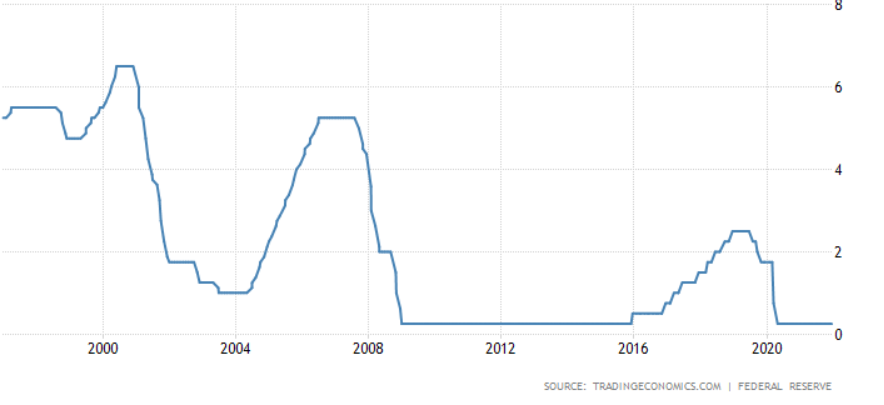
Train\_set : 모델의 학습을 위해 사용됩니다.

Test\_set : 생성된 모델의 예측 성능을 평가하는데 사용됩니다. 학습과 튜닝에는 이용되지 않고, 미래의 타겟 값이 관측되지 않은 데이터라 가정하여 예측이 잘 되는지 평가하는데 사용됩니다.

모델의 검증해야 하는 이 test set이 train set의 데이터와 다른 특성을 지니고 있으면 모델 성능 문제 뿐만 아니라, 모델 평가에도 적절하지 않다고 판단됩니다. 기업 부도와 경기 국면같은 거시경제적 요인은 동시에 고려되어야 합니다. 그래서 기간을 기준으로 train test split을 진행하는 경우 경기 국면이 모두 반영되게끔 나눠야합니다. 선행 연구 다수에도 기간을 기준으로 분리를 한 것으로 확인됩니다. 하지만 저희가 설정한 기간에서는 07년부터 20년까지의 기간으로 미국의 경기 순환 10주기가 종료된 2009년 6월이 포함되어 있습니다. 하지만 아직 11주기가 확정되지 않은 상태이고 저희가 설정한 기간과 부합한다고 보기는 어렵다고 판단하였습니다. 따라서 미국 기준 금리 추이를 확인한 결과, 08년 이후 급격한 금리 인하가 발생하였고 16년 이후로 금리가 상승하다 코로나 19의 충격으로 다시 급격하게 인하한 것을 확인할 수 있었습니다. 미국 금리가 신흥국 기업 부도 위험이 높아지게 되는데 한국의 주권 시장은 MSCI에서 아직 신흥국으로 분류하고 있습니다. 따라서 미국 금리 변동이 국내에 미치는 영향이 크다고 판단하였습니다. 이를 적용하여 기간 기준이 미국 금리의 변동 추이를 반영하지 못하는 한계를 극복하고자 임의 추출을 통해 train, test를 분리하여 model을 검증해보는 시도를 해보았습니다.

<미국 금리 표> <https://ko.tradingeconomics.com/united-states/interest-rate>



<기업경영분석><http://ecos.bok.or.kr/flex/visualStat2015.jsp>

