

기술

과학·기술

[마소 395호] 인문학도, 데이터 사이언티스트가 되다

오세용 기자

입력 2019.02.01 06:00

소프트웨어 전문지 마이크로소프트웨어 395호는 데이터과학을 주제로 담았습니다. 데이터과학에 대한 개론, 학습 방법, 실무 적용 사례, 학계 등 마소 395호 주요 기사들을 IT조선 독자에게도 소개합니다. **[편집 자주]**

필자가 처음 만난 데이터 사이언티스트는 사람의 마음을 읽는 마법사였다. 고등학교 2학년, 우연히 학교로 날아온 서울대 데이터마이닝 캠프 공문을 보게 됐고, 흥미를 느껴 참가했다. 평소 사람에 대한 호기심이 많아 사람이 어떤 관심, 어떤 생각을 가지고 살아가는지 항상 궁금했다. 그런 필자에게 빅데이터를 통해 사람의 생각과 욕망을 읽을 수 있다는 데이터마이닝은 마치 마법 같았다.

필자는 그날 캠프에서 빅데이터에 단숨에 매료됐다. 서울대 데이터마이닝 캠프는 한국BI데이터마이닝학회에서 고등학생을 대상으로 개최하는 캠프로 분류, 예측부터 클러스터링, 데이터 시각화 등 데이터마이닝과 관련된 개괄적인 내용을 배울 수 있는 캠프다. 또한 데이터가 실제 우리 삶 속에서 어떻게 가공, 활용되는지 데이터 마이너가 되기 위해 어떤 역량을 갖춰야 하는지 등을 알 수 있었다.

데이터로 사람의 마음을 읽고, 다양한 솔루션을 도출해 낼 수 있다는 기대를 했다. 필자는 반드시 데이터 분석가가 되겠다는 다짐을 했다. 학교로 돌아와서도 데이터 분석 관련 서적을 읽으며 정보를 찾아봤다. 이후 대학교에 진학했고, 신학과에서 인문학 공부를 하며 데이터 분석을 하고 싶다는 마음은 더욱 강해졌다.

컴퓨터 과학 공부와 함께 원래 목표였던 데이터 분석 스킬을 기르기 위해 노력했다. 먼저 빅데이터 분석 동아리에 들어가 여러 기법을 배웠고, 작은 딥러닝 프로젝트를 진행했다. 하지만 진짜 현업에서 가치를 창출할 수 있는 데이터 분석을 경험해보고 싶다는 욕심이 생겼고, 이 욕심을 채울 수 있는 곳을 찾았다.



엠로보 단체사진(사진 맨 오른쪽이 필자 유예진). / 마이크로소프트웨어 395호

필자가 선택한 곳은 스타트업이었다. 2017년 12월 데이터 저널리즘 스타트업 엠로보에 입사했다. 실제 현업에서 사용할 수 있는 머신러닝 프로젝트를 진행했다. 부족한 필자가 머신러닝 프로젝트를 전임하게 된 것은 큰 행운이었다. 주어진 첫 미션은 딱 두 가지 키워드였다. 기업공시 그리고 머신러닝이었다.

미션이 굉장히 러프해서 주제를 구체화하는 것 자체가 큰 난관이었다. 미션과 동시에 자유롭게 쓸 수 있는 GPU까지 선물 받았다. 어깨는 더 무거워졌다. 그렇게 머신러닝 프로젝트의 막이 올랐다.



DART 전자공시시스템(출처: dart.fss.or.kr). / 마이크로소프트웨어 395호

2가지 키워드 '기업공시'와 '머신러닝'을 두고 필자는 기업공시로 할 수 있는 다양한 머신러닝 프로젝트를 고민했다.

기업공시란 주주, 채권자, 소비자 등 기업 이해관계자에게 기업가치를 평가하고 판단하는 기업 정보 자료를 제공하는 것을 말한다. 그리고 금융감독원에서는 상장법인 등이 공시서류를 인터넷으로 제출하고, 투자자 등 이용자는 제출 즉시 인터넷을 통해 조회할 수 있도록 하는 종합적 기업공시 시스템인 전자공시 시스템을 운영하고 있다.

그렇다면, 기업공시로 할 수 있는 머신러닝이 뭐가 있을까?

첫 프로젝트였고 자율적으로 주제를 선정할 수 있는 만큼 가장 재밌고 흥미로운 것을 하고 싶다는 욕심이 들었다. 역시 '돈'이 연관돼야 프로젝트가 흥미로워질 것이라는 생각이 제일 먼저 들었다. 도메인도 금융인 만큼 무조건 주가와 관련된 프로젝트를 해야겠다고 결심했다. 이 프로젝트를 통해 무언가 실질적인 효과를 이끌어 낼 수 있다면 좋을 것 같았다. 필자가 엠로보에 입사한 이유이기도 했다.

엠로보에서는 공시를 실시간으로 기사로 바꿔 메일링하는 공시봇 서비스를 했다. 이 서비스에 함께 실려갈 수 있는 결과물이 나온다면 프로젝트가 더욱 유익하고 효과적일 것 같았다. 이런 생각을 바탕으로 프로젝트 최종 목표를 기업 공시 이벤트가 발생하면 실시간으로 주가 상승 신호를 알아내 투자에 이용하는 것으로 결정했다.

유예진 필자의 '인문학도, 데이터 사이언티스트가 되다'에 대한 자세한 내용은 '마이크로소프트웨어 395호(<https://www.imaso.co.kr/archives/4654>)'에서 확인할 수 있다.

IT Chosun 뉴스레터를 **구독신청** 하세요