

# Data Scientist Portfolio

---

이정해(LEE JEONGHAE)

2021

# 이정해 Jeonghae Lee

- Contact

[Jeonghae.Lee@cass.city.ac.uk](mailto:Jeonghae.Lee@cass.city.ac.uk)

- Github

<https://github.com/jhlsmart92>

- LinkedIn

<https://www.linkedin.com/in/jeonghaelee/>

- Education

한동대학교 경영경제학부

울산과학기술원 융합경영대학원 에너지 상품거래 및 금융공학

Business School of City, University of London, Energy Trade and Finance

국가평생교육원 컴퓨터공학과 (재학)

- IT skills

Python, Machine Learning, Deep Learning, Java Script

- Work

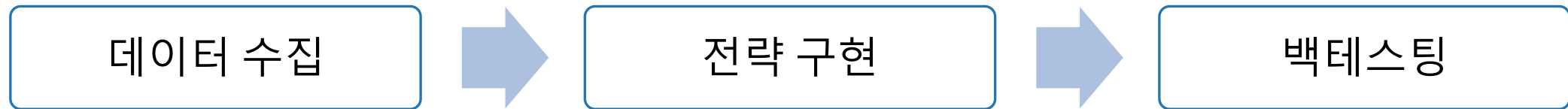
에너지경제연구원 에너지수요관리 연구팀 연구보조원

통계청 제6차 OECD 세계포럼 준비기획단 사무보조원

# Experience related with Data Science

Skills	Project Name	Institution
Python, Crawling	울산과학기술원 졸업 프로젝트: 파이썬을 이용한 알고리즘 트레이딩	UNIST ECTFE
Python Visualization, API	전기차 충전소와 아파트 가격간의 관계 분석	아시아 경제
Machine Learning, Deep Learning, Crawling	국내 주가 정보 수집 및 인공지능을 이용한 예측	아시아 경제
Deep Learning, Crawling, Natural Language Processing	뉴스 데이터를 이용한 주가 정보 예측	아시아 경제
Deep Learning, Crawling, Natural Language Processing	자연어 처리를 이용한 클라우드 펀딩 성공률 예측	아시아 경제
Statistical Analysis, Time Series Analysis	런던 시립대 경영대학원 졸업 프로젝트: The Relationship Between WTI Price, Won-Dollar Exchange Rate and Korean Domestic Gasoline Price	Business School of City, University of London
Regression Model	NASDAQ IPO Price Analysis Using Regression Analysis	Business School of City, University of London

# PYTHON 을 이용한 알고리즘 트레이딩 구현



- 데이터: 네이버 종목데이터, CompanyGuide 재무데이터
- 전략: 증권 및 재무 데이터 수집 후 PER, PBR, PCR, PSR, ROE, 모멘텀, F-SCORE 등을 구현
- 백테스팅: 2년간 투자를 가정했으며, 1년 단위로 리밸런싱
- 결론: 저 PER 단독 전략이 가장 우수한 성과.  
전략을 혼합하는 것 보다 하나의 지표를 이용한 전략이 더 좋은 성과를 보임.  
시장이 급변하는 상황(판데믹 등)에서는 유효하지 않을 수 있음

울산과학기술원 융합경영대학원 졸업 프로젝트  
2020년 12월

PYTHON 을 이용한  
알고리즘 트레이딩 구현

UNIST ICT&CS CAPSTONE PROJECT

김민우

2020 DECEMBER 17  
UNIST ICT&CS  
2019-2020 1학기

# PYTHON 을 이용한 알고리즘 트레이딩 구현

데이터 수집 후 데이터프레임화

	PBR	PCR	PER	PSR	총현금흐름	per 순위	ROA	ROIC	부채비율	영업이익 률	유동비율	roa 순위	마법공식 순위
Unnamed: 0													
A002620	0.75	0.6	0.61	6.64	78	1	141.5	4.2	18.1	9.1	250.7	3	1
A088390	0.76	0.61	0.64	2.47	-47	2	58.4	-9.9	41.7	-33.3	175.2	7	2
A078070	0.61	1.19	1.21	15.06	28	8	47	3.6	3.4	20.2	2614.9	12	3.5
A192400	1.67	2.46	2.69	2.67	884	14	75.1	29.3	36.8	14.4	202.4	6	3.5
A001800	0.9	0.72	0.76	9.98	919	3	41.2	0.2	38	4.4	118.7	18	5
A148780	1.52	2.51	2.66	0.5	37	13	45.8	184.3	61.2	18.6	226.4	14	6
A102280	0.47	0.97	0.99	1.3	-126	5	33.9	-21.7	106.7	-19.7	203.5	27	7
A069920	1.25	2.67	2.8	2.69	6	15	36.3	-8.7	28.3	-3.8	291.1	21	8
A090470	2.02	3.43	3.52	0.46	788	23.5	45	191.5	41	17.6	195.9	15.5	9
A100090	0.55	0.82	0.92	0.72	-133	4	25	-7.9	112.7	-17	38	43.5	10

Table5 : 저PER 고ROA

	2017/12	Dec-18	Dec-19	Jun-20
IFRS(연결)				
매출액	963761	968126	1057464	471784
영업이익	45747	24222	36055	14541
당기순이익	45464	16450	31856	9300
자산	1781995	1806558	1945122	2018045
부채	1034421	1067597	1181465	1257564
자본	747574	738960	763658	760481
영업활동으로인한현금흐름	39224	37643	4198	7167

Table 1: 재무제표 데이터 데이터프레임(종목코드 A005930)

	Dec-16	Dec-17	Dec-18	Dec-19	Jun-20
유동비율	258.5	218.8	252.9	284.4	300
부채비율	35.9	40.7	37	34.1	32
영업이익률	14.5	22.4	24.2	12.1	13
ROA	9	15	13.8	6.3	5.9
ROIC	20.3	33.2	30.4	13.2	13.5

Table 2: 재무비율 데이터프레임(종목코드 A005930)

	Dec-16	Dec-17	Dec-18	Dec-19	Jun-20
PER	13.18	9.4	6.42	17.63	NaN
PCR	6.85	6.12	4.01	7.42	NaN
PSR	1.46	1.62	1.16	1.65	NaN
PBR	1.48	1.76	1.1	1.49	1.37
총현금흐름	423491	623891	691786	494337	253122

Table 3: 투자지표 데이터프레임(종목코드 A005930)

# PYTHON 을 이용한 알고리즘 트레이딩 구현

전략별 백테스팅 및 리밸런싱

	주식포트폴리오	현금포트폴리오	종합포트폴리오	일변화율	총변화율
Unnamed: 0					
2018-06-01	99719612	280388	100000000	NaN	0
2018-06-04	99973212	280388	100253600	0.002536	0.002536
2018-06-05	100629981	280388	100910369	0.006551	0.009104
2018-06-07	102787021	280388	103067409	0.021376	0.030674
2018-06-08	102084548	280388	102364936	-0.006816	0.023649
...	...	...	...	...	...
2020-05-25	92841124	280388	93121512	0.007601	-0.068785
2020-05-26	92269026	280388	92549414	-0.006144	-0.074506
2020-05-27	93239976	280388	93520364	0.010491	-0.064796
2020-05-28	90676557	280388	90956945	-0.02741	-0.090431
2020-05-29	91889262	280388	92169650	0.013333	-0.078303

Table9: 저PER기준 백테스팅 결과



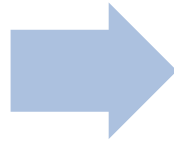
Chart2: 리밸런싱 시행 여부에 따른 백테스팅 결과 비교



Chart3: 각각의 전략에 따른 2년간의 백테스팅 결과

# 전기차 충전소와 아파트 가격 간의 관계분석

데이터 수집



시각화

- 데이터: 공공데이터포털 전기자동차 충전소 정보 API, 국토교통부의 아파트 공시지가 데이터
- 시각화: matplotlib, folium
- 결론: 전기차 충전소 분포와 아파트 공시지가 사이에 양의 관계가 있음을 알 수 있음  
그러나 이러한 관계는 공시지가가 높은 지역일 수록 강하게 나타남  
이는 인과관계가 아닌 단순 상관관계

# 전기차 충전소와 아파트 가격 간의 관계분석

## 주제 선정 및 데이터 수집(Sql)

### 전기차 충전소 많으면 집값이 비싸다...이유는?

머니투데이 | 유희석 기자

VIEW 7,140 | 2019.04.24 06:33

美 충전소 밀집 지역 집값, 주변 평균의 2.6배 ↑  
부유층 전기차 선호...충전소 접근성 프리미엄도



미국 캘리포니아주 투아레메에 있는 전기차용 충전소 모습. /AFPBBNews=뉴스1

미국에서 전기차용 충전소가 많은 지역 집값이 그렇지 않은 곳보다 훨씬 비싼 것으로 나타났다. 미국의 온라인 부동산 중개업체 리얼터닷컴이 미국 내 전기차 충전소 1만9743곳의 위치 정보를 분석한 결과, 충전소가 많은 상위 20개 지역 평균 집값은 78만2000달러로 주변 지역보다 1.5배 높았다. 미국 전체 평균 주택가격 대비로는 2.6배나 비싼 수준이다.

Google 광고

이 광고 그만 보기

이 광고가 표시된 이유

파일(F) 편집(E) 보기(V) 도구(T) 도움말(H)

새 데이터베이스(N) 프로젝트 열기(P) 데이터베이스 연결(A)

데이터베이스 구조 데이터 보기 Pragma 수정 데이터베이스 셀 수정하기(C)

테이블(T): test\_2

	index	구	충전소 개수	평당 가격
필터	필터	필터	필터	필터
1	1	종로구	12	1282
2	2	중구	32	1279
3	3	용산구	12	1618
4	4	성동구	25	1687
5	5	광진구	13	1550
6	6	동대문구	41	1108
7	7	중랑구	12	847
8	8	성북구	99	1036
9	9	강북구	8	864
10	10	도봉구	28	778
11	11	노원구	31	971
12	12	은평구	30	905
13	13	서대문구	16	1197
14	14	마포구	53	1474
15	15	양천구	164	1348
16	16	강서구	66	1024
17	17	구로구	10	908
18	18	금천구	16	804
19	19	영등포구	35	1357
20	20	동작구	5	1408
21	21	관악구	19	982

1 - 20 of 25 특정 레코드 행으로 가기: 1

모드: 문자열

현재 데이터 타입: 문자열 / 숫자 1 자 적용

원격(R)

신원 연결할 ID를 선택하세요

DBHub.io 로컬 현재 데이터베이스

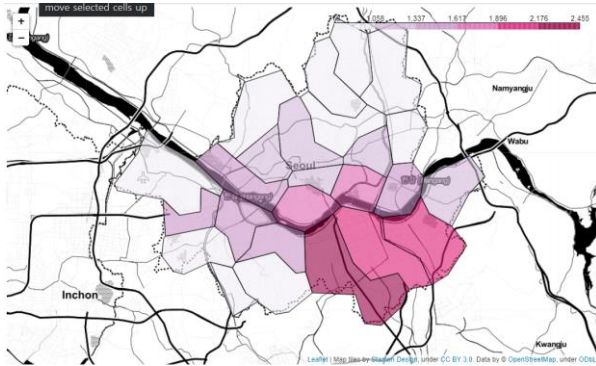
이름 마지막

SQL 로... 플... DB 스키... 원...

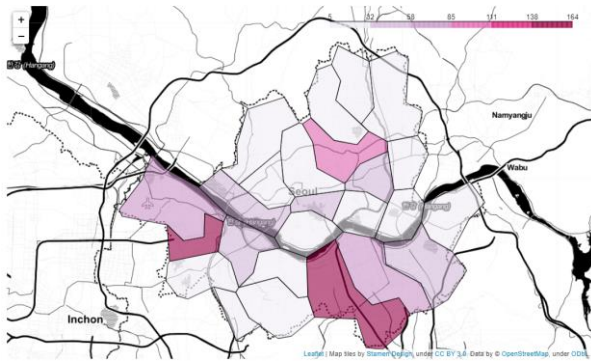
UTF-8



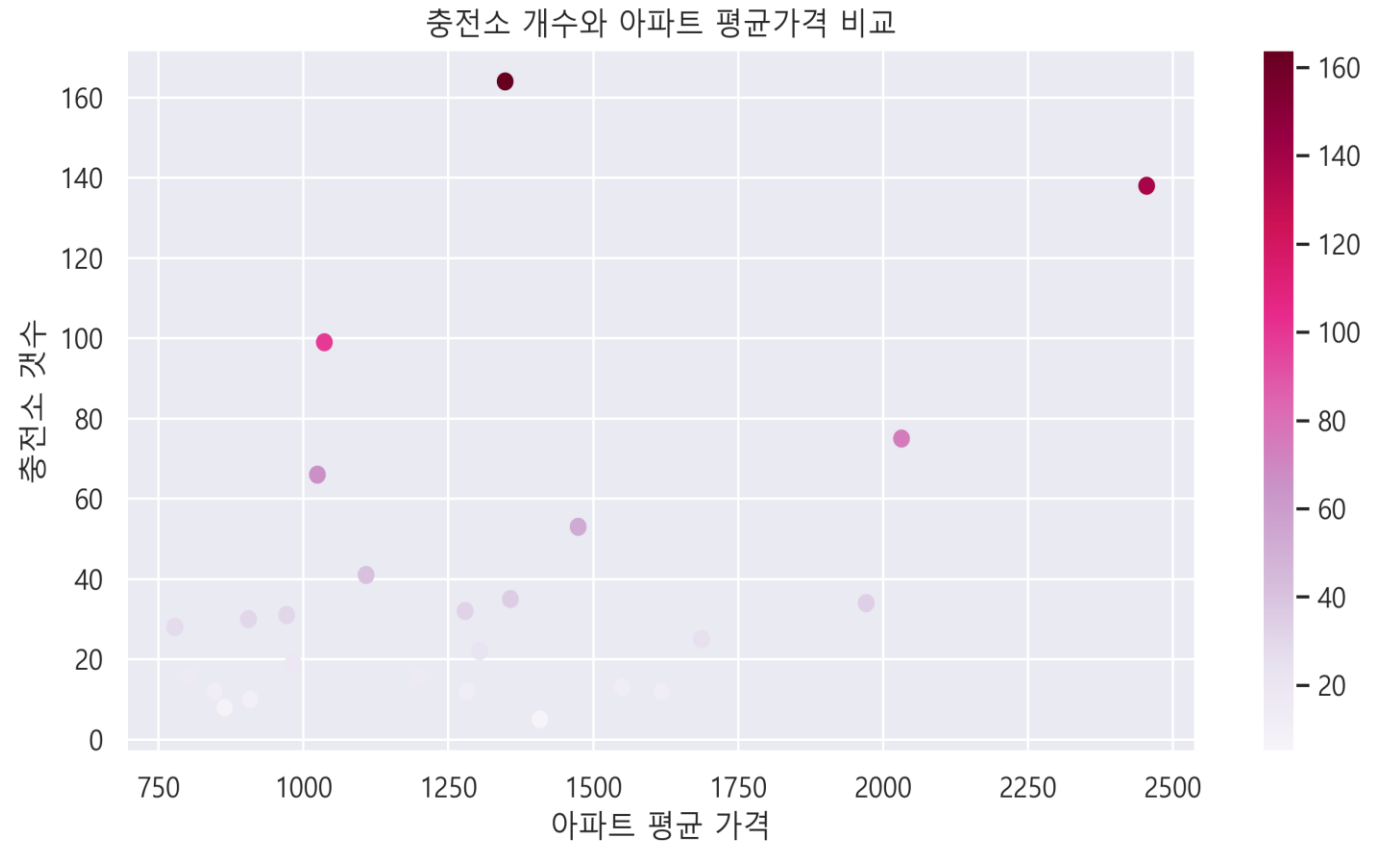
# 전기차 충전소와 아파트 가격 간의 관계분석



서울시 구별 전기차 충전소 분포



서울시 아파트 가격 현황



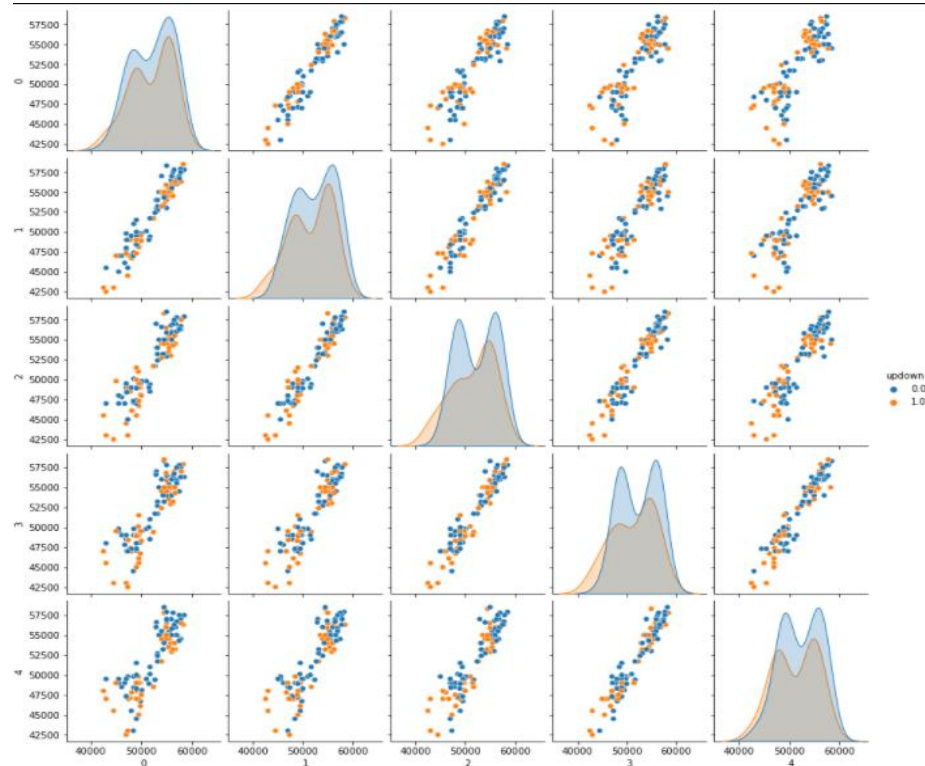
# 국내 주가 정보 수집 및 인공지능을 이용한 예측



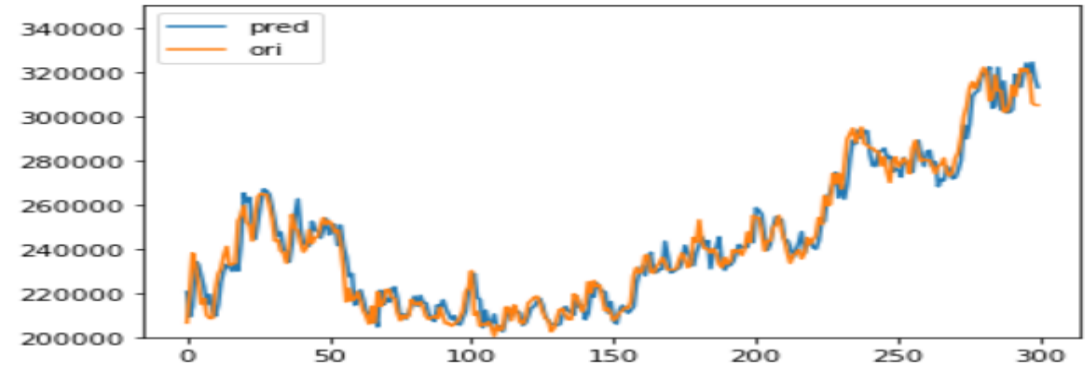
- 데이터: 네이버 주식(시가, 종가, 저가, 고가, 거래량)
- 내용: 머신러닝과 딥러닝을 이용하여 이전 3일 동안의 데이터를 학습, 다음날의 종가를 예측  
특정 주식을 1주 선택하여 이평선 전략을 수행
- 결론: 딥러닝과 머신러닝 모두 단기적인 예측을 성공적으로 수행

# 국내 주가 정보 수집 및 인공지능을 이용한 예측

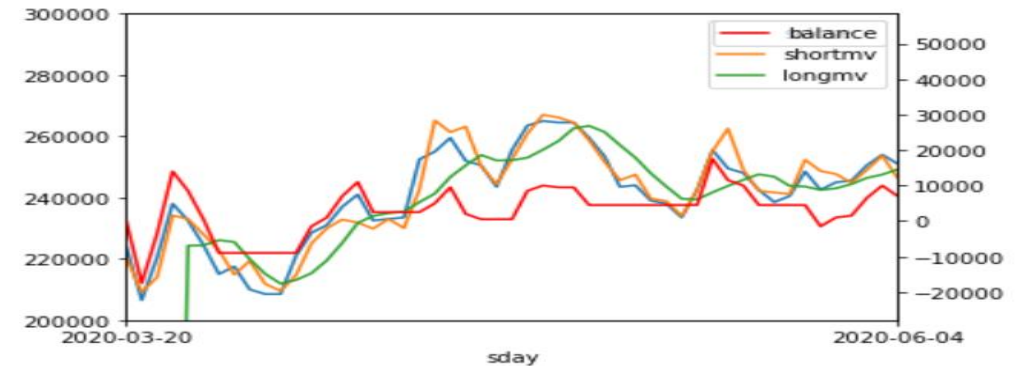
## 머신러닝 KNN 활용 국내주가 예측



상승일 하락일 분포



예측 주가와 실제주가 비교



이평선 전략 수행 시 잔고상황

# 국내 주가 정보 수집 및 인공지능을 이용한 예측

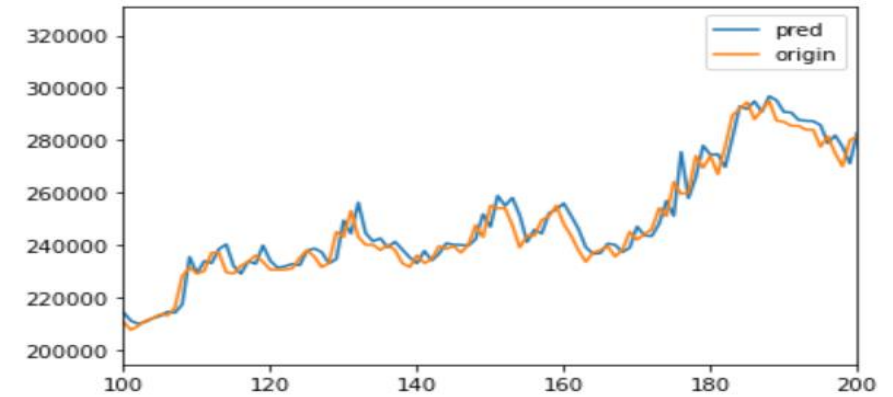
## 딥러닝(LSTM) 활용 국내주가 예측

```
Model: "sequential_4"
Layer (type)                Output Shape              Param #
-----
lstm_2 (LSTM)                (None, 64)                17920
dense_1 (Dense)              (None, 1)                 65
-----
Total params: 17,985
Trainable params: 17,985
Non-trainable params: 0

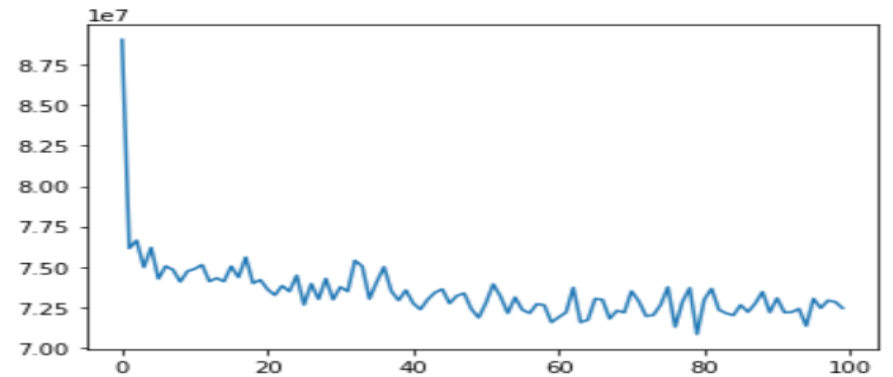
model.compile(loss="mse",optimizer="adam",metrics=['CosineSimilarity'])
hist=model.fit(X_train,Y_train,epochs=100,batch_size=10)

Epoch 1/100
599/599 [=====] - 2s 1ms/step - loss: 79623146.1600 - cosine_similarity: 1.0000
Epoch 2/100
599/599 [=====] - 1s 1ms/step - loss: 83707372.8067 - cosine_similarity: 1.0000
Epoch 3/100
599/599 [=====] - 1s 1ms/step - loss: 78010020.4200 - cosine_similarity: 1.0000
Epoch 4/100
599/599 [=====] - 1s 1ms/step - loss: 77443303.4667 - cosine_similarity: 1.0000
Epoch 5/100
599/599 [=====] - 1s 1ms/step - loss: 61050356.8767 - cosine_similarity: 1.0000
Epoch 6/100
599/599 [=====] - 1s 1ms/step - loss: 77043572.9800 - cosine_similarity: 1.0000
Epoch 7/100
599/599 [=====] - 1s 1ms/step - loss: 85512091.3167 - cosine_similarity: 1.0000
Epoch 8/100
599/599 [=====] - 1s 1ms/step - loss: 82543147.2733 - cosine_similarity: 1.0000
Epoch 9/100
599/599 [=====] - 1s 1ms/step - loss: 60227744.2433 - cosine similarity: 1.0000
```

LSTM 모델 형태

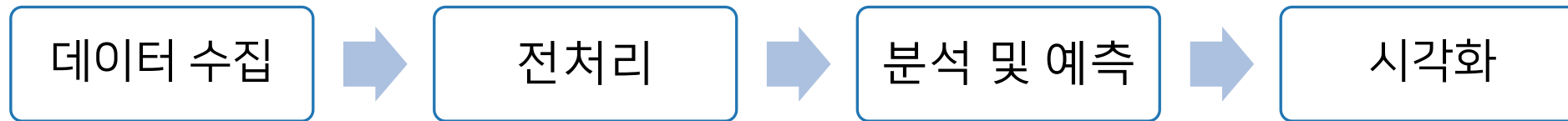


예측 주가와 실제주가 비교



Loss 감소 형태

# 뉴스데이터를 이용한 주가 예측



- 데이터: 네이버 주식(시가, 종가, 저가, 고가, 거래량) 및 특정 종목의 뉴스데이터 헤드라인
- 내용: 기존의 주가정보 예측을 자연어 처리를 통해 보강  
해당 종목의 뉴스 헤드라인 단어를 수집 후 형태소 기반으로 Bag of Words 구축  
특정 기간동안 빈번히 언급된 단어들을 변수로 활용(1000개)
- 결론: 자연어 변수 추가 후에도 단기적인 예측을 성공적으로 수행  
기존 주가정보만을 활용한 모델에 비해 Loss 가 급격히 감소하는 추세를 보임

# 뉴스데이터를 이용한 주가 예측

빈도수 기반 상위 1000개 단어 변수에 추가

[219]:

	idx	sopen	shigh	slow	sclose	svol	광 주	영 화	이 용	카 카 오 독	...	정 치	무 대	논 란	스 캔	이 외	또 는	단 말	아 동	슈 퍼	구 성
0	2017-01-01	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	2017-01-02	175500.0	176000.0	171500.0	172000.0	42861.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	4.0	0.0	0.0	0.0	0.0	0.0	2.0
2	2017-01-03	172000.0	173500.0	170000.0	173000.0	34550.0	4.0	0.0	2.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	3.0	0.0	0.0	2.0
3	2017-01-04	172500.0	172500.0	171000.0	171500.0	42407.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	44.0	0.0	0.0	0.0	0.0	4.0
4	2017-01-05	173000.0	173500.0	171000.0	173000.0	25519.0	3.0	0.0	2.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

5 rows x 1006 columns

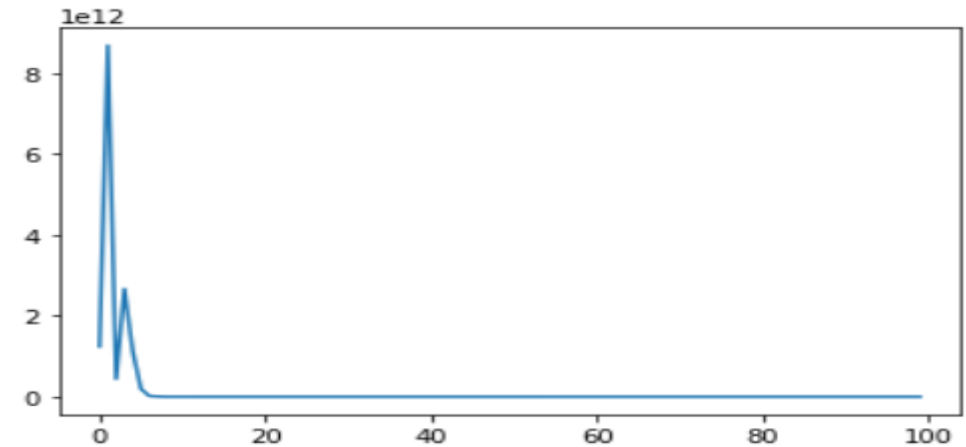
주가 변수와 자연어 변수로 이루어진 데이터프레임

Model: "sequential"

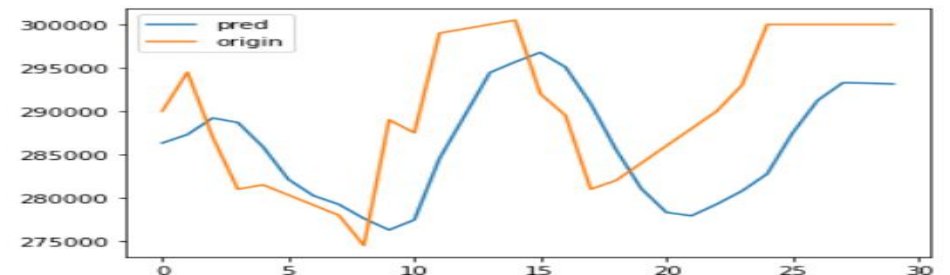
Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 64)	273920
dense (Dense)	(None, 1)	65

Total params: 273,985  
Trainable params: 273,985  
Non-trainable params: 0

모델형태

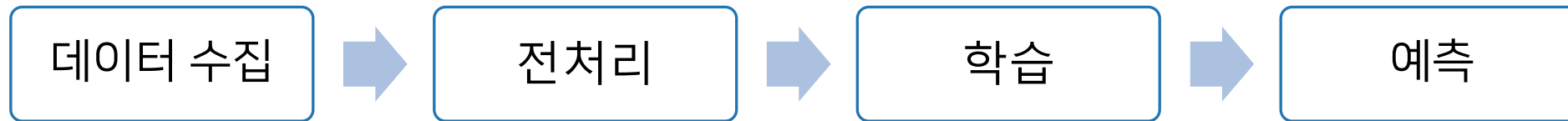


Loss 감소 형태



예측 주가와 실제주가 비교

# 자연어 처리를 이용한 크라우드 펀딩 성공률 예측



- 데이터: 국내 크라우드 펀딩 사이트(Wadiz)의 각 카테고리 크롤링
- 내용: 각 제품들의 제목과 제품설명을 취합하여 피쳐로 활용, 앵콜 여부를 레이블로 활용  
형태소 분석기를 통해 Bag of Words 구성  
카테고리별 피쳐와 레이블을 취합하여, 상위 빈도의 단어로 LSTM 모델 학습
- 결론: 83% 정확도로, 크라우드 펀딩의 제품 소개 콘텐츠에는 노이즈가 많아 정확한 예측에는 난항  
자연어 변수와 더불어, 댓글 수, 서포터 수 등 다양한 변수 필요

# 자연어 처리를 이용한 크라우드 펀딩 성공률 예측

## 데이터 수집 및 전처리

새 데이터베이스(N) 데이터베이스 열기(O) 변경사항 저장하기(W) 변경사항 취소하기(R) 프로젝트 열기(P) 프로젝트 저장하기(V) 데이터베이스 연결(A) 데이터베이스 닫기(C)

데이터베이스 구조 데이터 보기 Pragma 수정 SQL 실행

데이터베이스(T): wadiztable 모든 열에서 필터링

	idx	ncategory	ntitle	ncontent	nsupport	namount	npercent	
	필터	필터	필터	필터	필터	필터	필터	필터
1	186 테크-가전		[세균/악취/곰팡이까지 한방에~	...	217	19,013,000	3802	https://www
2	187 테크-가전		[7.9억 영글] 업그레이드로 돌...	...	185	10,236,200	341	https://www
3	188 테크-가전		[2.4억 영글] 유인+흡입+감전!	...	140	6,223,500	622	https://www
4	189 테크-가전		[압도적평점 4.9점!] 후기로 ...	...	112	9,113,700	911	https://www
5	190 테크-가전		[1195% 영글 편당] 우리 집의...	...	42	3,486,300	348	https://www
6	191 테크-가전		[4차 영글][누적 1억 돌파 예...	...	14	4,854,000	970	https://www
7	192 테크-가전		[제습만을 위한 진짜 제.습.기...	...	12	8,789,000	878	https://www
8	193 테크-가전		[올 여름 캠핑과 차박은 ANKE...	...	552	173,463,000	5782	https://www
9	194 테크-가전		[티셔츠 속에 COOL한 시원함...	...	1,767	72,708,710	7270	https://www
10	195 테크-가전		[글로벌 4억 영글] 빠진 내 ...	...	439	66,517,001	6651	https://www
11	196 테크-가전		<불 없이 조리한다? > 한 뼉 ...	...	475	62,018,000	12403	https://www
12	197 테크-가전		[3분 OK]타치 한번에 커피와 ...	...	859	161,802,000	16180	https://www
13	198 테크-가전		[편당 요청해도] 반바지로 돌...	...	911	174,451,060	34890	https://www
14	199 테크-가전		[모니터발청대 끝판왕] 블루...	...	68	5,923,000	118	https://www
15	200 테크-가전		[교체/유지 비용 없이] 평생 ...	...	287	36,068,000	3606	https://www
16	201 테크-가전		책상 위 스마트기기를 한번에 ...	...	431	21,259,700	1062	https://www
17	202 테크-가전		[악청-피치 썩! 카메라 피치식...	...	529	29,143,000	2914	https://www
18	203 테크-가전		아이패드와 연신- 태블릿과 랩...	...	341	35,751,000	3575	https://www
19	204 테크-가전		[14770% 영글] 손가락 하나...	...	270	16,408,800	3281	https://www
20	205 테크-가전		의류 속 찌든 얼룩 타파! 초음...	...	259	15,115,000	3023	https://www
21	206 테크-가전		[젠더어2세대] 100W 고효율 ...	...	144	19,863,000	1986	https://www

데이터베이스 셀 수정하기(C)

모드: 문자열

1 [영글] 번기청소 단 한번 설치면 평생 끝! 세균/악취/곰팡이까지 한방에~

현재 데이터 타입: 문자열 / 숫자  
40 자

적용

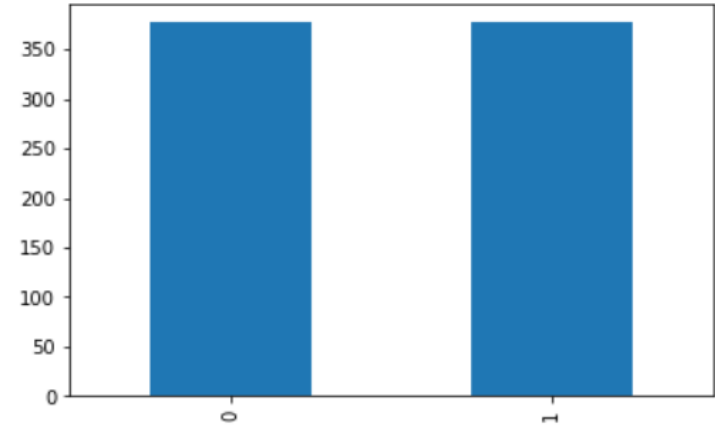
원격(R)

신원 연결할 ID를 선택하세요

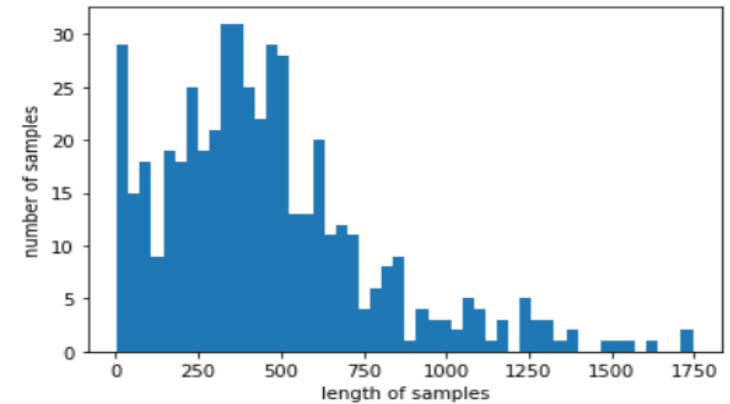
DBHub.io 로컬 현재 데이터베이스

이름 마지막 수정 크기

각 회사별 제품 카테고리, 제목, 상품소개를 크롤링하여 DB 구축



0: 실패, 1: 성공 (SMOTE 적용)



컨텐츠 Padding

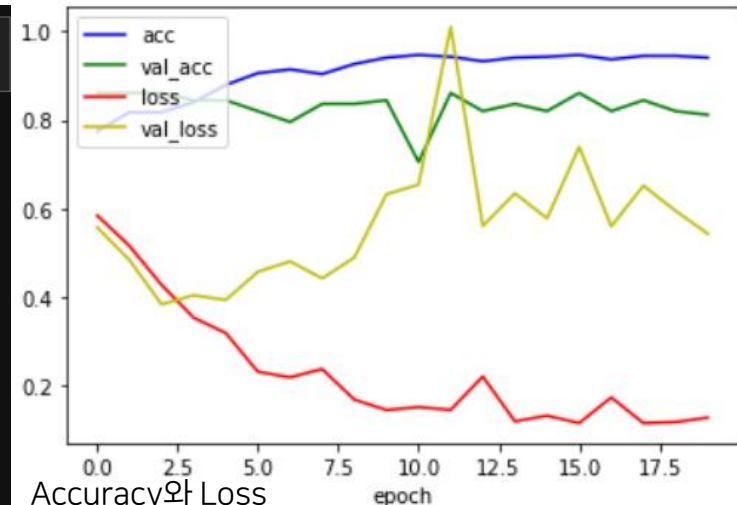


# 자연어 처리를 이용한 크라우드 펀딩 성공률 예측

## 모델 학습

```
model.compile(optimizer='rmsprop', loss='binary_crossentropy', metrics=['acc'])
history = model.fit(x_train, y_train, epochs=20, callbacks=None, batch_size=60, validation_split = 0.2)
```

Epoch 1/20  
9/9 [=====] - 31s 3s/step - loss: 0.5838 - acc: 0.7741 - val\_loss: 0.5567 - val\_acc: 0.8607  
Epoch 2/20  
9/9 [=====] - 28s 3s/step - loss: 0.5154 - acc: 0.8172 - val\_loss: 0.4835 - val\_acc: 0.8607  
Epoch 3/20  
9/9 [=====] - 25s 3s/step - loss: 0.4271 - acc: 0.8172 - val\_loss: 0.3831 - val\_acc: 0.8607  
Epoch 4/20  
9/9 [=====] - 26s 3s/step - loss: 0.3531 - acc: 0.8398 - val\_loss: 0.4040 - val\_acc: 0.8443  
Epoch 5/20  
9/9 [=====] - 27s 3s/step - loss: 0.3185 - acc: 0.8789 - val\_loss: 0.3932 - val\_acc: 0.8443  
Epoch 6/20  
9/9 [=====] - 26s 3s/step - loss: 0.2310 - acc: 0.9055 - val\_loss: 0.4570 - val\_acc: 0.8197  
Epoch 7/20  
9/9 [=====] - 25s 3s/step - loss: 0.2185 - acc: 0.9138 - val\_loss: 0.4803 - val\_acc: 0.7951  
Epoch 8/20  
9/9 [=====] - 26s 3s/step - loss: 0.2372 - acc: 0.9035 - val\_loss: 0.4419 - val\_acc: 0.8361



Accuracy와 Loss  
Loss 가 하락하는 반면, Validation Loss는  
상승하므로 과적합임을 알 수 있음

## LSTM 모델 학습

# 자연어 처리를 이용한 크라우드 펀딩 성공률 예측

## 실제 펀딩 제품 성공률 예측

```
def sentiment_predict(new_sentence):  
    new_sentence = tknWord(new_sentence) # 토큰화  
    new_sentence = [word for word in new_sentence if not word in stopwords] # 불용어 제거  
    # print(new_sentence)  
    encoded = tokenizer.texts_to_sequences([new_sentence]) # 정수 인코딩  
    pad_new = pad_sequences(encoded, maxlen = max_len) # 패딩  
    score = float(loader.predict(pad_new)) # 예측  
    print(score)  
    if(score > 0.5):  
        print('성공')  
    else:  
        print('실패')
```

```
# 앵콜펀딩 [13018%앵콜] 3만원대 스포츠웨어 이시국에 헬스장 공용복 입으세요? https://www.wadiz.kr/web/campaign/c  
sports_1 = '''
```

부담되는 값비싼 스포츠브랜드, 오래 입을 수 있는 퀄리티, 기능성을 모조리 때려 박은 스포츠웨어 싸구려 중국산이

반바지는 하체 근육을 잡아주지 못 할까...? 운동 선수들이 레깅스를 입는 이유 어떤 운동이든, 어떤 종류이든

...

```
sentiment_predict(sports_1)
```

0.6453437209129333

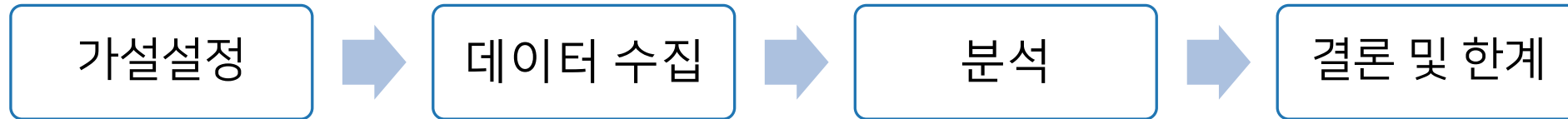
성공

정확도는 83% 로 다소 낮음

성공확률 0.5 이상일 경우 성공으로 설정

실제 최근 앵콜이 이루어진 제품 콘텐츠를  
입력한 결과 6개의 제품 중 5개의 제품 예측 성공

# The Relationship Between WTI price, Won-Dollar Exchange Rate and Korean Domestic Gasoline Price



- 데이터: 석유공사 유가정보 서비스, 한국은행
  - 가설설정: 코로나 사태로 인해 WTI 가격 하락으로 휘발유 가격이 하락했으나, 이 기간 중 원-달러 환율이 상승해 휘발유 가격 하락을 제한했을 것
  - 방법론: 그레인저 인과관계, 공적분 검정, 벡터오차수정모형, 충격반응검사, 분산분해, 더미변수
- 결론: WTI와 환율이 휘발유 가격에 영향을 미치나, 이 영향은 제한되고 있음을 알 수 있음  
현재 약 60%에 달하는 유류세의 영향으로 설명력이 제한 및 왜곡  
원유를 전량 수입에 의존하는 한국으로서는 유류세가 증가하면, 휘발유 사용량이 감소하여  
원유 수입이 감소하고 경상수지 흑자 개선에 기여  
그러나 과도한 세금이 소비자에게 부담이 될 수 있다는 점에 유의

# The relationship between WTI price, Won-Dollar exchange rate and Korean Domestic gasoline price

Data Description				
Variable	Definition	Unit	Symbol	Period
Foreign Exchange rate	Won-Dollar exchange rate	Log( Won/Dollar )	EXR	Monthly
Oil price	West-Texas Intermediate spot price	Log( Dollar/BBL )	WTI	Monthly
Gasoline price	Average gasoline price sold in Korean gas stations	Log( Won/Liter )	GAS	Monthly

최근 20달러 환율, 휘발유 월별 가격을 수집  
이 외에 코로나시점과 금년간의 원유(WTI), 원-달러 환율,  
금융위기와 코로나 팬데믹 시점의 더미 데이터 포함

Augmented Dickey-Fuller Unit Root Test			
variables	t-Statistic	Prob	result
EXR	-2.312206	0.1690	Non-stationary
GAS	-1.937516	0.3147	Non-stationary
WTI	-2.565417	0.1016	Non-stationary
dEXR	-10.81856	0.0000	stationary
dGAS	-11.55866	0.0000	stationary
dWTI	-11.93870	0.0000	stationary

## Stationary Test

1차분후 Stationary 함을 확인 가능

Granger Causality		
Null Hypothesis:	F-Statistic	Prob.
DEXR does not Granger Cause DGAS	8.01808	4.E-05***
DGAS does not Granger Cause DEXR	2.49246	0.0608*
DWTI does not Granger Cause DGAS	35.0869	9.E-19***
DGAS does not Granger Cause DWTI	0.48618	0.6922
DWTI does not Granger Cause DEXR	0.75888	0.5182
DEXR does not Granger Cause DWTI	4.86201	0.0027***

## Granger Causality

상호 인과관계를 가지는 변수들이 존재하므로 추가적인 테스트 필요

Cointegration test				
Trace test				
Hypothesized		Trace	0.05	
No. of CE(s)	Eigenvalue	Statistic	Critical Value	Prob.**
None*	0.0933	36.3734	29.7971	0.0076
At most 1	0.03179	12.4755	15.4947	0.1355
At most 2*	0.01864	4.59211	3.84147	0.0321
Trace test indicates no cointegrating eqn(s) at the 0.05 level				
Maximum Eigenvalue test				
Hypothesized		Max-Eigen	0.05	
No. of CE(s)	Eigenvalue	Statistic	Critical Value	Prob.**
None*	0.0933	23.8979	21.1316	0.0199
At most 1	0.03179	7.88337	14.2646	0.3907
At most 2*	0.01864	4.59211	3.84147	0.0321
Max-eigenvalue test indicates no cointegrating eqn(s) at the 0.05 level				

## 공적분 관계 분석

공적분 관계가 존재. 즉, 변수들간의 장기적 관계가 존재  
변수들간 장기적 관계가 존재하지 않을 경우, 벡터자기회귀 모형 활용,  
그러나 현재 장기적 관계가 존재하므로, 벡터오차수정모형을 활용

# The relationship between WTI price, Won-Dollar exchange rate and Korean Domestic gasoline price

$$\Delta Y_t = \alpha_0 + \alpha_1 Z_{t-1} + \sum_{i=1}^n \gamma_i \Delta X_{t-i} + \sum_{j=1}^m \delta_j \Delta Y_{t-j} + \varepsilon_t$$

Vector Error Correction Estimates			
Standard errors in ( ) & t-statistics in [ ]			
Cointegrating Eq:		CointEq1	
GAS(-1)	WTI(-1)	1	
		-0.42658	
	EXR(-1)	(0.02375)	
		[-17.9608]	
	C	-0.86086	
		(0.11615)	
		[-7.41170]	
		0.18508	
Error Correction:	D(GAS)	D(WTI)	D(EXR)
CointEq1	-0.0015	0.53158***	0.036919
	(0.02585)	(0.14141)	(0.03008)
	[-0.05798]	[ 3.75915]	[ 1.22753]
D(GAS(-1))	0.48437***	0.17729	-0.029718
	(0.07517)	(0.41121)	(0.08746)
	[ 6.44389]	[ 0.43114]	[-0.33979]
D(GAS(-2))	-0.32951***	0.5066	-0.281423***
	(0.07952)	(0.43502)	(0.09252)
	[-4.14374]	[ 1.16454]	[-3.04172]
D(GAS(-3))	-0.07708**	-0.37892	0.02342
	(0.06001)	(0.32829)	(0.06982)
	[-1.28440]	[-1.15424]	[ 0.33542]
D(WTI(-1))	0.11235***	0.27707***	-0.004545
	(0.01503)	(0.08220)	(0.01748)
	[ 7.47667]	[ 3.37057]	[-0.25995]
D(WTI(-2))	-0.04078***	-0.18918**	0.028444**
	(0.01617)	(0.08844)	(0.01881)
	[-2.52240]	[-2.13913]	[ 1.51220]
D(WTI(-3))	0.05428***	-0.18562**	0.04013**
	(0.01879)	(0.10282)	(0.02187)
	[ 2.88814]	[-1.80531]	[ 1.83518]
D(EXR(-1))	-0.09687*	-0.75432**	0.420042***
	(0.06086)	(0.33292)	(0.07081)
	[-1.59175]	[-2.26577]	[ 5.93232]
D(EXR(-2))	-0.15505***	-0.52953**	-0.167966
	(0.06420)	(0.35119)	(0.07469)
	[-2.41520]	[-1.50780]	[-2.24877]
D(EXR(-3))	0.13113**	0.27424	0.064582
	(0.06137)	(0.33576)	(0.07141)
	[ 2.13651]	[ 0.81678]	[ 0.90440]
C	8.17E-05	-0.00184	-0.000454
	(0.00056)	(0.00306)	(0.00065)
	[ 0.14595]	[-0.59906]	[-0.69771]
COV	-0.00278	-0.04053***	0.000717
	(0.00315)	(0.01726)	(0.00367)
	[-0.88113]	[-2.34908]	[ 0.19550]
FIN	0.001439	0.025482***	0.004064**
	(0.00174)	(0.00952)	(0.00202)

### Vector Error Correction Model

WTI 가격과 환율의 변화가 모두 휘발유 가격 변동에 큰 영향

또한 WTI 가격 변동과 환율 변동 모두 서로에게 큰 영향을 미치는 것을 확인

코로나 더미 변수는 WTI에 상당히 부정적인 영향

금융위기의 더미 변수는 WTI 가격 상승에 영향, 이는 금융위기 당시 원자재 가격 폭등과 연관

금융위기 변수는 환율에도 긍정적인 영향을 미침

이는 금융위기가 원화를 평가절하한 것과 맞물려 원-달러 환율 상승 견인

Variance Decomposition using Cholesky (d.f. adjusted) Factors			
Variance Decomposition of GAS:			
Period	GAS	WTI	EXR
1	100	0	0
2	92.6122	7.15088	0.236879
3	85.3023	11.5827	3.115042
4	79.6867	13.9825*	6.33085
5	78.2413	13.7875	7.97123
6	78.539	12.7049	8.756182
7	78.9398	11.8335	9.226684
8	79.0258	11.4097	9.564531
9	78.9421	11.2302	9.827697
10	78.8746	11.0287	10.09676*
Variance Decomposition of WTI			
Period	GAS	WTI	EXR
1	26.2943	73.7057	0
2	27.5046	70.1427	2.352733
3	29.5306	61.2668	9.202659
4	31.0536	53.896	15.05038
5	32.3053	48.9465	18.74824
6	33.2155	45.9735	20.81107
7	33.7896	43.9891	22.22137
8	34.2356	42.2275	23.53686
9	34.6617	40.4442	24.89412
10	35.0501*	38.7512	26.19869*
Variance Decomposition of EXR			
Period	GAS	WTI	EXR
1	0.69814	14.6521	84.64973
2	0.66138	14.8235	84.51517
3	1.31353	14.3077	84.37874
4	1.45234*	14.3974	84.15026
5	1.31512	15.3628	83.32209
6	1.18643	16.4938	82.31981
7	1.1019	17.3223	81.57586
8	1.03333	17.9324	81.03431
9	0.96678	18.5364*	80.49679
10	1.07377	12.0976	86.82865

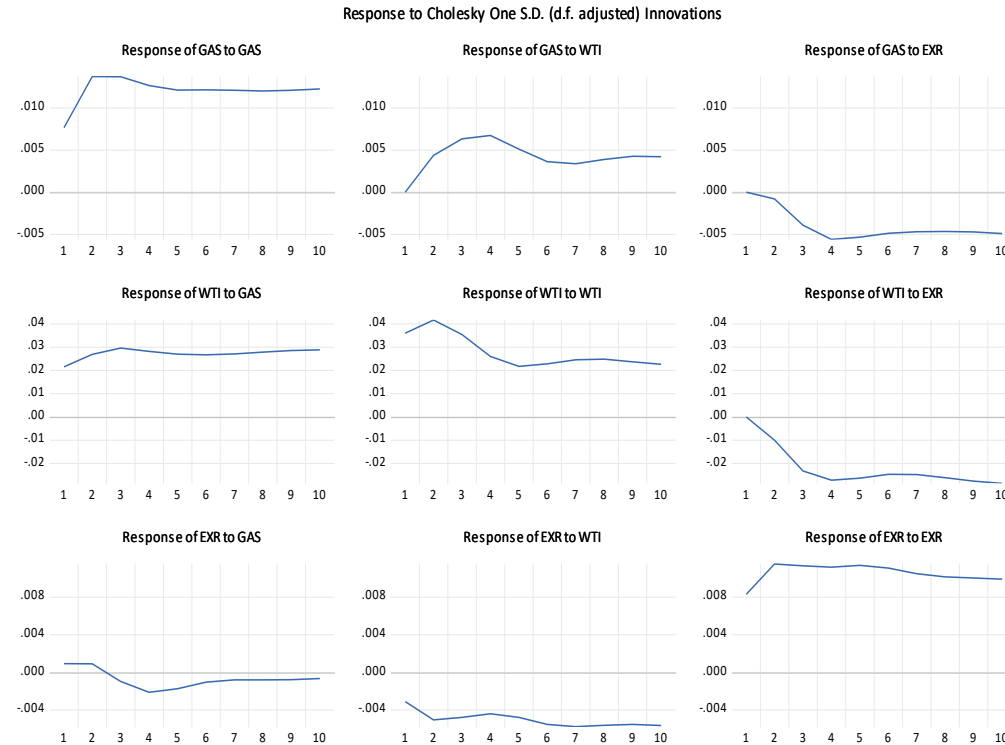
### Variance Decomposition

환율과 WTI는 휘발유 가격 변동에 대해 각각 10%와 14%만큼의 설명력을 가짐

타 변수들은 환율변화에 대한 설명력이 낮음을 알 수 있음

환율과 휘발유가격 변화는 모두 WTI가격변화에 대해 강한 설명력을 가짐

# The relationship between WTI price, Won-Dollar exchange rate and Korean domestic gasoline price



## Impulse response

WTI 가격과 환율의 변화는 상호간 부정적인 영향을 미치는 것으로 나타남

WTI 가격 변화는 휘발유 가격에 긍정적 영향을 미치나, 네 달 이후에는 그 영향이 감소

환율 변화는 네 달 까지 휘발유 가격에 음의 영향을 미치며 일정하게 유지

# NASDAQ IPO Price Analysis Using Regression Analysis

데이터 수집



분석

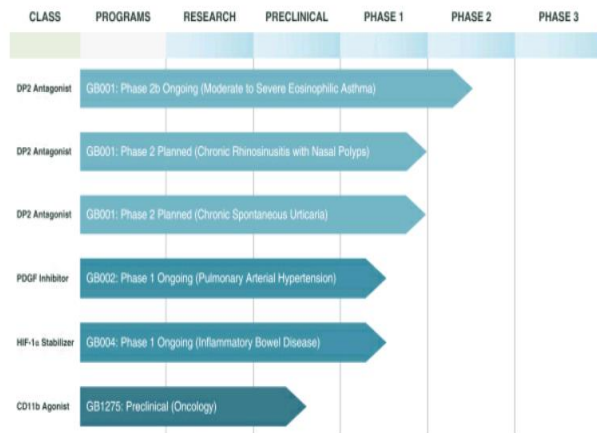


결론 및 한계

- 목적: 나스닥 바이오 종목 Gossamerbio IPO 성공여부 예측
- 데이터: Bloomberg NASDAQ IPO 데이터, 나스닥 바이오 섹터 시장종목 중 48개 기업의 6개 변수 이용  
48개 기업 중 가장 최근의 8개 기업은 Test Set 나머지 40개 기업은 Train Set
- 방법론: E-Views를 이용한 회귀분석
- 결론: GossamerBio를 회귀분석식에 적용한 결과 저평가 되어있음을 알 수 있음  
실제 IPO후 GossamerBio의 주가는 3% 상승

Business School of City, University of London  
2020년 5월

# NASDAQ IPO Price Analysis Using Regression Analysis



면역학 종양학 관련 다양한 임상시험 진행중  
수익금의 71%를 R&D에 투자

- 일반적으로 수익금 중 부채상환과 영업활동비용으로 지출되는 비율이 높은 경우, 부진을 겪는 경우 많음
- 2019년 17,250,000주를 \$16에 상장결정

Dependent Variable: UNDERPRICED\_INDICATOR  
Method: ML - Binary Probit (Newton-Raphson / Marquardt steps)  
Date: 03/24/20 Time: 00:32  
Sample: 1 40  
Included observations: 40  
Convergence achieved after 9 iterations  
Coefficient covariance computed using observed Hessian

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-4.877217	2.060241	-2.367305	0.0179
PRICE_UPDATED	20.45975	8.163791	2.506158	0.0122
STOCK_MARKET_SENTIMENT	14.52201	10.03779	1.446734	0.1480
IPOMKT	2.317269	1.165191	1.988746	0.0467
OVERHANG	0.003072	0.032681	0.094011	0.9251
RETURN_ON_ASSETS	-1.445124	0.669819	-2.157484	0.0310
TOTAL_ASSETS_TURNOVER	4.392923	2.358797	1.862357	0.0626
McFadden R-squared	0.598968	Mean dependent var	0.550000	
S.D. dependent var	0.503831	S.E. of regression	0.339139	
Akaike info criterion	0.901932	Sum squared resid	3.795506	
Schwarz criterion	1.197485	Log likelihood	-11.03863	
Hannan-Quinn criter.	1.008794	Deviance	22.07726	
Restr. deviance	55.05111	Restr. log likelihood	-27.52555	
LR statistic	32.97384	Avg. log likelihood	-0.275966	
Prob(LR statistic)	0.000011			
Obs with Dep=0	18	Total obs	40	
Obs with Dep=1	22			

## 회귀분석식

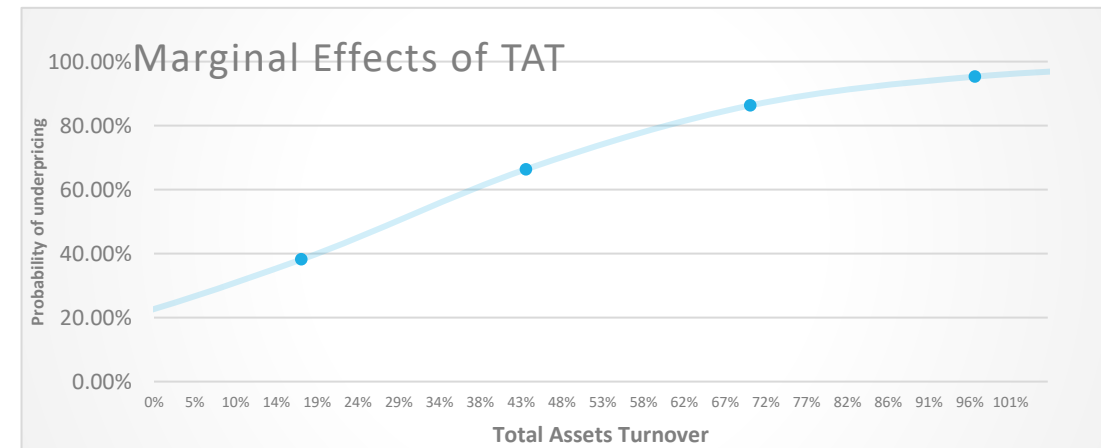
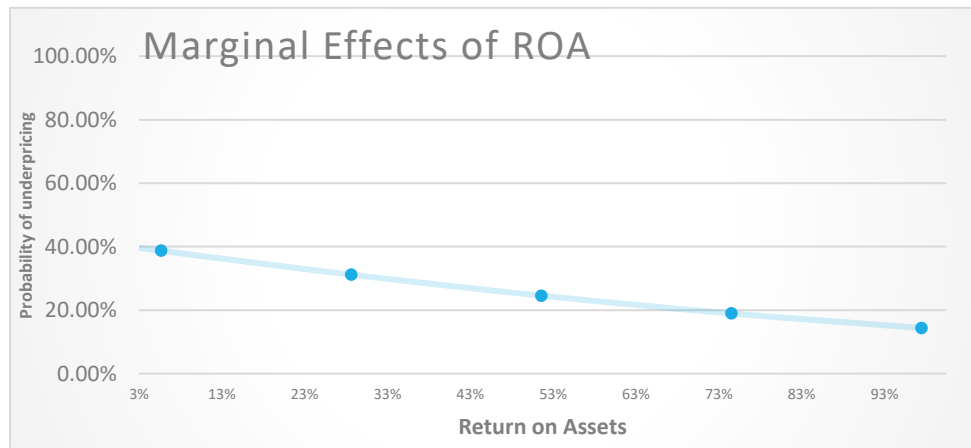
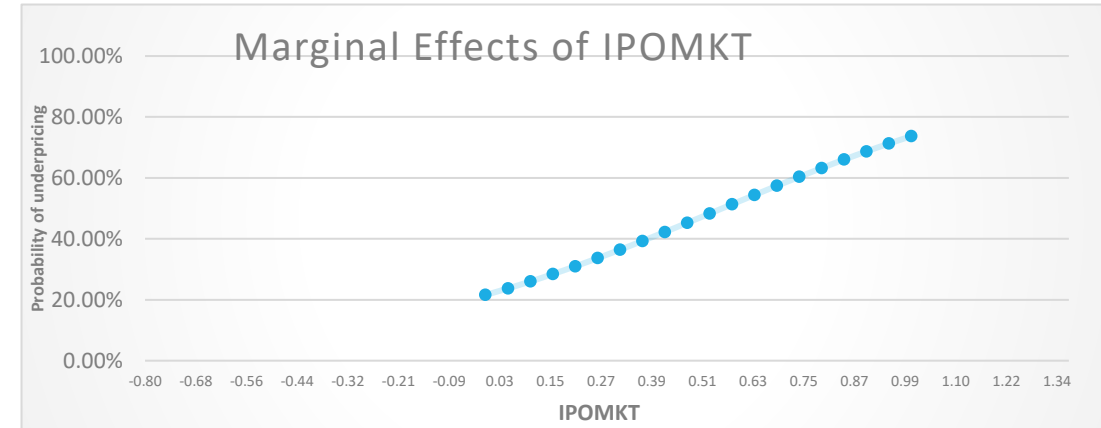
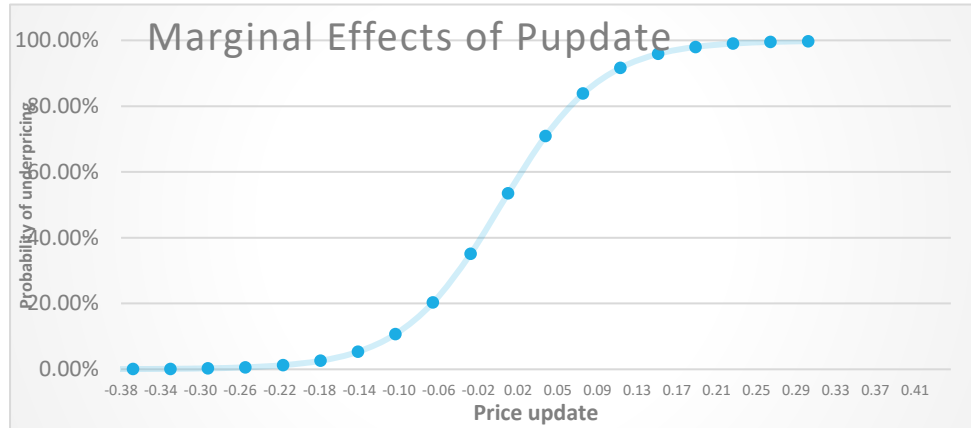
$$y = \alpha + \beta_1 * Pupdate + \beta_2 * SMKT + \beta_3 * ROA + \beta_4 * Overhang + \beta_5 * IPOMKT + \beta_6 * TAT + u_i$$

변수설명

- Price Update = 기업 공개가격 / (최고가 - 최저가)
- Stock Market Sentiment: 시장민감도
- IPO Market: 해당 IPO 시장의 Hot/Cold 여부
- Overhang: 개인투자자들의 잠재물량
- Return on Asset: 순이익/총자산
- Total Asset Turnover: 매출액/총자산



# NASDAQ IPO Price Analysis Using Regression Analysis



90% 유의수준하에서 4개의 변수가 유의미한 것으로 판단,

# NASDAQ IPO Price Analysis Using Regression Analysis

Underpriced Issues  
Overpriced Issues  
Total Issues

3
5
8

실제 기업 사례를 적용하기 전 최근 8개 기업으로 테스트 시행, 정확도는 87.5%

Prediction Table				
	Number Correct	Number Incorrect	%Correct	Type I/II Error
Underpriced Issues	2	1	66.67%	33.33%
Overpriced Issues	5	0	100.00%	0.00%
Total	7	1	87.50%	12.50%

Name	Pupdate	SMKT	ROA	OVERHANG	IPOMKT	TAT
Gossamer Bio Inc	0.1800	0.021916717	-0.3426	0.0267	1	-0.0004

LOGISTIC REGRESSION RESULTS							
$\alpha$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$	$\beta_6$	Logit Probability
-4.877	20.460	14.522	-1.445	0.003	2.317	4.393	87.37%

40개 기업 중 22개 기업이 기업 공개일 기준으로  
저평가되었으며,(공개일에 상승예상), 18개 기업이 고평가됨  
상기한 22개 저평가 기업 중 20개 기업이 실제로 저평가됨  
Type1에러는 13.64%, Type2에러는 27.8%

GossamerBio를 회귀분석식에 적용한 결과 저평가 되어있음을 알 수 있음.  
실제 IPO후 Gossamer Bio의 주가는 3% 상승

Thank You