[ 세미나 후기 ]

[ 회귀분석팀 : 서울시 아동보호전문기관 우선 입지 선정 ]

* 피피티 누가 만드신건가요.. 너무 잘 만들어서 내용이 눈에 훨씬 잘 들어온 것 같아요! 그리고 피드백을 반영해서 소득 수준 변수를 새로 만든것도 천만원단위로 끊어서 비율로 잘 만드신 것이 기억에 남습니다. 그리고 CCTV 데이터 누락된 부분에 대해서는 누락된 행정동의 데이터를 따로 가져와서 채워주는 노력까지.. 고생하셨어요..

그리고 입지선정이라는 주제를 어떻게 이끌어갈 수 있을지에 대한 인사이트를 얻을 수 있었던 발표였습니다. 수집한 데이터들로 위험도 지수를 산출하고, 대체기관 개수와 그 위험도 지수를 축으로 해서 여러 방법으로 클러스터링을 진행해, 가장 적합해보이는 클러스터링을 선택하는 과정까지 너무 스무스한 분석 흐름이었고 논리적인 흐름이었던 것 같습니다.. 덕분에 정말 많이 배운 것 같아요! 반경 내 수요 최대화를 실행해주는 MCLP 방법과 같은 입지선정 방법 같은 새로운 방법들을 배워서 좋았고, 마지막 최종 입지 선정에서 청소년 유해업소 근처 건물은 제외하는 세심한 분석까지.. 정말 보면서 감탄이 절로 나왔습니다.. 정말 고생하셨습니다!

이수정 : 앞부분이 저번 발표때의 피드백과 각종 데이터 전처리 부분이어서 흐름이 굉장히 중요한 부분이라 생각하는데, 그 점을 정말 잘 설명해주신 것 같습니다! 왜 데이터를 이렇게 전처리했으며, 왜 위험도 지수를 이렇게 산출했는지에 대한 흐름이 아주 깔끔하게 연결되어서 이해가 쉬웠던 것 같습니다. 고생하셨습니다~

진수정 : 클러스터링에 이렇게 많은 방법들이 존재하는지 처음알았고, 클러스터링 – 입지선정 – 최종 입지선정으로 이어지는 흐름이 너무 깔끔했다! 이 모든걸 수정이 너랑 은주가 이끌었다니.. 진짜 통계 잘하는구나.. 피셋 1년간 하느라 고생 많았어!! 나도 담학기까지 했을 때 너만큼 할수 있었음 좋겠다 ㅎㅎ

시계열분석팀 : 다사다난한 주제분석이었지만 성능이 꽤 잘나와서 정말 뿌듯했다! Svm의 accuracy 성능도 잘 나오고, 머신러닝의 랜덤포레스트, 딥러닝의 GRU모델의 RMSE 성능도 잘 나와서 다행이야.. 랜덤포레스트를 담당했을때.. 차분 – 시계열 교차검증을 랜덤포레스트에 적용 – 최적의 파라미터로 랜덤포레스트 적합해서 예측값 출력 – 예측값에서 차분을 풀어주는 이 과정을 코딩하는게 저 같은 코린이한텐 참 어려웠는데 피드백을 받으며 계속 수정한 결과 코드도 잘 나올 수 있었습니다..

이재현 : 피셋이 낳은 인재, 딥러닝 마스터 재현이형.. 형 뒤 졸졸 따라다니면서 딥러닝 배우려고 했는데 내 노트북으론 안 될 것 같아 다시 생각해보니..ㅎㅎ 발표의 앞부분이 분석의 흐름과 직결되는 부분이라 잘 설명해주는 것이 중요했는데 너무나 잘해줬다.. 고생많았어!!

이정우 : 20분을 넘기지 않기 위해 랩을 했는데, 알아듣기 힘들었을거 같아요..ㅋㅋㅋㅋ 혹시라도 나중에 다시 피피티를 보고 질문이 생기신다면.. 질문주시면 답해드릴게요 ! 그나저나 이 주제분석을 위해서 경제논문만 몇 개를 찾아본지 모르겠네요.. 어림잡아 10개정도 발췌독한거 같은데 정말… 순탄치 않았습니다..ㅎㅎ 그래도 제가 만진 모델이 성능을 좀 내니깐 기분이 좋았습니다. 첫 주제분석을 시계열 범벅인 비트코인으로 해서 많이 헤매기도 했지만 그만큼 얻어간것도 많았습니다.

선형대수학팀 : 정확한 예측을 위해 랜포를, 정밀한 해석을 위해 로지스틱 회귀를 적합시켜서 두 모델로 목적에 따라 적합시킨 것이 인상깊었습니다. 그리고 피피티 너무 센스넘쳐서 와…. 하며 봤어요! 어떻게 이렇게 잘 만드시죠..? 그리고 범주형 변수 인코딩들에 대해서 패키지에서 봤던 것들이라 반가웠습니다! 중간에 전체적인 흐름을 정리해서 연속형 변수, 범주형변수를 각각 어떻게 처리했고, 어떤 방식으로 모델링을 했는지를 정리해줘서 이해하기도 좋았습니다! 그리고 랜덤포레스트 해석을 할 때 importance plot은 알았는데 partial dependence plot과 lime같은 것들도 활용할 수 있다는 건 처음 알았습니다. 마지막 활용 방안까지… 정말 잘 들었습니다!

1. 전효림 : cramer’s V라는 걸 통해 다중공선성이 작은 변수를 선택하고, 남은 변수 세가지에 대해서 독립성검정을 진행 – 모두 독립이 아니어서 하나를 선택하는 이 흐름에 대해서 배울 수 있어 좋았습니다! 만약 저였으면 어떻게 할지 고민이 많았을 것 같은데 너무나 논리적으로 문제를 해결하셔서 대단하신 것 같습니다..
2. 김지민 : 랜덤포레스트에 중요도플랏 말고도 라임 같은 방법들이 존재하는지 처음 알았어요! 그리고 모델링을 9개나 진행하시다니.. 감탄했습니다.. 모델링부터 활용방안 – 의의와 한계까지 흐름이 아주 매끄러웠던 것 같습니다.!