

zero-base /

Life Changing Education

과제 (4월 27일)

—

- 시작하기 전에 미리 이야기를 드립니다.
- 이제부터는 머신러닝/딥러닝 과제를 여러분이 수행해야 합니다.
- 아마도 이 과정을 몇 번 해보면 데이터를 정리하는것, 머신러닝/딥러닝의 결과를 정리하는 것 말고는 없는데? 라고 생각할 수 있습니다.
- 사실입니다. 어떤 과정을 거치는지 잘 정리하는 것이 절반, 데이터를 확보 정리하는 것이 절반이라고 말 할 수 있을 정도입니다.



- 이제 여러분은 회사에 오랫동안 회자될 레전드 신입사원이 될 것입니다.
- 비록 어떤 데이터에 전문가도 아닐 수도 있고 여러분들이 모르는 분야에 대해 조사하고 결과를 얻어야할 수도 있습니다.
- 그러나 여러분이 멋지게 해 낼 것이라고 믿습니다.



- 여러분들이 이 과정을 충분히 즐기시길 바랍니다.

- 여러분들은 부산 해운대에서 오랜만에 친구를 만나기로 했습니다.



- 친구와 함께 즐거운 대화를 하던 중
- 옆 테이블에서 이런 이야기가 들려옵니다.



- 그것은 바로 전설의 투수 최동원에 대한 이야기였습니다.
- 어떤 이는 그는 롯데 최고의 투수라고 하고, 또 어떤 이는 한국 최고의 투수라고 합니다.
- 또 어떤 이는 지금 최동원이 한국에 있다면 아마 30억은 받았을 것이라고 하고, 또 다른 이는 아무리 그래도 아마 3억 정도일 것 이라고 합니다.

- 한 선수의 연봉을, 그것도 은퇴하고 고인이 된 선수의 연봉을 바로 가늠하는 것은 어렵습니다.
- 머신러닝을 익히고 공부하는 내용으로 이해해주기 바랍니다.



- 문제1 ) 우리 과제는 최동원의 롯데 시절 1983년부터 1988년의 데이터를 사용하도록 하겠습니다. 최동원의 당시 데이터와 함께 1983년부터 1988년 사이 투수들의 데이터(연봉포함)를 구하고, 2015년부터 2020년까지 프로야구의 투수들의 데이터를 구하시오.
- 모든 투수의 데이터를 구하기 어려울 수도 있습니다. 모두 구할 필요는 없습니다. 그러나 선발/마무리에 대해서는 고민하기 바랍니다. 왜냐하면 최동원은 선발로도, 혹은 마무리로도 활동을 했기 때문입니다.

- 문제2) 사실 가장 중요한 부분인데, 1980년대 최동원이 활동하던 투수들의 연봉 및 데이터를 2010년대 투수들의 연봉 및 데이터로 매핑(mapping) 즉, 변환하는 문제가 있을 것입니다. 이 문제를 효과적으로 해결하시오.
- 즉, 우리는 최동원이 현재 존재한다면 그 연봉을 알고 싶은 것이니, 최동원의 데이터만 현재 선수에 맞춰 변환할 것인지, 혹은, 당시 선수들의 데이터를 모두 변환해서 검토할 것인지 결정해야 할 것입니다.
- 이 문제는 간단하지 않습니다. 그러나 여러분들이 염두해 두어야 할 것은 “최동원 선수가 현역 선수라면 연봉이 얼마일까”라는 문제를 해결해야 한다는 것입니다.

- 문제 3) 이제 2번의 데이터를 기준으로 최동원의 성적이라면 연봉이 어떨지 예측하시오. 여기서 여러분들은 어쩌면 많은 고민을 해야하고, 다시 1번 문제로 돌아갈 수도 있습니다.



- 수고하셨습니다.
- 여러분은 최고입니다. 언제나 응원합니다.