우선 neural net이랑 naïve Bayesian등등 써봤는데 성능 엄청 안나옴.

그래서 포럼에서 스크립트 돌아다니면서 봤는데 대부분의 사람들이

xgboost라는 걸 씀. Xgboost는 extreme gradient boosting이라고 decision tree를 사용해서 learning시키는 건가봄 <https://en.wikipedia.org/wiki/Gradient_boosting>

여기 참고해 보시길. 그래서 나는 매틀랩 쓰니까 매틀랩 라이브러리 찾아다가 씀

<https://sites.google.com/site/carlosbecker/resources/gradient-boosting-boosted-trees>

이거 가져다가 씀, 모든 feature중에서 action\_type이 가장 큰 영향을 낀치다는 것을 알아냄, 그래서 이 action\_type을 parsing해서 17개의 feature로 바꿈 이때 성능이 매우 좋아짐, 또한 포럼을 돌아다니다 보니 time\_reamining < 3일 경우에 last\_moment라고 피쳐를 또 하나 만듬 이거 추가해서 또 좋아짐,

이 이후로는 별지랄을 해도 성능이 좋아지지가 않음, recommendation system이라 가정하고, base line estimation을 한 후 에 linear하게 mix해봤지만, 성능이 좋아지지 않음, 사람들 대부분이 python이랑 xgb classfier쓰길래, 나도 python에서 내가 만든 피쳐를 가지고 시험해 봄, 존나 성능 좋아짐, 피쳐 몇개를 제거하고 hyper parameter들을 validation을 통해 결정해서 냈더니 0.59 나옴 섹스