

## 05 | 주가예측 및 로보어드바이저

# 1 | 주가예측



# 1 주가예측은 불가능한가?

## 01 가격변동은 예측 불가능(Random Walk)

- “가격변화는 무작위이고 과거의 패턴을 따르지 않는다”라는 이론이 1900년대 중반까지 지배적
- “주가예측은 신의 영역이다”라고 평가

## 02 과거 패턴으로 미래 예측 가능

- 1988년 “주식시장은 무작위 하게 움직이지 않는다”는 것이 증명됨
- 최근 빅데이터 활용한 자동화된 종목추천 및 알고리즘 트레이딩이 각광받고 있음



## 2 추가 데이터의 기본적인 구조



- 날짜
- 몇 개월 밖에 안된 데이터도 있고 10년 이상 된 것도 존재함



- 종목
- 1천 개 이상의 종목이 있으며, 산업군으로 분류될 수도 있음



- 거래량
- 하루 거래된 총 주식수



## 2 추가 데이터의 기본적인 구조



- 고가
- 시장이 시작돼서 종료하기 까지 최고로 높은 거래가격



- 저가
- 시장이 시작돼서 종료하기 까지 최저의 거래가격



## 2 추가 데이터의 기본적인 구조



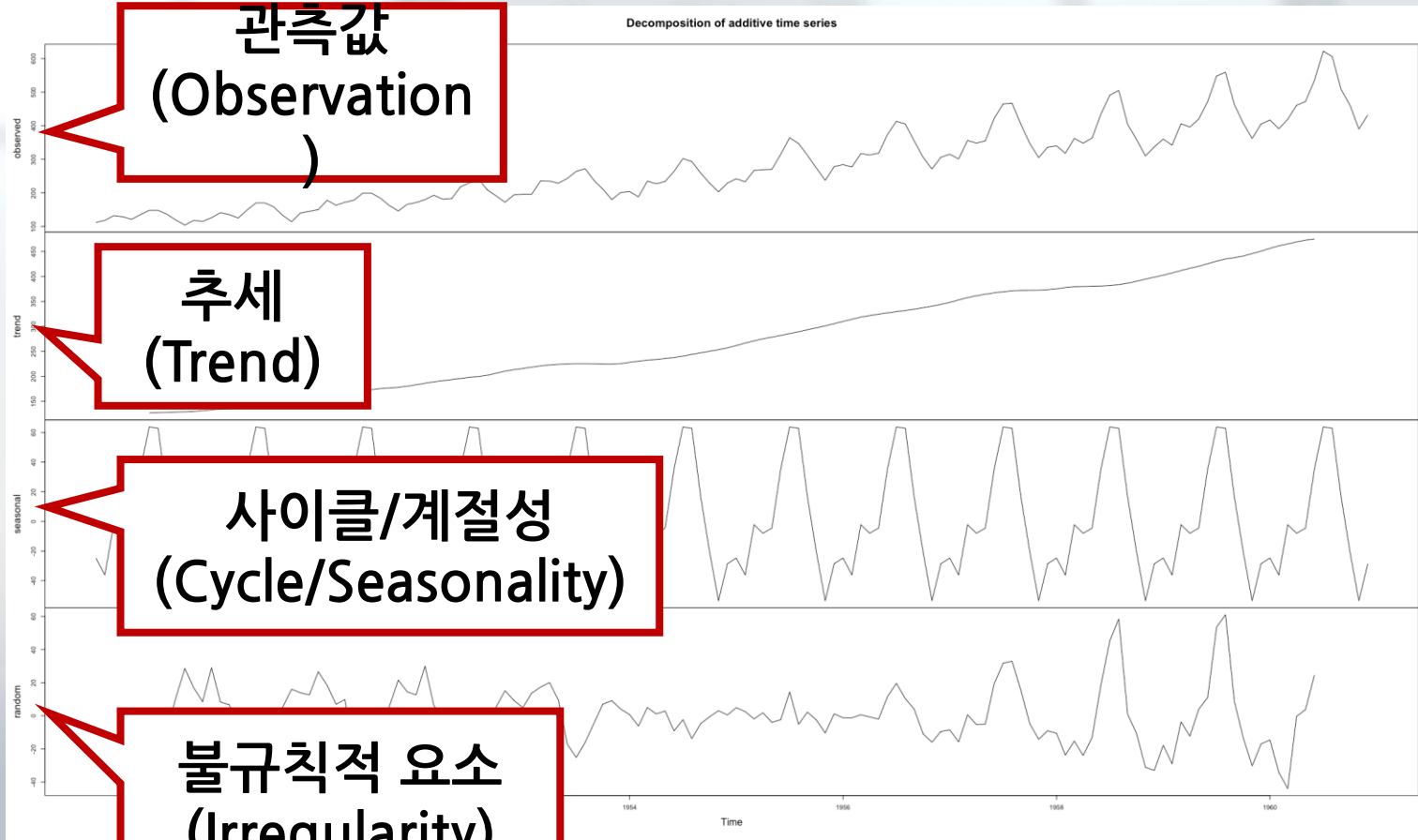
- 시가
- 시장이 시작된 시점의 첫 거래가격



- 종가
- 시장이 종료되는 시점의 최종 거래가격



### 3 일반적인 시계열 데이터의 구성



관측값  
(Observation)

추세  
(Trend)

사이클/계절성  
(Cycle/Seasonality)

불규칙적 요소  
(Irregularity)



## 4 < 추가예측 분야의 이슈

- 일반적 시계열 모형은 **오차가 커서 사용 어려움**
- 수천 개 종목에 대해 기업의 재무적 측면, 기업환경의 변화, 국내외 정치 경제적 영향, 투자자 심리, 가격자체의 변화패턴 등 **복잡한 요인 상이하게 존재**
- 다양한 요인이 실시간, 일단위로 변화하는 복잡한 환경  
→ 모든 정보를 **사람이 처리하기 어려움**
- 투자자들은 보다 **정확한 예측**을 기반으로 **저렴한 맞춤형 서비스**를 원함



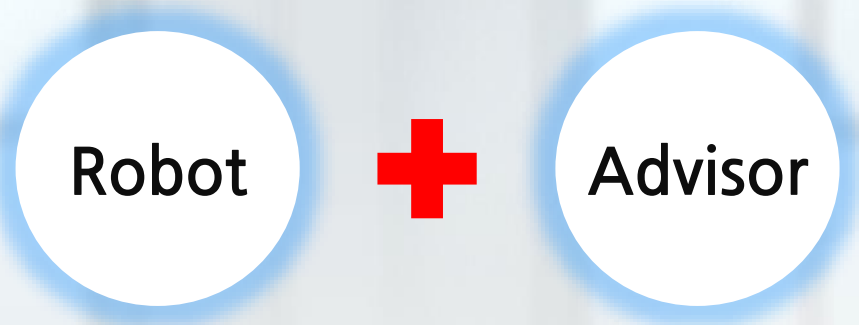
## 03 | 추가예측 및 로보어드바이저

# 2 로보어드바이저





# 1 로보 어드바이저의 출현



인공지능, 기계학습, 데이터 마이닝 기법 활용!

- 미국 대형 금융사들(피델리티, 메릴린치, 블랙록, 골드만삭스 등)이 활용
- 로보어드바이저 자문사 인수, 자체 개발 방식



# 1 로보 어드바이저의 출현



로보어드바이저  
운용자산만 210억 달러  
(약 24조원)

[미국 뱅가드(Vanguard) 社]

국내에는 아직 초기단계!



## 2 로보 어드바이저의 혜택 및 장점

- 다수 고객에게 **항시 서비스 가능**
- **일관된 품질**의 서비스 제공
- 시간이 지남에 따라 **정확도 향상** 가능
- 정형/비정형, 내부/외부 데이터 등 **다양한 정보 정밀 활용** 가능
- 사람이 **간과할 수 있는 패턴 고려** 가능
- 실시간 **급변 사태 대응** 가능
- 개인 투자성향에 맞추어 **개인화된 조언 제공** 가능



### 3 추가 데이터 입수

[Source Data]

일자	종목	시가	고가	저가	종가	거래량

다양한  
Technical Indicator

[Dataset]

종목	최근1주 평균시가	최근2주 평균고가	최근1개월 평균저가	최근6개월 평균종가	최근1년 평균거래량	ADX	bandPctB	종가기준 3% 상승여부
								Yes
								No
								No
								No



## 4 Technical Indicator를 활용한 설명변수

ADX  
aroon  
ATR  
BBands  
CCI  
chaikinAD  
chaikinVolatility  
CLV  
CMO  
DonchianChannel  
DPO

DVI  
EMV  
GMMA  
KST  
MACD  
MFI  
momentum  
OBV  
ROC  
RSI  
SAR

SMA  
sentiment  
TDI  
TRIX  
ultimateOscillator  
VHF  
volatility  
williamAD  
WPR  
Zigzag



## 5 R Script

```
install.packages("TTR")  
install.packages("quantmod")  
library(TTR)  
library(quantmod)  
x <- getSymbols("DWTI", src='yahoo',  
+ from="2013-01-01",  
+ to="2016-03-01", auto.assign=FALSE)  
x <- adjustOHLC(x, use.Adjusted = TRUE)
```



## 6 ROC(Rate of Change)

```
chartSeries(x, subset='last 3 months',  
  name=paste("DWTI"), line.type = "l", theme =  
  chartTheme("white",  
  up.col='green',dn.col='red'),  
  major.ticks="months", color.vol=FALSE)
```

```
addBBands(n=5,sd=1)
```

```
addSMA(n=40,col="blue")
```

```
addSMA(n=20,col="orange")
```

```
addROC()
```



# 6 ROC(Rate of Change)





# 7 News를 활용한 설명변수

## Valeant, Ackman must face U.S. insider trading lawsuit

REUTERS November 12, 2015



2015-11-12 부터 \$84에서 \$76으로 하락하더니 \$69까지 급락하는 모습 보임

Nov 18, 2015	69.75	75.26	69.33	72.60	72.60	11,844,400
Nov 17, 2015	72.22	74.00	69.95	70.32	70.32	13,442,800
Nov 16, 2015	74.49	75.69	72.25	73.32	73.32	10,267,200
Nov 13, 2015	74.54	76.30	72.09	75.41	75.41	14,632,000
Nov 12, 2015	76.81	77.00	73.69	73.77	73.77	17,453,100
Nov 11, 2015	84.62	86.11	78.60	78.90	78.90	16,933,800



## 8 애널리스트 분석결과 이용한 설명변수

### Aegean Marine: (ANW)

(Delayed Data From NYSE)

\$10.48 USD

+0.24 (2.34%)

Updated Sep 23, 2016 04:02 PM ET

Volume: 300,155

Open: \$10.23

Prior Close: \$10.24

Zacks Rank <sup>[?]</sup>:

Style Scores <sup>[?]</sup>:

+ Add to portfolio

Trades from \$3

1-Strong Buy 1

Value: A | Growth: B | Momentum: F | VGM: A

[View All Zacks Rank #1 Strong Buys](#)

- 종목별 거래추천정보 제공
- Strong Buy - Buy - Hold - Sell - Strong Sell로 구분된 추천정보



# 9 타깃변수

1주 기준 20% 상승 = Yes

1주 기준 20% 상승 = No



# 10 추가예측 및 빅데이터 분석

Input Data  
Period (3years)

Target Data Period  
(Day, Week, Month)

Modeling



O  
X  
X  
O  
X  
X

Deployment



O  
X  
O  
X  
X  
X



# 11 접근방법

## 현상

- 3,000개 종목 중 1개월 내, 전월 최종일 증가 대비 당월 1개월 내, 증가가 3% 이상 상승한 적이 한번이라도 있는 종목 : 600개
- 무작위 선택 투자 시, **20%는 3% 이상 가격 상승**

## 예측

- 3,000개 종목 중 1개월 내 전월 최종일 증가 대비 당월 1개월 내, 증가가 3% 이상 상승할 가능성 있는 종목 : 50개 예측
- 실제 3%이상 상승한 종목은 **40개**



## 12 모델링

```
library(caret)

stock <- read.csv("stock.csv")

model <-
train(class~.,data=stock,method="rf")
predict(model,type="class")
```



# 13 성과측정

## [Confusion Matrix]

Actual

		Actual	
		NO	YES
Prediction	NO	2390	560
	YES	10	40

$$\text{Accuracy} = (2390 + 40) / (2390 + 560 + 10 + 40) = 81.0\%$$

$$\text{Precision} = 40 / (10 + 40) = 80.0\%$$

$$\text{Recall} = 40 / (40 + 560) = 6.7\%$$

$$\text{Lift} = \text{Precision} / \text{Base} = 80 / 20 = 400\%$$

