**RNN VS CNN 주가 예측을 통한 주가데이터 성질 분석**

-RNN/CNN/LSTM-

소속 : 서강대학교 경제대학원 팀

이름 : 김원겸, 양병훈

|  |
| --- |
| **목차** |
| 1. **서론** 2. **데이터 전처리(삼성전자, 현대건설, SK하이닉스)** 3. **RNN** 4. SIMPLE RNN 모델 5. 심층 RNN 모델 6. LSTM 모델 7. **CNN**   1. SIMPLE CNN 모델  2. 심층 CNN 모델   1. **모델 테스트** 2. 현대 건설 주가 3. SK 하이닉스 4. **결론** |

**초록**

본고는 주가예측에 있어 RNN모형과 LSTM 모형 그리고 CNN 모형을 비교함으로써 각 모형의 특징을 알아보고자 한다 주가 데이터로는 삼성전자를 TRAINING 타겟이고, TEST 주가 데이터로는 SK하이닉스와 현대건설이 있다.

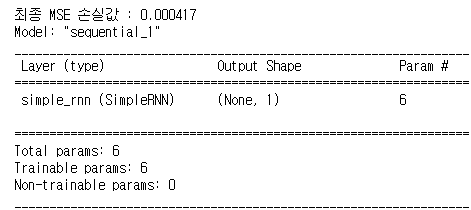
**I. 서론**

RNN모형은 입력된 기존 데이터에 새로운 입력데이트를 참고하여 새로운 데이터를 만드는 방식이다. 반면 CNN 모형은 입력된 데이터의 특징을 잡아내는 방식이다. 본고에서는 과거 입력과 새 입력의 조합으로 데이터의 특징을 잡아내는 방식과 합성곱의 연산으로 특징을 압축하는 방식 중 어느 것이 주가 예측에 적합한지를 살펴봄으로써 주가데이터의 성질을 알아보고자 한다.

**II.데이터 전처리**

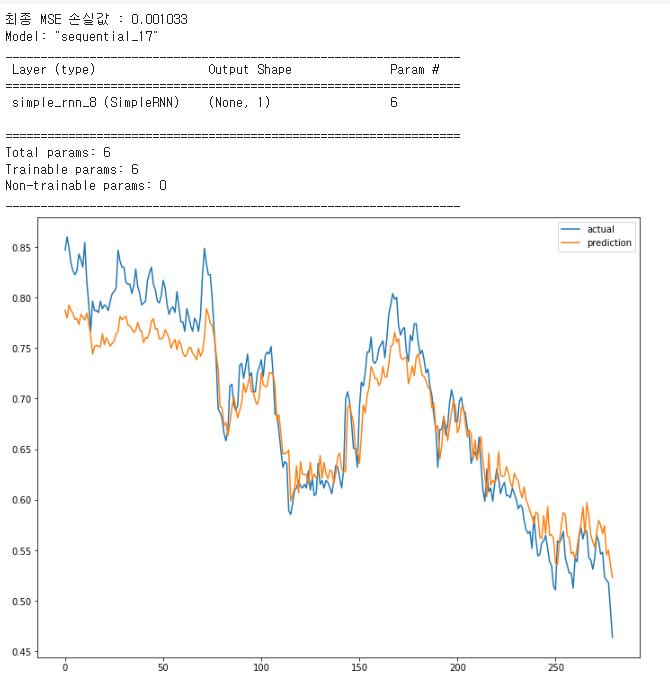
**III. RNN**

1. **SIMPLE RNN**
   1. **TEST SET 800**



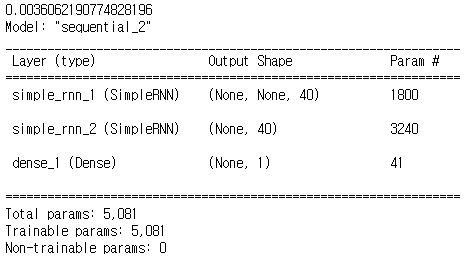


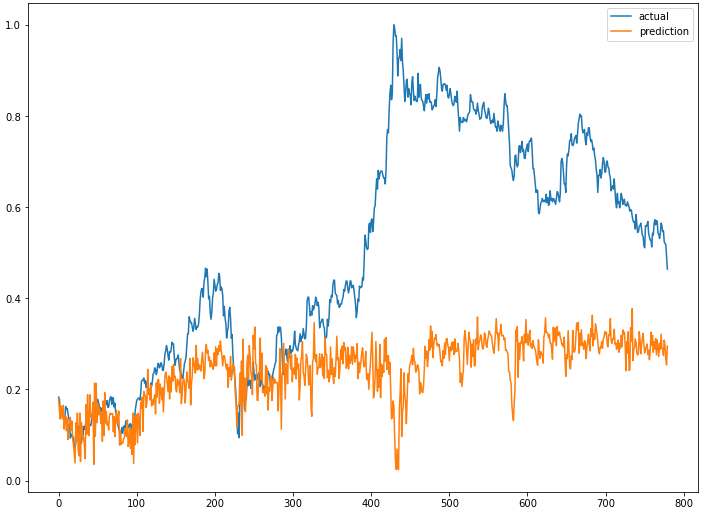
* 1. **TEST SET 300**



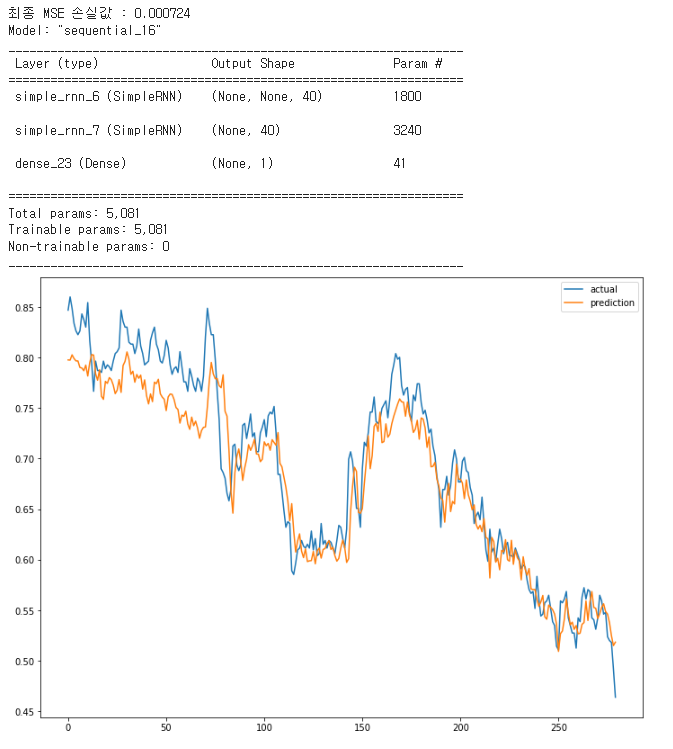
RNN의 경우 단일 LAYER에서는 Training set이 많으면 많을수록 좀 더 정확한 예측이 가능하다는 것을 살펴 볼 수 있다.

1. **심층RNN**
   1. **TEST SET 800**



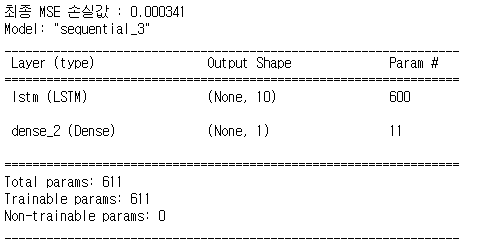


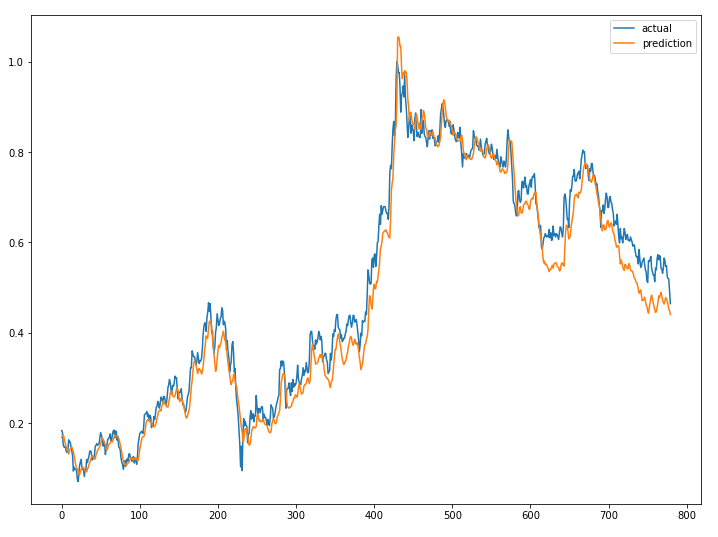
* 1. **TEST SET 300**



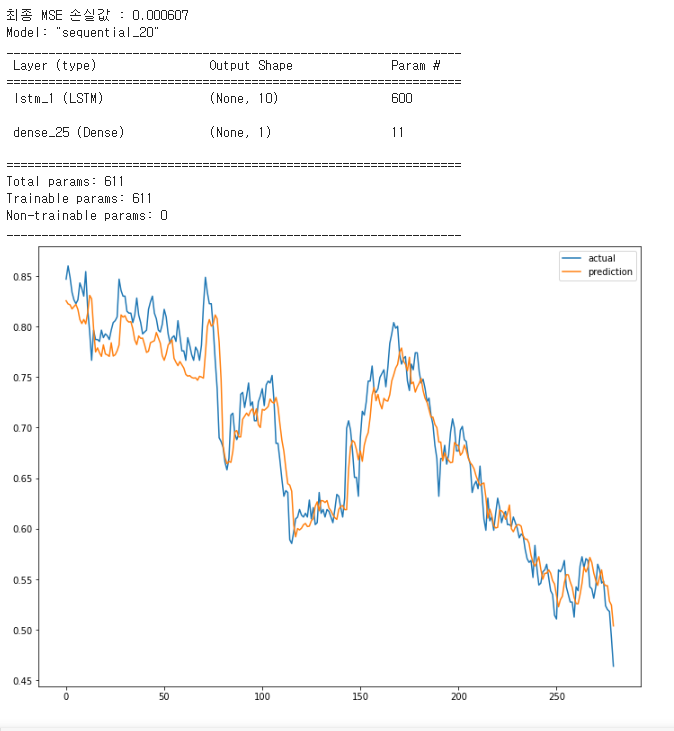
심층 RNN의 경우 LAYER의 조합에 따라 training set이 작을 때 Simple RNN에 비해 예측력이 매우 약하다는 점이 있다. 하지만 LAYER를 늘렸을 때 SIMPLE RNN에 비해 좀더 ACTUAL DAT에 가깝게 예측이 된다는 점을 알 수 있다.

1. **LSTM**
   1. **TEST SET 800**





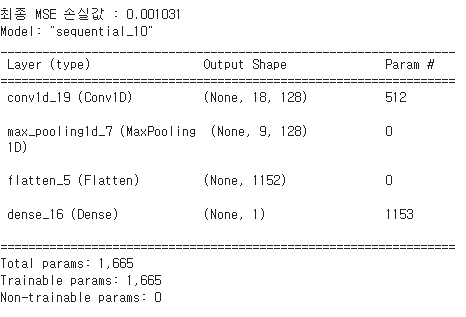
* 1. **TEST SET 300**

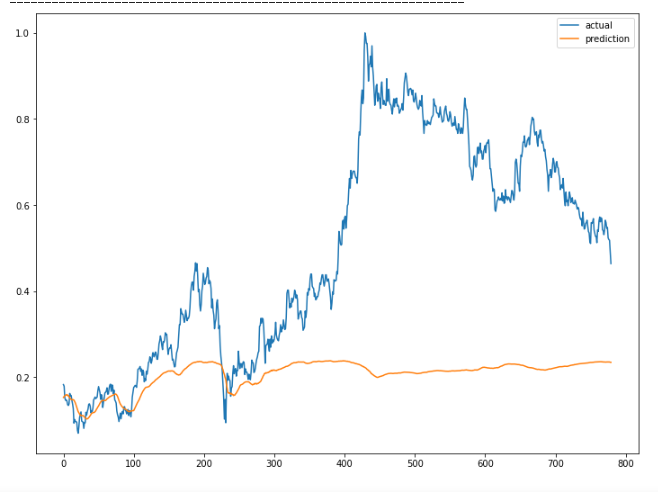


LSTM 모형은 RNN과 같은 순환학습을 하는 모형으로 차이점은 단기 메모리와 장기 메모리를 나눠 학습하여 예측하는 방식을 취한다. 마찬가지로 Training set이 적을 때도 RNN모형에 비해 주가데이터를 좀 더 정확하게 예측하고 있음을 볼 수 있다.

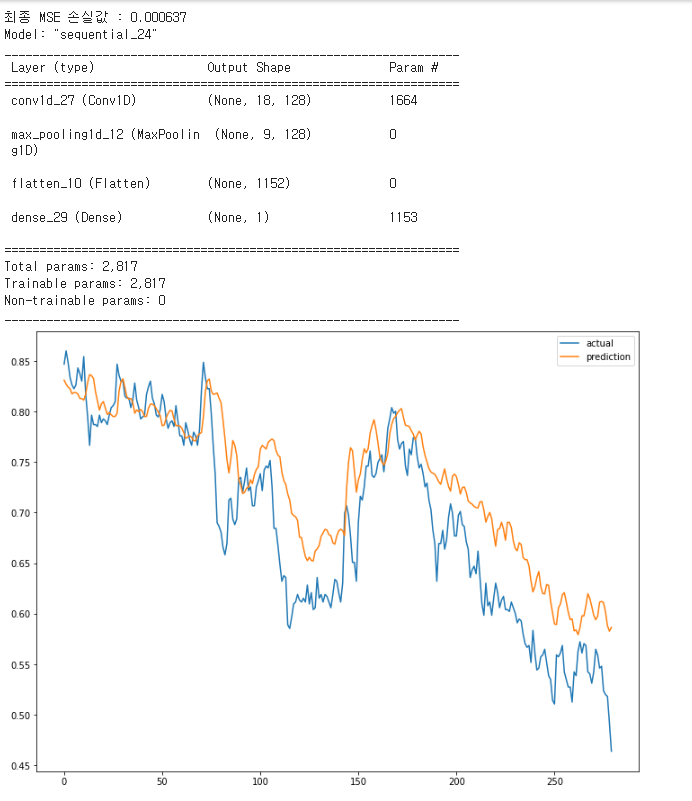
**IV. CNN**

1. **SIMPLE CNN**
   1. **TEST SET 800**



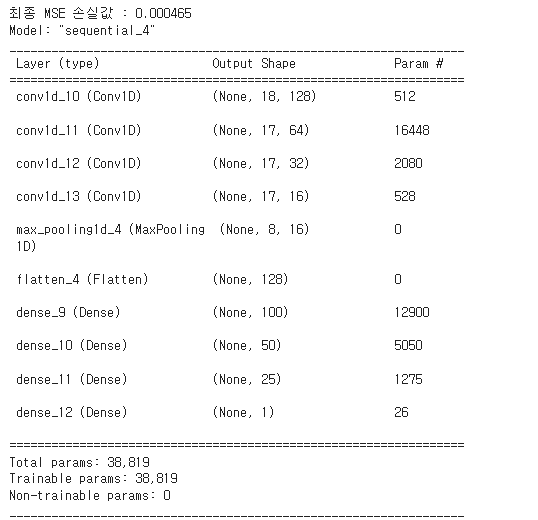


* 1. **TEST SET 300**



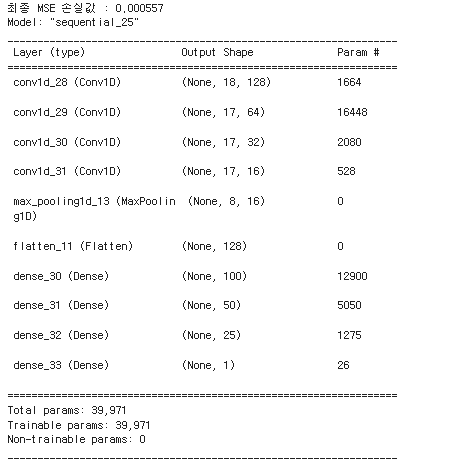
CNN 은 정해진 크기의 입력값으로 정해진 크기의 출력값을 내보낸다는 점과 특징을 압축한 다는 점(POOLING)에서 정보 손실의 발생이 크다. Filter size가 충분하지 않았던 탓인지 training set이 적을 때는 예측력이 매우 좋지 않았다. 하지만 training set이 늘어나고 부터는 성능이 매우 늘어난 것을 확인할 수 있다.

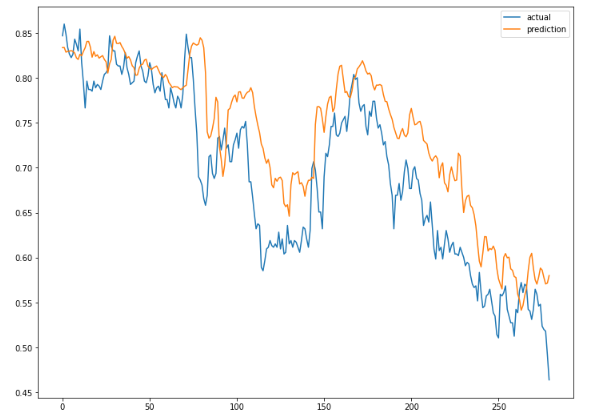
1. **심층 CNN**
   1. **TEST SET 800**





* 1. **TEST SET 300**

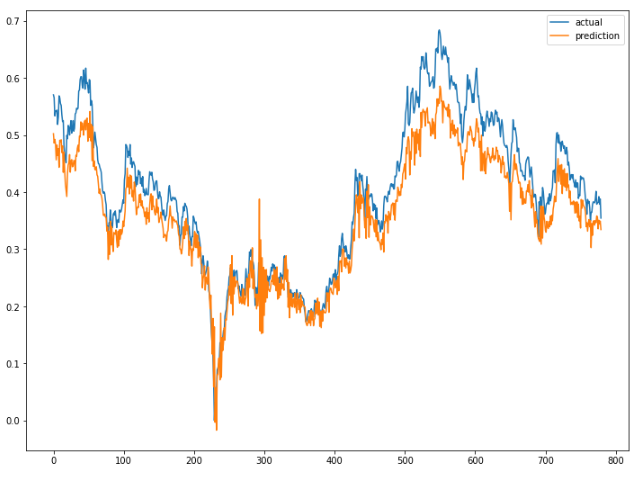




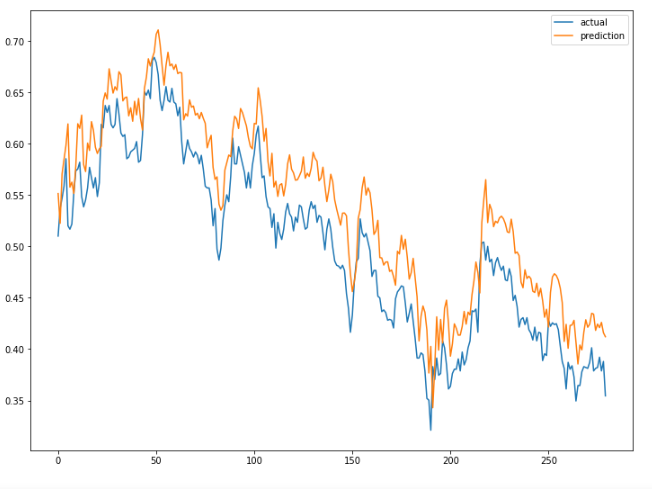
심층 CNN의 경우에도 training set이 적은 상황에서는 예측력이 매우 좋지 않았지만 늘어나자 성능이 매우 개선 되었다. RNN 모형과의 차이는 일관되게 ACTUAL DATA와의 범위가 유지 되었다는 점이다. 이 부분은 과적합 문제가 아닌지 판단된다.

**V. 모델 테스트**

1. 현대건설
   1. SIMPLE RNN
      1. TEST SET 800



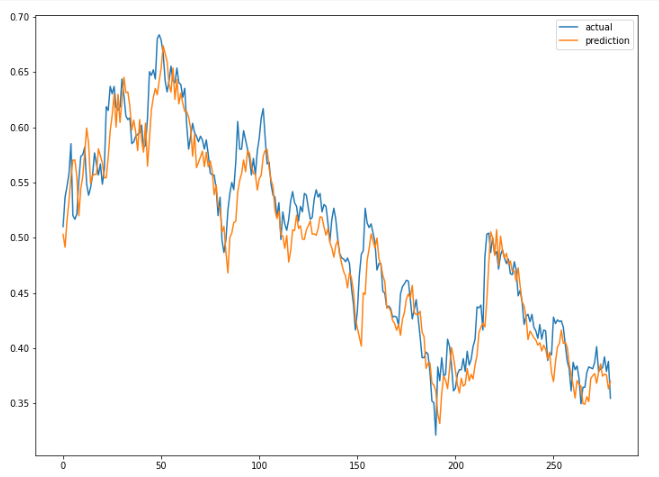
* + 1. TEST SET 300



* 1. 심층RNN
     1. TEST SET 800



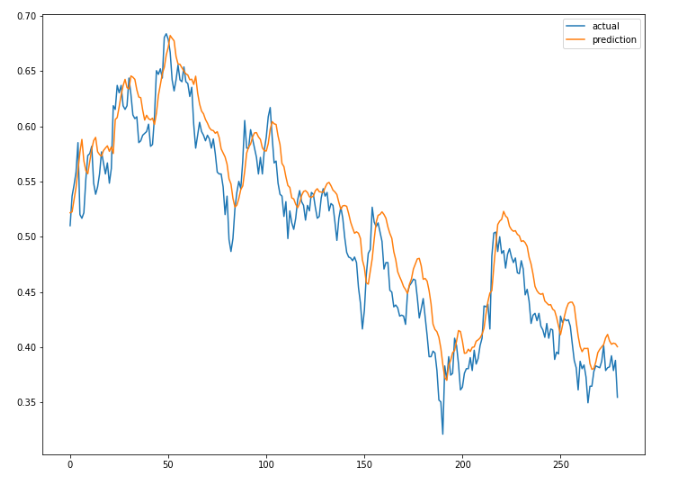
* + 1. TEST SET 300



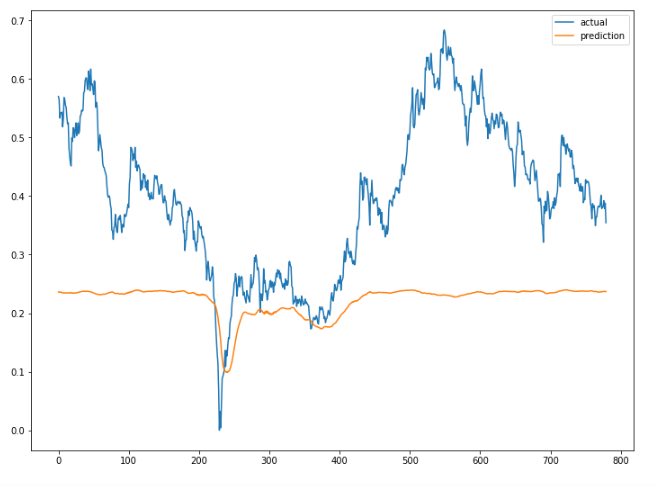
* 1. LSTM
     1. TEST SET 800



* + 1. TEST SET 300



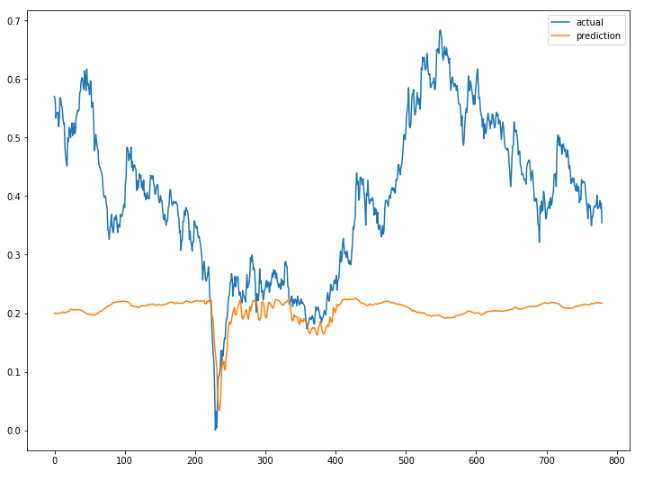
* 1. SIMPLE CNN
     1. TEST SET 800



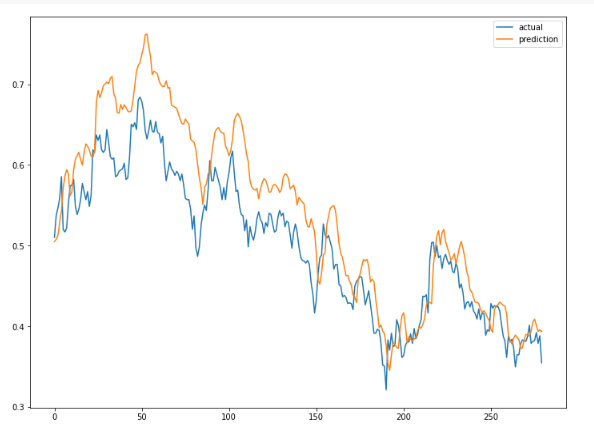
* + 1. TEST SET 300



* 1. 심층 CNN
     1. TEST SET 800

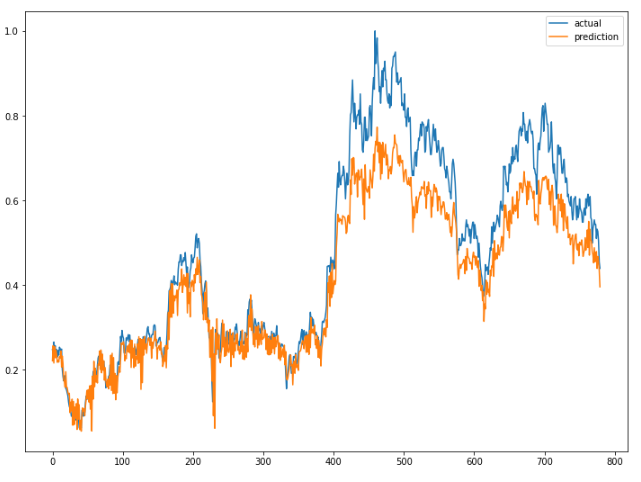


* + 1. TEST SET 300

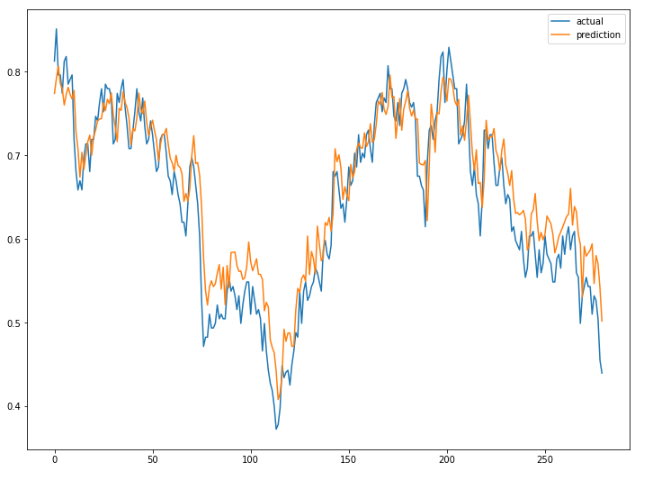


현대 건설 주가 데이터에서는 Training set의 적을수록 SIMPLE RNN의 예측력이 높았고 Training set이 많을수록 심층 RNN이 예측력이 좋았다.

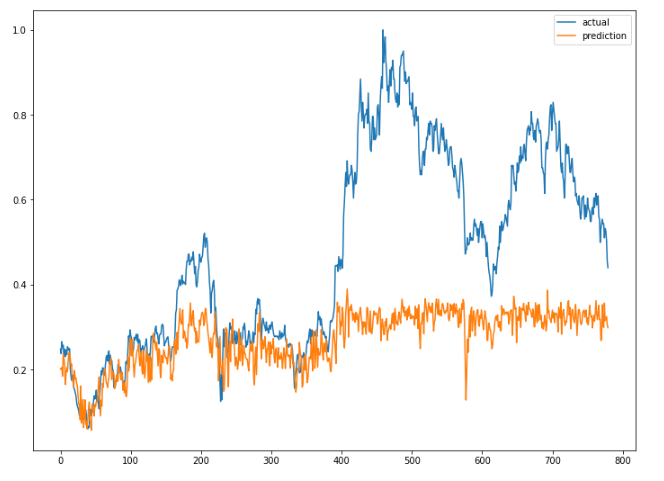
1. SK하이닉스
   1. SIMPLE RNN
      1. TEST SET 800



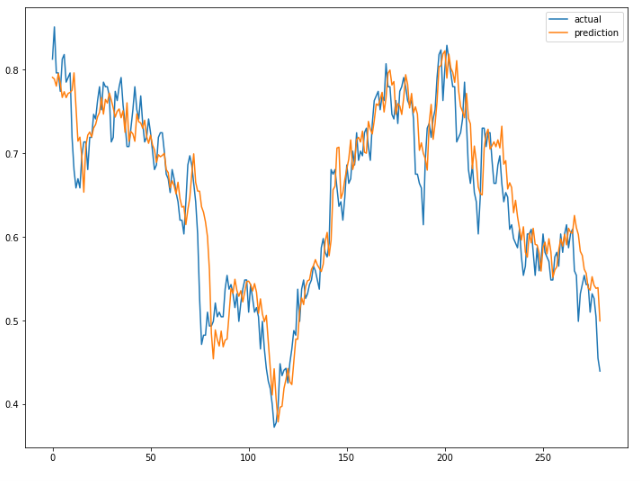
* + 1. TEST SET 300



* 1. 심층 RNN
     1. TEST SET 800



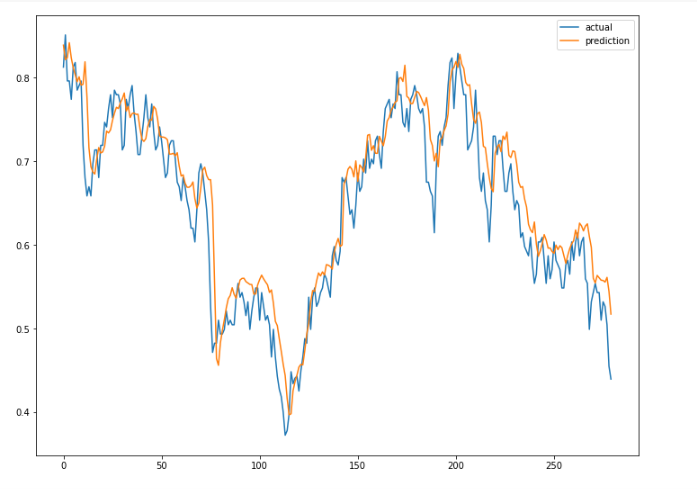
* + 1. TEST SET 300



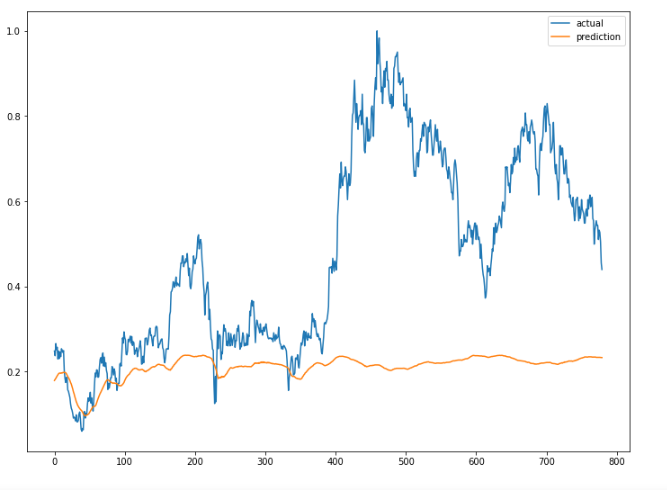
* 1. LSTM
     1. TEST SET 800



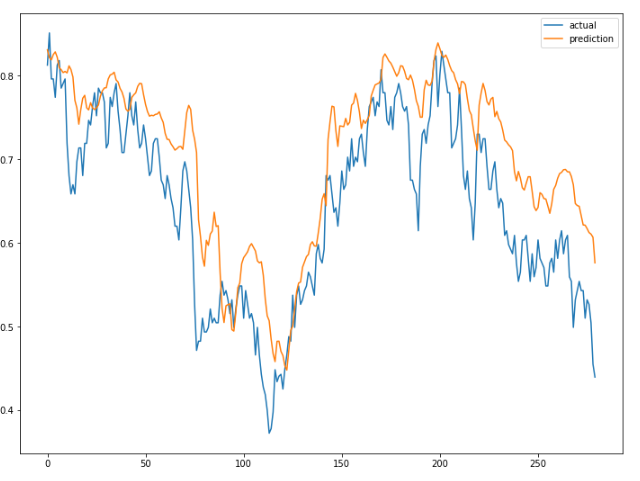
* + 1. TEST SET 300



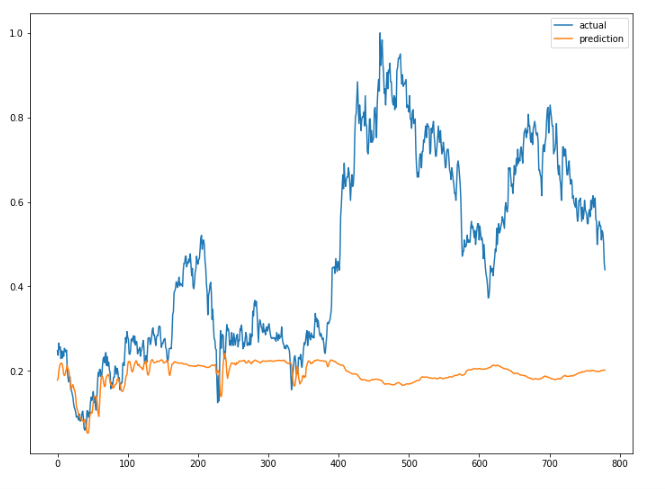
* 1. SIMPLE CNN
     1. TEST SET 800



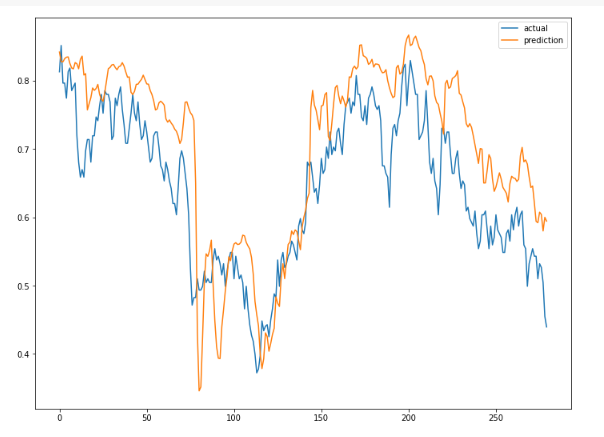
* + 1. TEST SET 300



* 1. 심층 CNN
     1. TEST SET 800



* + 1. TEST SET 300



SK 하이닉스 주가 데이터에서는 Training set의 적을수록 LSTM의 예측력이 높았고 Training set이 많을수록 SIMPLE RNN이 예측력이 좋았다.

1. **결론**

평균적으로 LSTM의 예측이 가장 ACTUAL에 근접했고, CNN보다는 RNN의 예측력이 뛰어났다.