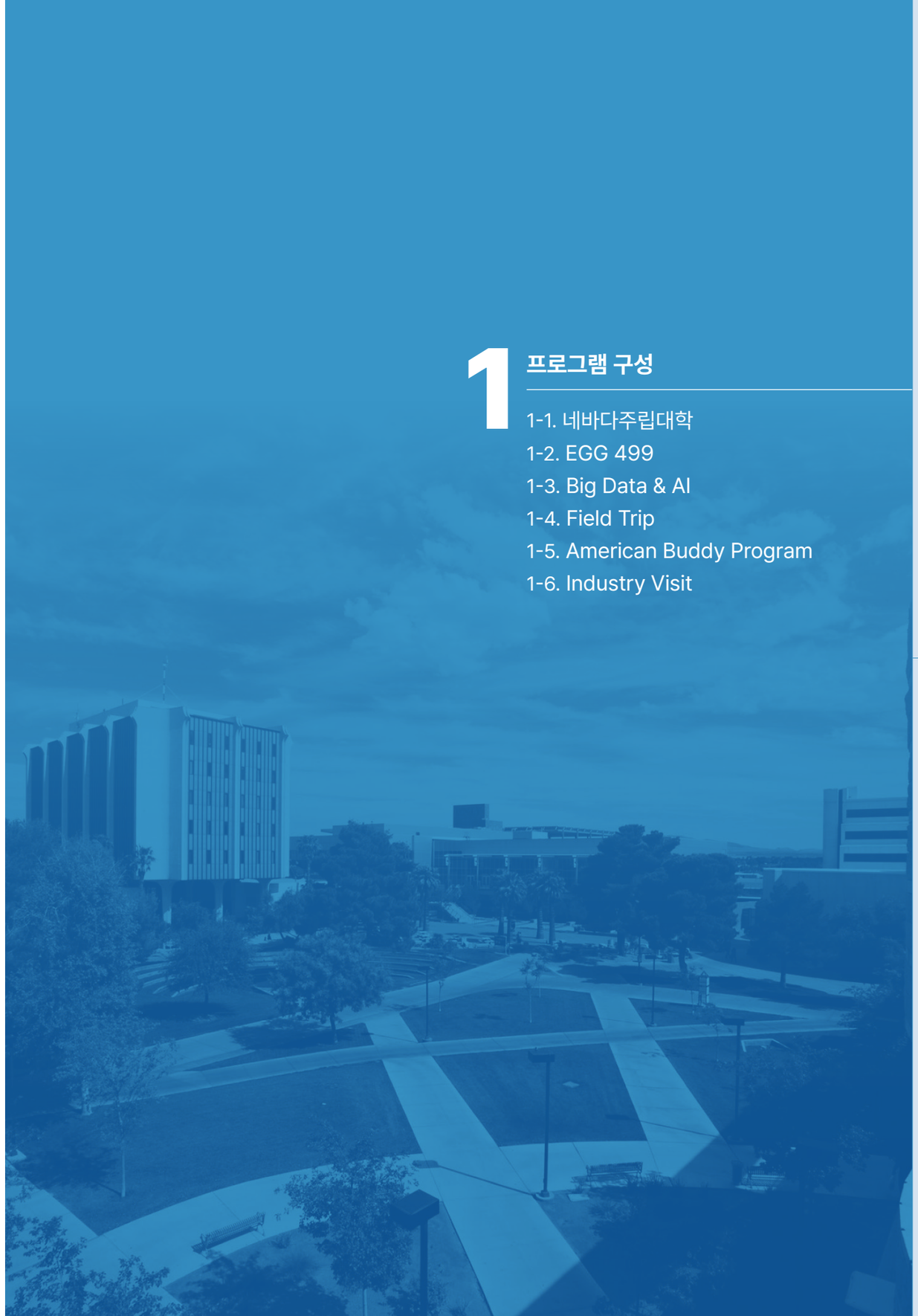
2023. 10.

경상국립대학교 빅데이터 혁신융합대학 사업단장 한 관 희

UNLV Summer Research Program

프로그램 전체 일정 6. 25SUN - 7. 22SAT		
교육 프로그램 6. 26MON ~ 7. 20FRI	EGG 499 Big Data & AI	
	비교과 프로그램	Field Trip American Buddy Program Industry Visit

목차 Contents



1 프로그램 구성

1-1. 네바다주립대학교
University of Nevada, Las Vegas, UNLV

네바다주립대학교(University of Nevada, Las Vegas, 이하 UNLV)는 미국 서부지역의 중요한 교육 및 연구 중심지 중 하나로, 학생들에게 다양한 학문 분야에서 높은 수준의 교육을 제공한다. 다른 명문대에 비해 역사가 짧은 편이지만 호텔과 관광산업이 발달한 라스베이거스에 위치하여 미국 내 호텔 경영학과 평가 순위 1, 2위를 다투고 있다.

UNLV는 아름다운 라스베이거스의 중심에 자리하고 있어, 학생들은 학문적인 도전과 동시에 도시의 다양한 문화 활동과 엔터테인먼트를 즐길 수 있다. 라스베이거스의 중요한 일원으로서 학생들에게 도시 생활의 혜택을 제공하며, 산업 및 비즈니스 파트너와의 협력을 통해 학생들에게 진로 기회를 제공한다. 또한 학생 중심의 교육을 통해 창의적인 생각과 문제 해결 능력을 개발하는 데 주력하고 있으며, 학생들이 성공적인 미래를 위한 기반을 마련할 수 있도록 지원하고 있다.

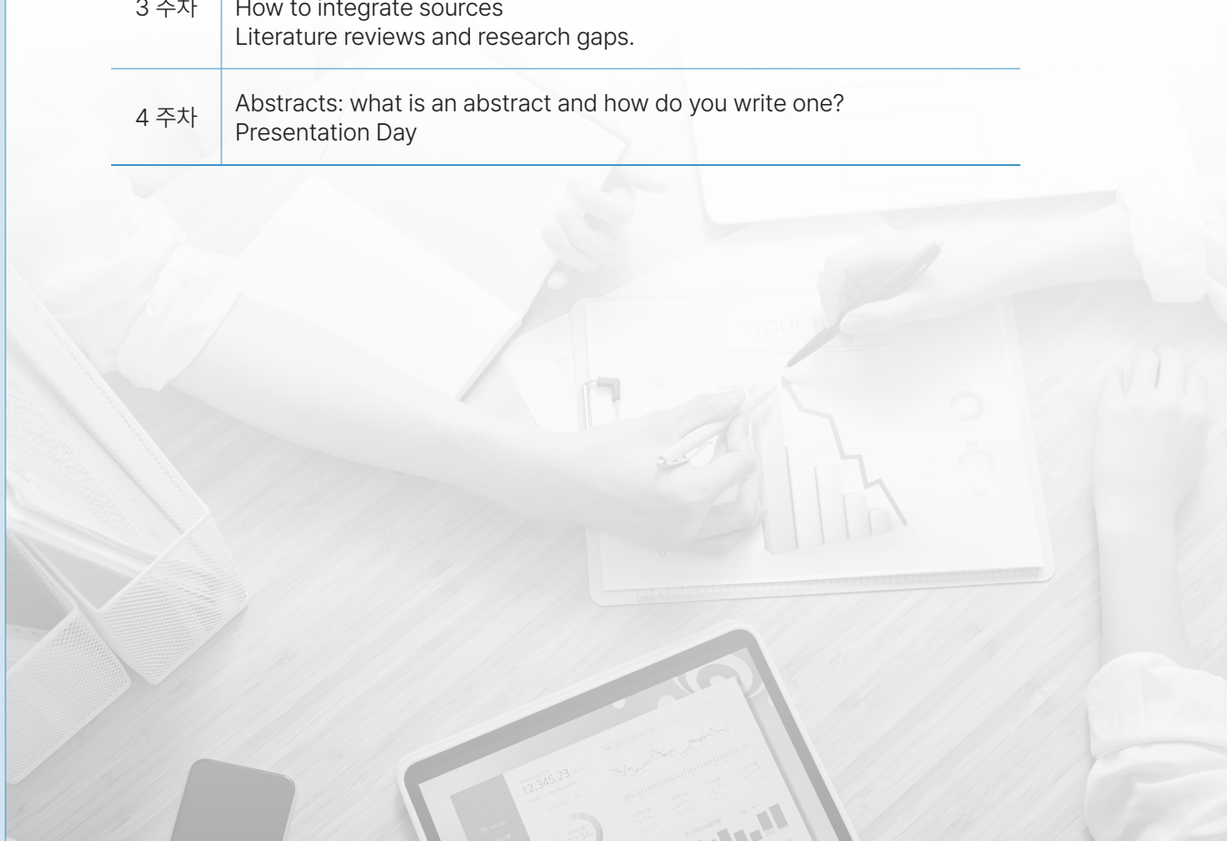
이번 해외 연구프로그램은 공과대학의 Engineering International Programs(EIP)와 협력하여 진행하였다. 이 프로그램은 다양한 국적과 배경을 가진 학생들이 협력하고 공부할 수 있는 환경을 제공하여 국제 학문 및 문화 교류 촉진, 학생들에게 글로벌 엔지니어링 문제를 해결하는 기술과 역량을 갖추도록 다양한 프로그램을 제공하고 있다.

EIP는 네 가지 프로젝트로 진행하는 빅데이터 교육과 다양한 Off Campus Program을 제공함으로써 학생들에게 풍부한 경험을 선사하여 학습 동기부여에 많은 도움을 주었다.

1-2. Research Internship for Engineering and Computer Science (EGG 499)

EGG 499 과정은 4주 동안의 집중 학습을 통해 연구 논문 작성에 필요한 핵심 능력을 키우는 과정이다. 강의를 통해 올바른 콤마 사용법, 문장 요약, 글쓰기 그리고 Google Scholar를 효과적으로 활용하는 방법을 배운다. 프로그램에 참여한 26명의 학생은 각자 작성한 연구 논문을 마지막 4주 차에 발표하는 것으로써 EGG 499 과정을 모두 이수하였다.

주차	강의 주제
1 주차	Syllabus day and introduction Review paragraph structure and commas
2 주차	Thesis statements and topic sentences How to format an essay with a thesis statement based on the body paragraphs
3 주차	How to do academic research How to integrate sources Literature reviews and research gaps.
4 주차	Abstracts: what is an abstract and how do you write one? Presentation Day



1-3. Big Data & AI

Big Data & AI 강의는 머신러닝의 기초부터 고급 주제까지 폭넓게 다루었다. 파이썬 데이터 구조부터 시작하여 KNN, SVM, 신경망과 같은 다양한 알고리즘을 이해하고, 선형 회귀와 정규화를 통해 데이터를 모델링하고 최적화하는 방법을 배웠다. 또한, Map Reduce와 Spark를 사용하여 대용량 데이터 처리와 Evaluation 및 Classification 주제를 통해 모델 평가와 분류 작업도 체계적으로 학습했다.

구성 항목	주요 내용
Introduction to ML	데이터 타입과 데이터 마이닝의 의의를 학습하고, 머신러닝의 개념과 다양한 종류, 예시를 파악한다.
Data Structures in Python	파이썬 데이터 구조인 숫자, 문자열, 리스트, 튜플, 딕셔너리, 넘파이 배열, 판다스 배열에 대해 학습한다.
Classification	머신러닝 학습 과정을 이해하고, 지도 학습의 분류와 회귀, 그리고 선형/비선형 분류에 대한 기초를 다진다. 모델 평가, 과적합과 과소적합에 대하여 학습한다.
KNN	KNN 알고리즘의 개념과 특징, 원리와 이에 따른 장단점을 학습하고, 결정 경계와 거리 측정, 피쳐 스케일링에 대해 이해한다.
Evaluation	혼동행렬, ROC 커브 등 모델 성능 평가 방법과 홀드아웃, 교차검증, 부트스트래핑 등 검증 기법을 배운다.
Linear Regression	데이터 간의 상관관계를 이해하고, 선형 회귀 분석 중 단순 회귀 분석과 다중 회귀 분석을 학습한다. 최소제곱법과 범주형 변수와 분류에서의 회귀분석을 익힌다.
Regularization on Linear Model	다중 회귀의 다중공선성 문제를 소개하며, 릿지 회귀 개념을 이해하고 활용 방법을 습득한다.
Map Reduce	대규모 데이터 처리를 위한 맵리듀스의 필요성과 기본 개념을 이해하며, 맵리듀스의 과정과 예시를 통해 활용 방법을 습득한다.
Spark	Spark의 기본 개념과 사용 의의를 이해하고, RDD의 역할과 기능을 습득한다.
Numerical Optimization	전역 최적화와 지역 최적화의 차이를 이해하고, Convex 함수와 gradient의 역할을 파악한다. 최적화 값을 효과적으로 탐색하는 방법을 습득하고, 학습률, step size와 같은 하이퍼파라미터에 대한 이해를 높인다.
SVM	초평면, 마진 등 SVM의 중요한 개념들을 학습한다. SVM의 원리를 이해하고 선형 데이터와 비선형 데이터에 적용할 경우를 살펴본다.
Neural Networks	퍼셉트론, 신경망의 기본 개념과 구조를 이해하고, 역전파 알고리즘, 경사하강법 등 신경망 학습과 딥러닝의 핵심 원리를 습득한다.

1-4. Field Trip

1) 라스베이거스 Strip and Downtown

라스베이거스는 크게 메인 스트립(strip)과 다운타운(downtown)으로 나뉘어진다.
약 6km 길이에 달하는 라스베이거스 스트립은 세계에서 가장 화려하고 아름다운 밤거리로 널리 알려져 있으며 세계적으로 유명한 호텔, 카지노, 레스토랑, 상점, 엔터테인먼트 시설로 가득 차 있어 라스베이거스의 주요 관광 명소 중 하나이다.

벨라지오(Bellagio) 리조트는 라스베이거스 스트립에서도 특히 관광객들이 발길이 끊이지 않는 곳으로 라스베이거스 시내 한복판의 거대하고 잔잔한 호수, 고전적이면서도 웅장한 호텔의 모습에 놀라지 않을 수 없다. 3.2헥타르 규모의 인공호수에서 펼쳐지는 분수 쇼는 다양한 음악과 물줄기를 통해 감동과 힐링을 선사한다.

다운타운 지역은 올드타운이라고 불리는 곳으로 라스베이거스의 초기 개척과 성장의 중심지로 빼놓을 수 없는 곳이다. 프리몬트 스트리트가 다운타운의 주요 거리이며, 머리 위의 전광판에서 화려하게 펼쳐지는 다양한 미디어아트와 음악, 스테이지와 길거리에서 펼쳐지는 다양한 공연들, 머리 위로 날아가는 짚라인까지 볼거리가 수두룩하다.

빼놓을 수 없는 볼거리로 『프리몬트 스트리트 전구쇼』가 있다. 450m 길이의 아케이드 천장에 1,600만 개의 발광다이오드가 장식돼 매시간 현란한 조명쇼를 볼 수 있다. 놀랍게도 프리몬트 스트리트 전구쇼는 LG전자에서 공급한 LED로 만들어졌다.

2) Campus Tour

▪ Student Library
도서관은 매우 현대적이고 조용한 분위기를 갖췄으며, 학업에 집중하기에 이상적인 장소이다. 또한, 도서관의 엄청난 자료와 열린 공간은 학생들에게 학습과 연구를 위한 훌륭한 환경을 제공해주기에 충분하다. 특히 반납된 책이 자동으로 정리되는 시스템이 인상적이다.

▪ Student Union
학생들이 모여서 다양한 활동을 즐길 수 있는 곳으로, 식당, 카페, 편의점 등 다양한 시설을 갖추고 있다. 학생들에게 다양한 활동을 제공함으로써 학교생활을 더욱 풍부하게 만들어준다.

▪ School Gym
수영장, 농구장, 웨이트 트레이닝 등 다양한 운동 시설과 프로그램을 통하여 학생들이 건강한 학교생활을 유지하고 즐길 수 있도록 돕고 있다.

UNLV 캠퍼스 투어는 학교의 모든 면을 알아보고 학생들의 생활과 학습 환경을 이해하는 데 매우 유용한 기회였으며, 이를 통해 학습 지원뿐만 아니라 학생들의 편의를 위해서도 다양한 시설과 서비스를 제공하는 학교라는 것을 알게 되었다.

3) National Park Tour

▪ 자이언 캐니언(Zion Canyon)
서부 유타(Utah) 주, 스프링데일(Springdale) 근처에 위치한 자이언 국립공원(Zion National Park)에는 여러 크고 작은 협곡이 있으며 대표적인 것으로 자이언 캐니언을 꼽을 수 있다. 1억 5천만 년에 거쳐 쌓인 퇴적층이 버진강(Virgin River)의 흐름에 의해 침식되어, 지금의 자이언 캐니언이 되었다. 자이언 캐니언은 그랜드 캐니언(Grand Canyon), 브라이스 캐니언(Bryce Canyon)과 함께 미국 서부 3대 협곡으로 꼽힌다.
자이언 국립공원의 대표적인 트래킹 코스로는 더 내로우즈(The Narrows)와 엔젤스 랜딩(Angel's Landing)이 있다. 더내로우즈 코스는 폭 6m, 높이 600m의 협곡 사이를 흐르는 버진 강을 따라 걷는 편도 26km짜리 코스이다. 물속을 걸으면서 올라가는 독특하고 모험적인 경험을 제공한다.

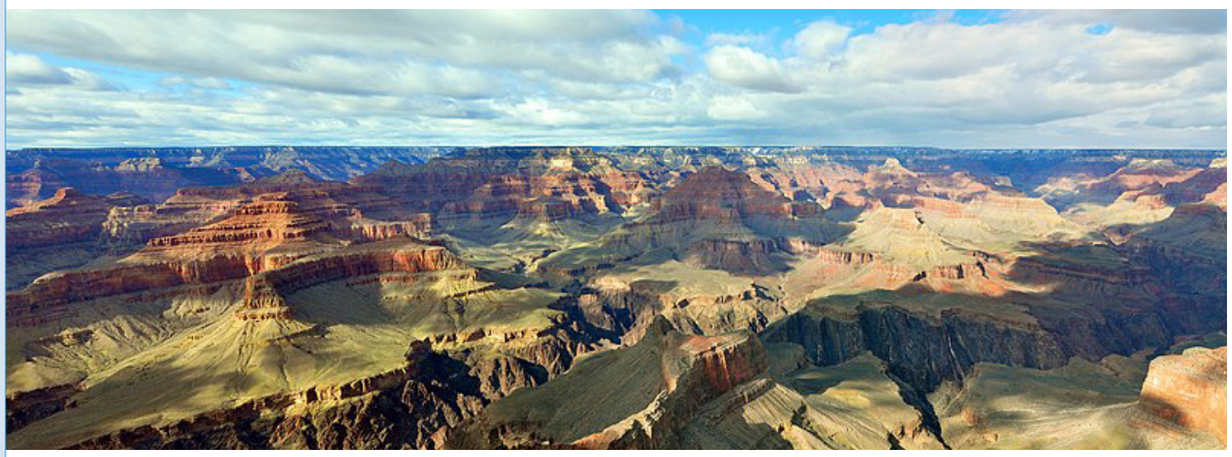
▪ 홀스슈 밴드(Horseshoe Bend)
다음으로 방문한 홀스슈 밴드는 콜로라도 강이 270도 굽어 흐르며 깎은 암벽의 모양이 마치 말발굽과 닮았다고 하여 홀스슈라는 이름이 붙여졌으며, 홀슈스 밴드의 밴드는 커다란 말발굽 모양의 암벽을 콜로라도 강이 띠를 두르고 있는 형태라 하여 칭하게 된 것이다.

▪ 앤텔로프 캐니언(Antelope Canyon)

앤텔로프 캐니언은 다른 캐니언과 달리 미국의 국립 공원이 아닌 나바호 부족이 관리하는 자치 구역에 속한다. 따라서 관광객이 마음대로 드나들 수 없으며 관광을 할 때도 인디언 현지 가이드와 동행해야 해서 엔텔롭 캐니언과 나바호 부족은 뗄 수 없는 사이이다. 앤텔로프 캐니언은 나바호 샌드 스톤이라는 잘 알려지지 않은 사암으로 형성된 좁고 깊은 협곡으로, 형태와 빛, 색깔이 각각 어우러져 독특한 아름다움을 감상할 수 있는 장소이다. 앤텔로프 캐니언은 사암으로 이루어져 있어 전반적으로 적색 계통의 색을 띠는 한편, 협곡 내부에 빛이 흘러 들어와 급류로 생긴 독특한 모양의 사암을 비추어 아름다운 광경을 만들어 낸다.

▪ 그랜드 캐니언(Grand Canyon)

'웅장한' 뜻의 Grand를 붙인 그랜드 캐니언은 이름만큼이나 웅장하고 거대한 협곡이다. 그랜드 캐니언은 수백만 년 동안 콜로라도 강의 침식작용에 의해 형성되었으며, 계곡의 깊이는 1,600m에 이르고 계곡의 폭은 넓은 곳이 30km에 이른다. 대체로 붉은색을 띠지만 지층 또는 지층군에서는 독특한 색들을 띠기도 한다. 이 계곡이 유명한 이유는 엄청난 규모와 아름다움이기도 하지만 다양한 바위층과 화강암, 석회암 등이 서로 다른 시기의 지질적 변화를 보여줌으로써 지구의 역사를 알려주는 장소이기 때문이다.



1-5. American Buddy Program

1) 코와붕가 베이 워터파크(Cowabunga Bay Water Park)

라스베이거스 최고의 워터파크로써 1960년대를 테마로 한 코와붕가 베이는 세계에서 가장 큰 인공 파도(22m)인 와일드 서프가 있어 더 유명하다. 또한 모든 연령대의 방문객들이 즐길 수 있는 다양한 놀이기구들이 밀집해 있으며 그 중 포인트 패닉(Point Panic)은 꼭 타봐야 하는 기구 중 하나이다.

2) NBA Summer League

2004년 라스베이거스 서머 리그가 처음으로 개최되었으며 현재는 3개의 서머 리그가 개최되고 있는데, 가장 널리 알려진 것은 30개 구단이 모두 참가하는 라스베이거스 서머 리그다. 서머 리그는 슈퍼스타들이 출전하는 리그가 아니며 주로 신인 선수나 유망주들이 경기에 참여한다.

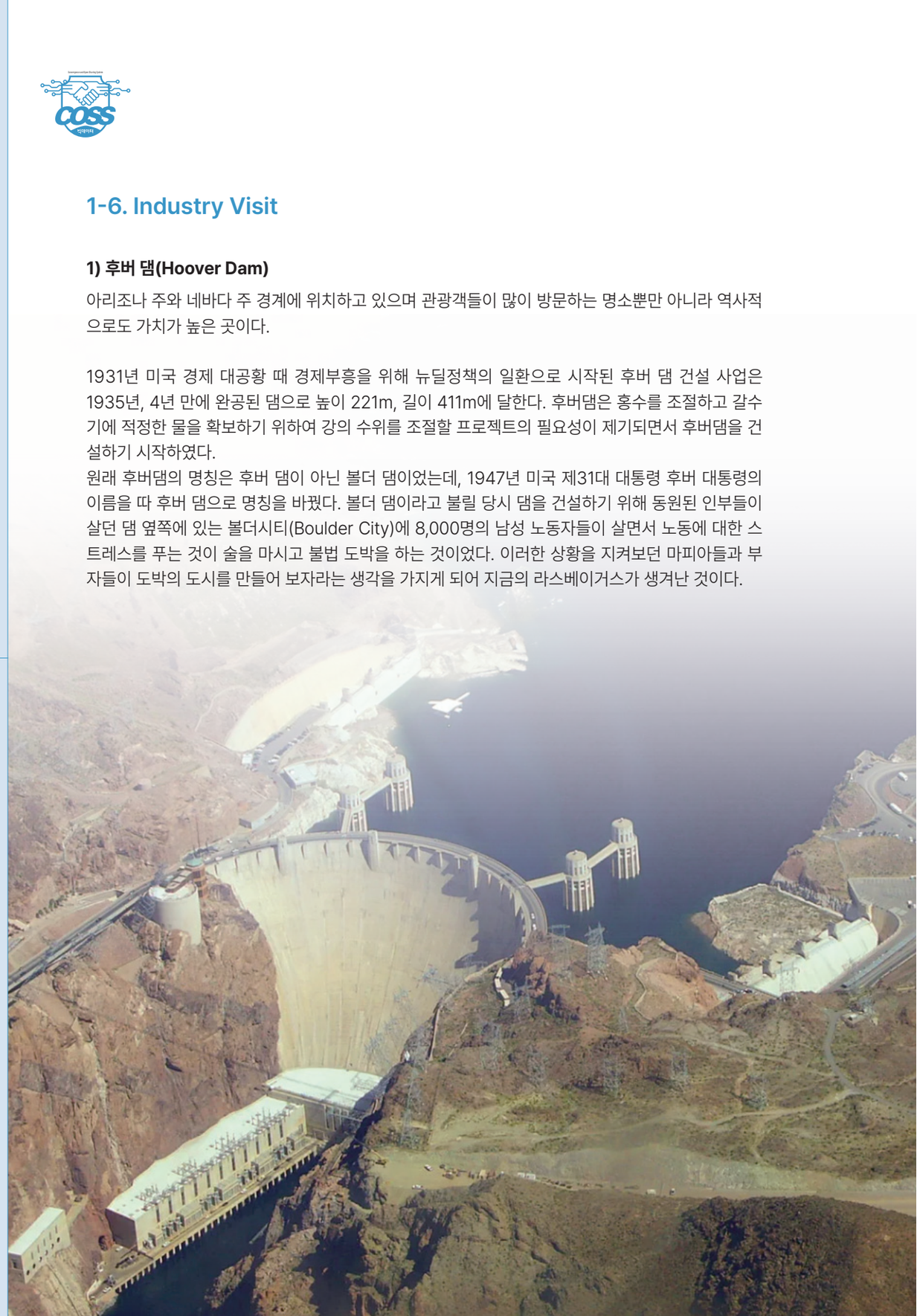
라스베이거스 서머 리그 (Las Vegas Summer League)는 매년 여름 네바다 주 패러다이스에서 개최되는데 해외 연구프로그램에 참여한 학생들은 토머스 & 맥 센터(Thomas & Mack Center) 농구장에서 Thunder vs Mavericks 경기를 관람하며 NBA의 뜨거운 열기를 경험했다.

1-6. Industry Visit

1) 후버 댐(Hoover Dam)

아리조나 주와 네바다 주 경계에 위치하고 있으며 관광객들이 많이 방문하는 명소뿐만 아니라 역사적으로도 가치가 높은 곳이다.

1931년 미국 경제 대공황 때 경제부흥을 위해 뉴딜정책의 일환으로 시작된 후버 댐 건설 사업은 1935년, 4년 만에 완공된 댐으로 높이 221m, 길이 411m에 달한다. 후버댐은 홍수를 조절하고 갈수기에 적정한 물을 확보하기 위하여 강의 수위를 조절할 프로젝트의 필요성이 제기되면서 후버댐을 건설하기 시작하였다.
원래 후버댐의 명칭은 후버 댐이 아닌 볼더 댐이었는데, 1947년 미국 제31대 대통령 후버 대통령의 이름을 따 후버 댐으로 명칭을 바꿨다. 볼더 댐이라고 불릴 당시 댐을 건설하기 위해 동원된 인부들이 살던 댐 옆쪽에 있는 볼더시티(Boulder City)에 8,000명의 남성 노동자들이 살면서 노동에 대한 스트레스를 푸는 것이 술을 마시고 불법 도박을 하는 것이었다. 이러한 상황을 지켜보던 마피아들과 부자들이 도박의 도시를 만들어 보자라는 생각을 가지게 되어 지금의 라스베이거스가 생겨난 것이다.



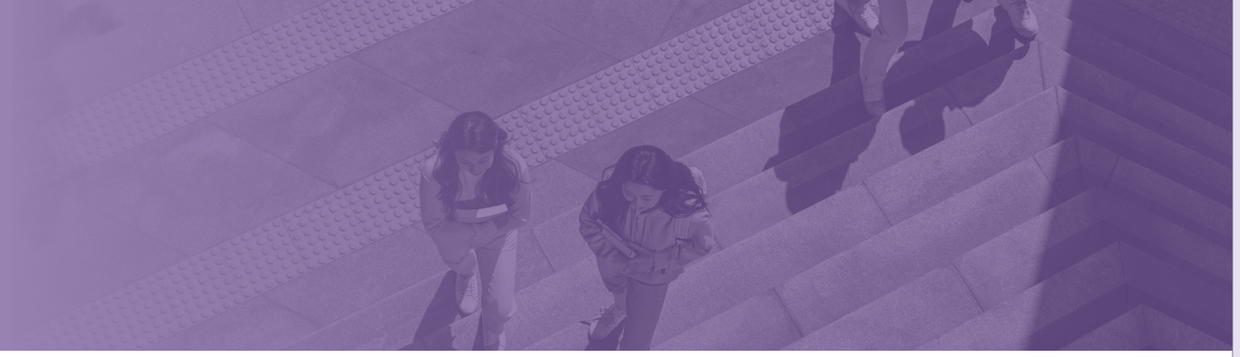
엄청난 규모의 후버 댐에 의해 생겨난 인공호수가 있는데, 이 인공호수 역시 미국에서 가장 큰 규모를 자랑하고 있으며 이름은 미드 호수(Lake Mead)이다. 길이 약 185km의 어마어마한 이 인공호수는 파도가 칠 정도로 큰 규모를 자랑한다.트레스를 푸는 것이 술을 마시고 불법 도박을 하는 것이었다. 이러한 상황을 지켜보던 마피아들과 부자들이 도박의 도시를 만들어 보자라는 생각을 가지게 되어 지금의 라스베이거스가 생겨난 것이다.

엄청난 규모의 후버 댐에 의해 생겨난 인공호수가 있는데, 이 인공호수 역시 미국에서 가장 큰 규모를 자랑하고 있으며 이름은 미드 호수(Lake Mead)이다. 길이 약 185km의 어마어마한 이 인공호수는 파도가 칠 정도로 큰 규모를 자랑한다.

2) 카쇼(Ka Show)

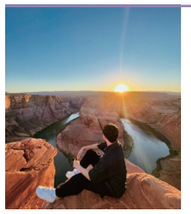
라스베이거스 3대 공연 중 하나인 카쇼는 일본어로 '불'이라는 뜻인 Ka를 인용하여 만든 쇼이다. 그래서인지 쇼는 동양적인 느낌을 풍기며 진행된다.

쌍둥이 남매의 모험을 영웅적인 이야기로 그려낸 카쇼는 4가지 신으로 이루어지는데 물, 공기, 지구, 불의 각 신에서 인간의 희로애락을 극적으로 나타내어 사랑과 전쟁을 추상적이고 몽환적인 연기로 표현하며 관객들의 감탄을 자아낸다.
카쇼는 블록버스터 급 무대와 신비로운 배우들의 곡예, 동양적인 요소가 많이 반영되어 있으며, 탄탄한 줄거리로 많은 사람에게 인기 있는 쇼이다. 단순 서커스 느낌의 공연이 아니라 대사가 없음에도 탄탄한 이야기와 기승전결로 누구나 쉽게 즐길 수 있는 공연이다.



2 참여 학생

Profile

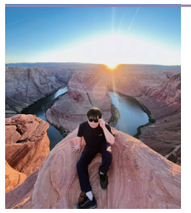


고진영 /경기과학기술대학교 컴퓨터모바일융합공학과

소중한 인연과 함께 좋은 일을 할 수 있어서 무척 행복했던 올해의 순간 중 하나는 바로 빅데이터 해외 연구프로그램에 참여한 것입니다. 새로운 문화와 경험을 만나며 성장할 수 있는 기회를 얻어서 정말로 자랑스럽게 생각합니다. 이 경험을 토대로 더 나은 미래를 향해 나아가고자 합니다.

김강현 /경기과학기술대학교 컴퓨터모바일융합공학과

좋은 사람들과 함께 지내며 같이 공부할 수 있는 기회가 되어서 너무 좋았습니다. 라스베거스에서의 소중한 경험은 앞으로 인생을 살아나가거나 공부를 하기 위한 밑거름이 될 것입니다.

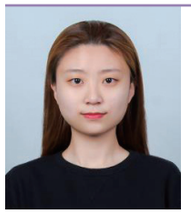


김도현 /경기과학기술대학교 컴퓨터모바일융합공학과

새로운 경험과 새로운 친구들, 새로운 문화를 체험할 수 있는 기회가 너무 새롭고 좋았습니다.
UNLV에서의 프로그램과, 라스베이거스에서의 한달은 잊지 못할 소중한 추억이자, 인생의 경험이 될 것입니다.

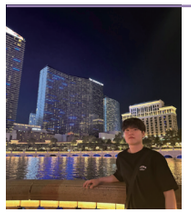
김민선 /경기과학기술대학교 인공지능학과

좋은 사람들과 함께 협동하여 멋진 경험을 쌓을 수 있었고, 미국이라는 나라에서 생활하며 공부할 수 있는 좋은 기회를 얻을 수 있어서 좋았습니다. 앞으로의 인생의 터닝포인트가 될 만큼 값진 시간이 되어 행복했습니다. 평생 잊지 못 할 시간이었습니다.



이재은 /경기과학기술대학교 컴퓨터모바일융합공학과

처음 가보는 미국과 처음 보는 사람들, 처음 해보는 것들로 가득 차 있던 즐거운 해외연수였습니다. 공부의 실력도 한 단계 올라갈 수 있는 좋은 기회였을 뿐만 아니라 다양한 경험들과 소중한 친구들과 인연들이 생겨 무엇보다 기쁩니다. 함께 가게 되어 진심으로 영광이었습니다!

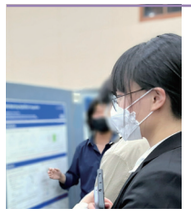


함성영 /경기과학기술대학교 컴퓨터모바일융합공학과

처음 라스베이거스라는 낯선 도시에 도착하여 여러 학생들과 함께한 미국 생활은 저에게 정말 뜻깊은 시간이었습니다. UNLV에서 대학 생활을 즐기며 다른 학생들과 협업하여 프로젝트를 진행할 수 있었던 시간들이 소중했습니다.

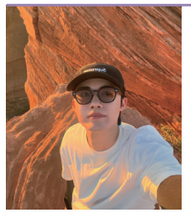
강동현 /경상국립대학교 수학과

새로운 환경을 경험하고 좋은 사람들과 함께 할 수 있어서 좋았어요. 가장 인상 깊은 경험이 됐어요.



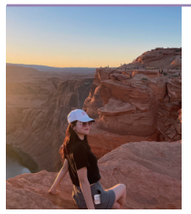
김가현 /경상국립대학교 물리학과

대학생의 신분으로, 이국적인 곳으로 멀리 떠나 특색있는 체험을 할 수 있었습니다. 다른 사람들과 다르게 원 소속학과에서 저 혼자 간 연수라 불안하고 힘든 점도 존재했지만, 다른 학과 소속의 새로운 사람들과도 친해지고 해외의 연구환경도 체험할 수 있었습니다.



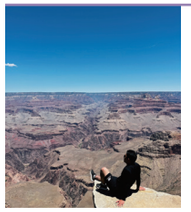
박건영 /경상국립대학교 산업시스템공학부

대학 생활에서 가장 의미 있었던 시간 중 하나입니다. 빅데이터 사업단에서 진행하는 해외 연구프로그램에 참여하여 4주간 미국에서 생활하며, 그곳의 최신 기술 동향과 문화를 몸으로 느낄 수 있었습니다.



신서빈 /경상국립대학교 산업시스템공학부

한 달간의 하계 빅데이터 해외 연구프로그램은 유익하고 가치 있는 경험이었습니다. 네바다주립대학교의 수업과 프로젝트를 통하여 데이터 분석과 머신러닝 기술의 역량을 향상할 수 있었고, 다양한 문화와 사람들을 접함으로써 사고의 폭이 넓어졌습니다. 더욱 열심히 노력하여 부족한 부분을 개선하고 발전시킬 것임을 다짐하는 귀중한 시간이었습니다.

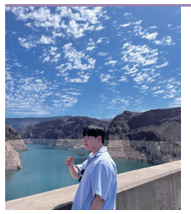


이인호 /경상국립대학교 산업시스템공학부

라스베이거스에서 보낸 이번 여름은 잊지 못할 경험이었습니다. 이번 프로그램은 다양한 측면에서 개인적인 성장과 학문적인 향상을 도모하는 기회였으며, 이를 통해 미래의 계획과 꿈을 키울 수 있는 계기가 되었습니다.

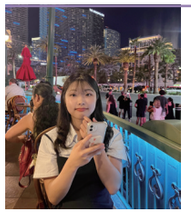
이지상 /경상국립대학교 정보통계학과

미국 라스베이거스에서의 해외 연구프로그램을 통해 소중한 경험과 학습 기회를 얻었습니다. 또한, 다양한 분야의 사람들과 프로젝트 진행을 함께 하면서 견문을 넓힐 수 있는 좋은 기회가 되었으며, 함께한 동료들과 즐거운 추억을 만들었습니다. 이 경험은 빅데이터 분야에서의 지식과 역량을 향상시키는데 도움이 되었고, 앞으로의 진로를 개척하는데 좋은 밑거름이 될 것이라 확신합니다.



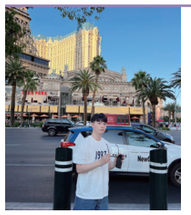
이태훈 /경상국립대학교 정보통계학과

통계학과 재학생으로서 빅데이터와 AI에 대한 관심이 있습니다. 다양한 공모전과 프로젝트에 참여하여 좋은 성적을 이뤘습니다.
이번 해외 연구프로그램을 통해 다양한 경험을 하고 학문적으로 한 층 더 성장할 수 있었습니다.



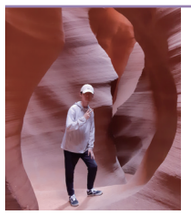
정인영 /경상국립대학교 정보통계학과

이 프로그램을 통해서 새로운 문화와 환경에 적응하면서 성장할 기회를 얻게 되었으며, 다양한 사람들과 교류하며 더 넓은 시야를 가질 수 있게 되었습니다.



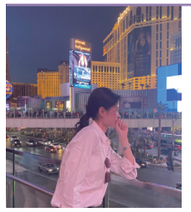
정현서 /경상국립대학교 산업시스템공학부

저는 미국이라는 나라에 로망이 있었던 편이었고, 죽기 전에는 한 번 꼭 가야겠다고 생각하고 있었습니다. 제가 지원한 '하계 빅데이터 해외 연구프로그램'은 '미국'에서 진행하는 프로그램이었습니다. 지원하지 않을 수가 없었습니다.



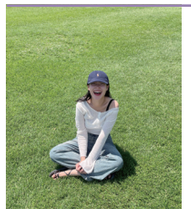
최규진 /경상국립대학교 수학과

빅데이터 해외 탐방 프로그램을 통해서 여러 사람과 같은 목표를 향해 공부하고 연구하는 활동을 할 수 있어서 좋았습니다.
이번 활동을 계기로 유창한 회화 능력을 갖춘 개척인으로 거듭날 수 있도록 노력하겠습니다.



강다영 /전북대학교 스마트팜학과

같은 분야를 좋아하고 비슷한 고민을 하며 함께 성장해나간 좋은 친구들을 만날 수 있어 행복했습니다. 첫 미국 여행을 해외 연구 프로그램으로 함께하며 좋은 추억만 남기고 갑니다.



권지현 /전북대학교 통계학과

같은 분야에 흥미를 가지고 있는 사람들과 프로젝트를 진행할 수 있어 뜻깊은 시간이었어요. 대학 생활 중 잊지 못할 추억이 될 거 같습니다.

김나영 /전북대학교 컴퓨터공학부

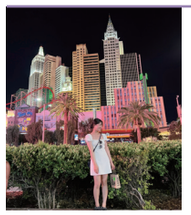
좋은 사람들과 한 달 동안 같이 공부하며 좋은 경험을 해서 행복했어요. 올해 들어 한 일 중에 가장 잘한 일이 해외 빅데이터 연구프로그램에 참여한 것입니다.

김미현 /전북대학교 컴퓨터공학부

미국에 다녀와서 시야도 넓어지고, 공부 욕심이 많이 생겼습니다. 한 달간 하루하루 너무 행복하였고, 이런 환경에서 살기 위해 지금 더 열심히 활동하고 싶은 의지가 생겼습니다.

김재현 /전북대학교 소프트웨어공학과

미국 좋네요. 다음에 또 올랍니다.



박세현 /전북대학교 소프트웨어공학과

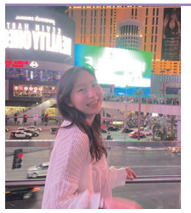
다양한 사람들과 여러 경험을 하고, 많은 것을 배운 뜻깊은 시간이었습니다. 올해 들어 한 일 중에서 가장 잘한 것을 뽑으라면 해외연수 프로그램에 참여한 것을 뽑을 것 같습니다. 절대 잊지 못한 경험이었습니다.

배소연 /전북대학교 기계시스템공학과

다양한 시각을 가진 사람들과 새로운 문화와 환경에서 깊이 있는 공부를 해 본 경험은 흔치 않을 것으로 생각해요.
네바다 교육프로그램 덕분에 이번 여름 방학이 4년간의 대학 생활 중 단연 특별한 시간이 되었습니다.

서수빈 /전북대학교 문헌정보학과

올해의 유일한 인문대생 참가자였습니다. 저는 심리학과 통계학을 복수전공하고 있고, 간혹 컴퓨터 공학 수업을 선택 이수한 경험이 있어 원활하게 참여할 수 있었습니다. 학교에서 이론으로 배운 내용을 실전에서 직접 적용할 수 있어 흥미로웠습니다. 뿐만 아니라 UNLV 교수님과 함께 하며, 한국에서는 접하지 못했던 새로운 연구 분야를 경험할 수 있어 특별했습니다.



연효진 /전북대학교 분자생물학과

새로운 곳에서 새로운 사람들과 함께 나아갔던, 쉽게 경험할 수 없는 값진 시간이었습니다. 해외 연구프로그램은 학습적 성취뿐만 아니라 제 개인의 성장에도 큰 영향을 주었습니다. 일상에서 벗어나 낯선 곳에서 생활하고 프로젝트를 진행하며 다양한 활동에 참여하는 도전적인 경험을 통해 다른 문화를 체험하고 자신의 사고를 다듬어 가는 소중한 시간이었습니다.

이용환 /전북대 컴퓨터인공지능학부

“Viva Las Vegas turnin' day into nighttime. Turnin' night into daytime If you see it once You'll never be the same again”
_ Viva Las Vegas, Elvis Presley

엘비스 프레슬리가 말한 것처럼 라스베이거스를 본 저의 삶은 이전과 달라졌습니다. 여러분께 저의 이야기를 말씀드리고자 합니다.

3 개인 감상문

Review

순번	소속	학년	이름
1	경기과학기술대학교 컴퓨터모바일융합공학과	3	**고진영**
2	경기과학기술대학교 컴퓨터모바일융합공학과	3	**김강현**
3	경기과학기술대학교 컴퓨터모바일융합공학과	3	**김도현**
4	경기과학기술대학교 인공지능학과	2	**김민선**
5	경기과학기술대학교 컴퓨터모바일융합공학과	3	**이재은**
6	경기과학기술대학교 컴퓨터모바일융합공학과	3	**함성영**
7	경상국립대학교 수학과	3	**강동현**
8	경상국립대학교 물리학과	4	**김가현**
9	경상국립대학교 산업시스템공학부	4	**박건영**
10	경상국립대학교 산업시스템공학부	4	**신서빈**
11	경상국립대학교 산업시스템공학부	4	**이인호**
12	경상국립대학교 정보통계학과	4	**이지상**
13	경상국립대학교 정보통계학과	4	**이태훈**
14	경상국립대학교 정보통계학과	4	**정인영**
15	경상국립대학교 산업시스템공학부	4	**정현서**
16	경상국립대학교 수학과	4	**최규진**
17	전북대학교 스마트팜학과	3	**강다영**
18	전북대학교 통계학과	4	**권지현**
19	전북대학교 컴퓨터인공지능학부	4	**김나영**
20	전북대학교 컴퓨터인공지능학부	4	**김미현**
21	전북대학교 소프트웨어공학과	4	**김재현**
22	전북대학교 소프트웨어공학과	3	**박세현**
23	전북대학교 기계시스템공학과	4	**배소연**
24	전북대학교 문헌정보학과	4	**서수빈**
25	전북대학교 분자생물학과	4	**연효진**
26	전북대학교 컴퓨터인공지능학부	3	**이용환**

고진영
경기과학기술대학교 컴퓨터모바일융합공학과

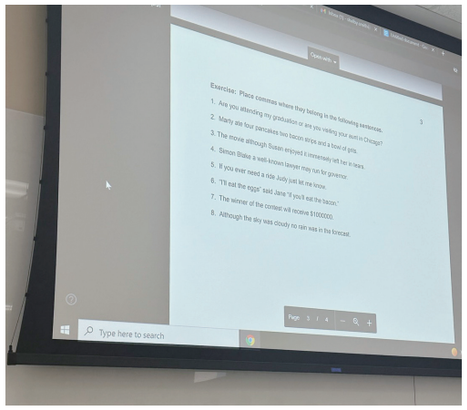
꿈 같았던 순간

빅데이터에 관심이 많던 나는 좋은 기회로 미국에 가게 되었다. 기대도 많이 됐지만 그만큼 걱정도 컸다. 대략 16시간의 비행을 통해 라스베이거스에 도착했다. 저녁에 도착하여 한국에서는 보지 못한 버스에 캐리어를 싣고 숙소로 이동하였다. 이동하는 동안 창밖의 풍경을 보며 내 가슴은 계속 방방 뛰고 설레는 마음으로 가득했다. 미국에서의 모든 순간이 새로웠고 신기했다.

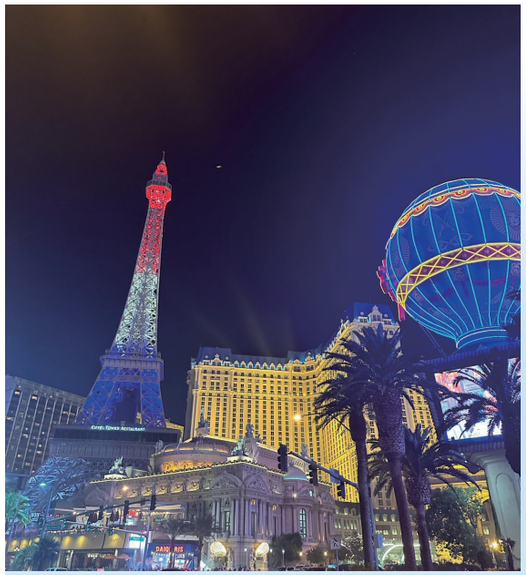
UNLV에서의 수업은 매우 흥미로웠다. 새로운 언어로 한 달 동안 공부한다는 생각에 걱정이 가득하였지만 같이 온 친구들이 많은 도움을 주었다. 내가 생각했던 것과 같이 수업은 자유로웠고, 학생들 대부분이 한국인이라 수업 분위기는 한국과 비슷하였다. 알아듣기 힘든 부분들도 많았지만, 교수님들께 질문을 하면 친절하게 답변해 주셔서 수업에 잘 참여할 수 있었다. UNLV 캠퍼스는 매우 컸고 특히 헬스장과 수영장 등 여러 시설 또한 너무 좋았다. 학교 점심시간에는 햄버거를 많이 먹으러 다녔다. 미국에서의 첫 햄버거는 짜릿하였다. 한국에서 나는 햄버거를 매우 좋아하는 학생이었다. 그래서 미국 본토의 햄버거 맛이 매우 궁금하였고 환상도 컸다. IN-N-OUT 햄버거를 먹었는데 한국에서의 상상을 충족시켜 주는 맛이었다. 매우 맛있었다. 단점은 조금 짜다는 것이다. 미국 음식은 대체로 많이 짰다.

한 달 동안 프로젝트를 진행하게 되었는데 우리 조의 주제는 '생존 분석'이었다. 생존 분석이란, 시간에 따른 사건 발생 여부를 분석하는 통계적 방법이다. 처음 프로젝트를 시작했을 땐 아주 어려웠다. 각자 많은 양의 논문을 찾아보며 팀원들끼리 서로의 정보를 공유하고, 토의하는 과정이 많았다. 또, 매주 발표 준비를 하다 보니 한 주 한 주가 금방 지나갔다.
여러 유전 정보를 받아와 새로운 데이터가 들어왔을 경우 환자의 생존과 사망을 예측하며 치유되는 데 걸리는 시간을 추정하기 위해 '콕스 비례 위험 모델'을 사용하였지만 아쉽게도 성공적인 하이퍼파라미터를 찾지 못했다. 하지만 팀원들과 함께 노력하는 과정은 나에게 뜻깊은 시간이었다.

여러 투어 프로그램 중 캐니언 투어가 가장 기억에 남는다. 캐니언까지 가는 모든 순간이 이쁘고 아름다워서 계속 사진만 찍었던 것 같다. 나는 자연을 보며 감탄이나 감동을 받은 적이 한 번도 없었는데 여러 캐니언과 홀스슈 밴드의 풍경은 지금까지도 기억에 남는다.



이번 프로그램으로 인해 식견이 넓어졌고 사고방식 또한 달라졌다.
네바다주립대학교에서 영어로 된 수업을 들으며 유익한 시간을 가졌고 새로운 환경과 새로운 사람들 속에서 나 자신을 다시 한번 돌아보며 부족한 부분을 채우는 좋은 계기가 되었다. 우리와는 다르게 자유로운 수업 환경이 매우 인상적이었고 미국의 노래와 매체, 뉴스로만 보던 모습들이 눈앞에 펼쳐지니 하루하루가 설레었다.



미국에서 다양한 경험을 하면서 시간이 아깝다는 생각을 처음으로 하게 되었다. 라스베이거스에서의 경험들로 앞으로 두려워하지 않고 도전하며, 세상을 바라보는 시야와 생각의 폭도 넓어진 것 같다.

빅데이터혁신융합대학사업의 해외 연구프로그램은 정말 좋은 기회였고 좋은 경험을 하게 해 주셔서 정말 감사하다는 말을 꼭 전하고 싶다.

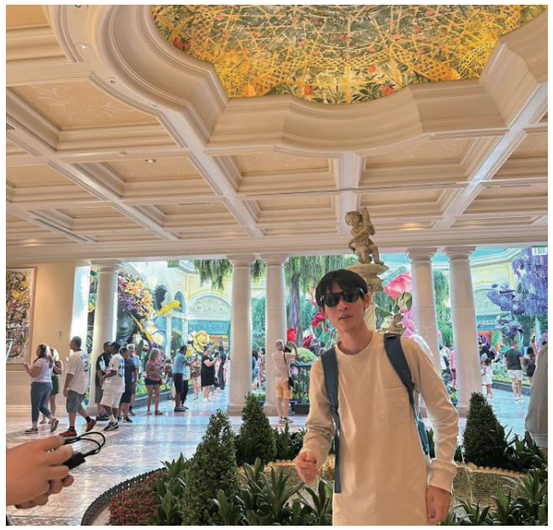
김강현
경기과학기술대학교 컴퓨터모바일융합공학과

라스베이거스에서 찾은 나의 꿈, 인공지능

처음에는 단순히 미국에서 공부하면 영어도 배울 수 있고 재밌겠다는 가벼운 마음으로 지원하게 되었다. 하지만 미국에서의 경험은 나의 이런 생각과는 전혀 달랐다. 결과적으로 UNLV에서 보낸 4주는 흔히 하는 해외 유학 경험이 아니라, 나의 인생 전환점을 찾아내는 중요한 시간이 되었다.

라스베이거스에서 한달을 보내며 상상 이상의 문화적 충격과 언어의 장벽을 느꼈다. 음식은 정말 한국인 입맛에 맞지 않게 너무 짜고 달았고, 영어는 우리가 학교에서 듣던 듣기 방송의 친절하고 부드러운 영어가 아니었다.
학교에서 라스베거스의 스트립이나 다운타운 등 유명 관광명소를 소개해 주었는데, 라스베이거스 특유의 분위기는 정말 좋았다. 밤의 도시라 불릴 정도로 밤과 낮의 분위기가 확연히 달랐는데, 밤에 본 스트립은 낮과 다르게 화려한 조명과 불빛으로 가득 차 있었

고 다양한 공연이 준비되어 있었다. 그중 어린 나이에도 불구하고 정말 프로처럼 드럼을 잘 치는 아이의 연주를 계속 듣고 있었던 기억이 난다.
수업에서 다룬 머신러닝과 딥러닝은 어려웠지만 상당히 특색있었고 재밌었다. 다양한 알고리즘을 프로젝트에 적용하는 과정에서 수없이 많은 난관에 부딪혔다. '뇌암 환자의 생존 분석에 대한 연구 프로젝트'는 처음 겪어보는 일이였기에 정말 큰 도전이었다. 학교에서 배운 이론과 실제 데이터 분석 사이의 큰 괴리가 너무 심했다. 데이터는 전혀 정제되지 않은 데이터였다. 유전자 데이터가 2만개나 되었는데 이렇게 많은 데이터를 만져본 건 처음이라 딥러닝에서 사용할 수 있는 데이터로 전처리하기는 너무 힘들었다. 딥러닝 모델 성공을 위해선 최적의 하이퍼 파라미터를 찾아야 했지만 우리는 데이터 전처리에서 너무 많은 시간을 사용해서, 하이퍼 파라미터 튜닝을 할 시간이 부족하여 계속 실패하게 되었으며 그것으로 인해 스트레스가 심했다. 그런 어려움 속에서도 포기하지 않고, 끈질기게 문제점을 찾아내려고 노력했다. 계속 시도하고 교수님께 조언을 구하며 도전했지만, 문제를 해결하지 못했다. 하지만 그 과정에서 문제를 해결하는 방법을 배우게 되었다.



이번 경험을 통해 문제를 해결하는 방법과 논문을 깊이 있게 읽고 분석하고 이해하는 방법, 그리고 팀원들과의 협업을 통해 큰 프로젝트를 성공적으로 마무리하는 방법을 배웠다. 이러한 경험들은 나에게 있어서 귀중한 자산이 될 것이라고 장담한다. 프로젝트를 통해 진로를 미리 체험해보며, 앞으로 해야 할 공부가 나의 적성에 맞다는 것 또한 알 수 있었고, 이로 인해 나의 인생 목표와 방향성을 확고히 잡을 수 있었다. 나는 대학원에서 인공지능을 전공하여 관련 분야의 전문가가 되고 싶다는 확고한 의지를 갖게 되었다

한 달 동안의 LasVegas, 장대했던 우리의 여정

김 도 현
경기과학기술대학교 컴퓨터모바일융합공학과

나는 언제나 호기심이 넘치고 새로운 도전을 즐기며 앞서 나가려 노력한다. 평범한 것을 싫어하며 영상 편집, 코딩, 봉사활동, 아르바이트와 같은 다양한 활동을 경험 해오고 있으며, 강한 추진력도 갖고 있다. 어느 날, 교수님께서 강의실 문을 여시며 "미국 갈 사람 있니?"라고 물으셨을 때, 나는 주저하지 않고 손을 들며 "저요!"라고 외쳤다. 이 기회를 놓치면 미국에 가볼 기회가 없을 것 같았다. 면접을 거쳐 최종 합격하여 라스베이거스로 떠날 수 있게 되었다.

라스베이거스 공항에 도착했을 때, 뜨거운 공기가 우리를 반겨주었다. 긴 비행으로 힘들게 도착한 라스베이거스는 정말 아름다웠다. 카지노의 도시에 걸맞은 명성답게 공항에서부터 펼쳐진 수많은 슬롯머신은 내가 미국에 있다는 사실을 실감나게 해주었다. 공항을 나와 Alexis Park Resort에 도착했을 때는 이미 새벽 1시였다. 10시간이 넘는 비행의 피로감은 서 있게 하는 것조차 힘들게 만들었다.

라스베이거스에 보낸 1달의 시간은 고난의 연속이었다. 제일 힘들었던 것은 언어가 아닌 음식이었는데, 입에 맞질 않아 직접 요리했다. 첫 주부터 요리를 하다 보니 어느새 다른 친구들의 끼니도 챙겨주는 나를 볼 수 있게 되었다. 맛있게 먹어준 친구들이 너무나 고마웠다!

한 달간의 수업에서 많은 것을 배웠다. 머신러닝과 딥러닝 분야에서 새로운 지식을 얻을 수 있는 좋은 기회였다. 데이터 분석, 모델 학습, 평가 등을 배우며 이 분야에 대한 이해가 높아졌다. 우리 팀이 하게 된 프로젝트는 이미지 기반의 암 사진 분류였다. 데이터 양이 너무 많아 조교 선생님이 전처리 하여 보내주기로 했는데, 예상보다 늦었고, 수십 수만 개의 데이터를 전부 분류하지 못하는 바람에, 프로젝트를 완성하지 못했다. 프로젝트를 성공하진 못했지만, 우리의 노력과 경험은 헛되지 않았다. 오히려 이런 실패는 미래의 성공을 위한 학습 경험이라고 생각하게 되었다. 프로젝트의 어려움을 극복하며 팀 협력과 문제 해결 능력을 좀 더 키워 이를 토대로 다음 프로젝트와 학습 과정에서 더 나은 결과를 만들어 내겠다고 다짐했다.

가장 기억에 남고 즐겼던 프로그램은 후버댐 투어이다. 후버댐 내부로 내려갔을 때, 거대한 수로 파이프가 우리 앞에 나타났다. 파이프 안의 물이 흐르고 있는 모습은 자연과 기술이 어떻게 공존하는지를 느끼게 해주었다. 장대했던 후버댐 투어를 통해 새로운 열정과 확신이 가득한 마음으로 UNLV로 돌아왔다.

한달동안 어려움이 많았지만, 그 어려움을 통해 더 많은 것을 배울 수 있었다. 이 기회를 주신 빅데이터 사업단과 UNLV 프로그램 관계자께 감사하며, 프로그램의 경험을 바탕으로 성장해 나가야겠다고 다짐한다.

김 민 선
경기과학기술대학교 인공지능학과

라스베이거스 UNLV에서의 시간은 나에게 뜻깊은 시간이었다. 캐나다 유학 생활에 이어 또다시 낯선 미국 땅에 발을 들였을 때, UNLV와 다양한 프로그램들에 대한 설레임으로 불안감을 별로 느끼지 못했다.

UNLV에서의 첫 주는 주로 캠퍼스를 돌아보며 학교생활에 적응하는 시간을 보냈다. 미국 대학 생활을 경험하고 전공 수업을 듣는 동시에 버디 프로그램과 라스베이거스를 즐길 수 있는 소중한 기회를 얻게 된 것은 큰 축복이었다. 라스베이거스의 밤 문화, 빛나는 스트립, 버디 프로그램으로 보낸 시간은 '이게 미국이구나' 하는 느낌을 받을 수 있었다.
라스베이거스에는 많은 호텔과 맛집, 그리고 여러 공연장이 많다. 미국의 문화를 체험하는 시간은 뜻깊었다. 이러한 경험을 통해 내면의 세계도 넓어지게 되었다. 한국에서 찰리푸스 티켓팅에 실패 했었는데, 라스베이거스에서 찰리푸스의 공연이 열려 운이 좋게 학교 친구들과 함께 콘서트를 관람할 수 있었다. 경쟁률이 낮고 공연장도 작아서 맨앞에서 관람했다. 미국에서 콘서트를 본다는 것은 색다른 경험이기도 하고 한국보다 저렴하다.

수업 중에서 가장 힘든 부분은 논문을 분석하는 것이었다. 논문을 처음 접하면서, 전부 영어로 쓰인 복잡한 내용을 이해하고 분석하는 과정은 정말 쉽지 않았다. 하지만 교수님과 동료 학생들의 도움으로 점차 이해할 수 있게 되었다. 프로젝트 수행에선 모든 것이 순조롭지 않았다. 하이퍼파라미터의 조절에 따른 딥러닝 모델의 성능 변화를 관찰하는 과제를 진행했는데, 여기서 여러 난관에 부딪혔다. 하이퍼파라미터의 세세한 조절과 검증 손실을 최소화하기 위하여 노력했고, 이것은 우리 팀에게 큰 도전이었다. 수업만 들었던 내가 직접 무언가를 도출하기 위해 논문을 찾고 방법을 연구하는 것은 내 인생의 가장 큰 도전이었으며, 이 프로젝트를 완성하고 싶은 욕구가 불타올랐다. 이러한 어려움은 결국 나를 더 성장할 수 있게 만들었다. 특히 팀원들과의 협업을 통해 문제 해결 능력과 커뮤니케이션 능력을 키울 수 있었다. 프로젝트의 결과물로 완벽한 성과를 내지 못했지만, 그 과정에서 얻은 지식과 경험은 앞으로의 나의 학문적 향방에 큰 도움이 될 것이라고 확신한다.

마지막으로, 이런 기회를 제공해준 빅데이터혁신융합대학사업단과 UNLV, 함께 고생한 팀원들, 그리고 지도해주신 교수님께 깊은 감사의 마음을 전하고 싶다. 이 경험은 나의 인생에서 잊지 못할 소중한 추억으로 남을 것이다.

이 재 은
경기과학기술대학교 컴퓨터모바일융합공학과

나의 전환점

교수님의 추천과 함께 “재은이는 당연히 갈 거지? 신청한다~” 라는 말을 뒤로, 급작스럽게 신청된 프로그램이었다. 학교의 대표라는 자리까지 도맡게 된 상태에서 새로운 사람을 만나 처음 가보는 타지에서의 생활은 꽤나 걱정스러웠다. 해외를 가는 것이 처음은 아니지만, 미국은 항상 꿈에만 그리던 곳이었으니 더 긴장되었다. 영어로 수업을 진행한다고 하는데 내가 과연 잘 해낼 수 있을지도 고민했다. 물론 그런 고민보다 ‘내가 미국을 간다!’라는 두근거림이 더더욱 컸다. 내가 생각한 대로 자유로운 분위기일까? 내가 보던 미국 만화들 같은 문화일까? 나의 라스베이거스 생활은 즐거운 상상과 함께 시작되었다.

경기과학기술대학교와 다른 학교의 가장 큰 차이점은 바로 이번 프로그램을 함께한 학생 수가 적었다는 점이다. 처음에는 너무 적은 것이 아닌가 생각했지만, 그것이 큰 장점이었다는 것을 깨닫는 데에는 시간이 오래 걸리지 않았다. 6명이라는 인원수는 우버에 딱 맞는 인원수였고 덕분에 우리는 어디를 가던 6명이서 함께 다니고 함께 즐기며 돈독한 사이로 지낼 수 있었다. 또 인원수가 적은 덕분에 외부의 다른 프로그램으로 함께 참가하게 된 다른 대학생과 같은 조로 묶여 함께 다니기도 했다. 미국생활 경험이 많은 학생은 정말 유용한 정보를 많이 알려 주었다. 가장 좋았던 팁은 찰리푸스의 공연이 라스베이거스에서 열린다는 소식을 전달받고 한국에서는 꿈도 못 꿨을 앞에서 15번째 자리에서 찰리푸스의 공연을 직관했다는 것이다. 독립기념일을 앞두고 시먼 아울렛에서 굉장한 세일을 한다는 소식도 그 학생에게 전해 들어서 다른 학교의 사람들에 비해 다양한 정보로 미국을 즐길 수 있었다.

우리 팀의 프로젝트 주제는 세포 이미지를 분석하여 암세포 여부를 구별해내는 것이었다. 항상 이미지를 분석하는 프로젝트에 관심이 있었기에 나는 고민 없이 선택할 수 있었다.
수업은 진심으로 재미있었다! 솔직히 수업이 재미있어 수업 시간을 기다리기도 했다. 학교에서는 배울 수 없었던 기본적이고 기초적인 개념을 다지고 배워갔다. 개념이 확실하게 잡히니 학교에서는 그저 외우기만 했던 부분들이 이해되고 쌓여나가기 시작했다. 물론 영어로 수업을 들으며 막히는 부분도 있었다. 하지만 계속해서 질문하고, 더 집중하며 수업 내용을 하나씩 덧붙여갔다. 라스베이거스에서 정말 열심히 공부한 덕에 한국에서 가져간 노트를 전부 채워버려 새로운 공책을 마련하기도 했다.
학우들과 함께 모여 밥을 먹고 함께 밤을 새우며 프로젝트를 진행하면서 소중한 인연을 만들어 나갔다는 점이 나에게는 가장 행복했다. 인생에서 한

번쯤은 꼭 가야 한다는 그랜드 캐니언을 함께 본 친구들을 쉽게 잊을 수는 없을 것이다. 사진을 찍어주고 관광지를 다니고, 자이언 국립공원에서 수영도 하며 생긴 우정은 끈끈하다. 이 프로그램으로 얻은 가장 큰 것은 이 소중한 인연들이었다.

이 프로그램에 참여하기 전에는 흥미는 있었지만 내가 정말 잘 해낼 수 있을지에 대한 막연한 두려움으로 준비를 망설이곤 했다. 하지만 이번 해외 연구프로그램를 기반으로 마음을 다잡았다. 내가 정말로 이 분야에 대해 더 깊이 공부하는 것을 즐길 수 있는 모습을 보고 대학원과 함께 나의 미래를 설계하고 있다. 이 프로그램으로 얻은 또 다른 감사한 점은 바로 나의 미래에 대해 깊이 고민하고 다짐하게 해 주었다는 점이었다. 더욱 정진하고 앞으로 나아가고 싶다. 또다시 라스베이거스에, UNLV에 갈 때, 전문가가 되어 멋진 발전된 모습으로 발을 딛고 싶다. 나에게 이런 포부를 품게 하며 내 인생의 전환점을 마련하게 한 이 프로그램에 참여하게 되어 참 감사하다.

함성영
경기과학기술대학교 컴퓨터모바일융합공학과

빅데이터 해외 연구프로그램을 통하여 다양한 경험을 할 수 있었다. 처음에는 외부 활동과 강의로 바쁘게 채워진 생활에 적응하는 데 어려움을 겪었다. 그러나 시간이 지날수록 새로운 친구들과 함께하는 활동과 공부는 나에게 큰 즐거움을 주었다.

국립공원에 방문한 것은 내게 큰 감동을 주었다. 자이언 캐니언과 홀슈밴드, 앤틸롭 캐니언, 그랜드 캐니언의 아름다움과 웅장함을 직접 경험하면서, 환경 보호와 지속 가능한 관리의 중요성을 깨닫게 되었다. 아침 일찍부터 돌아다녔는데도 피곤함을 잊을 만큼 아름다운 곳이었다. '죽기 전에 꼭 가봐야 하는 장소로 꼽는 이유가 있구나!'라는 생각이 들었다. 미국의 자연과 문화를 직접 체험하면서 새로운 시각을 얻을 수 있었다.

미국의 스포츠하면 가장 먼저 농구가 떠오른다. 미국에는 NBA가 있기 때문이다. 라스베이거스에서 NBA 경기를 볼 수 있는 기회가 있었는데, 너무 즐거웠다. 사람들의 열기가 완전 뜨거웠고 즐기는 모습들을 보며 나 또한 열심히 호응하며 즐기게 됐다. 라스베이거스에서 또 하나의 공연을 봤는데 바로 찰리푸스 콘서트였다. 거의 맨 앞에서 볼 수 있었고 콘서트를 처음 봤는데 보길 잘했다는 생각이 들었다. 미국의 콘서트 문화를 볼 수 있는 좋은 기회였다.

우리 팀의 프로젝트는 암 예후 예측을 위한 딥러닝 모델을 개발 것이었다. 이 프로젝트를 위해 논문을 조사하고 실제 데이터를 사용하여 모델을 구축하는 과정은 매우 도전적이었고, 함께 노력하는 팀원들과의 협력을 통해 성공적으로 마무리할 수 있었다. 프로그램에서는 다양한 주제의 교육 강의를 듣고, 팀 프로젝트를 수행하면서 실무 지식을 배웠다. 머신러닝과 딥러닝 교육이 가장 인상 깊었고, 실제 응용과 코딩 능력을 향상할 수 있었다. 이런 교육이 나중에 진로를 선택할 때 큰 도움이 될 것 같다. 아쉬웠던 점은 프로그램 일정이 밀려있어 더 많은 시간을 각 주제에 충분히 할애하지 못했다는 점이다. 프로그램 내에서 추가적인 심화 교육의 기회를 제공한다면 더욱 풍부한 학습 경험을 얻을 수 있을 것으로 생각한다.

프로그램 참가로 머신러닝과 딥러닝 분야에 대한 지식과 실무 경험을 쌓을 수 있었다. 이를 토대로 앞으로는 이 분야에서의 직무 경력을 쌓고, 프로그램을 통해 만난 새로운 친구들과 협력하며, 앞으로의 미래를 준비하고자 한다. 이 프로그램의 다양한 활동과 교육은 내게 귀중한 경험과 지식을 줬다. 특히, 머신러닝과 딥러닝 수업을 통해 새로운 기술과 방법론을 배우고 이를 실제 상황에 적용하는 기회를 얻었다. 또한, 영어로 의사소통하는 능력이 향상되었다. 초반에는 어려움을 겪었지만, 시간이 지남에 따라 더욱 능숙하게 의사를 표현하고 이해할 수 있었다. 이러한 언어 능력은 국제적인 프로젝트나 협업에서 중요한 역할을 할 것이다.

이 프로그램은 내게 큰 영감과 동기부여를 주었으며, 앞으로의 성장과 미래를 준비하는 데 큰 역할을 할 것이다.

강 동 현
경상국립대학교 수학과

사고(思考)가 확장되는 시간

정해진 일만 반복하는 것이 너무 지루했다. 그래서 색다른 변화를 찾던 와중에 이번 프로그램을 접했다. 새로운 환경에서 전문적인 것들을 배우기를 기대하며 프로그램에 신청했다.

보통 머신러닝을 공부하면 어떤 것들이 있고 어떻게 사용하는지만 공부한다. 하지만 학교에서는 이것들을 수학적으로 다루고 수업을 진행했다. 수식들을 보면서 의미를 이해하고 직접 코드를 짜면서 정말 구체적으로 배우고 있다고 생각했다. 선형함수로 이진 분류를 하는 수업이 있었다. 이진 분류 각각의 label 값들의 절댓값의 차이가 작을수록 분류의 성능이 높아지는 경향이 있었다. 왜 이런 현상이 일어나는지 고민을 했지만, 이유는 알 수 없었다. 교수님께 여쭌 후 내용을 이해할 수 있었다. 선형함수를 이용해 역치를 기준으로 분류하다 보니 선을 긋기 좋은 값일수록 분류를 잘한다고 했다. 어떤 의미인지 이해하고 보니 당연했다. 만약 내가 고민을 계속했다면 한 달은 고민만 하고 있었을 것 같았다. 교수님이 명쾌하게 답을 주시는 것을 보고 확실히 내가 여기 와서 조금씩이라도 배우고 있다는 게 실감이 났고 그것이 너무 좋았다.

영어로 대화하는 것은 너무 힘들었다. 영어 성적이 좋지도 못하고 말하기는 더 못했다. 하지만 밥은 먹어야 하고 한인 타운이 있는 것도 아니기 때문에 어쩔 수 없이 가게에 들어갔다. 줄을 서면서 앞에 있는 사람을 유심히 관찰했다. 어떻게 주문하는지, 어떤 말을 쓰는지를 듣고 준비하고 있었지만, 인생은 실전이었다. 내 차례가 되니 머리는 하얘지고 아무것도 기억나지 않았다. 그래도 태연하게 손으로 가리키며 어떻게든 주문했다. 식당 점원이 내가 어떤 말을 하는지 이해하지 못해 반문하는 순간 말문이 막혔다. 하지만 이런 낯선 상황도 지속해서 노출되다 보니 익숙해졌다. 정확한 영어는 아니더라도 내 의사를 전달하는 방법을 터득했다. 영어에 익숙해진 탓인지 영어 논문을 읽는 것도 수월해졌다. 그래서 왠지 모를 자신감마저 들었다.

어디선가 이런 이야기를 들었던 것 같다. 사고(思考)가 확장되는 순간은 새로운 변화에 적응하는 순간이라고. 이번 한 달이 그런 시간이었다.

김 가 현
경상국립대학교 물리학과

사막의 열기만큼 뜨거웠던 4주동안의 연수

대학생 신분으로 국외 배낭여행 혹은 교환학생 등 해외로 나가서 새로운 문화를 접하고 배우는 것과 자신을 부각할 수 있는 역량을 기르는 것, 둘 다 많은 학생들이 바라고 있지만, 일정상 혹은 현실상의 문제로 양립하기는 어려움이 있다. 그러나 이번 빅데이터사업단 네바다주 연수는 이 두 경험을 학교로 동시에 가능하게 하여 나처럼 데이터 연구 및 분석 분야를 지망하는 학생들에게 해외문화를 접하고, 최신 연구 동향을 익히고 기술을 배울 수 있는 기회가 되었다.

방학 4주라는 긴 시간 동안 네바다주립대학교(UNLV)의 시설을 이용하며, 현재 가장 대두되고 있는 빅데이터와 인공지능 분야의 이론을 익히고 연습하며, 직접 연구프로젝트를 수행할 수 있었던 것은 값진 경험이었다. 실제 현재 연구 대상이 되는 데이터를 접하고 현직 연구자들의 피드백을 받을 수 있었으며 저널 검색 혹은 연구방법론, 과학 에세이 작성 방법 등 다양한 스킬을 배웠다. 또한 배움의 기회뿐만 아니라, 휴식과 여가로 구성된 다양한 관광 및 체험기획 또한 머나먼 곳에서 경험을 쌓는 좋은 기회가 되었다.

UNLV 2023 Summer Research Program은 크게 두 가지의 수업으로 이루어져 있다. 하나는 빅데이터와 AI에 대해서 전반적인 지식을 배우고 PBL 수업을 통해 연구를 수행해보는 'Big Data & AI'이며, 나머지는 과학 에세이 작성법에 대해서 배우고 실습하는 'EGG 499'으로써 일주일에 Big Data & AI 10차시, EGG 499 2차시가 배정되었다. 오전에 진행된 이론강의는 전반적으로 머신러닝 등의 기계학습에 대해서 이론적 기반을 다지며 수식적으로 통계학습 및 분석에 대해서 배울 수 있었으며, 오후수업은 초기에는 파이썬 실습을 수행했고, 후반에는 팀 프로젝트 작업이 이루어졌다.

나는 우리 학교 수학과 두 명의 학생들과 한 팀을 이루었는데, image segmentation을 주제로 road segmentation에 대하여 같은 데이터로 각자 기계학습을 통하여 모델을 제작하는 것을 계획했으며, 둘째 주 피드백에서 모두가 하나의 모델을 형성하는 것 대신, supervised learning, unsupervised learning, semi-supervised learning을 통하여 모델을 제작하는 것을 권유받았다. 내부 논의 끝에 나는 supervised learning을 통하여 모델링하기로 하였으며 팀원들끼리 서로 피드백 및 현황을 오후 일과 뒤에 도서관에서 논의하며 프로젝트를 수행했다. 이후 연구와 학습을 지속하며 시행된 결과물을 확인했으며, 연구 활동을 수행할 때 필요한 정보를 얻는 방법을 배웠고, 수업을 바탕으로 더욱 효율적으로 프로젝트를 진행해 나갔다. 이러한 노력에 힘입어 마지막 주에 우리 학교와 경기과학기술대학교의 빅데이터사업단 단장님께서 참여하신 최종발표회에서 성공적으로 결과물을 발표할 수 있었다.

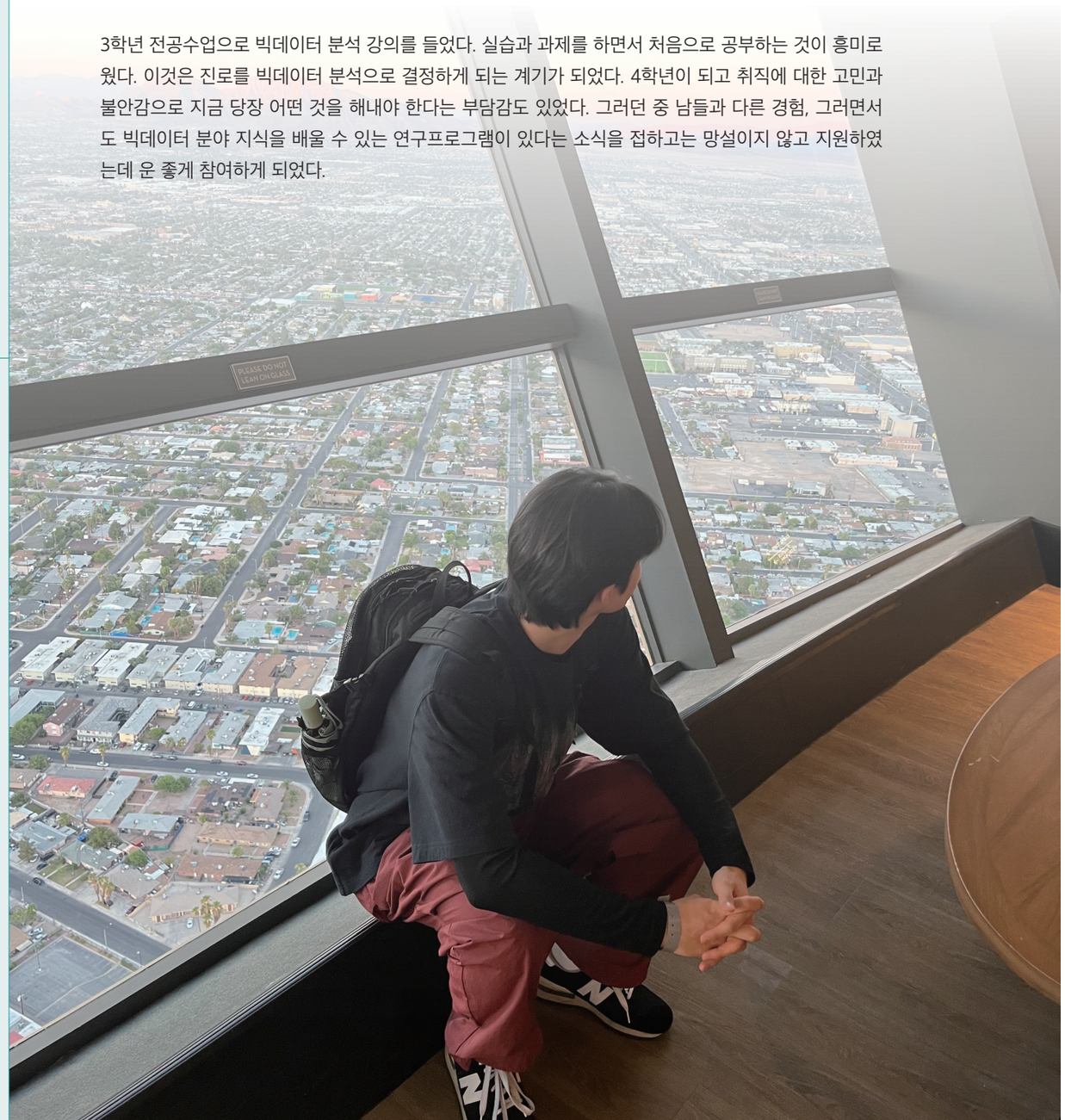
4주간의 연수를 돌이켜볼 때, 선발 학생들의 빅데이터와 인공지능 분야 수강 시간 차이 등으로 인해 이론 학습 및 팀 프로젝트 진행이 다소 매끄럽지 못했다. 또한 홍보가 있었음에도 각 학교의 대다수 학생이 이번 연수 프로그램에 대해 알지 못했다는 점과 프로그램과 지원 기간을 인지한 학생들도 제약조건으로 인해 지원을 주저했다는 사실이 안타까웠다.

학기 중에 시도했던 작은 용기가 이번 여름 동안 잊을 수 없는 경험과 기억으로 돌아왔다. 내년에는 더 많은 학생들이 참여하여 뜻깊은 경험을 함께 하길 적극 추천한다.

박건영
경상국립대학교 산업시스템공학부

모든 것이 미국 서려웠다

3학년 전공수업으로 빅데이터 분석 강의를 들었다. 실습과 과제를 하면서 처음으로 공부하는 것이 흥미로웠다. 이것은 진로를 빅데이터 분석으로 결정하게 되는 계기가 되었다. 4학년이 되고 취직에 대한 고민과 불안감으로 지금 당장 어떤 것을 해내야 한다는 부담감도 있었다. 그러던 중 남들과 다른 경험, 그러면서도 빅데이터 분야 지식을 배울 수 있는 연구프로그램이 있다는 소식을 접하고는 망설이지 않고 지원하였는데 운 좋게 참여하게 되었다.



인천에서 출발하여 약 22시간 만에 라스베이거스에 도착했다. 입국심사를 마치고 공항 문을 나가는 순간 한국에서 겪어보지 못한 건조함과 열풍이 아직도 생생하게 기억난다. 현지 시각 23시, 온도는 섭씨 40도였다. 숙소에 도착하여 짐을 풀고 잠이 들었다. 시차 적응도 제대로 하지 못해 비몽사몽인 상태로 학교까지 걸어갔다. 가로수 대신 선인장이 있는 낯선 풍경이 사막에 있다는 것을 상기시켜 주었다.

우리는 네바다주립대학교(UNLV)에서 4주간 영어로 진행되는 빅데이터 이론 수업, 실습, EGG 수업에 참여했다.

빅데이터 이론 수업은 인도계 교수님이 진행하였고 배운 것을 바탕으로 오후에는 실습을 진행했다. 이 수업 시간에 배운 것을 토대로 팀을 구성하여 프로젝트를 진행하였는데 우리 팀이 진행했던 연구 주제는 “뇌종양 생존분석”이었다. 4주간 총 2번의 중간 발표와 마지막 발표를 하였고, 발표를 통해 연구의 방향성이라든지 기술적인 부분을 질의응답 하여 만족할 만한 결과로 연구를 끝마칠 수 있었다.

EGG 수업은 영어로 논문을 쓰기 위한 문법이라든지 논리적 글의 흐름을 배우는 수업이었다. 이 수업에서는 프랑스, 독일계 학생도 포함되어 있었고 캐나다계 교수님이 강의하셨는데, 앞서 언급한 각 팀의 연구 주제로 마지막 주차에 논문을 써보는 것으로 수업은 마무리되었다.

학습활동 외에도 버디 프로그램을 통해 또래 외국 친구들과 놀러 다니며 20대 미국 문화와 그들의 생활을 좀 더 가까이서 체험할 수 있었고, 1박 2일간 진행되었던 자이언 캐니언, 홀스슈 밴드, 엔텔로프 캐니언, 그리고 그랜드 캐니언 투어를 통해 처음으로 자연에 압도되는 경험도 하였다. 마지막으로 후버댐 투어를 통해 80년 전에 그와 같은 거대한 댐을 건설할 수 있었던 미국의 자본력과 추진력에 경외감마저 들었다.

이번 프로그램을 통해 학문적인 성장과 견문을 넓히고 무엇보다 걱정되었던 영어 회화에 자신감이 붙었다. 또한 같은 분야에 흥미를 느끼는 사람들과의 생활을 통해 부족한 점들을 깨닫게 되었으며, 그 부분을 채우기 위해 열심히 학업을 이어가고 있다.

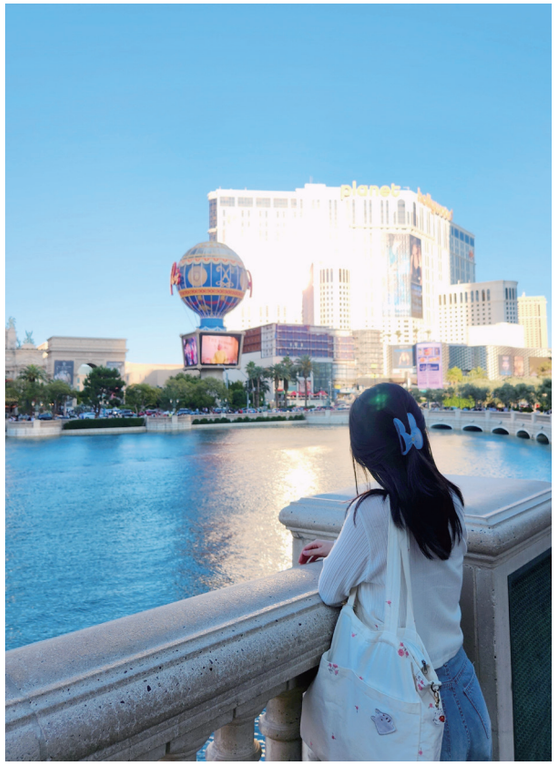
마지막으로 프로그램을 기획하고, 전폭적으로 지원해 주신 사업단에 감사드립니다.

신 서 빈
경상국립대학교 산업시스템공학부

내 인생의 한 페이지가 될 수 있게

하계 빅데이터 해외 연구프로그램에 참여하여 미국에서 공부하며 생활하는 기회가 생겼다. 외국을 가는 것은 처음인데 한 달 동안 머물러야 하는 일정이기 때문에 신청하기까지 많은 고민을 하였다. 영어를 능숙하게 구사하는 편이 아니었고, 그곳의 문화와 환경에 대한 이해도 부족했기 때문에 막연한 두려움과 불확실함이 있었다. 하지만 대학 생활 중 네바다주립대학교에서 수업을 들으며 빅데이터 프로젝트를 진행하는 것은 나의 역량을 키우고 발전할 다시없을 기회라 생각하였다. 또한 지금까지 살아왔던 환경에서 벗어나 새로운 환경에서 도전하고 배우며 성장할 기회를 놓칠 수 없다고 생각해 결국 신청하게 되었다.

미국에 도착하였을 때, 길거리의 풍경부터 사람들의 모습, 심지어 표지판과 신호등까지 모든 것이 낯설었다. 하지만 이 모든 것을 경험으로 받아들이고 잘 해낼 수 있을 것이라 다짐을 하며 적응하는 데 노력하였다. 4주 동안 라스베이거스에 머무르며 대학교에서 수업을 듣고 프로젝트를 진행하였다. 수업 이외의 시간에는 자유롭게 사람들과 어울려서 놀러 다니기도 하며 많은 경험과 추억을 쌓았다.



네바다주립대학교에서 들었던 교육 중 'EGG499' 수업이 많은 도움이 되었다. EGG 수업은 대학교 1학년 교양으로 들었던 글쓰기 수업을 영어로 듣는 것과 같은 느낌이었다. 이 수업을 통하여 영어로 작문하는 능력을 향상할 수 있었으며, 앞으로 프로젝트를 진행하거나 논문을 조사하고 내용을 정리해야 할 일이 있을 때 유용하게 활용할 수 있을 것으로 생각한다. 또한 'Big Data & AI' 수업은 산업공학 전공 수업과 데이터 사이언스 복수 전공을 통하여 배운 내용과 중복되는 내용이 많아서 비슷한 느낌이었다. 그동안 배웠던 개념들을 통합적으로 학습할 수 있어서 전체적인 흐름을 체계적으로 이해하고 상기시키는 데 큰 도움이 되었다.

기억에 남는 순간 중 하나는 미국의 독립기념일이었다. 우연히 날짜가 맞아 프로그램 도중 독립기념일을 보낼 수 있었다. 독립기념일은 미국 독립 선언이 채택된 것을 기념하는 날로, 공휴일이라 뜻밖의 자유 시간을 즐기기 위하여 동료들과 함께 Strip을 돌아다니며 구경하였다. 뉴욕뉴욕 호텔 건물 위에 있는 롤러코스터를 탔으며, 미라지 호텔의 'Volcano'라고 불리는 화산 쇼를 관람하기도 하였다. 이후에는 라스베이거스에서 독립기념일을 기념하여 열리는 가장 큰 이벤트인 불꽃놀이를 구경하였다. 불꽃놀이를 제대로 구경하기 위해 좋은 자리를 찾아가려 하였으나 거리는 사람들로 꽉 차서 발 딛는 것조차 힘들었다. 할 수 없이 제자리에서 지켜볼 수밖에 없었는데, 시야에 제한이 있었지만 불꽃놀이의 규모가 상당히 커서 구경하는 재미가 있었다. 사람들과 섞여서 구경하는 것도 낭만적이었다. 온종일 돌아다니기도 하였고 인파 속에서 움직이기 어려워 숙소까지 돌아오기도 힘들었다. 하지만 사람들에 섞여 라스베이거스의 분위기와 매력을 온전히 느낄 수 있었고, 다양한 경험을 하여 좋은 추억으로 남은 하루였다.

다음으로, Field Trip으로 국립공원 투어를 갔던 것이 잊지 못할 순간이었다.

다녀왔던 국립공원은 라스베이거스가 위치한 네바다주가 아닌 유타주, 애리조나주에 있었기 때문에 개인적으로 방문하기 어려웠으나, 프로그램으로 진행하여 가이드분의 안내를 받으며 모두 투어에 참여할 수 있었다. Zion National Park, Horseshoe bend, Antelope Canyon, Grand Canyon을 방문하였다. 이곳에서 캐년들을 구경하며 사진을 남기고, 트래킹, 물놀이 등 다양한 액티비티도 즐겼다. 이 투어를 통하여 계속된 프로젝트와 수업들로 인하여 쌓인 피로를 풀었다. 그뿐만 아니라 자연 속에 담긴 시간의 흐름과 신비로움을 몸소 느끼며, 투어를 오기 전에 했던 기대를 보란 듯이 뛰어넘어 인상 깊고 소중한 경험으로 남았다.

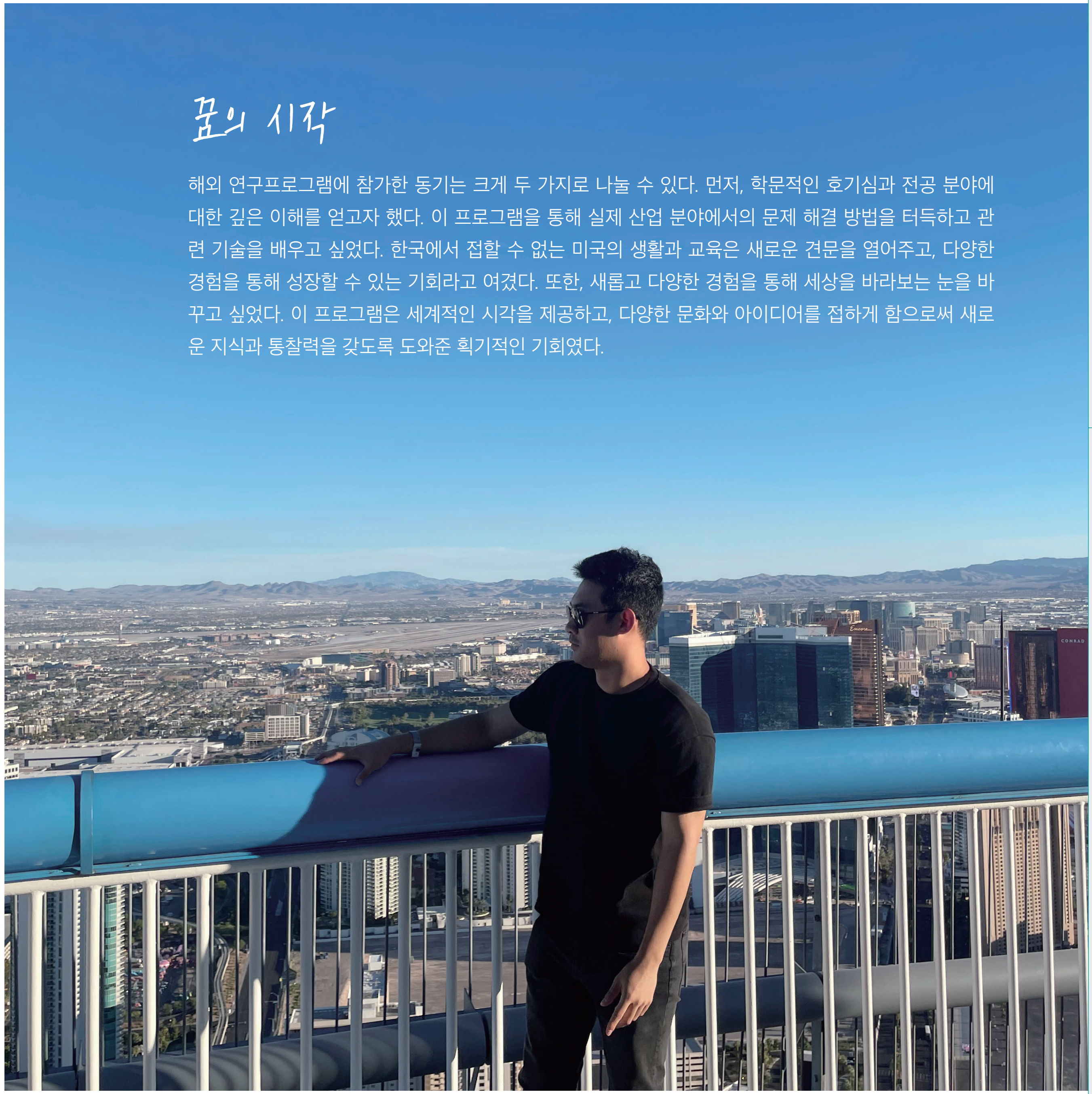
이번 하계 빅데이터 해외 연구프로그램은 굉장히 유익하고 귀중한 경험을 안겨 주었다. 네바다주립대학교에서의 수업은 데이터 분석과 머신러닝 기술의 역량을 향상할 수 있었고, 학술적 글쓰기도 배울 수 있었다. 또한 프로젝트를 진행하며 팀원들과의 협력이 중요하다는 것을 깨달았으며, 문제를 해결하는 능력을 키울 수 있었다. 한 달 동안 현지에 녹아들어 사람들을 접하고 교류를 하며 다양한 문화와 사고방식을 이해하고 수용할 수 있는 마음을 가지게 되었다.

아쉬웠던 점이 있다면, 실제 사용되는 영어 어휘와 발음을 이해하고 구사하는 데 어려움을 겪었다는 것이

다. 4주 동안 프로그램에 참여하며 영어로 발표를 하고 현지인들과 대화를 나누면서 영어에 익숙해졌으나, 회화를 더 많이 공부하고 갔으면 이 기회를 효과적으로 활용할 수 있었을 것이라는 생각에 아쉬움이 남았다. 또한 네바다 주립 대학교에서 공부하며 더 넓은 세상에 대한 열망이 강해졌고, 데이터 사이언티스트로서의 진로를 위하여 영어로 교류하고 기술을 습득하는 것이 필수라는 것을 깨닫게 되었다. 이로 인하여 영어 회화 공부의 필요성을 강하게 느껴 프로그램 이후 우리나라에 돌아와서도 영어에 소홀하지 않고 실력을 더욱 높이기 위하여 노력하고 있다. 정제되지 않은 일상 영어를 팟캐스트, 유튜브 등에서 꾸준히 듣고 있으며, OPIc 고득점 획득을 목표로 회화 연습도 하고 있다.

한 달이라는 시간은 짧았지만, 프로그램에 참여하며 쌓은 경험은 나에게 자신감과 도전 정신을 크게 키워주었다. 이번 경험을 통해 배운 것들과 목표를 통해 나의 부족한 부분을 발견하고 열심히 노력하여 개선하고 발전시킬 것임을 다짐하였다.

이 인 호
경상국립대학교 산업시스템공학부



꿈의 시작

해외 연구프로그램에 참가한 동기는 크게 두 가지로 나눌 수 있다. 먼저, 학문적인 호기심과 전공 분야에 대한 깊은 이해를 얻고자 했다. 이 프로그램을 통해 실제 산업 분야에서의 문제 해결 방법을 터득하고 관련 기술을 배우고 싶었다. 한국에서 접할 수 없는 미국의 생활과 교육은 새로운 견문을 열어주고, 다양한 경험을 통해 성장할 수 있는 기회라고 여겼다. 또한, 새롭고 다양한 경험을 통해 세상을 바라보는 눈을 바꾸고 싶었다. 이 프로그램은 세계적인 시각을 제공하고, 다양한 문화와 아이디어를 접하게 함으로써 새로운 지식과 통찰력을 갖도록 도와준 획기적인 기회였다.

이 지 상
경상국립대학교 정보통계학과

라스베이거스에서의 학문과 문화 체험

빅데이터와 딥러닝이 미래산업에 큰 영향을 줄 것이라는 생각에, UNLV 빅데이터 해외 연구프로그램에 참가하기로 마음먹었다. 이번 기회를 통해 라스베이거스와 미국 서부의 문화, 학술 분위기, 그리고 뛰어난 연구자들과의 교류를 체험할 수 있었다.

프로그램 초반, 나는 영어로의 의사소통에 상당한 어려움을 느꼈다. 특히, 프로젝트에 관한 발표를 진행할 때, 내 생각을 영어로 정확하게 전달하기가 쉽지 않았다. 그럼에도 불구하고, 열심히 준비하고 같이 참여한 친구의 도움을 받으면서 그 난관을 함께 극복하였다. 이러한 경험은 나에게 영어에 대한 두려움을 없애주었으며, 다양한 배경을 가진 사람들과 협업하는 법도 배울 수 있었다.

프로그램 동안 한국에서 하기 힘든 경험을 쌓았다. 라스베이거스의 날씨부터 생활은 낯설지만, 점차 적응해 나갔다. 겪어보지 못한 새로운 문화와 환경은 처음 느껴보는 신선함이었다. 1달간의 미국 대학 생활은 새로운 수업 방식을 경험할 수 있는 기회를 주었고 학문적 이론과 실전을 동시에 접할 수 있는 기회였다.

프로그램에서 진행한 프로젝트는 'Deep-Learning Based Survival Analysis Using Genomic Data'이다. 해당 프로젝트는 환자 데이터와 유전 정보를 통해 생존 분석을 수행하는 것이다. 해당 프로젝트는 사전에 접해본 적이 없다. 꽤나 도전적인 작업이다. 수행 과정에서 영어로 적힌 논문을 읽고 팀원들과 협력하여 문제를 해결하는 과정이 훗날 비슷한 문제 해결에 중요한 경험이 될 것임을 확신했고 이를 통해 어려운 상황에서도 문제를 해결하고 협력할 수 있는 능력을 키워나갈 수 있을 것으로 기대된다.

프로그램에서 기억에 크게 남을만한 활동은 미국의 국립 공원을 탐방한 경험이다. 그랜드 캐니언과 후버 댐의 아름다움은 어떤 단어로도 형용할 수 없는 광경이었다. 한눈에 들어오지 않는 풍경을 통해 자연의 광활함을 느낄 수 있었다. 이 경험은 자연환경의 소중함과 보호의 중요성을 더욱 강조해주었다.

프로그램 참가 전후로 가장 크게 달라진 것은 자신감이다. 짧다면 짧고, 길다면 긴 1달 동안의 미국 생활은 다른 문화와 사람들을 이해하고 존중하는 시야를 넓히는 소중한 시간이었다. 앞으로 이러한 경험을 바탕으로 국제적인 경력을 쌓고, 글로벌한 환경에서 업적을 이루어내고자 하는 꿈을 꾸게 되었다.

미국 해외 연구프로그램은 삶과 경험에 새로운 장을 열어주었다. 더 큰 꿈을 향해 나아가기 위한 첫걸음이었다. 프로그램을 통해 세상은 결코 좁지 않으며, 아는 것이 많이 부족하다는 것을 알게 되었다. 이번 기회를 발판 삼아 새로운 꿈을 키우고 더욱 발전시켜나가고자 한다.

연구의 주제인 '암 환자 유전자 데이터 분석'은 나에게 큰 도전이었다. 이론적인 지식을 넘어 실제 데이터와 마주하면서, 많은 시행착오와 실험을 거쳐야 했다. 나와 팀원들은 수많은 데이터 중에서 의미 있는 정보를 추출하고, 그것을 바탕으로 신뢰할 수 있는 예측 모델을 만들기 위해 노력하였다. 이 과정에서 혼자서는 불가능해 보이는 일도 팀원들의 도움과 협력이 있다면 극복할 수 있다는 것을 느낄 수 있었다.

프로그램의 일정은 상당히 알찼다. 매일매일 연구와 발표 준비로 바쁘게 지냈으며, 라스베이거스 스트립 거리와 국립공원 방문, 후버댐 탐방과 같은 여러 투어 프로그램을 통해 미국의 역사와 문화를 체험할 수 있었다.

프로그램 참가 전과 후의 내 모습은 확연히 다르다. 이전에는 이론 위주의 학문적 지식만을 추구했다면, 이제는 현장에서의 경험과 실제 문제 해결 능력을 중요하게 생각하게 되었다. 앞으로 이 경험을 바탕으로 빅데이터와 딥러닝 연구를 열심히 하려고 한다.

결론적으로, '빅데이터 해외 연구프로그램'은 나의 인생에서 빼놓을 수 없는 중요한 경험이었고, 내가 얻은 지식, 경험, 그리고 소중한 추억은 앞으로의 학술적 연구와 삶의 방향에 큰 도움이 될 것이라 믿는다. 마지막으로 이러한 기회를 주신 경상국립대학교 빅데이터 혁신융합대학 사업단 관계자분들께 진심으로 감사의 인사를 드리며 글을 마무리한다.

이 태 훈
경상국립대학교 정보통계학과

새로운 경험과 행복했던 기억

군 복학 후 선배의 권유로 'Python'이라는 프로그래밍 언어를 배웠다. 이후 데이터 분석, AI에 대한 공부를 하였다. 그리고 다양한 공모전과 연구를 하던 중 '하계 빅데이터 연구프로그램'을 알게 되었다. '미국'에서 연구프로그램이 진행되며 프로그램에서 소개된 프로젝트들이 매우 흥미로웠다. 평소에 빅데이터와 AI에 많은 관심이 있었기 때문에 프로그램 참가 신청을 한 결과, 최종 선발이 되었다. 미국이라는 나라가 더 낯설게 느껴져 걱정 반 설렘 반으로 프로그램을 준비했다.

해리 리드 공항에 도착했을 때부터 긴장할 수밖에 없었다. 입국심사를 받던 중 세컨더리룸(2차 입국조사실)으로 끌려갔다. 그곳에서 모든 짐과 돈 검사를 받았다. 현금이 작아서인지 심지어 카드에 얼마 들고 왔는지까지 물어봤다. 처음엔 너무 당황해서 횡설수설하다가 네바다주립대학교의 프로그램 초청장을 보여주자 내보내 주었다. 프로그램에 참여한 26명 중 나만 세컨더리룸을 경험했다.

빅데이터 분석 프로젝트 진행을 위해 다양한 강의를 수강했다. 물론 영어로 진행되었기 때문에 처음부터 완벽하게 이해하긴 어려웠다. 담당교수님은 어려운 부분이 있으면 예시를 들어 설명을 잘 해주셨고, 여러 개념이나 이론의 중요성에 대해서도 말씀해 주셨다. 이미 어느 정도 알던 내용이라 복습하는 기분이 들었고 부족했던 몇몇 개념을 학습할 수 있었다. 빅데이터나 기계학습에 대한 사전 지식이 없다면 따라가기 조금 힘든 강의였다.

이론 수업뿐만 아니라 실습수업도 진행했다. 실습수업에서 직접 Python을 이용해 이론 수업에서 진행했던 개념을 적용해 볼 수 있었다. 라이브러리를 사용하지 않고 코딩을 해야 했기 때문에 코딩 실력 향상에 많은 도움이 되었다. 이 수업 역시 Python을 처음 접한 학생에게는 어려웠을 것 같다.

4명이 한 조가 되어 “Deep Learning-Based Survival Analysis Using Genomic Data”를 주제로 선정하여 프로젝트를 진행했다. 생존분석이 다들 생소하고 유전자 데이터도 생소했지만 다양한 논문을 참고하며 지식을 습득할 수 있었고, 프로젝트를 잘 마무리할 수 있었다. 또 프로젝트 발표를 영어로 진행했기 때문에 영어로 말하는 것에 자신감을 얻을 수 있었다.
처음에는 연구 프로젝트만 진행하는 줄 알았지만 다양한 문화 프로그램이 포함되어 있었다. Water Park, NBA Summer League, National Park 등이 있었다. 특히 National Park Tour는 나를 가장 설레게 했다. 죽기 전에 봐야 한다는 Grand Canyon을 실제로 볼 수 있었기 때문이다. Grand Canyon뿐만 아니라 Zion Canyon, Antelope Canyon, Horse shoe bend 등 다양한 관광지 경험을 할 수 있었다. 자연의 힘은 정말 대단했다. 살면서 한 번 해볼까 말까 한 경험을 할 수 있었다.

다양한 문화 체험과 수업 모두 나에게 도움이 많이 되었다. 하지만 프로그램이 장점만 있을 수는 없었다. 날씨는 매우 더웠으며 거리를 걷는 사람을 찾기가 어려웠다. 또, 프로젝트가 수업내용만으로 해결할 수 없었다. 관련 분야의 선행 학습이 필요했다. 다음 연구프로그램에서는 영어뿐만 아니라 빅데이터와 관련된 지식도 평가를 했으면 좋겠다.

프로그램을 통해 좋은 사람들과 함께 연구와 미국 문화를 즐기고 의미있는 시간을 보내며 스스로 한국에서 놓치고 있었던 부분에 대해 알게 되었다. 항상 눈치 보고 조급하게 살아왔는데 여유롭고 긍정적인 마음을 가질 수 있었다. 다른 나라의 문화를 접하고 경험하는 것이 좋은 동기부여가 되었다. 끝으로 프로그램을 준비해 주신 빅데이터 사업단, UNLV 관계자분들, 함께 프로그램에 참여한 모든 분께 감사 인사를 전한다.

정 인 영
경상국립대학교 정보통계학과

한 달, 짧지만 소중한 시간

졸업을 앞두고 GPP 같은 해외 탐방 프로그램에 참여하지 못한 게 아쉬움으로 남아 있었다. 때마침 빅데이터 사업단에서 해외 연구프로그램 참여학생 모집 공고문을 보게 되었다. 미국에 가본 적이 없기도 하고, 미국 학교에서 연구프로그램을 통해 빅데이터 관련 기술을 습득하는 것이 도움이 되고 좋은 경험이 될 것 같아서 프로그램을 신청하게 되었다.

UNLV에서 4주 동안 유전체 데이터를 이용한 딥러닝 기반 생존 분석을 주제로 프로젝트를 수행하였다. 1주 차에는 자료수집과 전처리를 진행하였다. 다운받은 데이터를 결합한 후, 결측값을 제거하고 같은 이름을 가진 gene은 분산이 큰 것으로 선택하여 데이터 전처리를 수행하였다. 3주 차에는 생존 분석에 적합한 feature를 선택하는 작업을 수행하였다. 우리는 feature selection을 위해 Cox-PH 모델을 활용하여 feature selection을 수행하였다. 이 프로젝트를 통해 직접 데이터를 전처리하고 분석하고, 모델 학습과 모델 평가를 수행하면서 생존 분석에 대해서 더 잘 이해하게 되었다.
버디 프로그램으로 미국에서 5번째로 높은 건물인 스트라토스피어 타워에 갔다. 스트라토스피어 타워의 놀이기구를 타기 위해서는 엘리베이터를 타고 106층까지 올라가야 했다. 꼭대기에 가면 전망대가 있고 3

종류의 놀이기구를 탈 수 있었는데, 당시에는 바람 때문인지 엑스스크림만 운행하고 있어서 놀이기구 한 가지밖에 타지 못했다. 엑스스크림은 탑승 후 앞, 뒤로 왔다 갔다를 반복하는 놀이기구였다. 놀이기구를 타고 스릴을 즐기면서 풍경을 바라보니 더욱 재미있게 느껴졌다. 3가지의 놀이기구 중 빅샷이 가장 타고 싶었는데 타지 못해서 아쉬웠다.

투어 프로그램 중 가장 기억에 남는 프로그램은 마지막 필드트립인 내셔널 파크 투어였다. 자이언 캐니언에서의 트래킹을 통해 협곡을 따라 흐르는 강물을 뚫고 가서 자연의 아름다움을 가까이에서 느낄 수 있었다. 또한 엔테롭 캐니언에서는 자연 속에서의 평온과 경치의 아름다움을 동시에 누릴 수 있었다. 이곳을 방문하여 대자연의 위대함과 함께 평화로운 시간을 보내는 것은 정말로 소중한 경험이었다.

이 프로그램을 통해서 새로운 문화와 환경에서의 적응을 통해 성장할 기회를 얻게 되었으며, 다양한 사람들과 교류하며 더 넓은 시야를 가질 수 있게 되었다. 특히, 영어로 수업을 듣고 발표하는 과정은 외국어 능력을 향상시키는 데 큰 도움이 되었다. 그리고 팀 프로젝트를 통해 팀원들과 함께 과제를 수행하며 소통하고 협력하는 능력을 기를 수 있었다. 또한 현재 나에게 부족한 부분이 많이 있다는 것을 다시 한번 느끼게 되었고 나를 되돌아보는 시간이 되었다. 이 분야에 관심이 있는 많은 학생이 이러한 기회를 놓치지 않고 경험해 보았으면 좋겠다.

정 현 서
경상국립대학교 산업시스템공학부

모든 것이 처음이었다

운이 좋았다. 쟁쟁한 경쟁률을 뚫고 미국에 갈 수 있는 기회를 얻게 되었다. 미국은 어떤 곳일까? 내가 살면서 미국에 갈 일이 있을까? 해외여행 얘기가 나오면 내가 항상 생각하던 것들이다. 그만큼 미국이라는 나라에 대한 로망이 컸다. 평범한 대학 생활을 하던 중, '하계 빅데이터 해외 연구프로그램'홍보를 접하게 된 나는 한 치의 고민도 없이 지원할 수밖에 없었다.

하계 빅데이터 해외 연구프로그램은 빅데이터에 관련된 연구를 직접 경험하고, 빅데이터 역량을 얻기 위한 프로그램이었다. 나는 딥러닝을 활용한 유전체 데이터 분석과 생존분석에 대한 프로젝트를 맡게 되었다. 처음에는 다소 어려워 보였으나, 노력하고 이해하려는 노력 끝에 복잡한 내용을 어느 정도 이해하게 되었다. 처음 만나는 팀원들과 실제 암과 관련된 데이터를 다루고 문제를 해결하는 과정에서 큰 성취감을 느낄 수 있었다.

가장 기억에 남는 날은 스트립(The Strip)에 처음 방문했던 날이다. 서울만 가도 입을 벌리고 구경하는 나에게 펼쳐진 스트립의 모습은 말 그대로 환상의 도시였다. 미국은 무엇이든 크게 만든다는 말은 사실이었다. 각종 호텔과 큰 건물, 즐길 거리 들로만 가득 찬 거리는 라스베이거스가 왜 "Sin City"라고 불리는지 이해가 될 수준이었다. 여러 지역에서 온 관광객들과 현지인들 사이를 걸어 다니니, 라스베이거스가 나를 환영해주는 기분이 들었다.

미국에서 눈에 보이는 모든 것이 처음이었다. 미국의 대학교, 미국의 문화, 미국의 음식, 그리고 미국. 나의 눈에

보이는 모든 것이 처음이었고, 내가 딛는 걸음 하나하나가 처음 밟는 땅이었다. 인생 살면서 한 달이나 해외에서 지내는 것은 상당히 귀한 경험이라고 생각한다. 한 달을 지내면서 해외의 문화를 몸으로 느끼고, 현지인들과 교류하는 것은 나의 세상에 대한 식견을 넓혀주는데 전혀 부족하지 않았다. 빅데이터에 대한 교육과 실습을 진행하면서 빅데이터 역량이 향상됨은 물론이고, 학문적인 동기부여 요소로도 크게 작용하였다. 특히 도움이 되었다고 생각하는 것은 영어회화 구사능력이다. 누군가의 도움 없이 현지인들과 대화하는 것은 나의 영어 스피킹에 대한 자신감을 크게 불어넣었다.

물론 이런 낯선 환경에 노출되는 것에 마냥 장점만 있는 것은 아니었다. 총기 소지 국가에서는 주위를 항상 둘러보며 언제 들이닥칠 위험을 항상 대비해야 했다. 아시아인에 대해 인식이 좋지 않은 사람들을 마주할 때면 정신적으로 힘든 부분도 분명히 존재했다. 날씨 또한 매우 더워, 열기에 내성이 없는 나는 상당히 고생했다. 물론 이렇게 몸소 느낀 장점이든 단점이든, 나에게는 모두 소중한 추억이 될 것이다.

이 글을 보는 사람들에게 하고 싶은 말이 있다.

전도체는 전기를 흐르게 하는데 일부 전기가 저항에 의해 소실된다. 비슷하게, 우리가 다른 사람과 후기를 공유할 때도 정보나 감정이 소실될 수밖에 없다. 후기를 전달하고 전달할수록 손실되는 정보와 감정이 많아진다. 글이나 사진으로는 어느 정도 전달할 수는 있지만, 분명 한계가 존재한다. 저항 없이 정보와 감정을 받아들이기 위해서는 직접 참여하는 방법밖에 없다. 프로그램에 참여할지 고민이 된다면, 일단 지원하라. 그러면 '경험'이라는 초전도체를 발견할 수 있을 것이다.

최규진
경상국립대학교 수학과

해외연구프로그램을 통한 미국 여정 : 의사소통과 연구의 기회

이번 빅데이터 해외 연구 프로그램에 참여한 이유가 무엇이라 묻는다면 다음 두 가지로 답할 수 있을 것 같다. 먼저 의사소통 능력 향상이다. 현지인들과의 직접적인 대화를 통해 의사소통 능력을 키우고 그들을 통해서 미국의 다양한 문화를 경험하고 싶었다. 다음은 연구 경험이다. 해외 연구프로그램의 프로젝트 활동을 통한 연구 경험은 향후 대학원 진학 후, 학업 및 연구에 큰 도움이 될 것으로 생각했다.

4주간 라스베이거스에서의 생활은 하루하루가 새로웠다. 그들의 팁 문화는 생소했고 오직 미국에서만 먹을 수 있는 음식들은 라스베이거스의 더운 여름을 버티게 해주는 원동력이 되어주었다. 미국에 오기 전에 인종 차별 또는 언어의 차이에 대한 걱정들이 있었지만 그들의 배려 덕분에 현지에서 생활하는 데 큰 어려움은 없었다. 프로그램 활동 중에서 가장 충격적이었던 부분은 그랜드 캐니언이었으며 그랜드 캐니언의 광활한 풍경은 왜 그랜드 캐니언이 죽기 전에 가야 할 장소인지 실감할 수 있었다.

4주간의 해외 연구 프로그램에서 주축이 되는 활동은 영어 문법 수업과 머신러닝/딥러닝 수업이었다. 영어 문법 수업은 프로젝트 활동을 연구 논문으로 만들기 위해 필요한 영문법에 대해 알려주었다. 해당 수업을 통해 올바른 쉼표 사용법, 주장하는 글쓰기, google scholar를 활용한 논문 탐색에 대해 배울 수 있었다. 영어 문법 수업을 통해 연구 논문을 만드는 과정에서 나의 회화 실력에 경각심을 느낄 수 있었다. 현지 교수님의 말 속도를 따라가기 힘들어 수업의 내용을 이해하는 데 어려움이 있었으며 영작을 하는 과정에서도 표현 교정 과정을 하나하나 확인해야 했었다. 머신러닝과 딥러닝 수업은 여러 머신러닝 모델의 알고리즘, 모델 평가 방법, 딥러닝 학습 과정에 대한 이론적인 내용 학습과 이를 바탕으로 구현하는 과정을 통해 이해할 수 있었다. 머신러닝과 딥러닝 수업을 통해 동아리 활동을 통해 배운 데이터 분석에 대한 지식을 상기시킬 수 있어서 좋았으며 프로그램 이후에도 여러 공모전을 통해서 데이터 분석에 대한 감을 잃지 않도록 노력할 것이다.

해외 연구프로그램을 마치고 가장 먼저 한 것은 영어 발음 강좌 수강이었습니다. 유창한 회화 능력을 키우기 위해 유튜브 영상 섀도잉 강좌를 수강하고 있으며, 목표했던 의사소통 능력 향상을 위해 노력할 것이다. 데이터 분석에 대한 감을 잃지 않기 위해 다양한 공모전에도 참가하였으며 현재 참가 중인 '2023 빅콘테스트'에서 역량을 발휘해 좋은 성적을 얻을 수 있도록 최선을 다할 것이다.

프로그램의 장단점 및 개선점 등 이번 해외 프로그램을 통해서 한국에서는 경험할 수 없었던 다양한 미국의 문화를 경험할 수 있었다. 그리고 현지인들과의 대화를 통해 그들처럼 영어를 유창하게 하고 싶다는 열망을 얻을 수 있었다. 다만, 프로젝트를 완성하기에는 4주라는 시간이 짧았으며, 다른 도시의 경험도 해봤으면 좋았을 것 같다.

강다영
전북대학교 스마트팜학과

더 넓은 세상으로 나아가자

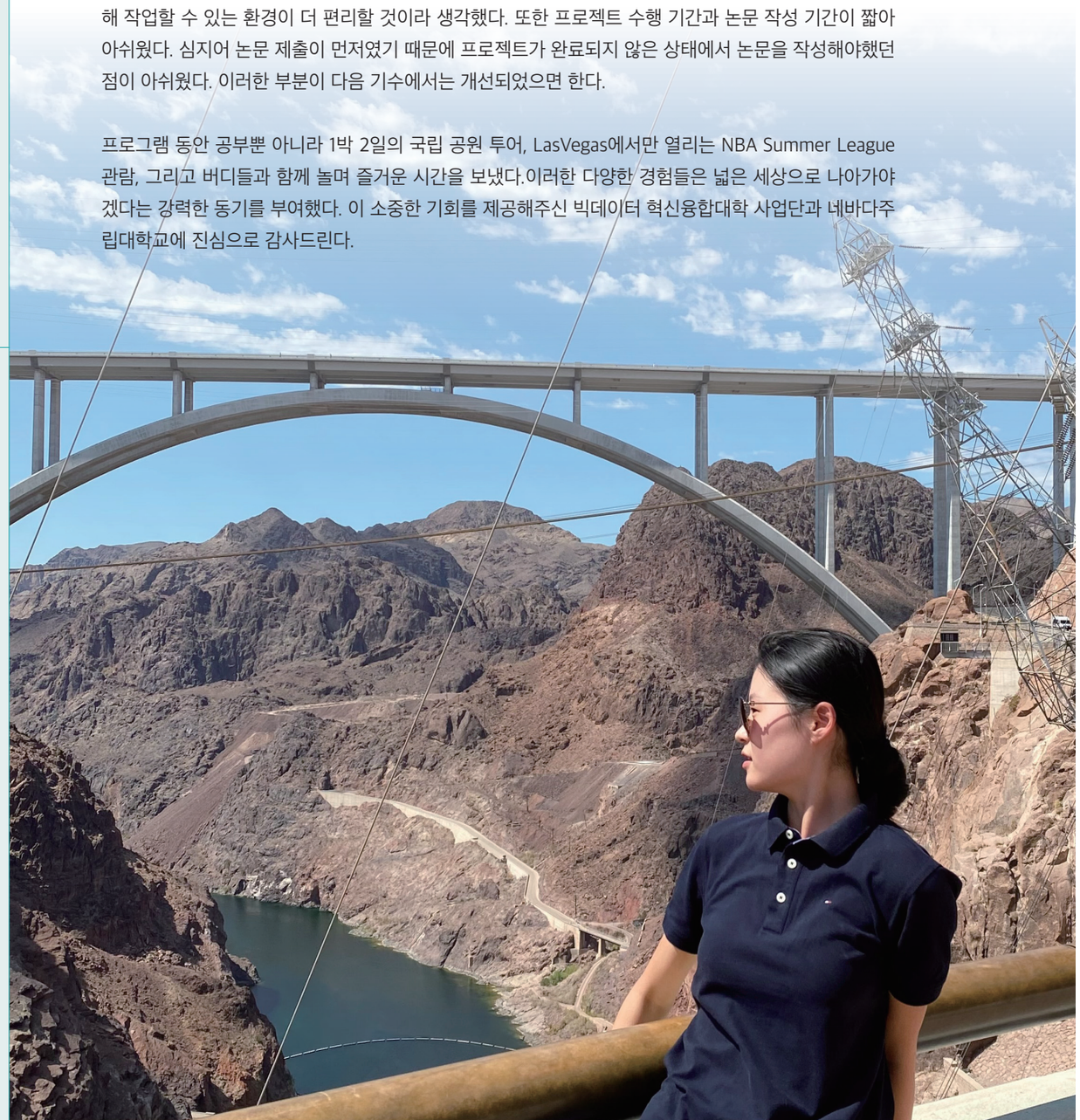
스마트팜학과에서는 생산량 예측, 환경 조절 등의 분야에서 인공지능이 많이 활용되고 있다. 이로써 빅데이터나 인공지능에 대해 자주 접하게 되면서 학부연구생으로 약 9개월 간 학과 연구실에서 활동하며 모델 개발 경험을 쌓게 되었고, 이를 통해 머신러닝에 대해 큰 흥미를 가지게 되었다. 어느날 교내 공지를 훑어보던 중, 우연히 미국에서 빅데이터 연구 프로그램 참여 학생을 선발하는 공고를 발견했다. 이 프로그램은 미국 네바다주립대학교에서 진행되며, 빅데이터 분야의 전문 교육과 연구 기회를 제공한다는 점에서 좋은 기회라고 생각했다. 미국을 처음 가보는 것도 물론 이유였지만, 더불어 인공지능 분야에서 선두주자인 미국에서 공부할 수 있는 기회가 주어진다는 게 훨씬 컸다. 따라서 이 프로그램을 통해 다양한 미국의 문화를 경험하고, 교육을 받아 빅데이터 분야에서의 역량을 향상하고자 참여하게 되었다.

4주간 생활하며 재미있었던 몇몇 일화가 있다. 노스프리미엄 아울렛 나이키 매장에서 쇼핑을 하던 도중, 흑인 분이 다가와서 직원이냐고 물어보며 상품에 대한 질문을 했다. 아니라며 고개를 저었는데, 이유는 모르겠지만 왠지 모를 뿌듯함이 느껴졌었다. 숙소 세탁방에서는 새로 발급받은 카드가 동작하지 않은 적이 있었다. 옆에서 자매 관계의 할머니분들이 카드 충전을 헤매고 계셔서 도와드렸는데, 내가 오류로 버둥대는 모습을 보시고는 본인들이 카드를 각자 발급받았다며 하나를 주셨다. 친절한 행동에 정말 감사했고, 빨래가 끝나면 옷을 챙겨서 가버리겠다고 웃으며 농담하셨던 순간이 아직도 생생하다. 마지막으로 전북대 학생들과의 에피소드이다. 총 10명의 학생과 함께 시간을 보내면서 서로 친해져 대부분의 활동을 함께 즐겼다. 기간 중 2명의 학생이 생일을 맞아 축하 파티를 열기도 하고, 평소에도 함께 모여 놀러 다니거나, 방에서 보드 게임이나 좀비 게임을 하며 재미있는 시간을 보냈다. 이런 사소한 즐거움들이 이번 프로그램을 더 기억에 남는 추억으로 만들었다.

본 프로그램의 연구 주제는 크게 4가지로 생존 분석, 이미지 분류, 이미지 분할 그리고 빅데이터였다. 기존에 다뤄 보았던 이미지 분류와 분할, 데이터에 대한 도메인 지식이 필요한 생존 분석을 제외하고 평소에 관심이 있던 빅데이터를 주제로 선정하였다. 구체적인 주제는 'Scalable Machine Learning for Bigdata'로 MapReduce를 구현한

머신러닝 모델을 직접 구현하는 것이었다. Sai의 오후 수업에서 배운 이론을 기반으로, MapReduce를 구현한 논문들을 찾아보고 실행해보며 pyspark도 익혀나갔다. 박사님 중 spark를 연구하시는 분이 없으셔서 전문적인 도움을 받기 어려웠지만, 이는 더 열심히 공부할 수 있는 계기가 되었고 팀원들과 함께 이해해가며 알고리즘 개발 방향성을 찾아갔다. 최종적으로 완성된 코드를 Sai와 공유하면서 좋은 피드백을 받고, 최종 발표에서는 교수님으로부터 칭찬을 받아 뿌듯했다.
프로젝트 수행 중 아쉬운점도 분명히 있었다. 연구를 위해 팀마다 할당된 GPU 서버가 학교 내에서만 사용 가능했다. 도서관의 개방 시간에 따라, 주말과 같은 휴일에 따라 제한된 부분이 있었고, 원격 서버를 통해 작업할 수 있는 환경이 더 편리할 것이라 생각했다. 또한 프로젝트 수행 기간과 논문 작성 기간이 짧아 아쉬웠다. 심지어 논문 제출이 먼저였기 때문에 프로젝트가 완료되지 않은 상태에서 논문을 작성해야했던 점이 아쉬웠다. 이러한 부분이 다음 기수에서는 개선되었으면 한다.

프로그램 동안 공부뿐 아니라 1박 2일의 국립 공원 투어, LasVegas에서만 열리는 NBA Summer League 관람, 그리고 버디들과 함께 놀며 즐거운 시간을 보냈다.이러한 다양한 경험들은 넓은 세상으로 나아가야겠다는 강력한 동기를 부여했다. 이 소중한 기회를 제공해주신 빅데이터 혁신융합대학 사업단과 네바다주립대학교에 진심으로 감사드린다.



권 지 현
전북대학교 통계학과

다시 없을 대학 시절의 값진 추억

대학 생활의 2년을 코로나로 보내고 나니 어느덧 4학년이 되어 있었다. 호기심이 많고 다양한 활동에 참여하길 좋아하는 나는 대학 시절에 어학연수를 꼭 하리라는 다짐을 예전부터 품고 있었다. 하지만 갑자기 터진 코로나로 어학연수의 꿈은 사라진 채, 대학 생활을 마무리해가고 있었다. 그러던 중 학교 홈페이지를 보다 우연히 이번 프로그램을 알게 되었다. 어학연수에 뜻이 깊었던 나는 망설임 없이 바로 지원서를 작성했고 운이 좋게도 이번 '하계 빅데이터 해외 연구프로그램'에 선발되어 네바다주립대에서 한 달 동안 교육을 받을 기회를 얻게 되었다. 외국에서 전공과목을 수강할 수 있다는 점과 한 달간의 외국 생활이라는 점이 나에게 크게 다가왔고 대학 생활 중 마지막 기회일 것이라는 생각에 이번 프로그램이 더욱 기대되었다.

4주 동안의 프로그램 동안 가장 몰두했던 일은 바로 프로젝트였다. 프로젝트는 4개의 주제 중 하나를 선택하여 분석해보는 방식이었다. 나는 평소 해본 적이 없던 딥러닝 분야에 관심이 생겨 'Project 4 : Deep Learning-Based Survival Analysis Using Genomic Data'를 선택하였다. 프로젝트 4는 코드 구현보다 사전 지식이 더 중요한 문제였기 때문에 생존분석과 유전학 데이터와 관련된 논문들을 찾아보며 기초 지식을 습득하는 데 힘썼다. 처음 도전해보는 분야라 시행착오도 많았고, 해결해야 할 문제들도 많았다. 엎친 데 덮친 격으로 중간에 노트북 충전기까지 고장이 나서 노트북 충전기가 올 때까지 프로젝트에 참여하는 데 어려움이

많았다. 다행히도 실력파 팀원들을 만나 프로젝트는 잘 마무리 할 수 있었고 프로젝트를 진행하며 그저 수업만 듣는 것만이 아닌 학생이 주도적으로 프로젝트를 이끌어 가며 추가로 더 공부해야 했기에 스스로 성장할 기회가 되었다는 생각이 들었다.

미국에서의 한 달간의 생활은 나에게 있어 매우 도전적인 일이었다. 미국에서 영어로 수업을 듣는 것도, 모든 주문을 영어로 하는 것도, 문제가 생기면 해결해야 하는 것도 모두 혼자 힘으로 해야만 했다. 처음엔 힘들었으나 차츰 적응해 나갔고 끝 무렵에는 한국으로 돌아가기 아쉬울 정도로 라스베이거스에 정이 많이 들렀다.

한 달간의 미국 생활은 그 나라의 문화를 경험해 볼 수 있는 값진 시간이었다. 미국이라는 먼 나라까지 와서 한 달을 지내는 것은 대학생 때 말고는 할 수 없는 활동일 것이다. 또한 버디 프로그램과 관광 일정, 프로젝트 등의 일정이 이미 짜여있어 데이터 관련 경험뿐만 아니라 외국인과의 교류, 관광까지 모두 경험해 볼 수 있었다. 같은 분야에 흥미를 느끼고 있는 사람들과 프로젝트를 진행해 보며 나 스스로 부족한 점에 대해 깨닫고 많이 배워갈 수 있는 시간이었으며 영어를 끔찍이 싫어했는데 한 달 동안 라스베이거스에서의 생활 후 영어 회화에 흥미가 생겨 영어를 더 배워보고 싶다는 생각마저 들었다. 이번 프로그램은 나에게 다시 없을 대학 시절의 소중한 추억이 되었다.

김나영
전북대학교 컴퓨터공학부

한 달, 내가 성장했던 기간

빅데이터 해외 연구프로그램에 참가한 동기는 빅데이터와 인공지능에 대한 깊은 관심과 열정이었다. 또한, 내가 지금까지 배웠던 인공지능에 대한 개념들을 직접 프로젝트에 녹여보고 싶었다. 이 프로그램을 통해 더 나은 빅데이터 기술과 개념을 이해하고 실전 경험을 쌓고자 했다.

4주간의 미국 생활은 나에게 다양한 환경과 경험을 제공했다. 특히, 워터파크, NBA 경기, 그랜드 캐니언 관광, 그리고 라스베이거스의 엔터테인먼트 쇼를 위한 기계들 관람 등 다채로운 활동을 즐길 수 있었고, 특히 그랜드 캐니언과 자이언 캐니언의 아름다운 풍경은 평생 잊지 못할 순간 중 하나였다. 이와 더불어 친구들과 함께 스트립 거리와 다운타운에 놀러가며 더윈뷔페, 핫앤쥬시, 고든램지버거등의 맛있는 음식들을 먹었고 최근 개봉했던 미션임파서블 영화를 문화의 날인 화요일에 영화관에서 보는 등 다양한 문화들을 접할 수 있었다.

프로그램에서는 빅데이터 분석을 위해 GPU를 활용하고, 컴퓨터 비전 개념을 프로젝트에 적용하는 등 실제 업무에서 필요한 스킬을 향상시킬 수 있었다. 또한 영어로 수업을 듣고 논물을 쓰며 발표하는 경험을 통해 영어로 듣고 말하는 것에 대한 자신감을 가질 수 있었다. 또한, 담당해주신 교수님들의 조언과 수업은 내가 미래에 대해 결정을 내리는 데 큰 도움이 되었다.

프로그램의 장점 중 하나는 미국에서만 누릴 수 있는 다양한 활동과 관광 기회가 주어졌다는 점이다. 그러나 날씨가 무척 더웠기 때문에 이동이 힘들었고 프로젝트 진행할 때 데이터셋을 너무 늦게 주셔서 프로젝트 진행이 지체되었던 점이 아쉬웠다. 또한, 프로그램의 강의 내용이 좀 더 최신 기술을 다룬다면 더욱 가치 있었을 것 같다.

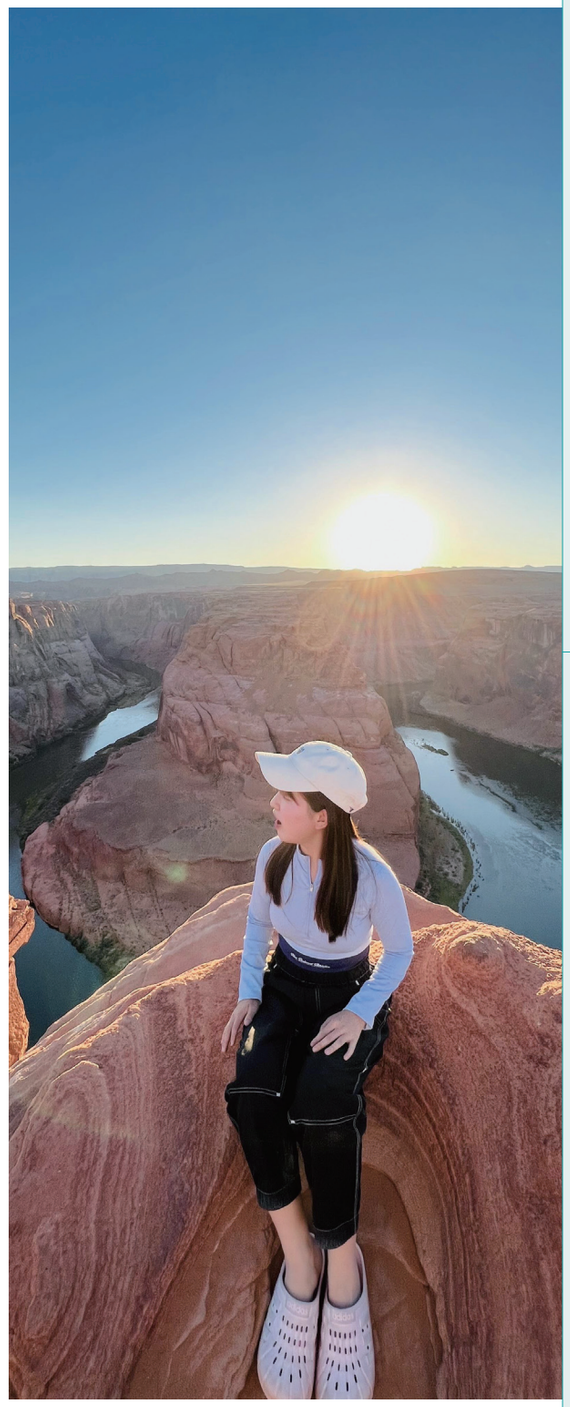
이 프로그램을 참가하기 전에는 이 인공지능에 대한 개념을 어떻게 활용할지 고민이 많았는데 참가하고 나서 감이 잡혔다. 앞으로는 연구프로그램에서 얻은 지식과 경험을 활용하여 더 나은 프로젝트를 진행하고 싶다. 연구프로그램은 내 미래에 대한 큰 계기가 되었으며, 앞으로도 계속해서 발전하고 성장하기 위해 노력할 것이다.

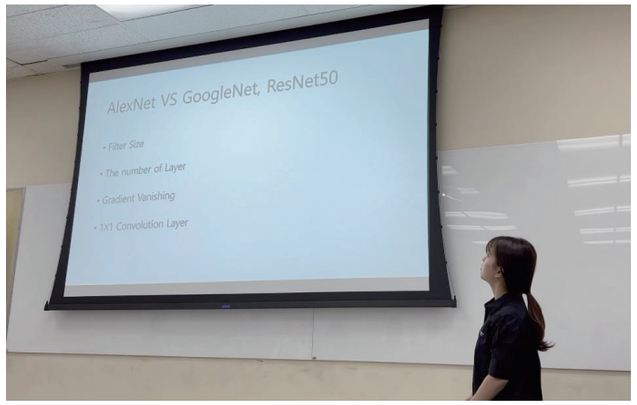
김 미 현
전북대학교 컴퓨터인공지능학부

한 달, 꿈같던 시간

네바다주립대학교에서의 연구프로그램이 3주가 넘어선 시점이었다. 평일 아침에는 대학교 수업에 출석하고, 오후부터 자는 시간까지 암 이미지 분류 모델을 구축하는 작업에 몰두하고 있었다. 이런 바쁜 일상에서 피곤함에 시달리던 중, 학교에서 주최하는 캐니언 투어에 참여하게 되었다. 그때까지는 캐니언에 대해 잘 알지 못했고, 그저 학교에서 일러주는 일정에 따라 몸을 맡기기로 했다. 그러나 이런 기대 없던 일정이 내게 미소를 머금게 하는 행복한 순간으로 남게 되었다.

머신러닝의 이론을 배우고, 그 개념을 활용하여 과제를 하다가 마지막 2주는 주어진 프로젝트에 힘을 썼다. 우리 팀이 맡은 주제는 암 이미지 분류였으며, 주어진 데이터를 받고, 암이면 1, 암이 아니면 0으로 분류하는 모델을 구축하는 것이었다. 매일매일 팀원끼리 코드를 짜고, 공부를 하고, CNN의 유명한 아키텍쳐를 이용해서 정확성을 구하고, 비교 분석하였다. 이러한 과정에서 컴공의 대표적인 진로 방향인 프론트 엔드, 백엔드보다는 인공지능, 머신러닝 대학원과 데이버 분석에 관심이 크게 생겼으며, 진로 방향성이 점차 확고해졌다. 데이터를 분석하는 데에 큰 재미를 느꼈으며 모델마다 결괏값이 다르게 나와 어떠한 데이터에 어떠한 모델을 적용시켜야 하는지 분석하는 논문을 읽고, 구축하는 과정을 앞으로도 하고 싶어졌다. 그래서 데이터 사이언티스트라는 직업에 관심이 생겼으며, 이러한 방향으로 국내에서 석사를 진학하고, 할 수 있다면 박사는 해외에서 하는 것

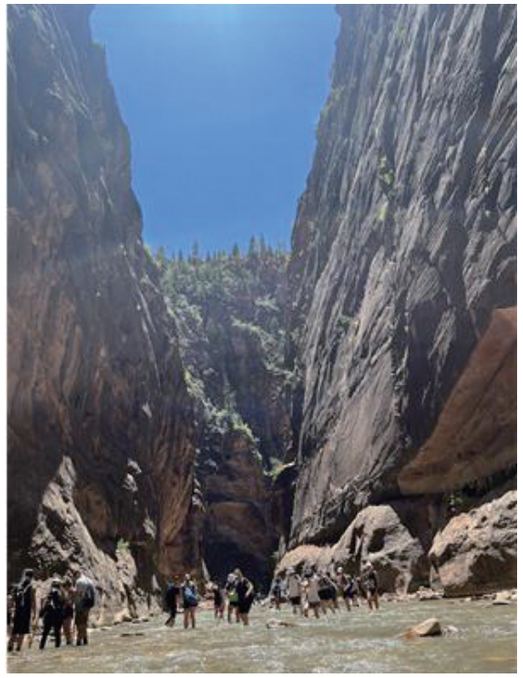




까지 목표가 생겼다. 또한 수업과 발표를 영어로 진행하고, 자료를 준비하면서 논문을 많이 봐서 그런지 영어 실력이 늘었다. 마지막 날, 영어로 발표할 때는 하나도 떨리지 않았으며, 귀도 뚫려서 한국에 와서 영어 토익 리스닝이 잘 되었다. 이 흐름을 놓치지 않기 위해 꾸준히 영어 공부도 하는 중이다.

한국에서는 취미 생활이 밥, 카페, 전시회 이 정도가 전부였다. 그러나 미국인들은 운동, 문화를 더 다양하게 즐기는 느낌이었다. 학교 gym에도 각종 체육시설이 다 들어있으며 자유롭게 이러한 시설을 편하게 누리는 게 얼마나 큰 행복인지를 알게 되었다. 이것들을 놓치지 않고 미국에서의 삶 그대로 한국에서도 유지하고 싶어 삶을 대하는 태도가 변하게 되었다.

첫 번째로 삶의 주체성이 증가했다. 같이 간 학생들은 공부, 공모전, 운동, 심지어 놀기까지 적극적으로 참여하는 친구들이었다. 그들과 같이 힘을 모아 목표를 이루기 위해 함께 화합하는 것이 이 나이대에 할 수 있는 가장 행복한 경험 중 하나라는 것을 느꼈다. 그래서 미국 다녀온 이후에도 SW대회, 해커톤 대회 등에 팀을 이루어 온 오프라인으로 만나며 열심히 참여하고 있다. 또한 이러한 성향을 가진 친구들을 만나기 위해 나 역시도 주체성이 강한 활동들에 참여하면서 인간관계를 구축해나가야 함을 깨닫게 되었다.

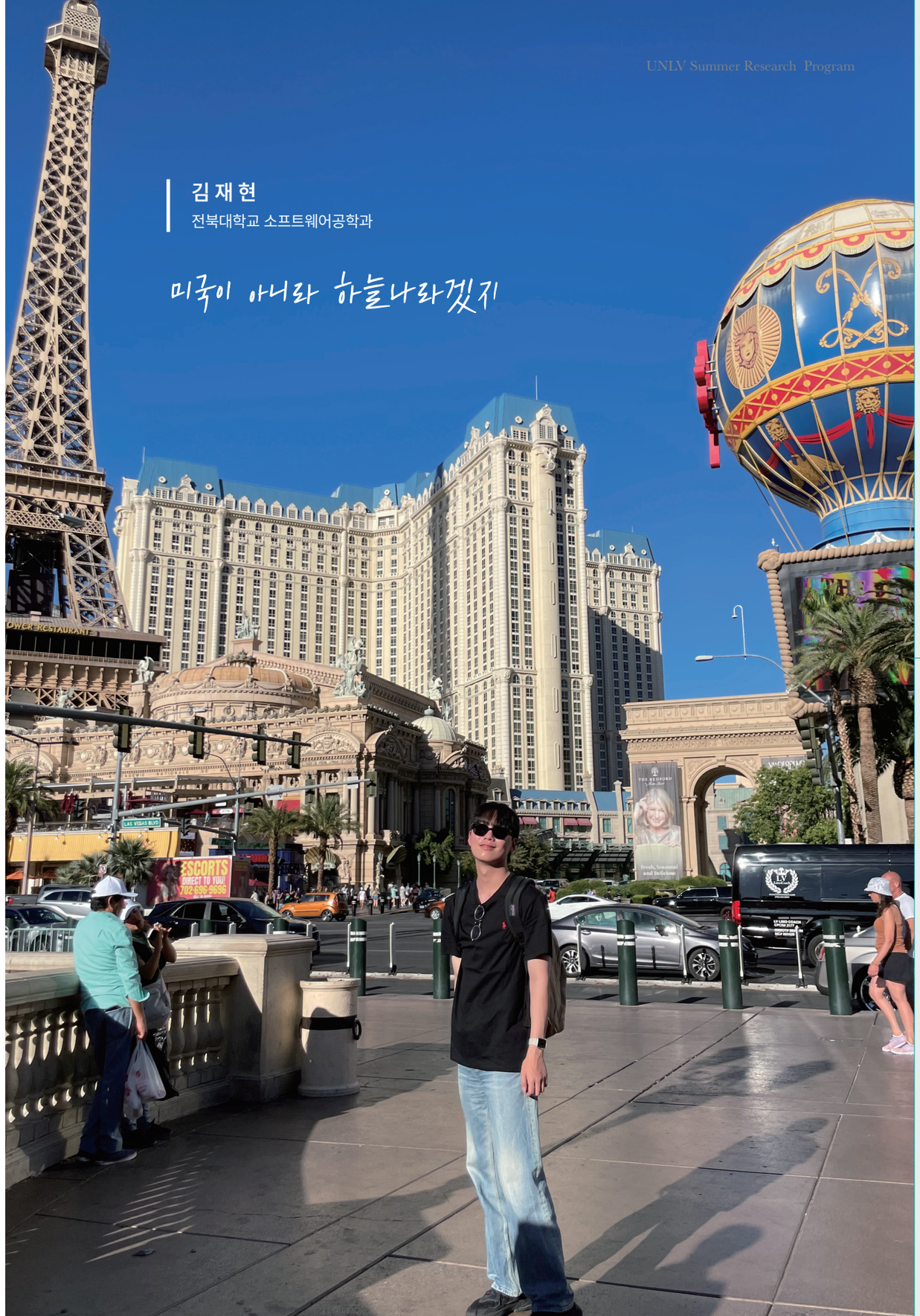


두 번째로는 운동이 교양의 일부라는 것을 알게 되었다. 네바다주립대학교의 체육시설은 정말 좋았다. 각종 유산소, 웨이트, 배드민턴, 농구장, 수영장까지 학생들이 체육시설을 즐길 수 있다. 원래는 헬스장을 끊고도 몇 번 안 가던 학생이었는데 네바다주립대학교 체육관에서 친구들과 근력, 유산소, 배드민턴하며 그 재미를 알게 되었다. 이 모든 게 일종의 교양이라는 인식이 강하게 생겼으며 땀 흘리는 재미도 생겨서 전주에 돌아와 필라테스와 헬스를 거의 매일 하고있다.

이러한 스스로의 변화와 만족 그리고 행복했던 기억들 덕분에 후배들에게 이 프로그램을 추천해주고 싶고, 기회가 된다면 또 가고 싶다.

김 재 현
전북대학교 소프트웨어공학과

미국이 아니라 하늘나라겠지

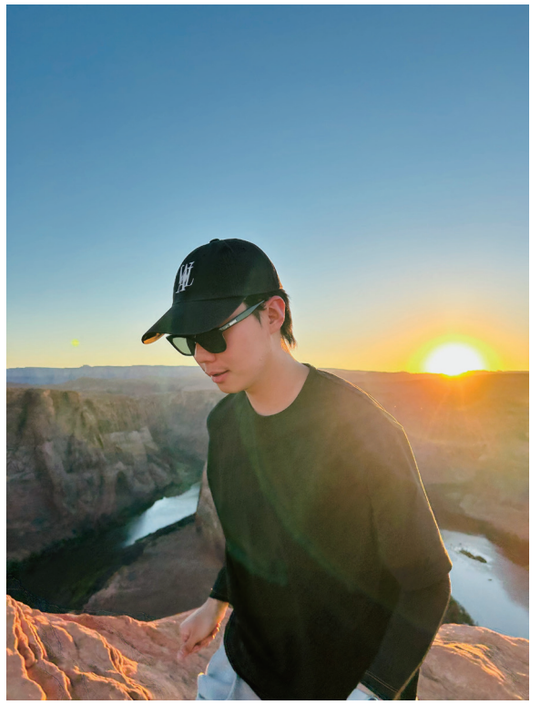


흔히 미국을 천조국이라 한다. 국방 예산에 천조를 쏟아부어 그렇다고 하는데 사실 다른 뜻도 있다. '천조(天朝)' 즉, 하늘의 왕조라는 뜻이다. 국제 질서의 중심 국가, 광활한 대지와 약 4억 명의 인구를 가진 나라에 가볼 수 있다는 건 정말 설레는 일이다. 소위 말하는 '아메리칸 드림'은 한국에 사는 내게도 버킷리스트에 담아볼 만한 그런 꿈같은 일이었다.

미국에 가게 된 건 올해 여름이었다. 4학년 2학기를 앞둔 나는 프로젝트 마무리와 취업 준비를 한다고 연구실에 묻혀있듯 살고 있었는데, 학교 홈페이지에서 생뚱맞게 미국에 보내준다는 프로그램이 있길래 무심코 신청한 결과 프로그램 대상자 선정 즉, 미국에 가게 된 것이다. 빅데이터와 인공지능을 토대로 진행되는 연구 프로그램은 네바다주 라스베이거스에 있는 네바다주립대학교에서 진행된다고 했다. 말로만 듣던 사막에 세워진 거대한 도시. 벨라지오 분수와 태양의 서커스가 유명한 라스베이거스로 향하는 설레는 마음으로 비행기에 올랐다.

라스베이거스는 쉽지 않은 도시다. 비행기에서 내린 우리를 맞이한 건 건조한 사막의 바람과 작열하는 태양이다. 섭씨 45도가 넘는 오후 4시에 한번 놀라고 길에서 흔한 대마초 냄새에 또 한번 놀랐다. 나는 원래 땀을 잘 흘리지 않는데 여기서는 이마에 땀이 항상 흥건했다. 생각보다 적응하기 힘든 환경에서 한 달 동안 지내야 한다는 사실이 부담스러웠다. 하지만 귀국 후 더 많은 추억을 쌓지 못한 것을 후회할 지도 모른다는 생각에 현지 적용을 위해 노력했다. 그러다보니 어느새 필요한 물건을 사려고 월마트 회원가입을 하고 있는 나를 발견하게 되었다.

우리는 굉장히 바쁜 시간을 보냈다. 매일 수업을 듣고 과제와 프로젝트를 하며 그 와중에도 열심히 놀았다. 라스베이거스 스트립에는 독특한 테마를 가진 호텔들이 많아 볼거리가 참 많다. 어딜 가도 카지노가 있고 음악이 함께한다. 기억에 남는 장면들이 몇 개가 있는데 그중 하나는 치즈케익을 사기 위해 미로 같은 호텔을 탐험한 기억이다. 마치 이상한 나라에 온 앨리스처럼 호텔 속에서 한참을 걸었고 주변 풍경은 이탈리아에 온 듯한 분위기와 심지어 호텔 천장에는 노을 지는 하늘과 구름도 있었다. 우리는 신나게 놀고 구경하고 먹으며 스트립을 즐겼고 롤러코스터도 탔다. 하지만 아쉽게도 멀리 보이던 관람차는 타지 못했다.



이번 미국행에서 가장 좋았던 경험을 말하라고 하면 당연히 국립공원 투어다. 1박 2일 동안 진행된 투어는 내게 절대 잊지 못할 기억으로 남아있다. 물속을 걷는 하이킹은 상상이 쉽진 않겠지만 자이언 캐니언에선 당연하다. 모두가 막대기를 하나씩 쥐고 협곡의 물길을 거슬러 올라간다. 유타주에서 애리조나주로 넘어가서 볼 수 있었던 건 말발굽을 닮은 거대한 협곡이었다. 말발굽을 닮아 홀스슈 밴드라 불리는 곳에서 우리는 노을과 함께 다 같이 사진을 찍었다. 밤에는 바비큐 파티를 했는데 교수님께서 김치찌개와 데킬라의 조합을 알려주셨다. 그리고 그날 태어나서 처음으로 은하수를 봤다. 다음날에는 앤텔롭 캐니언과 그랜드 캐니언에 가서 버킷리스트의 한쪽을 채울 수 있었다.

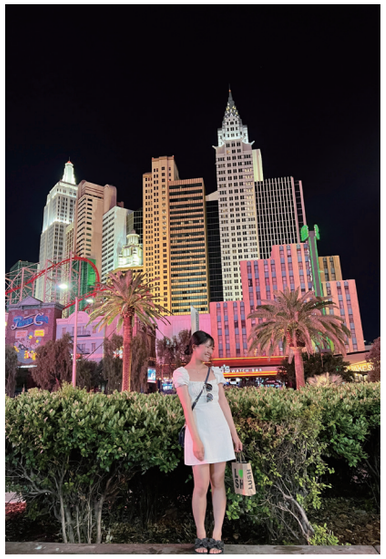
프로젝트와 논문이 마무리되면서 집에 갈 시간도 점점 다가왔다. 삭막해 보이지만 뜨겁고 화려한 도시에 정을 붙이고 처음 만난 친구들과 많은 추억을 쌓았다. UNLV 체육관에서 하는 수영, 버디들과 함께했던 볼링장과 노래방, 꼭 하고 싶었던 NBA 직관, 숙소에서 생긴 바퀴벌레 소동과 좀비 게임, 아무도 라이터가 없어 촛불을 못 켰던 생일파티와 어색하게 영어로 주문하던 햄버거 메뉴까지 그 모든 순간이 지금도 눈앞에 생생하다. 다들 날씨가 덥다고, 음식이 입에 맞지 않다고 투덜거리던 처음처럼 한국은 비가 많이 온다고, 습해서 가기 싫다고 투덜거렸지만 사실은 정든 라스베이거스를 떠나는 게 그저 싫을 뿐이란 걸 모두가 알고 있었다.

한국에 돌아온 나는 전처럼 프로젝트를 하고 취업을 준비하고 있다. 꿈같았던 한 달의 시간 때문에 내 삶이 극적으로 변할 거라고는 기대하지 않는다. 좋은 꿈도 그저 꿈일 뿐이지 않나. 하지만 내게 무언가 달라진 것이 있다. 사실 정확하게 무엇이 달라졌다고는 설명이 어렵지만 새로운 뭔가를 마주할 때 더 열린 마음으로 받아들인다던가, 일을 시작하기 전에 주저하던 성격이 그냥 부딪혀 보자는 성격으로 바뀌었다든지 뭐 그런 소소한 변화 말이다. 아! 영어에도 자신감이 좀 생겼다.

우리는 생소한 곳에서 계획대로 움직이며 현재를 통제할 때 비로소 자신의 삶을 산다고 생각한다. 과거의 후회와 미래의 걱정에서 벗어나 순수한 현재를 온전히 살아간다는 건 내가 삶의 주인이 되기 위해 필요한 과정 중 하나다. 지루하고 어지러운 일상에서 벗어나 낯선 곳에서 현재에 집중하는 것. 거기서 오는 경험과 고양된 몸은 내가 다시 일상을 살아가는 원동력이 된다. 이번 프로그램은 언젠가부터 멈춰있던 나에게 그런 의미였다.

박 세 현
전북대학교 소프트웨어공학과

미국에서의 잊지 못한 기억들



미국은 나에게 있어 다양한 것을 경험할 수 있고 엄청나게 크고 모든 것을 이룰 것만 같은 꿈같은 나라였다. 언제 한번 가볼 수 있을까 상상을 하던 그때 해외 교육프로그램을 발견하게 되었고, 선발되어 프로그램에 참여하는 기회가 생겨서 너무 행복했다. 실감이 나지 않았는데 네바다주립대학교(UNLV)에 도착을 하고 나니 '진짜 미국에 왔구나'는 것을 실감하였고 너무 신기했다. UNLV에서는 영어로 수업을 하여 전공 분야의 영어 지식과 말하기 실력이 많이 상승하였다. 그리고 한국에서는 혹시 영어를 사용하면 말이 틀릴까 많이 걱정하였는데, 이 프로그램으로 영어 소통에 대한 자신감이 많이 상승하여 좋았다.

UNLV에서는 빅데이터에 대한 공부를 매일 하고 프로젝트까지 있어서 어떤 식으로 코드를 작성하고, 이 이론은 어떤 것인지 알게 되는 계기가 되었고, 자료 검색을 어떻게 하는지와 논문 읽고 쓰는 방법을 알게 되어 대학원 생활에 많은 도움이 될 것으로 생각한다. 미국에 오기 전에 AI 쪽 분야를 공부하고 싶고 대학원을 가야 할지 고민을 하고 있었는데, 이 프로그램을 통해 어느 정도 진로가 잡히는 계기가 되었고, AI에 더 다양한 분야가 있다는 것을 알게 되었다. 미국에 가기 전에는 그냥 두루뭉술하던 미래가 뭔가 체계적으로 좀 잡혀서 이 프로그램이 나에게 긍정적인 영향을 끼쳤다고 생각한다.

다양한 문화생활을 했다. 그중에서도 영화관 방문과 학교 시설에서의 운동 경험은 특별한 것이었다. 대형 스크린과 풍부한 사운드 시스템은 영화를 더욱 생생하게 만들었고, 내가 좋아하는 배우와 감독의 작품을 큰 화면에서 감상하는 것은 정말로 흥미로웠다. 미국의 영화 문화

는 진정으로 멋지다고 생각했다. 미국의 학교 시설에서의 운동은 정말로 다양한 활동을 즐길 수 있는 기회를 제공했다. 체육관에서 운동하면서 다양한 스포츠와 운동을 시도할 수 있었다. 실내 클라이밍, 수영, 배드민턴, 농구 등 다양한 운동을 지속적인 즐길 수 있었다. 체육관을 한국으로 가져가고 싶은 만큼 너무 좋은 시설이라고 생각한다.

다양한 캐년들을 탐험하며 자연이 만들어낸 신비로운 풍경을 마주할 수 있었다. 자이언트 캐니언, 그랜드 캐니언, 홀슈밴드 등 각기 다른 모습의 캐니언들을 방문하면서 미국 자연의 창조력에 감탄을 금치 못했다. 또한, 미국의 넓은 땅은 정말로 웅장한 자연경관을 담아내고 있었다. 우리나라도 아름답지만, 미국의 넓은 영토는 훨씬 다양한 풍경을 연출했다. 이 모든 경험은 미국 생활의 소중한 추억이 되었으며, 특히 캐니언 투어와 그 속에서 느낀 감동은 앞으로 잊지 못할 소중한 기억으로 남을 것이다.

연구프로그램을 통해 많은 친구들과 사귀게 되었고 다양한 문화를 경험할 수 있었다. 이러한 순간들은 미국에서의 여정을 풍요롭고 의미 있게 만들어 주었으며, 미래에 대한 나의 열정과 역량을 더욱 키워주는 중요한 계기가 되었다. 이 특별한 경험들은 평생 잊지 못할 추억으로 남을 것이다. 이 프로그램을 신청할지 고민하는 친구가 있다면 무조건 해보라고 추천해주고 싶다.

배 소 연
전북대학교 기계시스템공학과

단연 특별한 방학, 네바다 교육 프로그램

학과 머신러닝 동아리에서 활동하면서 이 분야에 매력을 느낄 수 있었고, 이를 보다 깊이 공부하고 싶은 열망을 가지고 있었다. 그러던 중 빅데이터 사업단의 '네바다주립대학교 교육프로그램'을 알게 되었다. 한 달의 시간 동안 깊이 있는 공부를 할 수 있는 기회이자, 해외 대학의 수업을 들을 수 있는 기회였다. 해외 여행은 앞으로도 다녀볼 수 있겠지만, 해외 대학에서 수업을 듣고 미국의 문화를 직접 경험할 수 있는 기회는 흔치 않다고 생각했다. 마지막으로, 이 프로그램을 통해 미국에서의 학업 경험뿐만 아니라 다양한 문화와 사람들을 만나는 기회를 통해 여러모로 성장할 수 있다고 생각했다. 이러한 이유로 프로그램에 참여하게 되었다.

프로그램 진행 과정에서 가장 인상 깊었던 활동은 생존 분석 프로젝트였다. 이 주제를 선택한 것은 처음에는 쉽지 않았지만, 결과적으로 굉장히 가치 있는 경험이었다. 본교에서는 프로젝트 경험이 거의 없었고, 학부 연구생 등의 활동도 없어 논문 분석 및 활용 경험이 없었다. 하지만 이 프로젝트를 통해 다양한 논문을 분석하고 새로운 지식을 습득하는 데 많은 도움이 되었다. 팀원들이 프로젝트 진행과 학습 방향을 제시해 주어 프로젝트 진행에 많은 도움이 되었다. 팀 프로젝트였기 때문에, 활발한 토의가 이루어졌고 이를 통해 새로운 연구 아이디어를 도출할 수 있었다. 프로젝트의 전체 과정을 경험해 본 것이 처음이었기에 매우 인상적이었다.

또한 투어 프로그램 중 가장 좋았던 것은 그랜드 캐니언에서 급류를 타고 내려오는 활동이었다. 그랜드 캐니언은 꼭 가보고 싶었던 여행지 중 하나였는데, 이번 프로그램을 통해 광활한 자연경관을 감상할 수 있어 기뻤다. 흔히 진행하는 버스 투어만으로는 느낄 수 없는 특별한 체험을 했고, 이를 통해 그랜드 캐니언의 아름다움을 더욱 깊게 느낄 수 있었다.

이번 프로그램을 통해 얻은 경험은 4년간의 대학 생활 중 단연 특별한 것으로 생각한다. 여러 학과에서 다양한 시각을 가진 친구들을 만나 새로운 커뮤니티를 만들게 되었다. 또한 머신러닝이라는 새로운 분야의 교육을 통해 진로 결정에서 선택의 폭을 넓힐 수 있었다. 만약 2024년에 프로그램이 또 진행되어 머신러닝과 빅데이터에 관심이 있는 학생들이 참여한다면, 대학 생활 중 가장 특별한 방학이 될 것이라고 확신한다. 마지막으로 이번 프로그램에서 함께 교육을 들으며 동고동락한 3개 학교의 학생들, 좋은 프로그램을 기획해 준 빅데이터 사업단 교직원분들, 교육과 투어를 진행해주신 네바다주립대 교수님들과 박사님들께 감사의 인사를 전한다.

서 수 빈
전북대학교 문헌정보학과

새로운 시작의 출발, UNLV 교육 !

Information science 분야로 미국 석사 진학을 염두에 두고 있었기 때문에 빅데이터 사업단에서 진행하는 프로그램은 무엇보다 더 솔깃한 제안이었다. 미국 대학 문화 혹은 수업 방식 등에 대한 경험과 지식이 없는 상태에서 나홀로 대학원 진학을 고민하는 과정은 마치 '맨땅에 헤딩하기'와 같았다. 그렇지만 이번 프로그램에는 머신러닝에 대한 교육뿐만 아니라 직접 연구를 진행해볼 수 있었기 때문에 굉장히 유익한 시간을 보낼 것 같아 참여했다.

숙소에서 학교까지 거의 매일, 라스베이거스의 50도 햇빛 아래 땀을 뻘뻘흘리면서 긴 거리를 걸어다녔다. 첫 주에는 프로젝트보다는 현지 환경에 적응하는 기간을 가졌다. 영어 수업에도 적응하기 위해 귀를 쫑긋하고 경청했으며, 내용도 프로그래밍 입문과 같은 가벼운 내용을 주로 다뤘다.

틈틈이 시간 나는 대로 대학도서관에서 본교 학생들과 함께 프로젝트를 진행했다. UNLV 와이파이를 사용하면 학교에서 제공하는 서버를 이용하는 것에는 문제가 없었기 때문에 주로 도서관에서 나머지 공부를 하고 숙소로 돌아갔다. 그뿐만 아니라 문헌정보학과 학생으로서, 미국의 대학도서관은 어떨지 무척 궁금했다. 아쉽게도 방학이라 실제 사용하는 이용자들은 많이 없었지만, 우리는 UNLV 학생들과 동등하게 모든 시설을 이용할 수 있었다. 미국 대학 도서관답게 칠판이나 그룹 스터디 환경이 잘 조성되어 있어서 너무 편하게 도서관에서 발표 준비를 할 수 있었다. ULNV에서 좋은 프로그램 덕분에 좋은 교수님으로부터 많은 지도를 받기도 했지만, 실상은 이렇게 도서관에 앉아 우리끼리 머리를 굴리면서 이것저것 시도해봤던 시간이 정말 큰 성장을 할 수 있게 만들었다. 영어로 진행되는 프로젝트와 생소한 연구 방법 등에 좌절될 때가 종종 있었지만, 그럴 때마다 서로 의지할 수 있는 좋은 친구이자 동료들이 있어 더욱 아름답게 기억될 것 같다.

이번 프로그램을 통해 석사 유학의 방향을 결정할 수 있게 되었고, 굳게 의지를 다질 수 있게 되는 큰 계기가 되었다. 좋은 시기에 좋은 프로그램으로 지원받을 수 있어 너무 감사하다.

연효진
전북대학교 분자생물학과

새롭고 소중했던 28일

평소처럼 학교 공지 사항을 살펴보며 참여할 만한 흥미로운 프로그램을 찾고 있었다. 그러던 중 '해외 연구프로그램'이라는 글이 눈에 띄었다. 고등학교 때부터 대학교에 오면 다양한 해외 활동을 하겠다는 다짐을 해왔지만, 코로나 팬데믹으로 인해 그런 다짐을 오랫동안 미뤄야만 했다. 미국 라스베이거스에서 진행되는 이 프로그램은 한 번도 가보지 않은 미국에서 한 달간의 생활을 통해 나의 해외 경험에 대한 소망을 이룰 수 있을 것 같았다. 그뿐만 아니라, 빅데이터 및 인공지능에 대한 교육과 관련 프로젝트를 진행한다는 점은 더욱 프로그램에 참여하고 싶게 만들었다. 인공지능 및 빅데이터 기술은 다양한 학문 발전에 기여하고 있었고, 내가 전공하는 생물학 분야도 예외는 아니었다. 생물학과 인공지능이 결합한 연구들이 계속해서 진행되고 있는 시점에서 프로그램에 생물학적 데이터를 기반으로 인공지능 기술을 활용하는 프로젝트도 포함하고 있었다. 나는 생물학과 정보학의 결합 분야인 생물정보학을 전공으로 삼아 대학원 진학을 목표로 하고 있었고, 이 프로그램은 내 능력을 성장시킬 기회라고 생각해 지원하게 되었다.
미국에서의 4주간의 생활은 쉽게 잊지 못할 소중한 시간이었다. 일상에서 벗어난 낯선 곳에서의 생활은 무엇이든 경험이 되었다. 어떻게 보면 한국에서의 생활과 다를 바 없이 학교에서 수업을 듣고, 점심을 먹고, 프로젝트를 진행하는 일들이었지만, 미국이라는 새로운 환경에서 새로운 언어로 수업을 듣고, 새로운 음식을 맛보고, 새로운 주제의 프로젝트를 진행한다는 점에서 모든 것이 달랐다. 수업이 끝나고 다 같이 장을 보러 마트에 가는 사소한 일도 즐거웠다. 한국과 달리 저렴한 소고기 가격에 눈이 돌아 소고기를 한 움큼 사기도 하고, 인기 있는 미국 국민 과자를 찾아 이것저것 먹어보기도 했다. 먹기 전까진 전혀 알 수 없었던 미국 복숭아의 단맛에 충격받아 복숭아를 사기 위해 마트에 가기도 하는 등 일상에서도 즐겁고 재밌는 일들이 참 많았다.

'유전체 데이터를 이용한 딥러닝 기반 생존분석'. 우리 팀이 선택한 주제였다. '유전체 데이터', '딥러닝', '생존분석'. 어느 하나 흥미롭지 않은 키워드가 없었다. 하지만 '생존 분석'은 이전에 들어본 적도 없었고, 일반적으로 사용되는 통계 분석 기법과는 달랐기에 팀원 모두가 해당 분야에 대한 사전 지식을 쌓을 필요가 있었다. 프로젝트를 진행하며 생존 분석에 대한 기본 개념부터 관련 연구에 대한 최신 논문들을 찾기까지 팀원 모두가 노력하고, 공유하는 과정이 지속되었다. 모두 다른 전공을 가진 팀원이 서로 다른 배경지식 위에서 본인들이 이해한 것을 설명하고 모르는 것을 해결해나가며 의견들을 조율해가는 과정은 감사한 경험이었다. 그 과정 동안 서로를 존중하고 이해할 수 있도록 개개인이 노력한 것에 감사하고, 앞으로 내가 다양한 사람들과 협력하는 데 있어 좋은 경험이 될 수 있다는 점 또한 감사하다고 생각되었다. 이러한 노력 끝에 우리는 연구 주제 결정부터 실험 진행 및 결론 도출까지 프로젝트를 잘 이끌어 마무리할 수 있었다.

가장 기억에 남는 건 미국의 국립 공원 방문이다. 처음 Zion National Park에 도착했을 땐, 산맥 규모에 놀

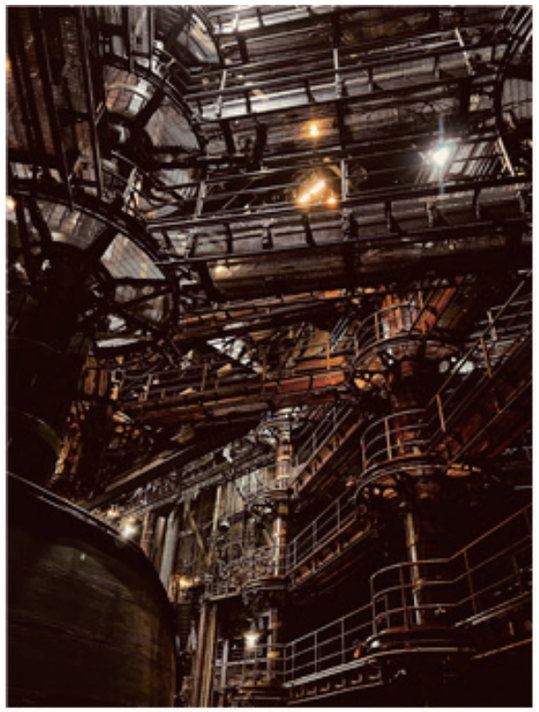
랐었다. 우리나라의 산처럼 푸르지 않고, 하얗거나 붉은색을 띠고 있어 전혀 다른 느낌이었다. 우리는 계곡에서 시원한 수영을 즐기면서 멋진 풍경을 볼 수 있었다. 높은 산맥 사이로 물살을 타고 흘러 내려가는 경험은 물과 하나된 듯 생생하고 신나는 경험이었다. 물놀이를 즐기곤 Horseshoe Bend로 향했다. 노을이 질 때쯤 도착해, 모두들 인생사진을 건진다며 사진 찍기에 열중이었다. 그만큼 말굽 모양의 신기한 지형과 노을이 참 예뻤던 곳이었고, 함께한 친구들과 마음에 드는 사진을 남길 수 있어 좋았었다. 이후에는 숙소에서 바비큐 일정이 있었다. 인생에서 한 번쯤은 봐야 한다는 풍경들을 보고 그리웠던 김치찌개와 삼겹살을 먹었다. 미국에 와서 가장 감격적인 순간이 아닐 수가 없었다. 저녁을 먹고 다같이 별 구경에 빠진 것도, 불멍하며 마시멜로를 구워먹던 것도 모두 좋은 추억이 되었다.

다음 날은 아침 일찍 Entelope Canyon 방문이 예정되어 있었다. 윈도우 바탕화면에서 한 번쯤 봤던 이곳을 실제로 본다는 생각에 들떠있었다. 다른 협곡들과는 다르게 땅 밑으로 펼쳐진 협곡이 신기했고, 뜨거운 햇빛을 피할 수 있던 점이 좋았다. 철 성분이 많아 붉은빛을 띤다는 모래들이 지형을 더 심미적으로 만드는 것 같았고, 머리 위로 햇빛이 들어올 때 더욱 예쁜 곳이었다. Grand Canyon에 처음 도착했을 땐 차로 이동하다 보니 높이를 가늠하기 어려웠지만, 내려서 가만히 풍경을 보니 '내가 보고 있는 게 맞나' 싶은 높이와 규모의 협곡이었다. 한눈에 내려다보이는 거리가 분당과 서울까지의 거리라는 말을 들으며 거리 감각이 사라지고, 사진으로는 담기지 않을 웅장함을 가진 곳이었다.

해외 연구프로그램은 새로운 곳에서 새로운 사람들과 함께 나아갔던 쉽게 경험할 수 없는 값진 시간의 연속이었다. 학습적 성취뿐만 아니라 개인의 성장에도 큰 영향을 주었고, 도전적인 경험을 통해 자신의 사고를 다듬어 가는 소중한 시간이었다. 새로운 사람들이 소중한 인연이 되어 프로그램을 마무리한다는 것 또한 감사한 경험이었다.

이 용 환
전북대학교 컴퓨터인공지능학부

Fabulous Trip in Fabulous Las Vegas!



빅데이터 해외프로그램에 참가한 경험은 매우 의미 있는 시간이었다. 이를 통해 다양한 경험과 인사이트를 얻게 되었다. 전공과 진로가 인공지능에 관련되어 있어서 빅데이터와 관련된 경험을 쌓고자 참가하게 되었다. 또한, 미국에 대한 호기심과 이해를 높이고 싶었다.

프로그램을 통해 4주간 다양한 경험을 하게 되었다. 라스베이거스와 주변 지역을 방문하면서, 사막 한가운데에서도 어마어마한 도시를 건설한 미국의 위력과 힘을 실감하게 되었다. 또한, 네바다주립대의 수준 높은 인공지능 강의를 들으며 미국 대학 교육의 질이 얼마나 높은지를 느낄 수 있었다.

프로그램 수행 내용 중에서는 생존 분석(Survival Analysis) 프로젝트를 진행하면서, 전공 분야와는 다른 주제에 대한 지식을 확장하는 경험을 했다. 또한, 개인적으로 좋았던 교육과 투어 프로그램은 매우 알찼다. 특히, 미국 서부의 자연을 느낄 수 있는 Grand Canyon, Zion Canyon 등의 장소는 잊지 못할 경험으로 남았다.
프로그램의 장점으로는 한 달이라는 짧은 기간에도 비교적 저렴한 비용으로 미국의 문화를 체험하고 인공지능에 대한 교육을 받을 수 있다는 것이었다. 그러나 한 달 동안 다양한 프로그램이 집중적으로 몰려있어서 몸과 마음이 지치는 순간도 있었고 그래서 프로그램의 완급 조절이 필요하다고 생각한다. 또 프로젝트 도메인이 Computer Vision에 집중되어 있어서 아쉬웠다. 다음 프로그램에서는 프로젝트 도메인 다양성을 더 고려한다면 보다 풍성한 프로그램이 될 수 있다고 생각한다.

참가 전후로 미국에 대한 인식의 폭이 확대되었다. 미국은 다양한 문화와 경험을 제공하는 나라임을 알게 되었고 이 경험을 토대로 앞으로는 AI 관련 공부를 더 깊이 할 것이다. 특히 이전에는 큰 관심이 없었던 북미 유학에 대한 관심이 증대되었다.

보빙사의 일원으로 미국을 방문해 조선으로 돌아온 민영익은 이렇게 말했다.

"나는 어둠 속에서 태어났다가 광명 속으로 들어갔습니다. 그리고 다시 어둠 속으로 돌아왔습니다. 아직 나는 내가 갈 길을 분명하게 내다볼 수가 없으나, 머지않아 찾아낼 수 있기를 바랄 뿐입니다."

나 역시 미국에서 광명을 보았다. 라스베이거스 고급 호텔들의 엄청난 규모, 놀라운 스케일의 공연, 압도적으로 거대한 자연환경 등 대국의 힘을 느꼈다. 그 속에서 내가 배워야 할 점들을 배우고 얻어왔다. 후버댐을 보면서는 공학자의 꼼꼼함을, 호텔의 대규모 극장을 보면서는 체계적인 시스템 관리의 필요성을 배웠다. 그렇기에 나는 민영익의 말을 조금 바꾸며 소감을 마치고자 한다.

"나는 광명 속으로 들어가 눈부신 세상을 보았습니다. 이제 그 한 줄기 빛을 가져와 무엇을 빛낼지 고민 중입니다."

4 팀별 프로젝트 완료 보고서

Team project reports

순번	주제		
	소속	학년	이름
1	**Survival Analysis**		
	경기과학기술대학교 컴퓨터모바일융합공학과	3	고진영
	경기과학기술대학교 컴퓨터모바일융합공학과	3	김강현
	경기과학기술대학교 인공지능학과	2	김민선
2	**Pathological Image Analysis for Cancer Classification**		
	경기과학기술대학교 컴퓨터모바일융합공학과	3	이재은
	경기과학기술대학교 컴퓨터모바일융합공학과	3	김도현
	경기과학기술대학교 컴퓨터모바일융합공학과	3	함성영
3	**Deep Learning Based Survival Analysis Using Genomic Data**		
	경상국립대학교 산업시스템공학부	4	이인호
	경상국립대학교 산업시스템공학부	4	신서빈
	경상국립대학교 정보통계학과	4	정인영
4	**Segmenting Road Traffic Objects Using CNN and RNN Architectures**		
	경상국립대학교 수학과	4	최규진
	경상국립대학교 물리학과	4	김가현
	경상국립대학교 수학과	3	강동현
5	**Deep Learning Based Survival Analysis Using Genomic Data**		
	경상국립대학교 정보통계학과	4	이지상
	경상국립대학교 정보통계학과	4	이태훈
	경상국립대학교 산업시스템공학부	4	정현서
	경상국립대학교 산업시스템공학부	4	박건영
6	**Cancer Classification**		
	전북대학교 컴퓨터인공지능학부	4	김나영
	전북대학교 컴퓨터인공지능학부	4	김미현
	전북대학교 소프트웨어공학과	3	박세현
7	**Deep Learning Based Survival Analysis Using Genomic Data**		
	경상국립대학교 정보통계학과	4	이지상
	경상국립대학교 정보통계학과	4	이태훈
	경상국립대학교 산업시스템공학부	4	정현서
	경상국립대학교 산업시스템공학부	4	박건영
8	**Scalable Data Processing: By applying MapReduce in Spark**		
	전북대학교 스마트팜학과	3	강다영
	전북대학교 소프트웨어공학과	4	김재현
	전북대 문헌정보학과	4	서수빈

Survival Analysis

KangHyun Kim
SeongJae Nam
JinYoung Ko
MinSeon Kim

Survival Analysis

KangHyun Kim, SeongJae Nam, JinYoung Ko, MinSeon Kim

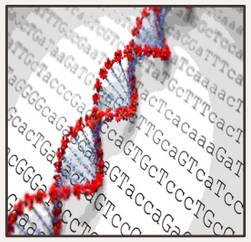
Table of contents

01 Introduction

How to reduce the number of features

1. **COX model**
—-> Reduce feature to about 2200

2. **Lasso + Cox Model**
—> Reduce feature to about 1200



Training Loss & Validation Loss plot

```
from DataLoader import load_data
from Train import trainCoxmodel

import torch
import numpy as np

import matplotlib.pyplot as plt

def plot_learning_curve(train_loss, val_loss,epoc):
    epochs = range(epoc)

    plt.plot(epochs, train_loss, 'r', label='Training loss')
    plt.plot(epochs, val_loss, 'b', label='Validation loss')
    plt.title('Training and Validation Loss')
    plt.xlabel('Epochs')
    plt.ylabel('Loss')
    plt.legend()

    plt.show()
```

Select Activation Function & Dropout_rate & Learning_rate

```
dtype = torch.FloatTensor
''' Net Settings'''
In_Nodes = 1202  ###number of genes
Hidden_Nodes = [600,300,150] ###number of hidden nodes
Out_Nodes = 1 ###number of hidden nodes in the last hidden layer
''' Initialize '''

Activation=['sigmoid', 'tanh','relu']
dropout_rate = [0.5,0.6,0.7, 0.8]
learning_rate = [0.001, 0.0001,0.0001,0.00001]

Num_EPOCHS = 20000 ###for training


''' load data and pathway '''

x_train, ytime_train, yevent_train = load_data("train_df0.csv", dtype)
x_valid, ytime_valid, yevent_valid = load_data("val_df0.csv", dtype)
x_test, ytime_test, yevent_test = load_data("test_df0.csv", dtype)

opt_loss = torch.Tensor([float("Inf")])
###if gpu is being used
if torch.cuda.is_available():
    opt_loss = opt_loss.cuda()
```

Calculate Loss & C_Index

```
hyper_group=[]
train_group=[]
val_group=[]
cindex_group=[]
for act in Activation: ### sigmoid, tanh
    for lr in learning_rate: ### 1e-3, 1e-4
        for dr in dropout_rate: ### 0.7, 0.8
            train_loss_list, val_loss_list, test_cindex = trainCoxmodel(x_train, ytime_train, yevent_train, ₩
                                                                        x_valid, ytime_valid, yevent_valid,₩
                                                                        x_test, ytime_test, yevent_test, ₩
                                                                        In_Nodes, Hidden_Nodes, Out_Nodes, ₩
                                                                        lr, Num_EPOCHS,act, dr)


            hyper_group.append([act,lr,dr])
            train_group.append(train_loss_list)
            val_group.append(val_loss_list)
            cindex_group.append(test_cindex)
            print(act, lr,dr)
#             plot_learning_curve(train_loss_list, val_loss_list,Num_EPOCHS)
```

First Case (36 ways)

Sigmoid, tanh, relu ⇒ Activation

0.001, 0.0001, 0.00001 ⇒ Learning Rate

0.5, 0.6, 0.7, 0.8 ⇒ Drop out Rate

600, 300 => number of nodes for each hidden layer

epochs = 20000

Cindex = About 83.3%

Second Case (48 ways)

Sigmoid, tanh, relu⇒ Activation

0.001, 0.0001, 0.00001 ⇒ Learning Rate

0.5, 0.6, 0.7. 0.8 ⇒ Drop out Rate

600, 300, 150 => number of nodes for each hidden layer

epochs = 20000

Cindex = About 81.7%

Third Case (24 ways)

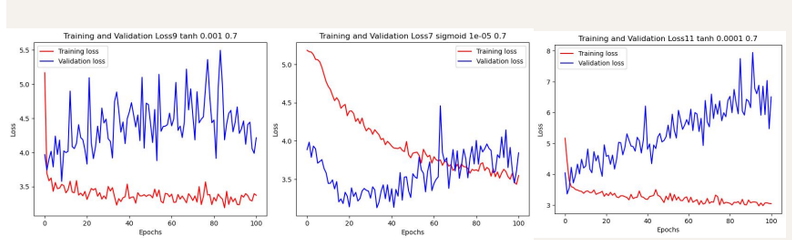
Sigmoid, tanh ⇒ Activation

0.001, 0.0001, 0.0001 ⇒ Learning Rate

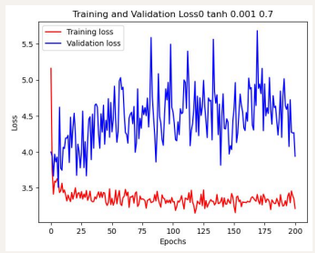
0.5, 0.6, 0.7, 0.8 ⇒ Drop out Rate

300, 150, 50 => number of nodes for each hidden layer
epochs = 20000

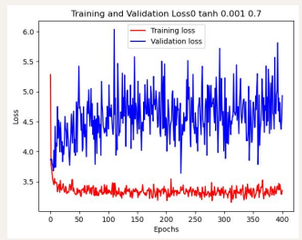
Q&A



Epochs = 40000



Epochs = 80000



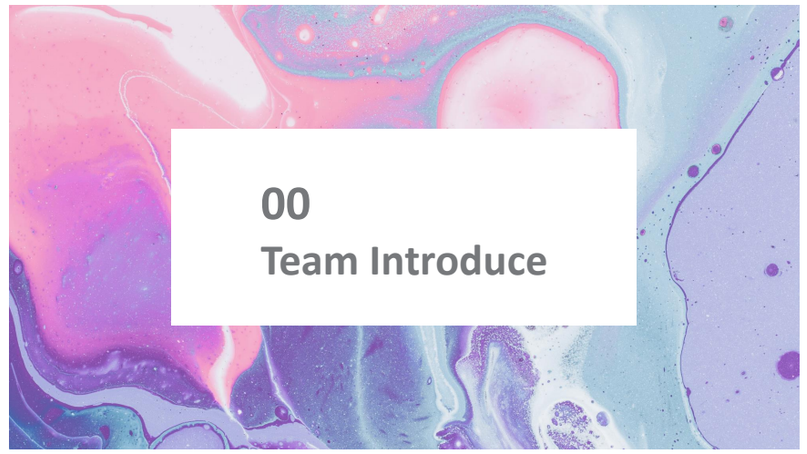
Summer Big Data Project

Jae Eun Lee
Do Hyun Kim
Seong Yeong Ham

Summer Big Data Project

Dark Mode
Jae Eun Lee, Do Hyun Kim, Seong Yeong Ham

Contents



00 Team Introduce
Introduce Team's Name / Member's do list

Jea Eun Lee
Coding
Code Analysis

Seong Yeong Ham
Error detection & Modify
Making PPT

Do Hyun Kim
Coding
Last Coding Organize

00 Team Introduce
Until Middle Presentation

Take a Data from Sai

- All preprocessing Data
- Patches done

01 Subject
Team's Subject

3. Pathological Image Analysis for Cancer Classification

01 Subject
Team's Objective

Team's Objective

Failed!! ☹

Development of a Data Integration and Computer Vision-Based Deep Learning Model for Cancer Prognosis Prediction & Recurrence.

00 10 Model
Using Model

First Try

CNN
- Sequential Model
- The Sequential model is a linear stack of layers,
  where you add layers one by one to build the neural network

ReLU(Rectified Linear Unit)

Conv2D

>> Simply CNN

00 10 Model
Using Model

```
# Define the model architecture
model = Sequential()
model.add(Conv2D(32, (3, 3), activation='relu', input_shape=(IMAGE_SIZE, IMAGE_SIZE, 3)))
model.add(Conv2D(32, (3, 3), activation='relu'))
model.add(Conv2D(32, (3, 3), activation='relu'))
model.add(MaxPooling2D(pool_size=(2, 2)))
model.add(Dropout(0.3))

model.add(Conv2D(64, (3, 3), activation='relu'))
model.add(Conv2D(64, (3, 3), activation='relu'))
model.add(Conv2D(64, (3, 3), activation='relu'))
model.add(MaxPooling2D(pool_size=(2, 2)))
model.add(Dropout(0.3))

model.add(Conv2D(128, (3, 3), activation='relu'))
model.add(Conv2D(128, (3, 3), activation='relu'))
model.add(Conv2D(128, (3, 3), activation='relu'))
model.add(MaxPooling2D(pool_size=(2, 2)))
model.add(Dropout(0.3))

model.add(Flatten())
model.add(Dense(256, activation='relu'))
model.add(Dropout(0.3))
model.add(Dense(2, activation='softmax'))
```


00 10 Model
Using Model

Adam(Adaptive Moment Estimation)
- Optimize Function

```
# Compile the model
model.compile(tf.keras.optimizers.Adam(learning_rate=0.0001), loss='binary_crossentropy', metrics=['accuracy'])
```


Failed!!

Forget a GPU…
why……

```
# Get the metric names
metric_names = model.metrics_names

# Evaluate the model on the validation set
val_loss, val_acc = model.evaluate_generator(val_gen, steps=len(val_gen))
print('val_loss:', val_loss)
print('val_acc:', val_acc)

# Get the labels of the test images
test_labels = test_gen.classes

# Make predictions on the test set
predictions = model.predict_generator(test_gen, steps=len(test_gen), verbose=1)

# Get the predicted labels as probabilities
y_pred = predictions.argmax(axis=1)

# Generate the confusion matrix
cm = confusion_matrix(test_labels, y_pred)

# Plot the confusion matrix
plt.figure(figsize=(8, 6))
sns.heatmap(cm, annot=True, fmt='d', cmap='Blues')
plt.title('Confusion Matrix')
plt.xlabel('Predicted Labels')
plt.ylabel('True Labels')
plt.show()

# Calculate the ROC AUC score
roc_auc = roc_auc_score(test_labels, y_pred)
print('ROC AUC:', roc_auc)
```


01 10 Model
Difference other model

Pytorch…!

It can be used for transfer learning with pretrained weights, making it suitable as initial weights for image segmentation tasks.

Sigmoid Activation Function Selection

Binary Cross-Entropy (BCE) Loss

widely used for binary classification tasks.

BCELoss computes the cross-entropy between the true labels and the predicted values.

01 10 Model
Difference other model

Adam (again)

Set Hyper Parameters

```
BATCH_SIZE = 16
NUM_EPOCH = 10
```

These values represent the batch size used during training and the number of epochs, respectively.

01 10 Model
Difference other model

```python
class SegmentationModel(pl.LightningModule):
    def __init__(self):
        super(SegmentationModel, self).__init__()
        self.resnet = torch.hub.load('pytorch/vision:v0.9.0', 'resnet50', pretrained=True)
        # 이진 분류를 위해 출력 노드 수를 1로 조정하고 sigmoid 함수를 적용
        self.fc = nn.Sequential(
            nn.Linear(2048, 1),
            nn.Sigmoid()
        )
        self.lr = 1e-5  # learning rate를 1e-5로
        self.criterion = nn.BCELoss()
```

The fc layer is added to adjust the output node count to 1 and applies a Sigmoid function to convert the problem into binary classification.

01 10 Model
Difference other model

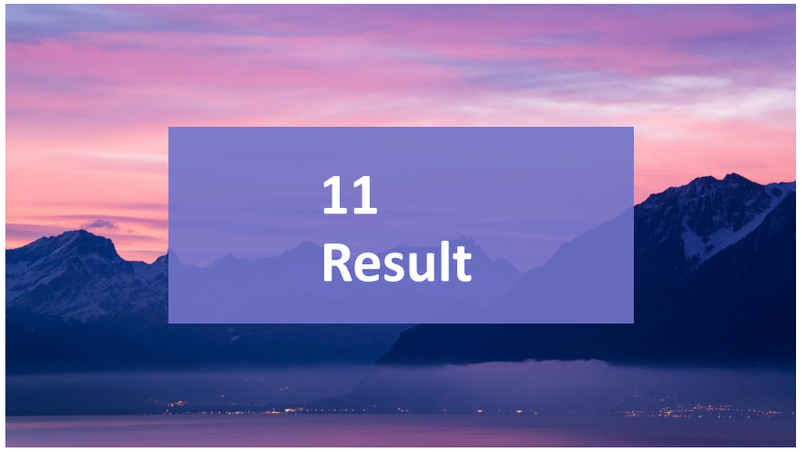
```python
# Trainer를 설정하고 모델을 학습 (Gradient Accumulation 설정)
trainer = pl.Trainer(max_epochs=NUM_EPOCH, accumulate_grad_batches=2)
trainer.fit(model, train_loader, val_loader)
```

Use gradient accumulation

The parameter accumulate_grad_batches=2 is set to perform gradient accumulation of 2, which helps effectively train the model with smaller batch sizes due to memory constraints.

Error

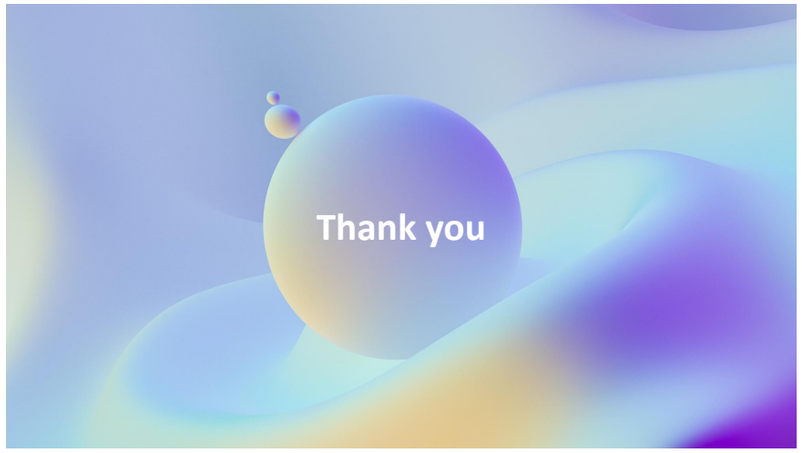
```
For debugging consider passing CUDA_LAUNCH_BLOCKING=1.
Compile with `TORCH_USE_CUDA_DSA` to enable device-side assertions.
```

00 11 Result
Project Result

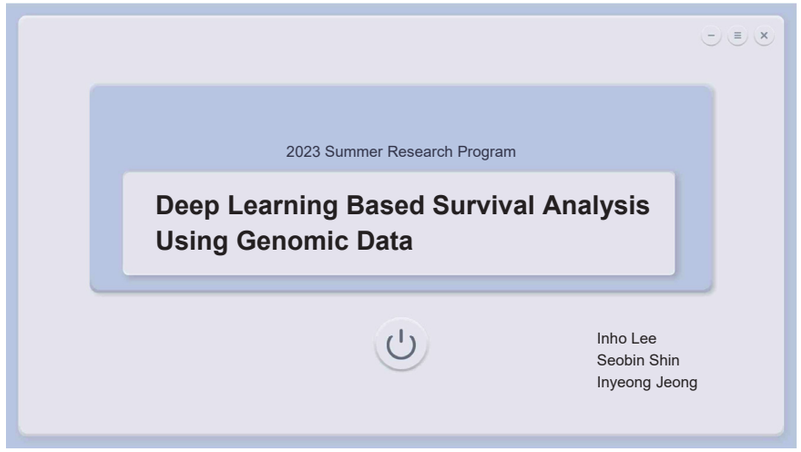
All failed...

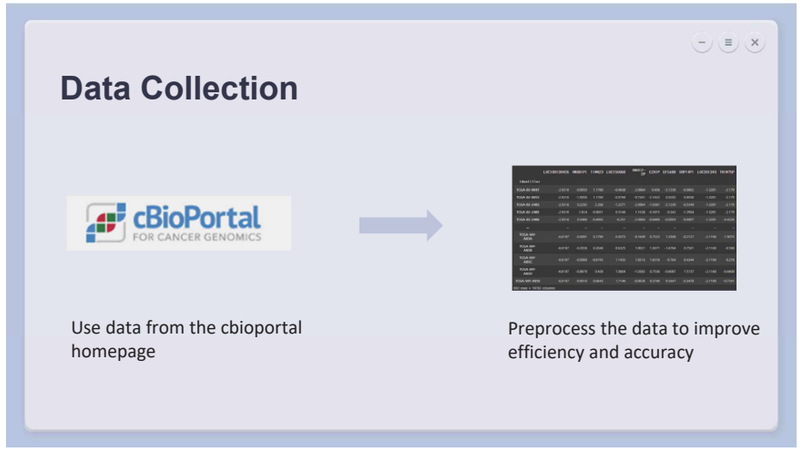
But it has a meaning



Deep Learning Based Survival Analysis Using Genomic Data

Inho Lee
Seobin Shin
Inyeong Jeong





Feature Selection

- Use the **Cox-PH** model to select the appropriate feature for survival analysis
- **Cox-PH**

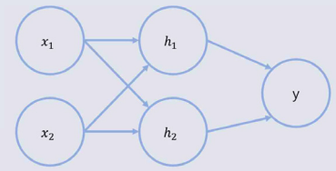
  h(t|X) = h0(t) * exp(b1X1 + b2X2 + ... + bp*Xp) (t = time)

  Use to evaluate the association between the survival time of patients and one or more predictor variables.
- Reduce the number of features **19761-> 3436**

Model Learning

Neural Network

- Feedforward Networks or Multilayer Perceptrons



Model Learning

Design of Neural Network

- **Activation function**: to compute the hidden layer values
  ⇒ ReLu, Sigmoid, Softmax
- **Cost function**: to optimize the model
  ⇒ Negative Log-likelihood Cost
- **Optimizer**: how to optimize the model
  ⇒ Adjusting Hyperparameter Values

Model Training

- K-Fold CV ⇒ k = 5
- Run model training according to the set hyperparameter value
- Calculation of Loss values according to iterations

Model Training

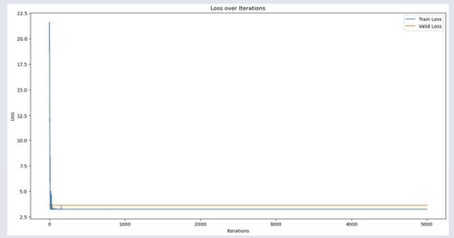
ReLU **Sigmoid** **Softmax**

Model Training (ReLU)

HYPERPARAMETERS:

dropout_prob = 0.9
activation = 'ReLU'
learning_rate = 0.001
epochs = 5000
batch_size = 64
KFold = 5

C-index = 0.5

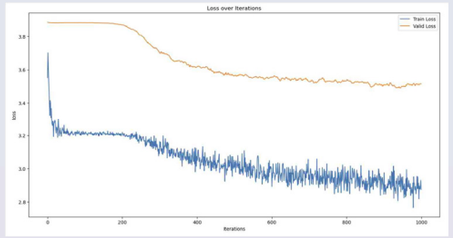


Model Training (Sigmoid)

HYPERPARAMETERS:

dropout_prob = 0.9
activation = 'Sigmoid'
learning_rate = 0.001
epochs = 1000
batch_size = 64
KFold = 5

C-index = 0.7562466859817505

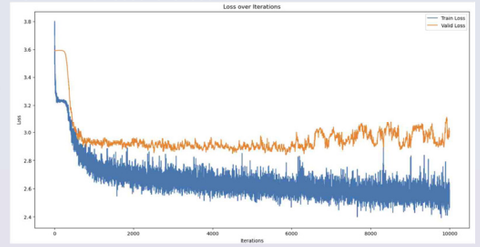


Model Training (Sigmoid)

HYPERPARAMETERS:

dropout_prob = 0.9
activation = 'Sigmoid'
learning_rate = 0.0007
epochs = 10000
batch_size = 64
KFold = 5

C-index = 0.7505518794059753

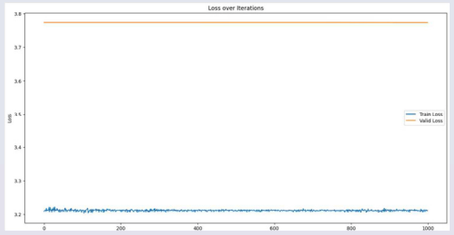


Model Training (Softmax)

HYPERPARAMETERS:

dropout_prob = 0.9
activation = 'Softmax'
learning_rate = 0.001
epochs = 1000
batch_size = 64
KFold = 5

C-index = 0.692832350730896

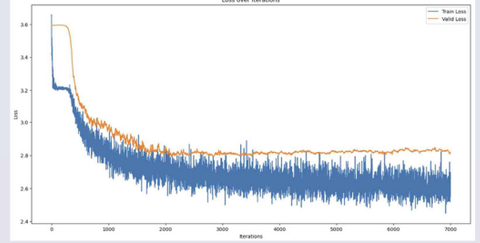


Model Training (Sigmoid)

HYPERPARAMETERS:

dropout_prob = 0.9
activation = 'Sigmoid'
learning_rate = 0.0007
epochs = 7000
batch_size = 64
KFold = 5

C-index = 0.810590386390686

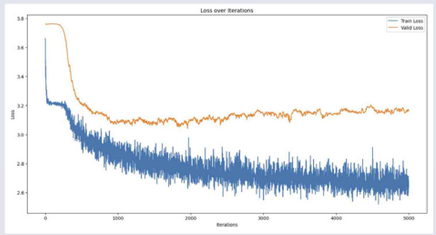


Model Training (Sigmoid)

HYPERPARAMETERS:

dropout_prob = 0.9
activation = 'Sigmoid'
learning_rate = 0.001
epochs = 5000
batch_size = 64
KFold = 5

C-index = 0.822630763053894



Evaluation

- Use **c-index** to evaluate models
- **C-Index**
  Survival data usually includes censored data
  ⇒ Use c-index instead of general indices

  C-Index ∝ model's accuracy

Evaluation

Sigmoid &

learning_rate = 0.007 &

epochs = 7000

best result during trial

Mean Validation Loss: 2.887149990218026

Standard Deviation of Validation Loss: 0.1787564699351698

Conclusion

References

[1] https://link.springer.com/article/10.1007/s12055-020-01108-7

[2] https://www.graphpad.com/guides/prism/latest/statistics/stat_censored_data.html

[3] https://www.nature.com/articles/s41598-019-43372-7

[4] https://ieeexplore.ieee.org/abstract/document/9794670

Segmenting Road Traffic Objects Using CNN and RNN Architectures

Gyujin Choi
Gahyoun Gim
Donghyeon Kang

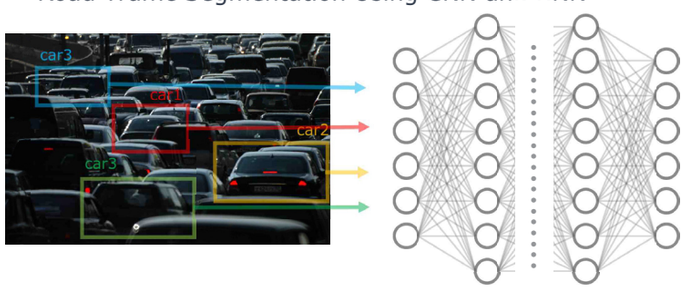
Segmenting Road Traffic Objects Using CNN and RNN Architectures

Gyujin Choi
Gahyoun Gim
Donghyeon Kang

Table of Contents

I. Topic

Road Traffic Segmentation Using CNN and RNN



II. Pros and cons of the model

Supervised Learning

- **Pros**
  - **Accuracy**
    - Since the model learns from labeled data, it's highly likely to produce accurate predictions.
  - **Interpretable Results**
    - The interpretation of the results predicted by the model is relatively clear, which aids in analyzing the relationship between predictors and outcomes.
- **Cons**
  - **Data Requirement**
    - Making model needs a large amount of labeled data, and labeling requires considerable time and effort.
  - **Overfitting**
    - Complex models may learn too well from the training data, reducing their ability to generalize to new data.

II. Pros and cons of the model

Unsupervised Learning

- **Pros**
  - **No need for data labeling**
    - can find hidden patterns or structures in unlabeled data.
  - **Useful with large datasets**
    - It is effective in extracting useful information from a vast amount of data.
- **Cons**
  - **Interpretation of results**
    - It is challenging to interpret and understand the results.
  - **Predictive power**
    - Without labels, it may have lower predictive power compared to supervised learning.

II. Pros and cons of the model

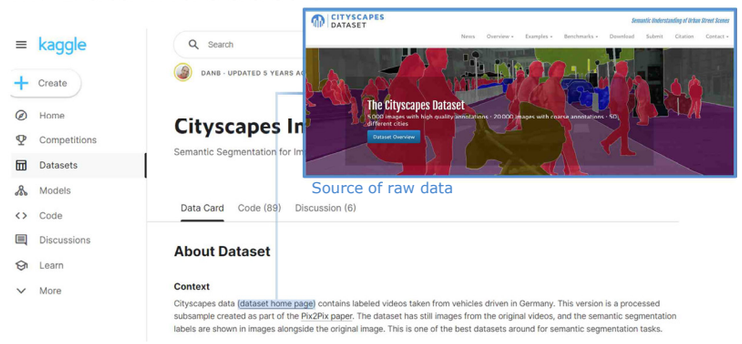
Semi-Supervised Learning

- **Pros**
  - **Efficient label usage**
    - Only a fraction of the large amount of data needs to be labeled, saving time and effort.
  - **Improved Accuracy**
    - It can achieve higher performance than using only supervised learning by utilizing unlabeled data.
- **Cons**
  - **Labeling errors**
    - Errors in labeling a portion of the data can affect the overall model performance.
  - **Increased Complexity**
    - The algorithm can be more complex and harder to implement compared to supervised learning.

III. Data introduction

kaggle

III. Data introduction

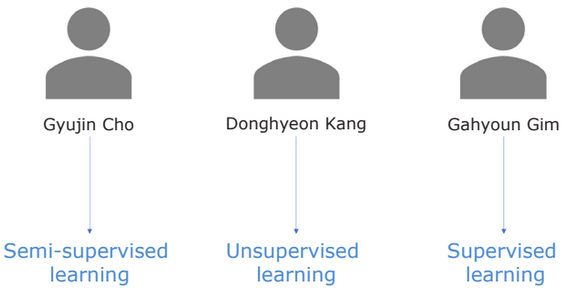


Source of raw data

III. Data introduction

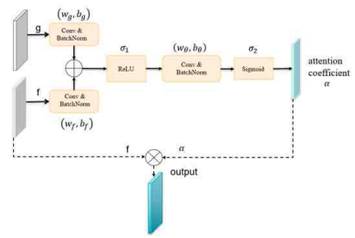
- dataset
  - **2975 training** images files
  - **500 validation** images files
- image file
  - 256x512 pixels
  - original photo on the left half of the image
  - t… …f.

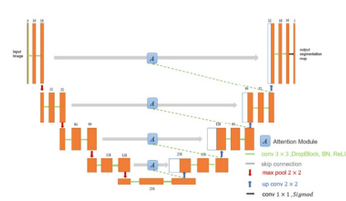
IIII. Team members' work



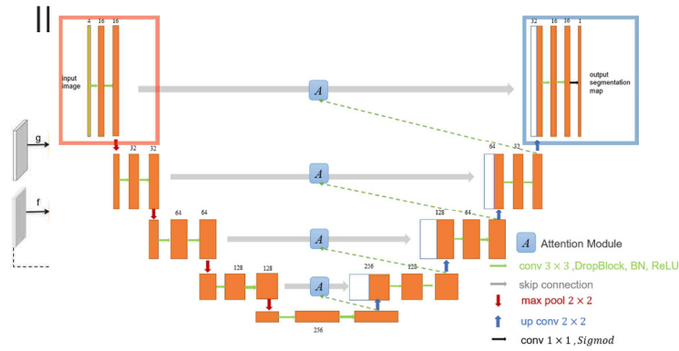
IIII-1. Supervised learning model

Guo, Y., Cao, X., Liu, B., & Gao, M. (2020). Cloud Detection for Satellite Imagery Using Attention-Based U-Net Convolutional Neural Network. Symmetry, 12(6), 1056. https://doi.org/10.3390/sym12061056





II



IIII-1. Supervised learning model

Introduction to model training methods

- # Contraction path × 3 times
  - Configuring a neural network
    - Convolution Layer
      - FILTERS = 32, KERNEL = 3, activation='relu', padding='same'
    - Pooling Layer
      - pool_size=(2, 2)

IIII-1. Supervised learning model

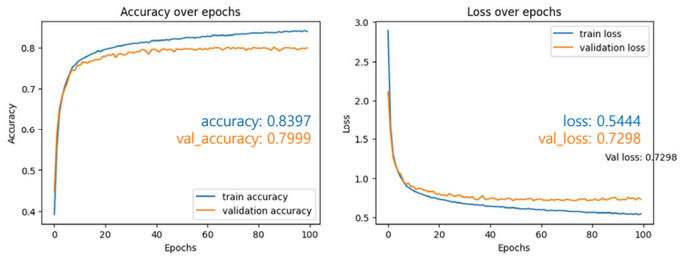
Introduction to model training methods

- # Expansion path × 2 times
  - Concatenate
    - Conv2DTranspose
      - FILTERS*2, (2, 2), strides=(2, 2), padding='same'
  - Convolution Layer & Pooling Layer
- output
  - activation='softmax'

- **Epoch: 100**

IIII-1. Supervised learning model

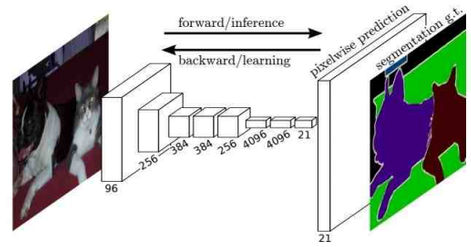
Results of the trained model



IIII-2. Semi-Supervised learning

Jonathan Long, Evan Shelhamer, Trevor Darrell, FCN, CVPR 2015

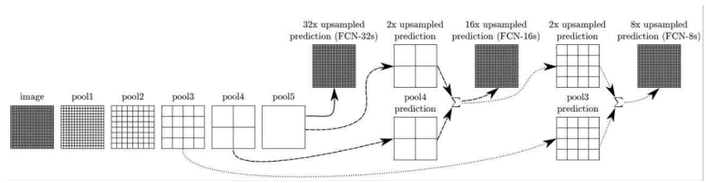
- Fully Convolutional Network



IIII-2. Semi-Supervised learning

Model Introduction(FCN)

- Fully Convolutional Network
  - Using FCNs structure
    - Convolution layer: kernel size(3,3), activation function(relu), padding(same)
    - pooling layer: kernel size(2,2)
    - Upsampling layer



IIII-2. Semi-Supervised learning

Introduction to model training method(Semi-Supervised learning)
Dali Chen, Yingying Ao, Shixin Liu, MDPI 2020 / Dong-Hyun Lee, research gate 2013

Algorithm : Semi-supervised learning method for image segmentation

Input: Training dataset $D = \{x_n, y_n, x_m;\ n = 1:N,\ m = 1:M\}$,
Updated dataset $D^* = \{x_n, y_n, x_m, y_m;\ n = 1:N,\ m = 1:M\}$,
Test dataset $\{x^*\}$

Output: The mask image y^*

STEP 1: Initialization

- Learning method: epoch, batch size, learning rate
- Divide train data into labeled and unlabeled data
- Initial training set D.
- Initial FCN model.

STEP 2: Update the FCN model parameters θ

STEP 3: Predict the pseudo-label y_m

STEP 4: Update training dataset D^*

STEP 5:

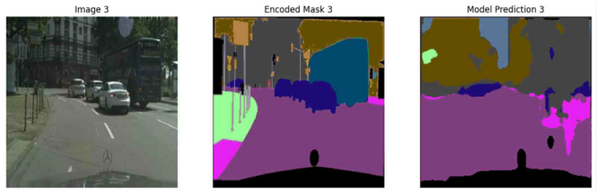
- **While** a stopping criterion is not met **do**
  - Update the FCN model parameters θ using the updated training dataset D^*
  - Return to **STEP 3**
- **end while**

STEP 6: Take x^* **as input and compute** y^*

- Epoch = 10
- Batch size = 32
- Learning rate = 0.001
- Stop criterion: repeat a certain number of time

IIII-2. Semi-Supervised learning

Result of trained model



Mean IOU: 0.16

IIII-3. Unsupervised learning

$$Var = \frac{Var_R + Var_G + Var_B}{3}$$

$$minlabels = (Var//10) + 3$$

IIII-3. Unsupervised learning

Advantages

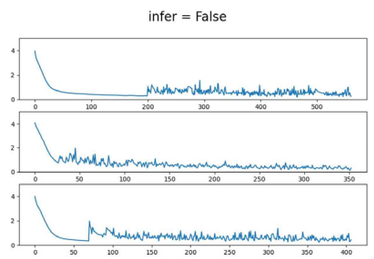
- No labeling

Disadvantage

- low accuracy

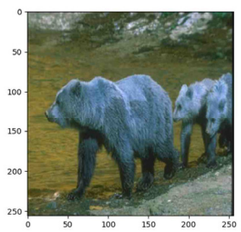
IIII-3. Unsupervised learning

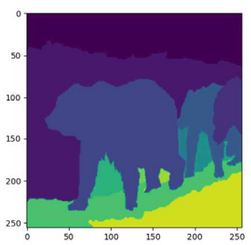
miou= [0.173, 0.024, 0.06]



IIII-3. Unsupervised learning

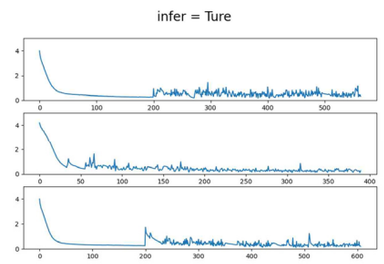
- BSDS 500 dataset



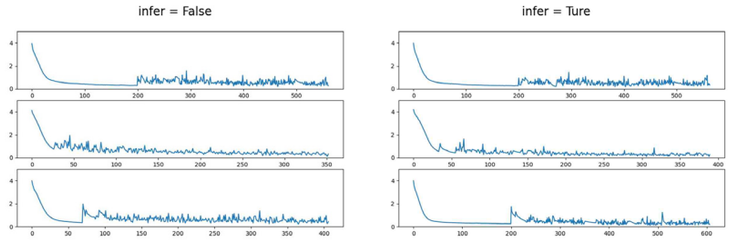


IIII-3. Unsupervised learning

miou= [0.043, 0.193, 0.015]



IIII-3. Unsupervised learning



Deep Learning Based Survival Analysis Using Genomic Data

Jisang Lee
Taehun Lee
Hyeonseo Jung
Geonyung Park

Deep learning based survival analysis using genomic data

Final Presentation

Team ESC

Statistics	Jisang Lee
Statistics	Taehun Lee
ISE	Hyeonseo Jung
ISE	Geonyung Park

Index

01. Data Preprocessing

Remove NaN data → Merge All Data ← 'OS_STATUS' and 'OS_MONTHS' for the label.

Merged_df

Row (Sample Patient) , Col (Gene,Label)

01. Data Preprocessing
-Feature Selection

covariate	exp(coef)	upper 95%	cmp to	z	p	-log2(p)
ZNHIT1	1.593133	0.0	0.804422	0.421153	1.247582	
ZNHIT2	1.112121	0.0	0.163742	0.869934	0.201021	
ZNHIT3	1.309286	0.0	0.411859	0.680443	0.555454	
ZNHIT6	1.271234	0.0	0.280002	0.779476	0.359424	
ZNRD1	1.323472	0.0	-0.112194	0.910669	0.135001	

Calculated each p-value by grouping
the genes of merged_df into 100 expressions.

```
# 피처 선택
top_features = cph.summary.sort_values('p', ascending=True)
top_features = top_features[top_features.p < 0.05]
top_features_df = pd.concat([top_features_df, top_features], axis=0)
print(f'0.05미만의 feature : {top_features_df.shape[0]}')
top_features_index = top_features['coef'].index.tolist()

Selected Genes: Index(['ZNF773', 'ZNF99', 'ZNF8', 'ZNF804B', 'ZNF750', 'ZNF
       'ZNF98', 'ZNF823', 'ZNF763', 'ZNF95'],
      dtype='object')
2.5059447758820644e-53
0.05미만의 feature : 3465
Selected Genes: Index(['ZP2', 'ZSCAN2', 'ZSWIM4', 'ZWILCH', 'ZSCAN20', 'ZYG
       'ZSWIM8', 'ZPBP', 'ZSCAN29'],
      dtype='object')
```

Only chose P-value <0.05
Reduce from about 190,000 features to about 3,000.

01. Data Preprocessing
-Feature Selection

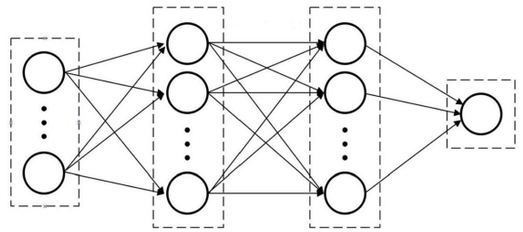
```
1 cph.print_summary()
```

Concordance	0.84
Partial AIC	2675.77
log-likelihood ratio test	413.06 on 62 df
-log2(p) of ll-ratio test	174.74

REASON FOR OUR FEATURE SELECTION

"MODEL" is statistically significant : P-Value < 0.05
"Features" is statistically significant : P-Value < 0.05

02. Model Structure



Deep Neural Network

```
def forward(self, x):
    x = self.fc1(x)
    x = self.activation(x)
    x = self.dropout(x)

    x = self.fc2(x)
    x = self.activation(x)
    x = self.dropout(x)

    x = self.fc3(x)
    x = self.activation(x)
    x = self.dropout(x)

    x = self.fc4(x)
    return x
```

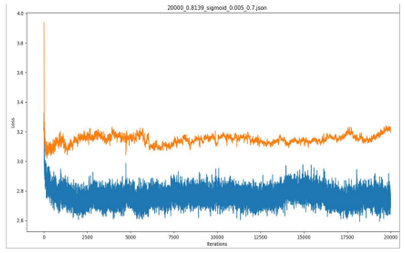
03. Hyperparameter Tuning

EPOCHS = 20,000

Dropout		Activation	Learning_rate	
0.50	0.60	Sigmoid	5e-3	1e-3
0.65	0.70	Hyperbolic-Tanh	5e-4	1e-4
0.75	0.80	ReLU	5e-5	1e-5
0.85	0.90		5e-6	1e-6

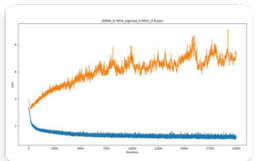
STEP 01 epochs 20,000
STEP 02 Dropout 0.70
STEP 03 Activation sigmoid
STEP 04 Learning Rate 5e-3

» Hyperparameter Tuning

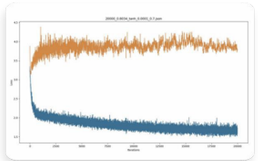


Total number of tuning: 192
The learning curve is selected as the best learned hyperparameter.

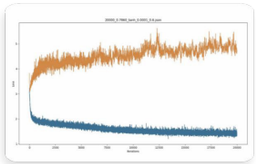
03. Hyperparameter Tuning
- Trial and Error



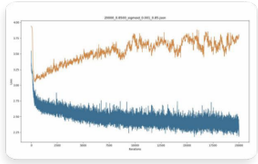
EPOCHS	20000
Activation	Sigmoid
Learning_rate	0.0001
Dropout	0.60



EPOCHS	20000
Activation	Tanh
Learning_rate	0.0001
Dropout	0.70

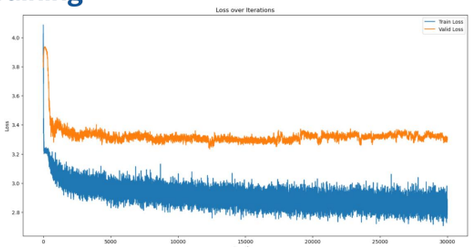


EPOCHS	20000
Activation	Tanh
Learning_rate	0.0001
Dropout	0.60



EPOCHS	20000
Activation	Sigmoid
Learning_rate	0.001
Dropout	0.85

04. Model Training

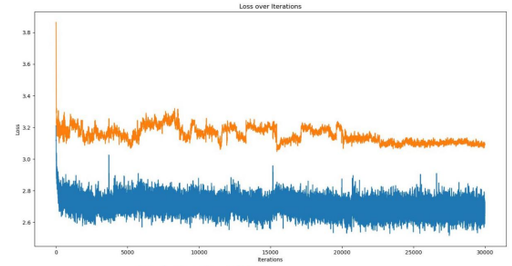



```
Epoch 29301/30000, Train Loss: 2.8656743998990373, Valid Loss: 3.324179172515869
Epoch 29401/30000, Train Loss: 2.8601923613193074, Valid Loss: 3.329929828643799
Epoch 29501/30000, Train Loss: 2.9184954709985185, Valid Loss: 3.307443857192993
Epoch 29601/30000, Train Loss: 2.790267427672651, Valid Loss: 3.318582534790039
Epoch 29701/30000, Train Loss: 2.814209087707657, Valid Loss: 3.3177595138549805
Epoch 29801/30000, Train Loss: 2.754281278657590, Valid Loss: 3.316178321838379
Epoch 29901/30000, Train Loss: 2.7933902439309160, Valid Loss: 3.3413915634155273
Best Epoch: 6701
Best Valid Loss: 3.1973493099212646
C-index: 0.791540801525116
```

05. Result

- 5 Fold - Cross Validation

1st Cross Validation

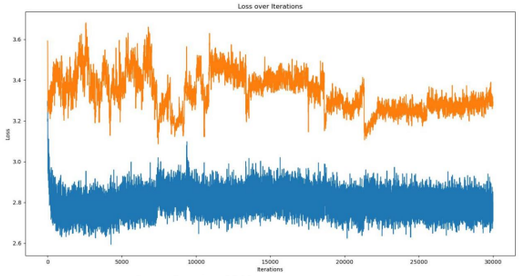



```
Best Epoch: 20310
Best Valid Loss: 3.242199420928955
C-index: 0.8008250594139099
```

05. Result

- 5 Fold - Cross Validation

2nd Cross Validation

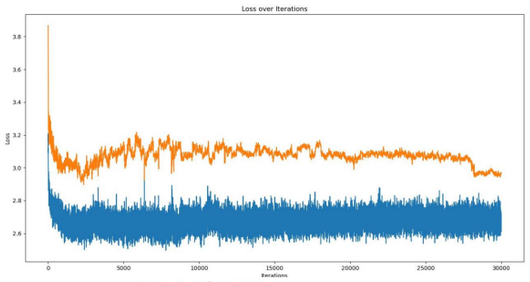



```
Best Epoch: 3768
Best Valid Loss: 3.0449931621551514
C-index: 0.843069314956665
```

05. Result

- 5 Fold - Cross Validation

3rd Cross Validation

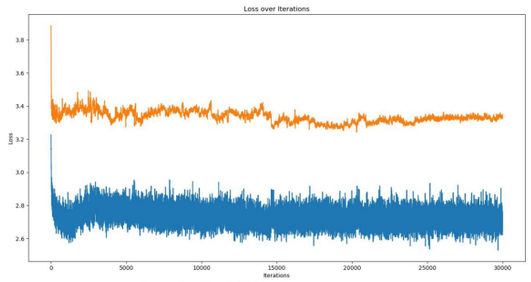



```
Best Epoch: 15453
Best Valid Loss: 3.04701828956604
C-index: 0.8442243933677673
```

05. Result

- 5 Fold - Cross Validation

4th Cross Validation

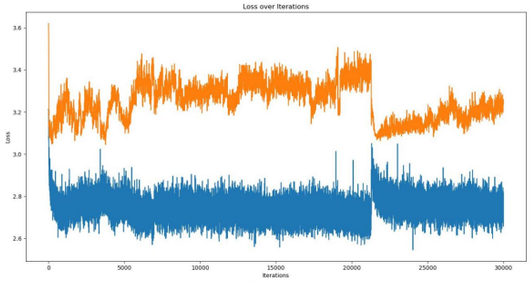



```
Best Epoch: 7449
Best Valid Loss: 3.087181568145752
C-index: 0.8496699929237366
```

05. Result

- 5 Fold - Cross Validation

5th Cross Validation




```
Best Epoch: 2356
Best Valid Loss: 2.8956408500671387
C-index: 0.8872937560081482
```

05. Result

Model Evaluation

Final C-Index : 0.845 ± 0.027

05. Limitation

- **Small Data Scale :** Small dataset weakens the model's reliability and generalization.
- **Biological Interpretation Difficulty :** Interpreting deep learning in biological terms is very challenging.
- **Time Constraints :** If there had been more time, we could have increased the number of iterations.

Cancer Classification

Nayeong Kim
Mihyun Kim
Sehyun Park

PROJECT 3
Cancer Classification

NAYEONG KIM
MIHYUN KIM
SEHYUN PARK

INTRODUCTION

Deep-Hipo: Multi-scale receptive field deep learning for histopathological image analysis

Sai Chandra Kosaraju [a 1], Jie Hao [b 2], Hyun Min Koh [c 3], Mingon Kang [a 4]

INTRODUCTION

- DATA SETS
- Training sets : about 70 thousands
- Validation sets : about 25 thousands
- Train sets : about 22 thousands

->We wouldn't train them all..

F1_SCORE & MATPLOT

```
from sklearn.metrics import f1_score

# Predictions for Validation Datasets
val_pred_probs = model.predict(validation_generator)
val_predictions = np.round(val_pred_probs)

# F1 Score calculation (for Validation dataset)
val_true_labels = validation_generator.classes
f1_val = f1_score(val_true_labels, val_predictions)

print("F1 Score on Validation Data:", f1_val)

# Predictions for Test Datasets
test_pred_probs = model.predict(test_generator)
test_predictions = np.round(test_pred_probs)

# F1 Score calculation (for Test dataset)
test_true_labels = test_generator.classes
f1_test = f1_score(test_true_labels, test_predictions)

print("F1 Score on Test Data:", f1_test)
```

```
import matplotlib.pyplot as plt

history = model.history.history

# LOSS
plt.plot(history['loss'], label='Train Loss', color='blue')
plt.plot(history['val_loss'], label='Validation Loss', color='orange')
plt.xlabel('Epochs')
plt.ylabel('Loss')
plt.legend()
plt.show()

# ACCURACY
plt.plot(history['accuracy'], label='Train Accuracy', color='blue')
plt.plot(history['val_accuracy'], label='Validation Accuracy', color='orange')
plt.xlabel('Epochs')
plt.ylabel('Accuracy')
plt.legend()
plt.show()
```

AlexNet

```
h dtype int32
         [[{{node Placeholder/_0}}]]
2256/2256 [==============================] - 19397s 9s/step - loss: 0.6013 - accuracy: 0.7189 - val_loss: 0.4870 - val_accuracy: 0.8279
Epoch 2/15
2256/2256 [==============================] - 18009s 8s/step - loss: 0.5947 - accuracy: 0.7191 - val_loss: 0.4902 - val_accuracy: 0.8279
Epoch 3/15
2256/2256 [==============================] - 10850s 5s/step - loss: 0.5946 - accuracy: 0.7191 - val_loss: 0.4864 - val_accuracy: 0.8279
Epoch 4/15
2256/2256 [==============================] - 9713s 4s/step - loss: 0.5944 - accuracy: 0.7191 - val_loss: 0.4794 - val_accuracy: 0.8279
Epoch 5/15
 184/2256 [=>............................] - ETA: 1:56:37 - loss: 0.5950 - accuracy: 0.7184


 204/2256 [=>............................] - ETA: 1:55:53 - loss: 0.5942 - accuracy: 0.7194
 205/2256 [=>............................] - ETA: 1:55:55 - loss: 0.5947 - accuracy: 0.7189
2256/2256 [==============================] - 11275s 5s/step - loss: 0.5944 - accuracy: 0.7191 - val_loss: 0.4923 - val_accuracy: 0.8279
Epoch 6/15
2256/2256 [==============================] - ETA: 0s - loss: 0.5944 - accuracy: 0.7191
2256/2256 [==============================] - 11048s 5s/step - loss: 0.5944 - accuracy: 0.7191 - val_loss: 0.5056 - val_accuracy: 0.8279
Epoch 7/15
2256/2256 [==============================] - 8626s 4s/step - loss: 0.5944 - accuracy: 0.7191 - val_loss: 0.4964 - val_accuracy: 0.8279
Epoch 8/15
2256/2256 [==============================] - 8678s 4s/step - loss: 0.5942 - accuracy: 0.7191 - val_loss: 0.5062 - val_accuracy: 0.8279
Epoch 9/15
2256/2256 [==============================] - 7908s 4s/step - loss: 0.5942 - accuracy: 0.7191 - val_loss: 0.4915 - val_accuracy: 0.8279
Epoch 10/15
2256/2256 [==============================] - 7679s 3s/step - loss: 0.5941 - accuracy: 0.7191 - val_loss: 0.4869 - val_accuracy: 0.8279
Epoch 11/15
2256/2256 [==============================] - 7909s 4s/step - loss: 0.5941 - accuracy: 0.7191 - val_loss: 0.4926 - val_accuracy: 0.8279
Epoch 12/15
2256/2256 [==============================] - 8016s 4s/step - loss: 0.5941 - accuracy: 0.7191 - val_loss: 0.4908 - val_accuracy: 0.8279
Epoch 13/15
2256/2256 [==============================] - 7932s 4s/step - loss: 0.5942 - accuracy: 0.7191 - val_loss: 0.4939 - val_accuracy: 0.8279
Epoch 14/15
2256/2256 [==============================] - 7821s 3s/step - loss: 0.5941 - accuracy: 0.7191 - val_loss: 0.4881 - val_accuracy: 0.8279
```

GoogleNet

```
         [[{{node Placeholder/_0}}]]
Epoch 1/15
72167/72167 [==============================] - ETA: 0s - loss: 0.4728 - accuracy: 0.77832023-07-18 06:51:42.289271: I tensorflow/core/common_runtime/execu
EBUG INFO) Executor start aborting (this does not indicate an error and you can ignore this message): INVALID_ARGUMENT: You must feed a value for placehol
h dtype int32
         [[{{node Placeholder/_0}}]]


72167/72167 [==============================] - 59620s 825ms/step - loss: 0.4728 - accuracy: 0.7783 - val_loss: 0.3528 - val_accuracy: 0.8588
Epoch 2/15
72167/72167 [==============================] - 50040s 693ms/step - loss: 0.4430 - accuracy: 0.7938 - val_loss: 0.4033 - val_accuracy: 0.8567
Epoch 3/15
```

ResNet50

```
Epoch 1/15
2256/2256 [==============================] - ETA: 0s - loss: 0.5787 - accuracy: 0.71882023-07-17 21:36:52.731135: I tensorflow/core/common_runtim
UG INFO) Executor start aborting (this does not indicate an error and you can ignore this message): INVALID_ARGUMENT: You must feed a value for p
dtype int32
         [[{{node Placeholder/_0}}]]
2256/2256 [==============================] - 19272s 9s/step - loss: 0.5787 - accuracy: 0.7188 - val_loss: 0.4802 - val_accuracy: 0.8241
Epoch 2/15
2256/2256 [==============================] - 16570s 7s/step - loss: 0.5617 - accuracy: 0.7260 - val_loss: 0.4206 - val_accuracy: 0.8313
Epoch 3/15
2256/2256 [==============================] - 14791s 7s/step - loss: 0.5571 - accuracy: 0.7270 - val_loss: 0.4701 - val_accuracy: 0.8220
Epoch 4/15
2256/2256 [==============================] - 17951s 8s/step - loss: 0.5525 - accuracy: 0.7296 - val_loss: 0.4184 - val_accuracy: 0.8321
Epoch 5/15
2256/2256 [==============================] - ETA: 0s - loss: 0.5486 - accuracy: 0.7314



2256/2256 [==============================] - 15681s 7s/step - loss: 0.5486 - accuracy: 0.7314 - val_loss: 0.4140 - val_accuracy: 0.8331
Epoch 6/15
2256/2256 [==============================] - 13239s 6s/step - loss: 0.5448 - accuracy: 0.7345 - val_loss: 0.4396 - val_accuracy: 0.8271
Epoch 7/15
2256/2256 [==============================] - 10896s 5s/step - loss: 0.5445 - accuracy: 0.7345 - val_loss: 0.3969 - val_accuracy: 0.8339
Epoch 8/15
2256/2256 [==============================] - 12667s 6s/step - loss: 0.5421 - accuracy: 0.7361 - val_loss: 0.4050 - val_accuracy: 0.8332
Epoch 9/15
2256/2256 [==============================] - 13117s 6s/step - loss: 0.5414 - accuracy: 0.7358 - val_loss: 0.4093 - val_accuracy: 0.8327
```

COMPARSION(TOP)

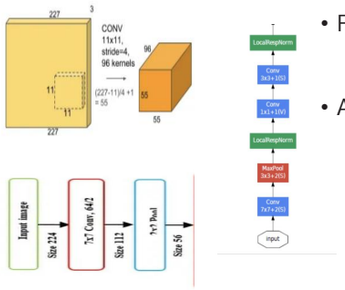
Architecture	validation	
	Loss	Accuracy
GoogleNet	0.3528	0.8588
ResNet50	0.3969	0.8339
Alexnet	0.4881	0.8279

- Accuracy : Googlenet > Resnet50 > Alexnet

AlexNet VS GoogleNet, ResNet50

- Filter Size
- The number of Layer
- Gradient Vanishing
- 1X1 Convolution Layer

AlexNet VS GoogleNet, ResNet50



- Filter Size
  - •Alexnet >> Googlenet, Resnet50
- Advantages of small filter size
  - •Reduce the number of parameters and computational complexity
    -> Computational efficiency
  - •Effective in detecting local pattern

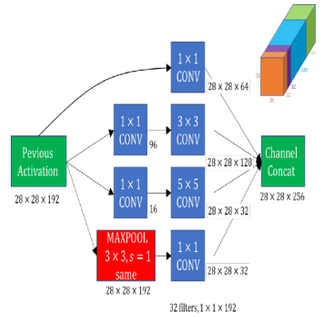
GoogleNet VS ResNet50

- Network Structure
- Number of Parameters
- Task-Specific Performance

AlexNet VS GoogleNet, ResNet50

- The number of Layer : Alexnet(11) << GoogleNet(24), ResNet50(51)
  - More layer
    - Advantage : Precision, Acurracy
    - Disadvantage : Overfitting, Training time, gradient vanishing
- Gradient Vanishing
  - AlexNet < Googlenet, Resnet50
    - Googlenet : Utilization of various filter sizes, Inception module
    - Resnet50 : Residual Learning

Inception Module



- Parallel usage of convolutional layers at multiple scales.
  - The network can effectively capture features at different levels of detail.
- Usage of 1x1 convolutional layers.
  - Reduce dimensionality and thereby decreasing computational complexity.

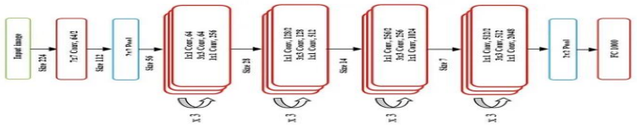
AlexNet VS GoogleNet, ResNet50

- 1X1 Convolutinal layer
  - Used in GoogleNet, ResNet50
  - GoogleNet : used to adjust the number of channels.
  - ResNet50 : employed within the Residual Block to perform dimensionality reduction and expansion.

GoogleNet

- Memory efficiency
- Width (22 layers)

ResNet50



- capture more complex and abstract features
- Depth (50 layers)

PARAMETERS

• GoogleNet
- The parallel convolutional operations within the Inception module.

• ResNet50
- A large number of parameters due to its numerous layers.

PERFORMANCE

- Task-Specific Performance

- Both models have been successful in various computer vision.

- Vary depending on the specific task or dataset.

- Recommend to evaluate and compare the performance of both models on the specific task of interest to determine which one performs better.

CONCLUSION

- Difference with ResNet50,GoogleNet VS AlexNet
- > 1X1 con layer
- Difference of ResNet50 VS GoogleNet
- > Inception Model

• Acurracy : Googlenet > Resnet50 > Alexnet

Thank you

For Listening

Survival Analysis

Jihyeon Kwon
Hyojin Yeon
Soyeon Bae
Yonghwan Lee

UNLV

Survial Analysis

JBNU

Dept of Statistic Jihyeon-Kwon
Dept of Molecular Biology Hyojin-Yeon
Dept of Mechanical Engineering System Soyeon-Bae
Dept of Computer Science and Artificial Intelligence Yonghwan-Lee

CONTENTS

Survival Analysis Using Genomic Data

UNLV

Introduction

UNLV

Survival Analysis

Predicting patient prognis

Evaluating treatment effictiveness

Identification survival predictor

Develope prediction model

Support decision-making in cancer management and treatment
Aid in the development of personalized treatment strategies

UNLV

Solution

feature selection — Identify the most relevant and informative features that have significant impact on the survival outcome

Cox-nnet — capturing nonlinear relationships between predictors and survival outcomes

UNLV

Data

Develop a survival prediction model

using gene expression data from LGG and GBM patients.

- LGG (Low Grade Glioma)
- GBM(Glioblastoma Multiforme)

UNLV

Project Aim

Comparing Feature Selection Method

Variance | **Cox - en** | **Similarity Score**

Modeling

Cox - nnet

UNLV

Main Obstcles of Cox-PH

1) HDLSS : High Dimensional, Low Sample Size Data($p \gg n$)

2) handling the highly nonlinear relationship between covariates

the conventional Cox-PH model assumes the linear contributions of covariates

UNLV

Process

Preprocessing Steps

Feature Selection

Modeling

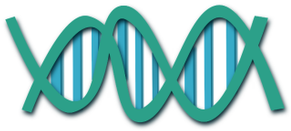
UNLV

Preprocessing Steps

1. Data Cleaning
2. Data Integration
3. Data Transformation
4. Data Encoding
5. Feature Selection
6. Data Splitting
7. Nomalization

UNLV Feature Selection

Feature selection in bioinformatics



- become an apparent need in many bioinformatics applications
- contain a large number of features or variables
- focus on the most relevant biological variables and extract meaningful insights

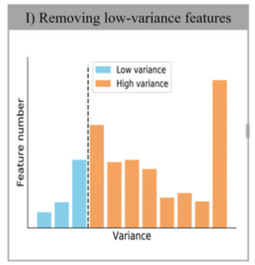
UNLV Feature Selection

advantages

1. to reduce dimentionality
2. to improve interpretability
3. to provide faster and more cost-effective models
4. to avoid overfitting and improve model performance

UNLV Feature Selection

Variance



removes the low variance features
from the dataset
that are of no great use in modeling

UNLV Feature Selection

Variance

In comparison to the considered mutual information filters, for both the variance and the carss filter, no categorization of the survival outcome and features, resulting in a loss of information, is required. An advantage of the variance filter over the carss filter is that it allows unbiased testing in a subsequent unregularized model, if the proportion of features to select is prespecified. This is because the variance filter does not take into account the survival outcome for feature selection, see also

variation filter primarily focus on the variability within each feature and does not directly consider its relationship with the target variable such as event or time.

Andrea Bommert and others, Benchmark of filter methods for feature selection in high-dimensional gene expression survival data, Briefings in Bioinformatics, Volume 23, Issue 1, January 2022, bbab354, https://doi.org/10.1093/bib/bbab354

UNLV Feature Selection

Variance

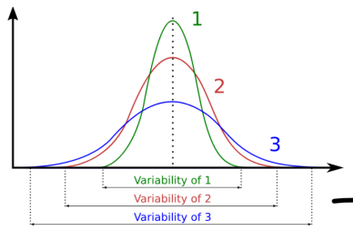
- It is applicable only on Numeric features
  We don't need to change the values because they are all numeric.
- We sorted each column in descending order for variance
- extracted only the top 25% of columns

	C14orf184	PRND	DUSP13	ROR2	STGD1	GC36	ABCC3	GRIN1	TUBB5	TCTE1	...	SRRM2-AS1	EHMT1-IT1	OR51B5	C15orf56	SEMA3D	TGI
0	-1.8663	0.3444	-0.8074	-0.1271	0.0105	-1.2257	0.5340	1.1275	-0.3087	-0.4544	...	-0.2133	-0.3970	-2.2017	0.3697	-0.3219	0.24
1	-1.8663	-0.0792	-0.3429	2.6934	2.8605	-0.9143	0.8302	0.7217	-1.6535	-1.0660	...	-1.1264	0.4206	-2.2017	-1.7053	-0.0508	1.16
2	-1.8663	1.7639	0.0606	-0.3858	1.0832	-1.7725	-1.3628	1.0106	-0.6268	-0.9831	...	0.2503	1.4604	-2.2017	-1.4632	-0.2370	-0.69
3	0.5988	-1.1133	-1.6228	1.6102	0.3975	-0.2319	-0.0561	0.5816	-1.6535	1.5515	...	0.5874	-0.2101	-2.2017	-1.3807	-0.6791	1.15
4	-0.5436	-0.7739	-0.2658	0.1155	-0.1367	0.7258	0.7033	-1.3216	-0.8549	0.2090	...	-0.6501	1.2114	-2.2017	0.0581	0.7251	-0.98
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
688	-1.9837	-0.2121	-0.7567	-0.8965	-0.5215	-0.8898	-1.0975	0.0065	0.0019	-0.2975	...	-1.0566	0.7261	-2.6205	-0.9382	0.9186	-0.80
689	-1.9837	-0.5625	-1.3565	0									19	-2.6205	0.1524	0.9213	0.22
690	-0.7267	-1.2253	-0.8716	-0									61	-2.6205	-1.1392	1.2792	-0.36
691	-1.9837	-0.2560	-0.0717	-0									55	-2.6205	-0.3367	0.2505	-0.86
692	-0.8073	-0.4160	-0.5597	-0.5881	-0.2493	-0.5010	0.7193	1.0921	0.7234	0.2650	...	0.6304	0.3587	-2.6205	1.0717	1.0076	-1.06

693 rows × 4940 columns

UNLV Feature Selection

Why select columns with high variance?



- they exhibit a wide range of values and indicate greater diversity
- more likely to contain more information and provide useful features

helps improve the performance of the model and captures important features that contribute to the variability in the dataset.

UNLV Feature Selection

Similarity Score

- Cosine-Similarity
  - method for measuring the similarity between vectors
  - measured by caculating the angle between two given vectors

UNLV Feature Selection

Cox-EN

elasticnet is a linear regression algorithm that combines Lasso and Ridge.

L1 Penalty: $R(w) := \frac{1}{2}\sum_{i=1}^{n}|w_i|$ — performs variable selection emphasizing the weight of important variables

L2 Penalty: $R(w) := \frac{1}{2}\sum_{i=1}^{n} w_i^2$ — reduces multicollinearity limiting the weights of variables

Elastic-Net Penalty:

$$R(w) := \frac{\varphi}{2}\sum_{i=1}^{n} w_i^2 + (1-\varphi)\sum_{i=1}^{n}|w|$$

A convex combination of L1 and L2 Penalty.

UNLV Feature Selection

Similarity Score

- Why is cosine similarity used?

	SLC6A7	WNT11	SLC35F3	MYH15	SCARNA2	SCRT1
0	1.3724	-1.3602	1.0665	1.5561	-1.0059	0.7365
1	0.3311	1.2562	-0.0234	-0.4776	-1.0641	-0.1162

Gene

- Relationship between features is not linear

- To capture the non-linear relationship between features, cosine similarity is used.
- To consider related or dependet feature together, select high similarity score feature

1. Calculate cosine similarity between features
2. Leave rows with at least 80% of similarity scores greater than 0 for each column
3. The number of feature after feature selection : 4,749

UNLV Feature Selection

Cox-EN

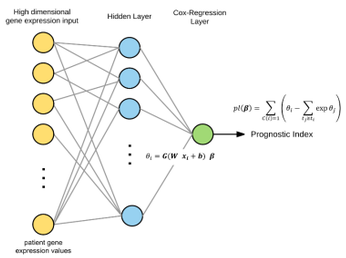
- If Coefficient = 0 (Remove Features)
- If Coefficient != 0 (Select Features)

coefficient is the weight

	ETV4	NUDT11	ERK	LAMP2	UBE2R2	DCAF8	UQCRB	FBXO21	VPS33B	CD79B	...	LINC01059	PARM1	BMPER	C3orf18	PTER	EXO1
0	-0.4221	0.1505	0.4421	0.0596	0.4707	-0.6780	-0.9751	1.0006	-0.5226	1.3109	...	1.0578	0.4246	0.6376	0.1435	0.7348	-0.167
1	0.1572	-0.3288	0.8236	1.3019	-0.0554	-0.4728	1.3939	-1.9952	-0.1080	-0.0937	...	-0.0519	0.9772	-0.1022	-1.8032	0.7630	-0.298
2	-0.1191	1.2676	1.2065	-1.7294	2.2173	-1.3839	2.1369	0.7450	0.3227	-0.1015	...	-0.7642	1.1538	-0.8520	-1.6975	-0.4469	1.614
3	0.7370	0.2926	1.8875	-1.0678	0.8670	-0.4598	-0.9267	1.2153	-0.6022	-0.5898	...	0.4318	-0.7051	-0.6896	-1.5898	-0.2594	-0.115
4	-0.1260	-1.8580	-0.0307	1.2720	0.0899	0.3444	1.0613	-1.3447	0.8643	0.2130	...	-0.1174	0.0765	-0.3796	-0.3473	0.2209	-2.871
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
689	-1.0958	0.0227	-0.4501	-0.5039	0.6577	0.1700	0.4749	-0.9099	0.8085	-1.6267	...	-0.6313	0.0649	0.7330	0.3259	-0.1539	-0.887
690	-1.2236	1.0234	-0.0357	-0.7977	2.1647	0.7220	0.7548	-0.2833	0.7866	-0.0310	...	-1.0278	0.5788	-0.8419	-0.4279	-0.8935	0.574
691	-0.3786	0.7146	-0.6880	-0.4221	693 rows × 3514 columns								-0.3212	-1.4954	0.1182	-0.7991	-0.366
692	-0.5748	0.8482	-0.5122	0.0187									-0.6663	-1.5466	-0.2799	-1.1087	-1.047
693	-0.9858	1.1262	0.2722	-1.5747	1.0528	0.3273	0.9979	-0.0751	1.2017	-1.1981	...	-0.5769	0.0425	1.6541	-0.8072	0.7371	0.050

UNLV Modeling

Cox-nnet



Article Source: Cox-nnet: An artificial neural network method for prognosis prediction of high-throughput omics data
Ching T, Zhu X, Garmire LX (2018) Cox-nnet: An artificial neural network method for prognosis prediction of high-throughput omics data. PLOS Computational Biology 14(4): e1006076. https://doi.org/10.1371/journal.pcbi.1006076

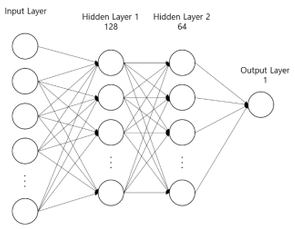
- Cox-nnet

Artificial neural network (ANN) framework that has been specifically developed for prediciting patient prognosis using high throughput transcriptomics data.

- Why Cox-nnet?
  - provides equivalent or better predictions compared to other methods.
  - takes advantage of ANNs to provide better insight by modeling complex non-linear relationships.

UNLV Modeling

Model Design



- Layer Architecture

1. Input Layer : Layer with features matching the dimension of input data
2. Hidden Layer 1 : Linear layer with 128 nodes
3. Hidden Layer 2 : Linear layer with 64 nodes
4. Ouput Layer : Linear layer with 1 node

UNLV Modeling

Model Design

- Activation Function

ReLU(Recified Liniear Unit)

To learn complex patterns by introducing non-linearity to the model

- Optimization Algorithm

Adam(Adaptive Moment Estimation)

To achieve fast and stable convergence to adjust the learning rate automatically

UNLV Modeling

Model Training

- Loss Function

Negative partial log likelihood

- Used in the Cox proportional hazards model to train the model
- Calculated for each sample considering the event occurrence and the predicted hazard ratio of each sample
- Decreses as the model predicts the actual event occurence time and event occurrence

UNLV Modeling

Model Training

- Hyperparameter Tuning
  - Dropout Rate : 0.8
    - set dropout ratio high to prevent overfitting because of HDLSS data.
  - Learning rate
    - Perform hyperparameter tuning using *Trainer* in *PyTorch Lightning*.
    - Use *lr_finder* to explore a range of learning rates and set a recommded learning rate.
  - Activation Function : ReLU

UNLV

Experiment & Result

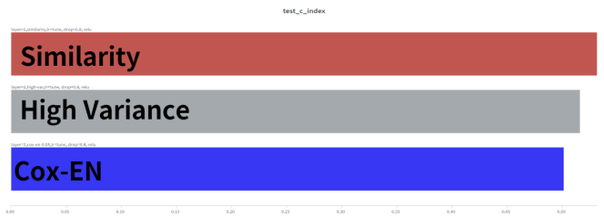
UNLV

Hyperparameter Table

Hyperparameter	Value
Learning Rate	PyTorch Lightning lr_find Function
Dropout Rate	0.8
Activation Function	ReLU

UNLV

Result

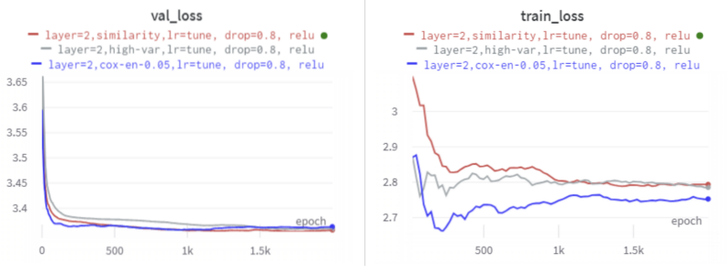


UNLV

Conclusion

UNLV

Learning Curve



UNLV

Conclusion

Adventage
Similarity score shows slight more improvement in the c-index compare to variance and cox-en, stable running curve

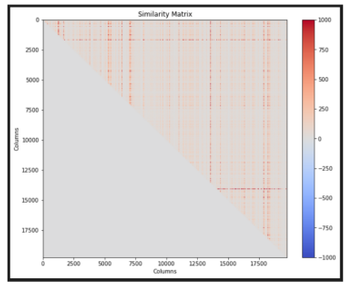
Limitation
The Similarity Score does not consider both labels (Month and Event)

Future Research
Consider both labels (Month and Event) cross-validation

Significance
Confirm that the similarity score is a meaningful approach, compared to the conventional methods

UNLV

Similarity Score based Feature Selection

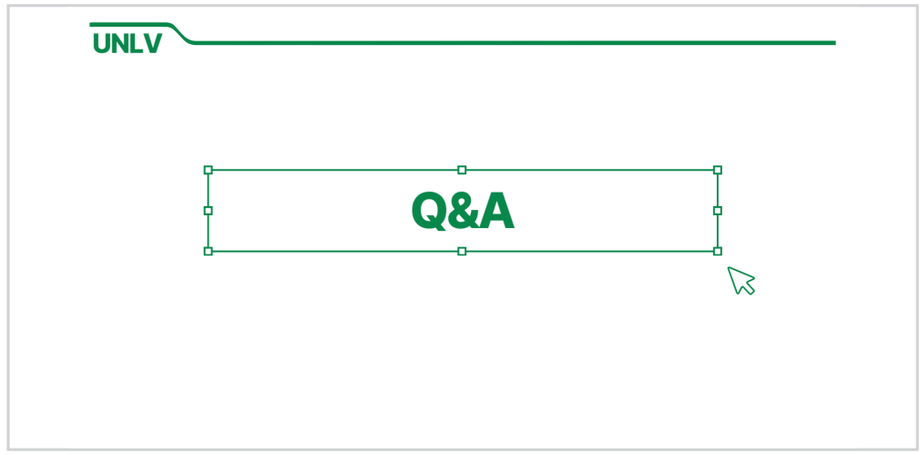


- The best c-index in 3 Feature Selections
- Don't use 2 labels(Only 1)
- Over 0 Score Threshold = 80%

UNLV

Conclusion

Capture the intrinsic structure of gene data,
Enabling the identification of meaningful features in the model



Scalable Data Processing:
By applying MapReduce in Spark

Dayoung Kang
Jaehyeon Kim
Subin Seo

Scalable Data Processing :
By applying MapReduce in Spark

#Big-data #MapReduce #ML

Dayoung Kang
Jaehyeon Kim
Subin Seo

Background Research

What is Big-data ?

: Any data set contains large volumes of information and complex data is called Big Data (BD). It has 4 characteristics. (4V's) [1]

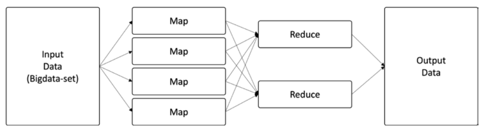
- Volume which is the quantity of data.
- Velocity is the speed of the data that during handling and generating.
- Variety refers to the range of data types and sources.
- Veracity is related to the truth of data which is important for precision in analysis.
- + **Value** is the importance of the data importance and this is a very significant feature in BD6.

➡ **BD is unlike traditional data, so it requires special processing to manage it.**

Background Research

How to process Big-data ?

: apply for MapReduce to process Big Data in parallel on multiple node.



Step1. Map
- Split input data to number of slices
- Apply specific function to each to generate intermediate results

Step2. Reduce
- Combine the intermediate results to make the final result.

About Data

: Customer information to predict who possible Defaulters are for Loans Product

```
train .head(10)
```

	Id	Income	Age	Experience	Married/Single	House_Ownership	Car_Ownership	Profession	CITY	STATE	CURRENT_JOB_YRS	CURRENT_HOUSE_YRS	Risk_Flag
0	1	1303834	23	3	single	rented	no	Mechanical_engineer	Rewa	Madhya_Pradesh	3	13	0
1	2	7574516	40	10	single	rented	no	Software_Developer	Parbhani	Maharashtra	9	13	0
2	3	3991815	66	4	married	rented	no	Technical_writer	Alappuzha	Kerala	4	10	0
3	4	6256451	41	2	single	rented	yes	Software_Developer	Bhubaneswar	Odisha	2	12	1
4	5	5768871	47	11	single	rented	no	Civil_servant	Tiruchirappalli[10]	Tamil_Nadu	3	14	1
5	6	6915937	64	0	single	rented	no	Civil_servant	Jalgaon	Maharashtra	0	12	0
6	7	3954973	58	14	married	rented	no	Librarian	Tiruppur	Tamil_Nadu	8	12	0
7	8	1706172	33	2	single	rented	no	Economist	Jamnagar	Gujarat	2	14	0
8	9	7566849	24	17	single	rented	yes	Flight_attendant	Kota[6]	Rajasthan	11	11	0
9	10	8964846	23	12	single	rented	no	Architect	Karimnagar	Telangana	5	13	0

Independent variables (X) Dependent variables (Y)

- Train_Data_shape : (252000, 13)
- It has **252,000 samples and 11 features.**
- Independent variables are used to predict of Risk_Flag which is dependent variables.
- Risk_Flag(Y) is binary clas (0 or 1) .

Preview

We try 2 way to Implement MapReduce !

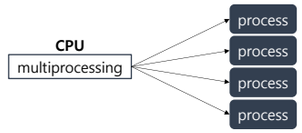
Python Multiprocessing VS PySpark

Python Multiprocessing

Using Python Muti processing library

Python

Modules for parallelism in Python: **Multiprocessing**



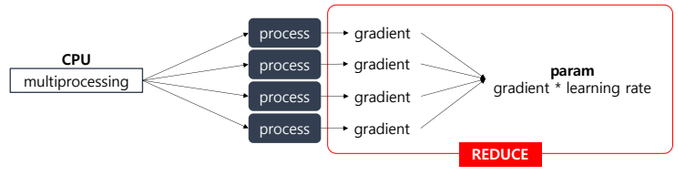
Pool Object

which offers a convenient means of parallelizing the execution of a function across multiple input values, distributing the input data across processes (data parallelism).

Python

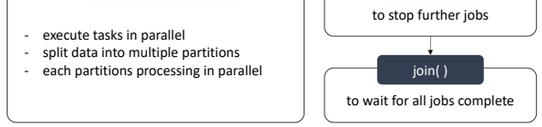
Reduce

- **Combines** the intermediate results and updates the model parameters
- Sum the **gradients** then multiplied by **learning rate** to return the **updated parameters**



Python

Multiprocessing: Pool Object



Python

MapReduce with multiprocessing

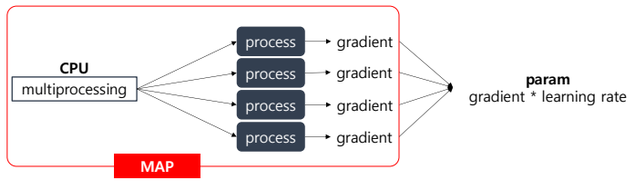
	Pycharm(M1)	JupyterLab(DIONE)
Pool Worker(CPU cores)	8	48
Processing TIme	4min 2.028sec	1min 3.902sec

- Since **multiprocessing** performs parallel processing based on the number of cores in the **CPU**, we were able to get faster results on the **DIONE** server with 48 CPU cores.

Python

Map

- Goal: Binary Classification with **Linear Regression**
- Calculate gradient by **Ordinary Least Squares (OLS)** for each partition



PySpark

Using Pyspark (RDDs)

Spark

Apache Spark™ is a multi-language engine for executing data engineering, data science, and machine learning on single-node machines or clusters.

- Batch/streaming data
- SQL analytics
- Data science at scale
- Machine learning

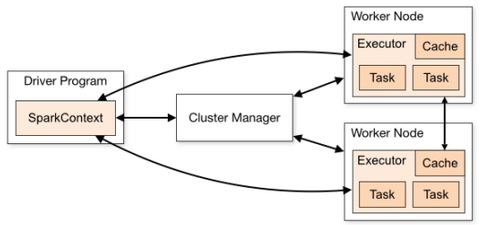
Spark

How does **Spark** do **distributed and parallel processing**?

- Clustering
- RDD & DAG

Spark

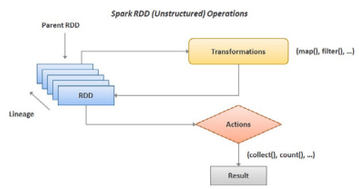
< Clustering >



Spark is capable of **in-memory operation** and **distributed processing**, which speeds up.
➡ Because, Cluster Manager allocates Worker nodes to CPUs on its own and performs distributed processing.

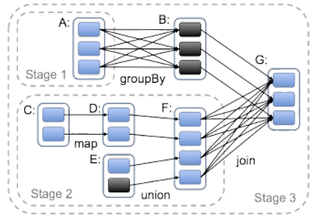
Spark RDD & DAG

< RDD >



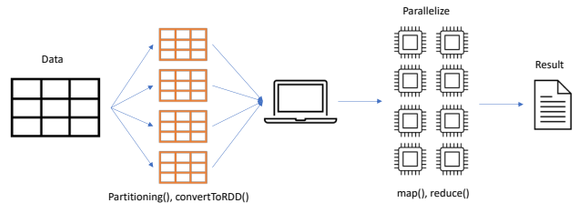
- RDD is Spark's fundamental data abstraction concept, which is a **distributed collection** of data elements divided into multiple partitions.
  ➡ Immutable & Read-only system!

< DAG>



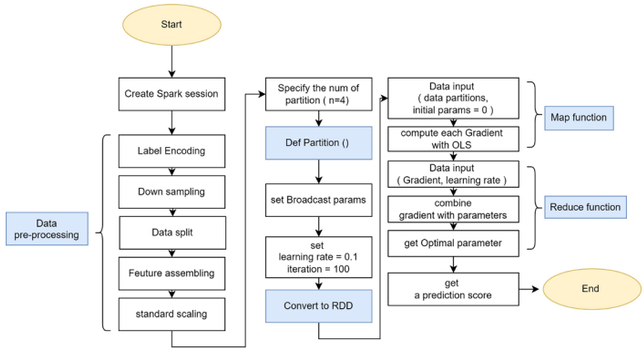
- DAG is a **Directed Acyclic Graph** that expresses dependencies between tasks.
  ➡ All transformation is recorded as a DAG.

Spark Process

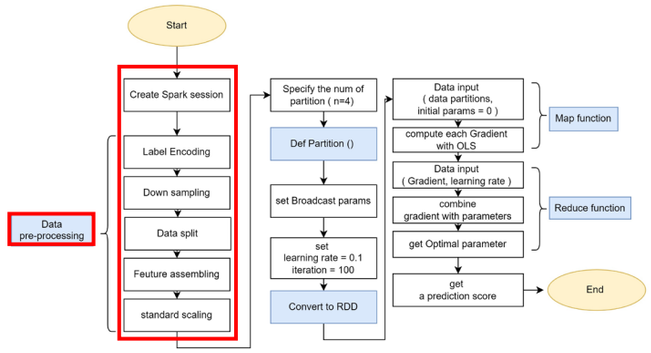


- The data is split into multiple **partitions** (num_partitions = 4).
- **The map function** calculates the gradients for each partition in parallel (worker = 8).
- **The reduce function** combines the gradients and updates the model parameters using a learning rate.
- This process is repeated for a certain number of iterations (num_iterations = 100).

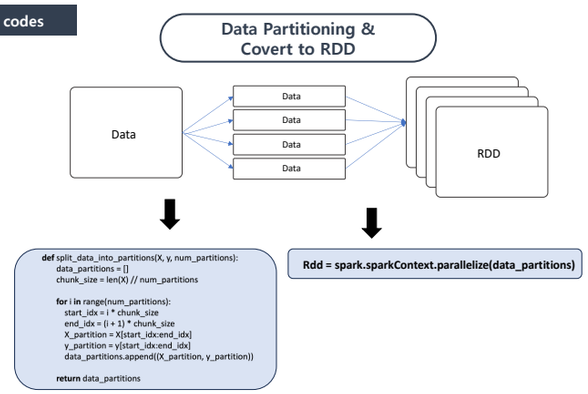
Flowchart



Current location



Function codes



Data Pre-Processing

Label Encoding

Categorical data to number

features
- "married_single"
- "profession"
- "house_ownership"
- "car_ownership"
- "city"
- "state"

Down Sampling

Class0
- 221,004 samples

Class1
- 30,996 samples

Each class 30,996 samples
Total 61,992 samples

Data Split

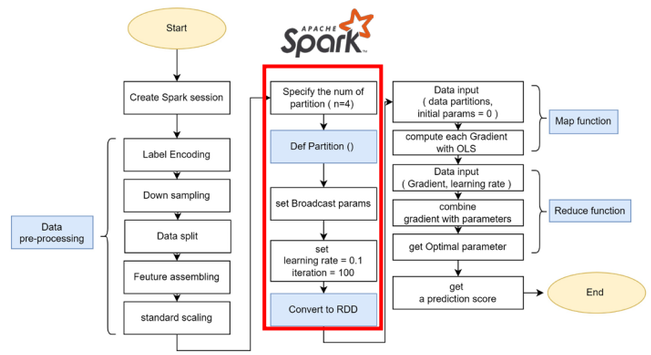
One Dataset file
- **Train** 64%
- **Validation** 16%
- **Test** 20%

Standard Scaling

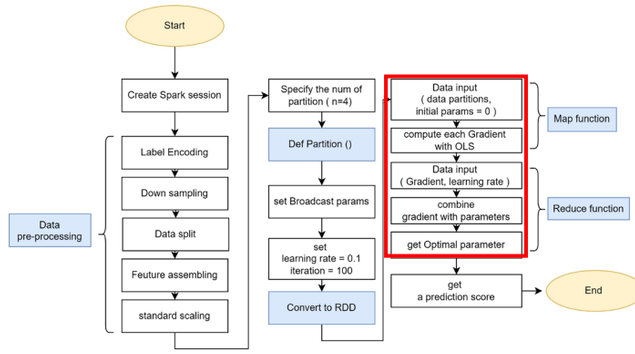
Calibrate scale differences

Features ex.
- Income(10,000,000)
- Age(14)

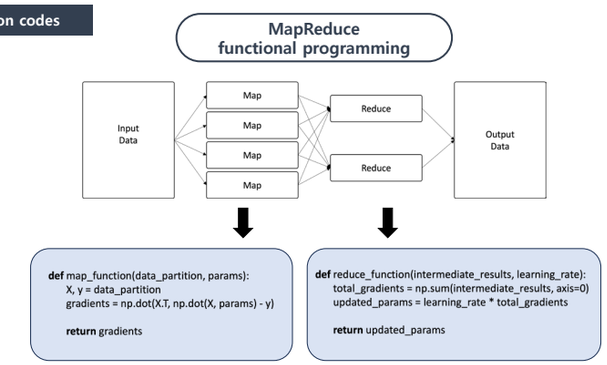
Current location



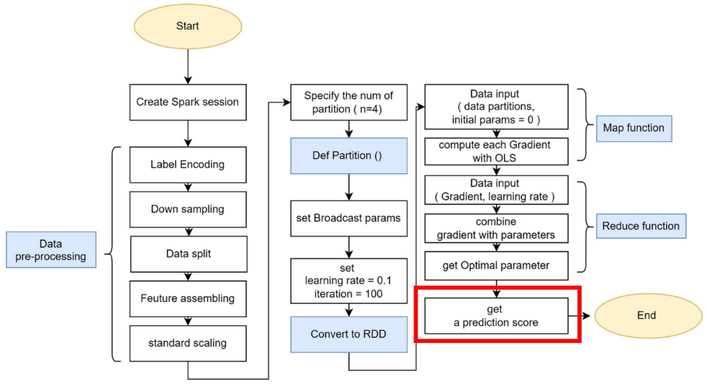
Current location



Function codes



Current location



Spark Result

Spark	
Elapsed Time	0.0 min 41.307 sec
Val-ACC	50.193
Test-ACC	50.191

- Scalable data can be distributed and processed in parallel through spark.
- And this method is effective for processing big data.

Discussion

- In this project,
  Our data set can be storage and processed with in a local storage, which has about 60,000 samples.

=> Therefore, we are going to test the model of this paper with a lot bigger data sets.
(Use large amounts of data that are not even stored on the local storage)

Reference

[1] Hiba Basim Alwan and Ku Ruhana Ku-Mahamud 2020 *IOP Conf. Ser.: Mater. Sci. Eng.* **769** 012007
[2] Matei Zaharia, Mosharaf Chowdhury, Michael J. Franklin, Scott Shenker, and Ion Stoica. 2010. Spark: cluster computing with working sets. In Proceedings of the 2nd USENIX conference on Hot topics in cloud computing (HotCloud'10). USENIX Association, USA, 10.
[3] Data Reference : https://www.kaggle.com/datasets/subhamjain/loan-prediction-based-on-customer-behavior?resource=download&select=Sample+Prediction+Dataset.csv
[4] Image Reference : https://subscription.packtpub.com/book/data/9781785889622/5/ch05lvl1sec39/linear-classification

2023학년도 빅데이터 혁신융합대학 사업단
해외 연구프로그램 결과보고서

| 발 행 일 | 2023년 10월
| 발 행 처 | 경상국립대학교 빅데이터 혁신융합대학 사업단
| 문 의 | T. 055)772-2775~7
| 주 소 | 52828 경상남도 진주시 진주대로 501
경상국립대학교 빅데이터 혁신융학대학 사업단
| U R L | http://smartbigdata.gnu.ac.kr

경상국립대학교 빅데이터 혁신융합대학 사업단
52828 경상남도 진주시 진주대로 501
T. 055)772-2775~7
http://smartbigdata.gnu.ac.kr