**프로젝트 2**

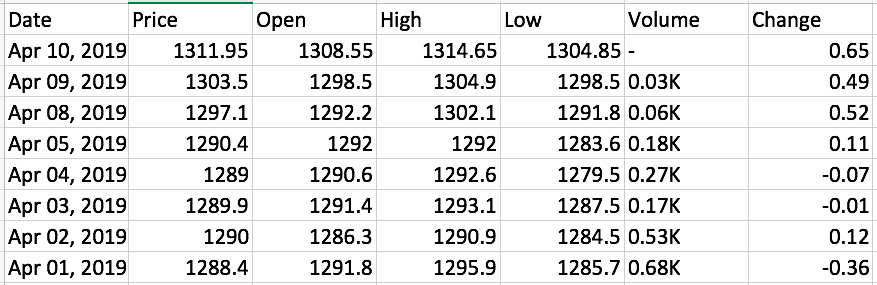
406.561 데이터기반학습(Learning From Data)

2021년도 1학기

**프로젝트 설명:**

프로젝트 2에서는 hidden markov model (HMM)을 활용하여 미래의 **금값을 예측**하는 기법의 개발을 목표로 한다. 개발 언어로는 Python을, 라이브러리로는 hmmlearn을 사용해야 한다.

* Gold price prediction using time-series data
  + 주어진 과거 데이터를 이용해 미래의 금값을 예측하는 문제
  + t-1시점부터 t-n시점의 데이터로 feature를 만들어 t시점의 Price(종가) 값을 예측한다. (향후 10일 동안)
* 데이터셋
  + 온라인 상에 공개 되어있는 data 사용
  + 2010년 1월 1일부터 2021년 4월 1일까지 daily data
  + Exchange rate (환율) 데이터
    - 원/달러, 원/호주달러, 원/유로, 원/홍콩달러, 원/파운드, 원/엔, 원/위안: Price, Open, High, Low, Change %
  + Commodities (원자재) 데이터
    - 금, 은, 구리, 백금, 브렌트유, WTI유, 가솔린, 천연가스:  
      Price, Open, High, Low, Volume, Change %
* **예시 자료**
  + 금값



* **프로젝트 목표**
  + **주어진 데이터를 이용하여 HMM 모델을 학습시키고 학습된 모델로 미래의 금값을 예측**
  + **예측된** **금값과 실제 금값의 차이를 통해 MAE(Mean absolute error)를 계산하여 개발된 모델의 성능 점수를 산출**

\text{MAE}(y, \hat{y}) = \frac{1}{n_{\text{samples}}} \sum_{i=0}^{n_{\text{samples}}-1} \left| y_i - \hat{y}_i \right|.

* + **모델의 performance 평가는 제출 이후 2주간의 실제 금값을 기준으로 MAE를 계산하여 이루어짐**

**지침:**

* 프로젝트는 팀 단위로 수행
* 제출할 결과물은
  + 보고서와 발표자료: 사용한 feature 목록 및 전처리 방법, 사용한 파라미터 그리고 성능 평가 결과 등을 포함한 ppt 형태의 파일과 3페이지 이내의 워드 파일
  + **발표 영상 : 팀 당 3분 30초 이내로 발표 영상을 제작하여 제출함**

**(powerpoint의 녹화기능을 사용)**

**(해당 발표 영상을 프로젝트 발표일에 zoom을 통해 재생할 예정)**

* + 소스코드와 .pkl 파일
  + 단, 소스 코드의 경우 추가적인 라이브러리를 사용하였다면 보고서에 추가적으로 기술해야 하며, test.py만을 실행 시 반드시 MAE 값이 출력되어야 함
  + **제출된 코드에 오류가 있다면 감점**
* 제출할 결과물들은 모두
  + LFD\_project2\_team00.zip 파일명으로 함께 압축하여 제출일(4월 28일) 23시 59분까지 ETL 사이트에 업로드
  + 제출일(4월 28일) 이후 변경 불가
* **제공된 데이터 이외의 데이터를 사용하는 것을 금지합니다.**

**주요 일정:**

* **프로젝트 결과물 제출일: 2020년 4월 28일 23시 59분까지**
* **프로젝트 발표일: 2020년 4월 29일**
* 제출 기한을 넘길 시, 매 24시간마다 전체 점수의 20% 감점

**평가 방법:**

* model saliency (절대 평가 - A, B, C): 40%
* performance (상대 평가 - S, A, B, C): 40%
* 발표: 10%
* 보고서 (소스 코드 포함): 10%