

ML DL을 활용한 외화 매입 시점 탐색

1. 관련 자료

- 1.1 인터넷 블로그 및 사이트
 - 1.1.1 [이상 감지 - ANOMALY DETECTION](#)
 - 1.1.2 [회귀분석을 이용한 시계열 데이터 이상 탐지](#)
 - 1.1.3 [시계열데이터의 이상탐지를 위한 패키지](#)
 - 1.1.4 [쉽게 이해하는 시계열데이터 비정상탐지](#)
 - 1.1.5 [이상탐지 Anomaly Detection](#)
- 1.2 관련 논문
 - 1.2.1 [SOM을 이용한 주기신호의 이상탐지 및 시각화](#)
 - 1.2.2 [다변량 시계열데이터 집단 상호간의 주성분 정보 비교를 통한 이상 예측](#)
 - 1.2.3 [시계열 모델 기반 트래픽 이상 징후 탐지 기법에 관한 연구](#)
 - 1.2.4 [시계열 분석을 이용한 Netflow 기반의 DDoS 공격 탐지](#)
 - 1.2.5 [시계열을 따르는 공정데이터의 모델 모수기반 이상탐지](#)
 - 1.2.6 [정규분포를 따르지 않는 비선형 데이터를 위한 모델기반 이상탐지](#)
 - 1.2.7 [하나금융경영연구소 해외 송금 환전 이용 현황 분석](#)
 - 1.2.8 [선물환율 결정요소를 이용한 인공지능경망 모형의 원/달러 환율 예측 유용성에 관한 연구](#)
 - 1.2.9 [스왑포인트 결정요소를 이용한 머신러닝 기반의 원달러 환율 예측 모형에 관한 연구](#)

2. 현물거래와 선물거래

- 2.1 외환거래는 일반적으로 거래약정이 체결된 이후 통상 2거래일 이내에 실제 자금결제가 이루어 지는데 이를 **현물환거래**라고 하며, 거래체결 후 3거래일 이후의 미래의 시기에 자금결제가 이루어 지는 외환거래를 **선물환거래**라고 한다. [위키백과: 선물가격](#)
- 2.2 논문에 의하면, 6개월과 3개월 뒤의 가격을 비교한 결과, **선물가격을 예측한 가격이 실제로 오차가 크다는 것**을 알 수 있다. [RMSE 비교 그래프](#)
- 2.3 따라서, **환율 가격 예측 모형**에서는 **선물가격**을 예측하기 보다는 **현물가격**을 예측하는 것이 더 바람직하다.

3. 실질적인 평가지표

- 3.1 회귀 문제의 경우 단순 **RMSE, MAPE** 등의 지표나, 분류 문제의 경우 **Accuracy, F1-score** 등의 일반적으로 머신러닝 모델을 평가하는 지표가 아닌 실질적으로 회사의 수익구조를 계산할 수 있는 평가 지표가 필요함. [평가지표 예시](#)
- 3.2 예시1 : 매수 시도 기간을 1년이라고 잡았을 때, 그 1년 동안 하루에 1달러씩 샀을 때의 가격 대비, 특정 시점에 매수를 몰아서 할 경우의 차이.
- 3.3 예시2 : 매수 시도 기간을 1년이라고 잡았을 때, 그 1년 중 가장 환율이 낮았던 날을 기준으로 220달러(시장이 1년에 220일 열린다고 가정)를 샀을 때와의 차이.

3. 매수 시스템에 대한 정보

- 기업에서 요구하는 매수 시스템이 일주일마다 일정금액을 매입하는지, 1년 동안 목표 금액을 매입하는지에 대한 정보를 알고 있어야 그에 맞는 보다 세밀한 모델 수립이 가능.

4. Random Cut Forest를 활용한 장기 예측 모형

- **4.1** 각 데이터별로 **변칙 점수**를 생성하고, 점수가 낮은 경우 **정상**으로 간주하고, 점수 값이 높은 경우 **변칙or이상**으로 간주한다.
- **4.2** 높고 낮음의 정의는 사용자에게 따라 다르게 설정하는데, 환율 모형의 경우 **값이 현저하게 낮은 지점 혹은 높은 지점을 선택하여 매입 포지션 지점을 "변칙"으로 놓고 학습을 진행하여** 포인트를 찾는다. 하지만 **일반적으로 평균 점수로부터 3의 표준편차를 초과하는 점수가 변칙으로** 간주된다.
- **4.3** 예시 : 1차원 시계열 데이터에 대한 변칙 감지 알고리즘의 적용 사례가 많습니다(예: 트래픽 볼륨 분석 또는 음성 볼륨 급증 감지). RCF는 임의 차원 입력으로 작업하도록 설계되었습니다. Amazon SageMaker RCF는 특징의 수, 데이터 설정 크기 및 인스턴스 수에 따라 조정됩니다.
- **4.4** [Random Cut Forest 설명 사이트](#)

5. 향후 모델 시도 방향

- **5.1 DNN(MLP)을 활용한 모델** : 심층 신경망(Deep Neural Network, DNN)은 입력층(input layer)과 출력층(output layer) 사이에 여러 개의 은닉층(hidden layer)들로 이뤄진 인공신경망(Artificial Neural Network, ANN)이다. 심층 신경망은 일반적인 인공신경망과 마찬가지로 복잡한 비선형 관계(non-linear relationship)들을 모델링할 수 있다.
- **5.2 LSTM(GRU)을 활용한 모델** : 일반적으로 DL 분야에서 Sequence Data에 좋은 성능을 보여온 LSTM(Long-shot-Term-Memory) 레이어와, GRU(Gated Recurrent Unit) 레이어를 사용하여 모델 개발.
- **5.3. Prophet을 활용한 모델** : Facebook에서 발표한 Time Series Python Library. [Prophet Official Paper](#)
- **5.4. Anomaly Detection(이상 탐지)을 활용한 모델** : 장기(6개월 이상)의 이상치(3표준편차를 벗어나는 지점)를 찾기 위해 사용하는

6. 모델 설계 단계시 참고 사항

- **6.1** Train Dataset과 Test Dataset를 12쌍을 만든다. -> **Window Sliding**
 - SET 01 : Train : 2018년 02월 01일 ~ 2019년 01월 31일 ----- Test : 2019년 02월 01일 ~ 2019년 02월 28일
 - SET 02 : Train : 2018년 03월 01일 ~ 2019년 02월 28일 ----- Test : 2019년 03월 01일 ~ 2019년 03월 31일
 - SET 03 : Train : 2018년 04월 01일 ~ 2019년 03월 31일 ----- Test : 2019년 04월 01일 ~ 2019년 04월 30일
 - SET 04 : Train : 2018년 05월 01일 ~ 2019년 04월 30일 ----- Test : 2019년 05월 01일 ~ 2019년 05월 31일
 - SET 05 : Train : 2018년 06월 01일 ~ 2019년 05월 31일 ----- Test : 2019년 06월 01일 ~ 2019년 06월 30일
 - SET 06 : Train : 2018년 07월 01일 ~ 2019년 06월 30일 ----- Test : 2019년 07월 01일 ~ 2019년 07월 31일

- SET 07 : Train : 2018년 08월 01일 ~ 2019년 07월 31일 ----- Test : 2019년 08월 01일 ~ 2019년 08월 31일
- SET 08 : Train : 2018년 09월 01일 ~ 2019년 08월 31일 ----- Test : 2019년 09월 01일 ~ 2019년 09월 30일
- SET 09 : Train : 2018년 10월 01일 ~ 2019년 09월 30일 ----- Test : 2019년 10월 01일 ~ 2019년 10월 31일
- SET 10 : Train : 2018년 11월 01일 ~ 2019년 10월 31일 ----- Test : 2019년 11월 01일 ~ 2019년 11월 30일
- SET 11 : Train : 2018년 12월 01일 ~ 2019년 11월 30일 ----- Test : 2019년 12월 01일 ~ 2019년 12월 31일
- SET 12 : Train : 2019년 01월 01일 ~ 2019년 12월 31일 ----- Test : 2020년 01월 01일 ~ 2020년 01월 31일
- **6.2** 모델 평가는 Test Dataset 12쌍에 대한 RMSE의 평균을 계산한다.(그리고 가장 높은 평균 1개, 가장 낮은 평균 1개 제외 -> 모델의 Generalization(일반화)를 확보할 수 있다.)
- **6.3** Feature는 **5.2 모델**과 **5.3 모델**의 경우, 종속변수만 이용. 그러나, **5.1 모델**과 **5.4 모델**의 경우 추가적인 독립변수가 필요(5.1 : 한국 콜금리, 미국 콜금리, 신용 스프레드, 비트코인 / 5.4 : 추후 자료 조사 후 필요한 변수 탐색 예정)하다.