

HIRA

빅데이터 기반 산·학·관 연계 맞춤형 인재 양성 교육

교육 소개

건강보험심사평가원이 보유한 빅데이터 인프라·분석기술의 공유·이전을 통해 빅데이터 기반 산업체 경쟁력 강화 및 맞춤형 청년 인재 양성 교육 지원 프로그램

[교육 대상] 보건 의료 분야에 관심있는 대학생 및 산업체

[운영 기간] 2020. 7. 22.(수) ~ 8. 28.(금)

[교육 과정]

I. 빅데이터 이론·실습 교육

- (이론) 빅데이터 이해와 활용
건강보험요양급여 관련 의료행위 코드이해
- (실습) SAS를 활용한 HIRA빅데이터 처리
분석자료 설계, HIRA빅데이터 통계산출

II. 팀별 프로젝트 수행

- (팀프로젝트) 산업 현장 수요를 반영한 프로젝트 주제 선정 및
분석-활용 방안 등 프로젝트 팀별 코칭
- ※ 산업체 1곳과 대학생(4~5명)을 한팀으로 팀별 프로젝트 진행

III. 교육 수료

- 팀별 프로젝트 발표회
- 교육 수료증 수여



질병 네트워크시스템 구축*

김희천 차장
(주)미소정보기술



1. 첫 만남

- 기대에 부른 학생들과 기업들의 첫 만남은 어색하지만 색다른 느낌으로 다가왔다. 참여 기업들에 대해서 미리 찾아보고 온 학생들도 있었고, 해보고 싶은 주제와 맞는 기업을 알고자 하는 학생들도 있었으며, 이미 설명을 들으면서 마음속으로 주제를 정하고 있는 학생들도 있었던 것 같다. 기업의 발표자들은 한정된 시간 안에서 기업의 비전과 업무 분야와 장점들을 전달하고 강렬한 인상을 남기기 위해 노력했던 것 같다.

2. 주제선정

- 학생들은 1주간 HIRA 빅데이터에 대한 설명과 분석법, 분석패키지에 대한 교육을 진행하였으며 기업들은 교육 담당자와 미팅하면서 진행하고자 하는 연구주제에 대한 조율이 이루어졌다. 빅데이터 전문기업이며 6개 상급종합병원 CDW 구축기업인 미소정보기술은 가장 자신 있는 분야이자, HIRA 빅데이터에 가장 잘 맞는 주제인 '질병 네트워크'를 제안하였다. 하지만 4주가 채 되지 않는 짧은 기간에 데이터로서도, 의료분야로서도 비전문가인 학생들이 이 주제를 잘 이해하고 따라와 줄 수 있을지 걱정되지 않을 수 없었다. 하지만 우려가 무색하게도 학생들은 굳은 의지를 표명했고 이것을 계기로 멘토로서의 마음가짐도 전환하게 되지 않았나 싶다.

3. 멘토링 과정

- 급할수록 돌아가라 했던가, 바쁜 일정에 마음은 급하지만 멘토링을 시작하면 꼭 하고 싶었던

* 「2020년 통계 HIRA 빅데이터 기반 산·학·관 연계 맞춤형 인재 양성 교육」 최우수 사례(팀명: :DNET)

교육이 있었다. 빅데이터의 의미와 접근법, 의료 데이터를 대하는 자세, 데이터를 표현하는 방법, 분석 솔루션의 활용법 등 알려주고 싶은 것들이 너무 많았다. 조원들과의 첫 미팅 때, 가벼운 인사와 앞으로의 각오를 들어보는 편안한 자리여야 할 그 시간에, 다짜고짜 빅데이터 개론을 펼쳤던 것 같다. 아이러니하게도 프로젝트 일정이 급함에도 숨을 고르고 주변 지식부터 끌어 올리려고 시작한 멘토링이, 빅데이터 기본 교육에서 숨 가쁘게 돌아갔던 것 같다.

- 빅데이터는 크기가 크다는 것만으로 정의할 수 없다. 어떤 필요에 의해서 데이터가 생성되었고, 어떤 목적에 의해서 수집되었으며, 어떤 기준에 의해서 분류되어 관리되고 있는지 그 이유를 먼저 알아야 한다고 알려주었다. 이번 교육을 통해 학생들이 다루어야 할 데이터는 HIRA 빅데이터인 건강보험 청구자료이다. 학생들은 이 데이터가 의료기관에서 이루어진 진료행위에 대해 건강보험 청구를 위한 데이터이며, 데이터의 속성과 구조 등을 교육을 통해 이해하고 있었다. 따라서 학생들에게 데이터 속성 전체를 이해하고 필요한 부분을 접근해 들어가는 Top-down 방식과 주제로부터 데이터를 검색하고 찾아 나가는 Bottom-up 방식을 모두 고려해볼 것을 제안하였다.
- 멘토링 방법은 4주간 빅데이터의 이해, 질병 네트워크의 정의, 분석 대상의 선정, 연구결과의 표현법에 대한 지도와 숙제, 결과와 피드백으로 진행되었으며 어렵지만 잘 따라와 준 조원들이 대견했다.
- 의료 데이터 및 용어에 대한 전문 지식 부족과 약 1억건 분량의 의료 빅데이터 활용 경험 부족, 질병 네트워크에 대한 이해 부족을 어려운 점으로 꼽았으나, 심평원 데이터의 구성요소와 용어에 대한 정의 학습, 분석에 대한 일정 기준 설정 및 재탐색(표본 줄이기), 질병 네트워크에 대한 기사, 논문 등을 탐색하여 적절히 학습해 나갔다. Top-down 방식 선택 이후 질병 네트워크 특성 이해를 위해 '데이터 분석'보다 '데이터 탐험'이라는 방향성을 제시해 주었으며 상병 선정을 위해 3가지 기준점을 찾아주었다. 첫 번째, 국민 관심 질병 목록을 기준으로 상병을 정의, 두 번째, 분석 단위 환자 기준으로 상병 일련번호를 10까지 고려, 세 번째, 환자 에피소드별로 데이터를 분류하는 것이 그것이다.
- 본 교육 과정에서 활용한 자료는 건강보험심사평가원의 건강보험 청구자료를 환자의 인구학적 특성을 기준으로 표본 추출한 후 민감정보에 대한 비식별처리가 진행된 자료이다. 2015년에 청구된 명세서를 기준으로 전체 환자의 3%를 추출하여 분석 대상으로 하고 있으며, 대상환자의 4개년도(2015~2018년)의 자료가 구축되어있었다. 이는 명세서 기준으로 1억 여건, 환자 기준 140만 여명이었다. 총 90개 주상병을 분석하였고 2차 분석 대상 질병을 선정하여 국내 질병 사망원인 중 2위(약 21%)를 차지하고 있는 "순환계 질환"을 선택하게 되었다. 이는 1위인 암(약 27%)이 부상병의 원인을 특정하기 어렵다는 점과 미리 예방이 용이하지 않다는 점을 고려한 결과이다. 또한 순환계 질환 중에서도 고혈압과 협심증은 다른 질병의 전조 증상이나 합병증인 특성을 지니고 있으므로 발생 비중이 3번째로 높은 뇌졸중(29,590명)을 분석 주상병으로

선택하게 되었다.

- 질병 네트워크의 특성상 주상병과 부상병의 선후관계, 인과관계 등을 고려해서 끊임없이 탐색해 나가는 것이 목적이나 한정된 시간과 인적 리소스를 생각할 때 주상병을 특정하고 하나의 네트워크를 구성하는 것으로 한정짓는 것이 중요했기 때문이다.
- 또한 몇가지 제한요인을 더 고려하였다. 요양기관을 상급 종합 병원과 종합 병원으로 제한하였다. 입원과 외래 진료를 모두 포함하되, 뇌졸중의 경우 외래는 4년간 8회 이상인 환자로 제한하였다. 다빈도와 관련성에 기반한 부상병을 선정하였고 이는 Bottom-up 방식을 활용하여 기준에 따른 표본 크기를 6,663명으로 산출하였다.
- 뇌졸중 최초 진단 일자 기준으로 이전 1년과 이후 1년의 전후 비교를 하였는데 이는 최초 진단 일자 전후 비교를 통해 새롭게 나타난 질병을 관찰하였고 전후 모두 나타난 질병이지만 빈도의 변화를 관찰하여 주상병과 관련된 부상병을 채택함과 동시에 한 방향이 과대평가 되는 것을 방지하기 위하여 전후 1년을 범위로 설정하게 되었다. 추후 환자 에피소드와 관련된 생활 패턴 데이터 등을 추가 활용하여 연구할 경우, 보다 심도있는 연구가 가능하며 질병에 대한 이해도를 높일 수 있을 것이라고 생각한 학생들의 의견을 듣고 대견하지 않을 수 없었다.

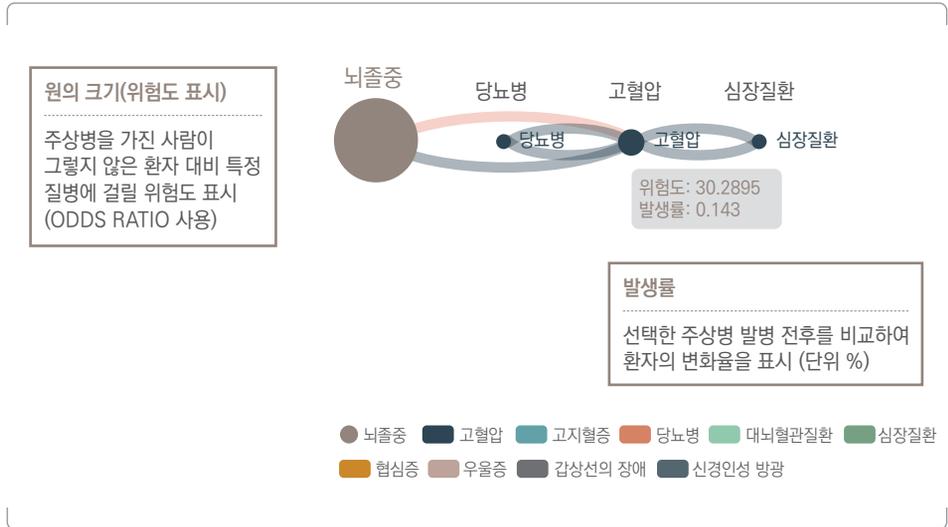
4. 업체 현장방문

- 3주차 주중에 미소정보기술 서울 본사를 방문하여 빅데이터 분석에 사용되는 솔루션 설명 및 BI 솔루션 Qlik Sense에 대한 3시간 분량 기초 교육을 받았다. 실제로 기업에서 빅데이터 분석에 사용되는 사례를 중심으로 현장학습 체험에 주안점을 두었으며 가능하다면 이번 교육에 도움이 될 수 있었으면 하는 바람이었다. 학생들은 매우 만족스러운 시간을 가졌고 교육 중 가장 기억에 남는 날이었다고 전하기도 했다.

5. 연구 및 기술지원, 최종 발표

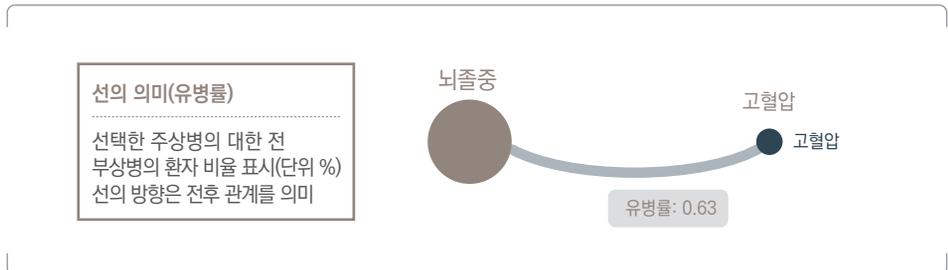
- 기업에서 멘토링 외에 기술적으로 도움을 줄 수 있는 부분은 미소정보기술 장점 중 하나인 시각화 부분에서의 웹 기술지원이었다. 기업부설연구소 선임인력과 컴퓨터공학과 학생의 협업으로 홈페이지를 구성하고 질병 네트워크 차트를 구현하였다.

[그림 1] 2depth까지 구현한 질병 네트워크의 일부 예시

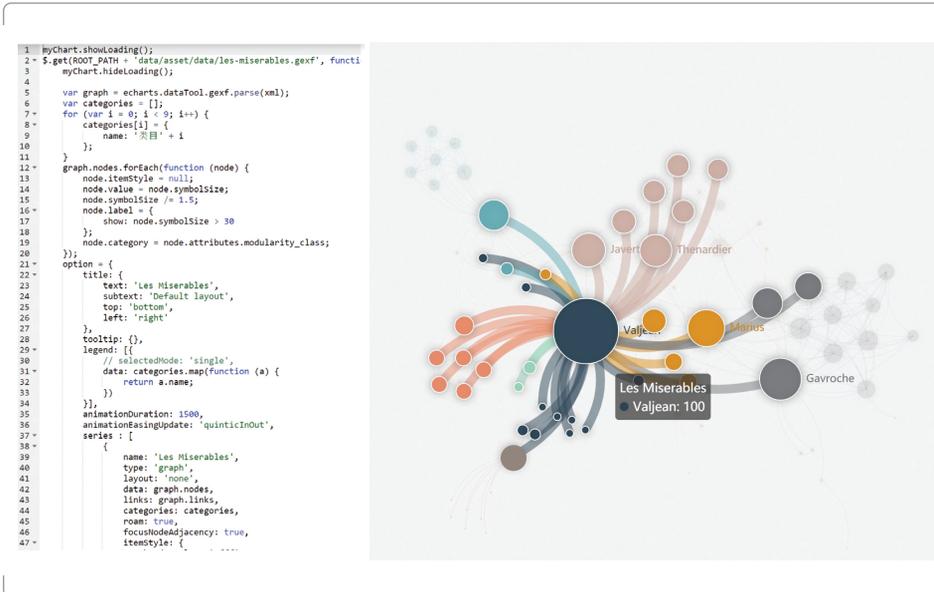


- 전자정부 프레임워크(e-gov framework)의 기본적인 틀을 제공, e-chart의 차트 표현과 일부 시각화 기능의 customizing, 웹 소스에 들어가는 데이터 형태 등을 지원하여 컴퓨터공학과 학생이 데이터를 정제, 적재하여 실제로 동작하도록 지원하였다.

[그림 2] 네트워크 차트에서 표현하고자 했던 부분을 정리하여 발표



[그림 3] 실제 동작하는 질병 네트워크 차트에 사용된 e-chart 예제 소스



- 정책상 연구 결과물이 폐쇄환경에서 외부로 나올 수 없으므로, 만든 결과물을 공개할 수는 없으나, 질병 네트워크 차트를 통해 (1) 직관적인 시각화에 의한 질병에 대한 관심 증대 (2) 정보 접근성 강화 (3) 국민들이 질병에 대한 관심 증가로 인한 신 의료기술 개발 촉진을 기대효과와 활용방안으로 꼽을 수 있었다.
- 최종 발표에서는 다른 조 학생들도 매칭된 기업들과 4주간의 결과물을 한자리에서 발표하였다. 훌륭한 조도 있었고 학생답게 풋풋한 느낌을 주는 조도 있었으나, 대체로 짧은 기간 동안 가진 역량을 보여주기 위해 많은 노력을 한 느낌을 주었다.

6. 멘토링 후기

- 4주간의 산·학·관 연계 교육은 심평원의 기획 의도에 걸맞게 참여 학생들에게 많은 기회와 시사점을 제공하였다.
- 첫 번째, HIRA 빅데이터를 배우고 연구할 수 있는 기회를 얻은 것 이다. 이는 비록 폐쇄된 환경에서 제공된 데이터만을 활용할 수 밖에 없었지만, 다기관 통합 빅데이터 플랫폼이 요원한 현 의료 데이터 상황에서 충분히 큰 메리트가 있는 기회였다. 과기부에서 추진하고 있는 10대 빅데이터 플랫폼의 헬스케어 분야가 의료 전체 도메인이 아닌 암으로 한정하여 시도하고 있으며, 참여한 11개 암센터(병원)의 자격을 갖춘 연구자만 IRB 승인을 통해서 다기관의 데이터를 분석할 수 있다는 점에서 HIRA 빅데이터의 청구데이터 분석 경험은 앞으로도 많은 인사이트를 주게 될 것이다.

- 두 번째, 의료 도메인에 전문지식이 없는 학생들도 연구에 필요한 최소한의 교육과정을 이수하여 분석 프로젝트에 참여할 수 있었다는 점이다. 3인으로 이루어진 조는 글로벌 행정학과 학생, 의과학융합전공 학생, 컴퓨터공학과 학생으로 이루어졌지만 전공을 떠나 단일 프로젝트로서 짧은 4주간의 교육 기간동안 의미 있는 고민들을 통해 자신들의 주장을 펼쳐나갈 수 있었다.
- 세 번째, 기업과의 연계 collaboration이다. 시작부터 1위를 목표로 하자는 강한 자신감과는 다르게 일정의 촉박함으로 인해 주 1회 대면 멘토링이 부족하여 상시 SNS Q&A를 진행하였다. 데이터 접근법과 분석 기법, 기술지원 등의 다양한 실무 경험을 겪으면서 분석을 결과물로 현실화하는 것에 자신감을 갖게 되었을 것이다.
- 안타깝게도 기업은 학생들과는 달리 얻을 수 있는 advantage는 크지 않았다. 시간적, 공간적 제약에 따른 현장 멘토링의 한계가 다소 아쉬운 점이다. 교육 이후 채용 의사도 적극 전달하였으나, 요즘 시기와 같은 고학력 시대에 공부를 더 이어나가거나 대기업, 공기업에 지원하고자 하는 의지에 응원의 박수만을 보낼 뿐이었다. 실제로 기업이 고민하고 있는 부분이나, 신규 사업으로 확장하고자 하는 아이템을 프로젝트 주제로 잡을 수도 있었으나, POC^[1] 혹은 BMT^[2] 개념의 교육을 진행하는 것은 우리가 따르기 때문이다.
- 산·학·관 연계 교육은 의도만으로도 매우 의미가 큰 사업이다. 원주, 나아가 강원 지역의 대표적인 공공기관으로서 건강보험심사평가원이 학생들에게 다양한 교육기회와 현장경험을 제공한다는 것만으로도 학생들의 지원 문의가 쇄도할 만하다. 더욱 신선한 교육 프로그램으로 서로 많은 것을 경험하고 얻을 수 있는 교육으로 거듭나길 기대해 본다.

용어설명

[1] POC(Proof Of Concept : 기술 검증, 개념 검증) : 신기술을 프로젝트에 도입하기 앞서 검증하는 목적

[2] BMT(Bench Marking Test : 실제 환경에서의 성능 시험) : 일반적인 성능 시험이 아닌 실제와 동일한 환경에서 비교시험을 반복하여 객관적인 성능을 평가하는 것