

0920 회의

2020년 9월 20일 일요일 오후 11:03

1. 모델링

1) LGBM

: 평균값으로 처리하기 -> 마무리

: 상대 feature를 추가해보기

관련 링크) https://github.com/dj-1087/BigContest-COIN/blob/master/LMST_EX.ipynb

2) LSTM

: 4개의 모델 사용

: best parameter search (grid search)

2. 모델 사용 근거 관련

: 시계열 모델 사용 조건 확인 (정상성 만족)

p값이 특정 변수 외에는 모두 정상성을 만족하고, 만족하지 않은 것조차 만족한다고 볼 수 있는 수준임. (by p-value 확인)

: 데이터 개수가 매우 적음 -> 트리 중에서도 부스팅 계열을 쓴 이유. 단일 트리모델을 쓰면 운에 의존할 가능성이 높아져서, variance를 높여줄수 있는

: 차원이 3개 이상이면 knn은 잘 안씀.

3. 앞으로의 방향

- 확실히 가능 : VAR(타율, 방어율) LSTM(타율, 방어율)
- 데이터 수집 방법 찾기
- 자신이 맡은 모델링 설명 -> 보고서 완성

참고)

```
0~163
0~19 / 20~39 / ....
0~19 : y_next => 20 ~ 39 mean
164/20= 4 + 20
ERA
140 ~ 159: y_next => null , 160 ~ 163 : y_next => null 163~
20+a => mean

20 경기 0~3 null

~163 + 20 concat VAR
LGBM
```

4. 역할분담

- 창건오빠 : LSTM 완성
- 상대오빠 : 최종 결과값 내는 코드 완성
- 민영언니 : 데이터 수집 모음
- 동현오빠, 윤정 : 보고서 작성 (ㄹ까지)
 - ㄱ. 문제 정의, 대회소개
 - ㄴ. 목표
 - ㄷ. 데이터 전처리 : 세이버메트릭스 추가
 - ㄹ. EDA + 모델 선정 검증 과정 : 여기는 왜 이 모델을 선정할 수밖에 없었는지 초점
 - ㅁ. 모델(VAR, LGBM, LSTM) : 정의, 필요과정, ~알고리즘 생성 ...
 - ㅂ. 결과