

# 빅 데이터를 활용한 주식 동향 분석

## (Stock trend analysis using big data)

최 원 우\*, 김 남 규, 한 창 재, 조 우 현

(Won-Woo Choi, Nam-Gyu Kim, Chang-Jae Han, Woo-Hyun Jo)

부경대학교 컴퓨터공학과

Abstract : Currently, many investors are participating in stock investments to obtain profits from market appreciation. However, stock analysis and investment still have a lot of difficult parts. This paper aims to reduce this accessibility by analyzing and visualizing stock trends using 'pandas' and 'plotly' based on past and present data.

Keywords : Stock Analysis, crawling, pandas, plotly

## I. 서 론

과거와 달리 개인의 소비 습관이 변화하며 시장에는 잉여 자본이 늘어나고 있다. 과거, 대부분의 사람들은 자금을 은행에 예탁하여 단리, 복리의 이자를 통해 돈을 모았다. 하지만 오늘날 시장은 저금리 현상이 지속적으로 이어지고 있다. 이에 따라 소비자는 더 이상 예금과 적금을 돈을 불러 모으는 형태로 사용하지 않고 단순히 안전하게 돈을 보관하는 '저축 보관 목적'으로 사용되고 있다. 또한 4차 산업혁명과 맞물려 '주식 투자'에 많은 개인투자자들의 관심과 참여도가 높아지고 있다. 과거 주식투자의 경우 회사의 주식을 매매하여 1차적인 이익 청구권 행사를 통한 배당 회수가 주된 목적이었으나, 최근에는 자유로운 주식 매수 매도를 통한 시세 차익이 성행하고 있다.

현 저금리 현상에 지속되는 가운데, 위험자산 투자는 가계 자산형성에 필수적인 요소로 자리 잡고 있으며 안정적인고 효율적인 투자를 위해 지속적인 연구가 필요하다. 따라서 본고의 목적은 시장에 형성되어 있는 주식들의 과거 및 현재의 데이터를 기반으로 주식 동향을 분석하고 가공하여 투자 입문자를 비롯한 다양한 개인들로 하여금 효율적인 투자 정보 제공에 도움을 주는데 초점을 둔다.

## II. 관련 이론

### 1. 주식 관련 내용

#### 1.1 주가

대표적으로 시가, 고가, 저가, 종가, 거래량이 있다. 시가는 장 시작 시 최초 거래된 가격을 의미한다. 고가는 하루 동안의 최고가를 의미한다. 저가는 하루 동안의 최저가를 의미한다. 종가는 마지막으로 거래된 가격을 의미한다. 거래량은 하루 동안 거래된 주식수를 의미한다.

#### 1.2 볼린저 밴드

주가의 변동이 표준정규분포 함수에 따른다고 가정하고 주가를 따라 위아래에 밴드를 표시하여 주가의 상대적인 높낮이를 알려준다. 주가의 20일 이동 평균선을 기준으로 상대적인 고점인 상단 밴드와 상대적인 저점인 하단 밴드로 구성된다.

#### 1.3 볼린저 밴드 지표 : %b

주가가 볼린저 밴드상 어느 정도에 위치하는지 보여주는 지표로 0이면 하단에 걸치게 되고 1이면 상단에 걸치게 된다. %b는 상한성과 하한선이 없기 때문에 상단 밴드위이면 1.0을 넘기고 하단 밴드 아래에 있으면 0.0보다 작은 수가 된다.

#### 1.4 MFI

주식 거래를 위한 자금이 유입되거나 유출되는 양을 측정해주는 지표이다. 이를 이용해 추세 전환시기를 예측하거나 시세의 과열 및 침체 정도를 파악하는데 유용하다.

#### 1.5 추세 추종 기법

%b와 MFI의 값을 기준으로 매수, 매도를 하는 기법이다.

매수 시점은 [%b > 0.8]이고 [MFI > 80]일 때이며 매도 시점은 [%b < 0.2]이고 [MFI < 20]일 때이다.

## 2. 추가 예측 관련 내용

### 2.1 텐서플로(Tensorflow)

구글이 개발하는 머신러닝, 딥러닝을 위한 오픈소스 플랫폼이다. 파이썬과 C++ API를 기본적으로 제공한다.

### 2.2 순환 신경망(RNN)

이전의 데이터를 통해 학습된 셀의 상태 정보가 다음 데이터를 이용하여 학습시킬 때 다시 사용되는 신경망으로 순차적인 데이터 처리에 적합한 모델이다. 입력과 출력 사이의 거리가 멀어질수록 학습 능력이 저하되는 단점이 있다.

$$h_t = \tanh(U_{x_t} + Wh_{t-1}) \quad (1)$$

식 (1)은 t 단계에서의 RNN Layer의 출력으로 tanh는 활성화 함수로 ReLU로 대체할 수 있다. x는 입력을 의미하고, U, W는 입력과 출력에 곱해지는 가중치이다.

### 2.3 장단기 기억(LSTM)

RNN의 단점을 극복하기 위해 고안된 것으로 은닉층의 메모리 셀에 이전 상태를 지울 것인지 결정하고 현재 state와 입력을 토대로 출력값을 결정한다.

$$p = W(h_{t-1}, x_t) + b \quad (2)$$

$$f_t = c_{t-1} \times \text{sigmoid}(p_f) \quad (3)$$

$$i_t = \text{sigmoid}(p_f) \times \tanh(p_i) \quad (4)$$

$$c_t = f_t + i_t \quad (5)$$

$$h_t = \tanh(c_t) \times \text{sigmoid}(p_0) \quad (6)$$

식은 위와 같으며 f는 망각 게이트로 이전 셀 상태에서 지울 정보를 학습시키는 용도이다. i는 입력 게이트로 새로운 데이터를 학습하는 용도이다. c는 셀 상태로 망각 게이트와 입력 게이트의 결과로 구한다. h는 은닉 상태로 출력을 의미한다.

## III. 시스템 설계 및 구현

### 1. 개발 시스템 구성도

본고에서 구현할 시스템은 그림 1과 같다. 크게

서버, 데이터베이스, 웹 세가지로 구성된다.

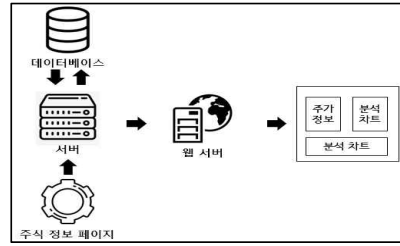


그림 1. 시스템 구성도

### 2. 주식 동향 분석 시스템 구성

시스템 구현에 사용하는 소프트웨어는 표 1과 같다. Python, Python 모듈 (Pandas, Numpy, Dash, Plotly, Tensorflow), MariaDB 등을 사용하였다.

표 1. 시스템 구성

	개발 모듈	기능
S/W 시스템	Python	전체적인 시스템 구현
	Pandas, Numpy	데이터 분석 및 처리
	Plotly	데이터 시각화
	Tensorflow	추가 예측 구현
	Dash	웹 페이지 구현
	MariaDB	추가 정보 저장

## 3. 시스템 설계

### 3.1 추가 정보 수집

그림 2는 추가 정보 수집 구성을 나타낸 것이다. Python의 'beautifulsoup' 모듈을 사용해 주식 정보 페이지에서 주식 정보 데이터를 Crawling하여 DataFrame에 추가한다. 이후 SQL을 사용하여 DB에 해당 정보를 저장한다.



그림 2. 추가 정보 수집

### 3.2 추가 분석 및 시각화

그림 3은 추가 분석 및 시각화를 나타낸 것이

다. 시각화는 'Plotly'를 사용하고 주가 예측에는 머신러닝 모듈 'Tensorflow'를 사용한다.

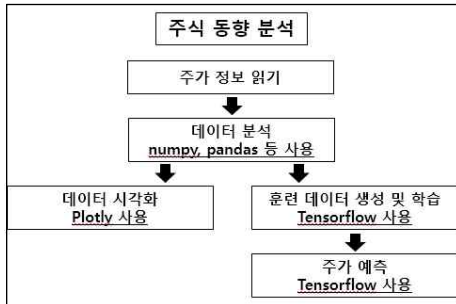


그림 3. 주가 분석 및 시각화

시각화로는 기본적으로 '시가, 저가, 고가, 종가'를 표현하기 위한 '캔들 차트'를 사용한다. 또한 '볼린저 밴드의 %b'와 '거래량 지표 MFI'를 사용하여 매수, 매도 시점을 나타내는 '추세 추종 기법'을 사용한다.

주가 예측을 위해 'Tensorflow'를 사용한다. 학습과 예측에 사용할 데이터는 DB에 저장된 주가들이다. 주가 데이터는 날짜별로 입력되어 있기 때문에 순차적인 데이터에 해당한다. 이에 적합한 LSTM Layer를 사용하여 학습 및 예측을 실행한다.

### 3.3 웹 페이지 구현

웹 페이지는 'Plotly'의 시각화 정보와 html 등을 쉽게 구현할 수 있는 Python의 'Dash' 모듈을 사용한다. Crawling으로 DB에 저장한 주식 정보와 'Plotly', 'Tensorflow' 등을 활용한 분석 및 시각화 결과를 확인할 수 있도록 구성하였다.

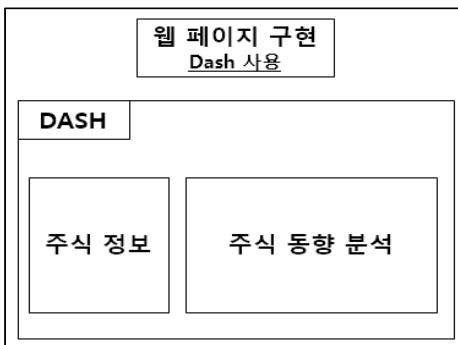


그림 4. 웹 페이지 구현

### 3. 실제 시스템 구현

그림 5는 실제 시스템을 구현한 화면이다. 주가 예측, 캔들 차트, 추세 추종 기법 등의 정보를 출력하여 투자할 수 있도록 한다.

상장 기업 전체에 대한 전일 종가와 거래량을 테이블로 표시한다. 캔들 차트, 추세 추종 기법, 주가 예측과 같은 세부적인 분석은 '드롭 다운 메뉴'를 통해 선택된 기업을 대상으로 행해진다. 그림 5, 6, 7, 8은 '드롭 다운 메뉴'에서 '삼성전자'를 선택한 결과이다.

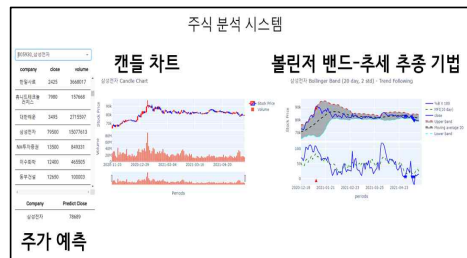


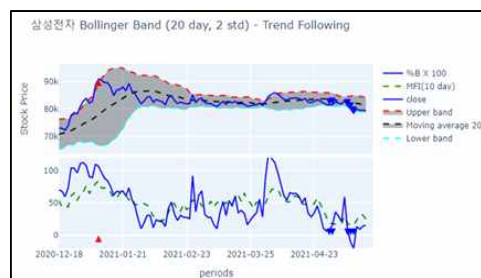
그림 5. 실제 구현 화면

company	close	volume
한일사료	2425	3668017
휴니드테크놀로지스	7980	157668
대한해운	3495	2715597
삼성전자	79500	15077613
NH투자증권	13500	849331
이수화학	12400	465505
동부건설	12650	100003
Company	Predict Close	
삼성전자	78689	

그림 6. 주가 예측



그림 7. 캔들 차트



#### IV. 결 론

본고에서는 분석 및 시각화를 통한 주식 동향 분석 시스템을 제안한다. crawling을 활용한 정보 수집과 Python의 라이브러리 ‘Pandas’, ‘Numpy’를 이용한 데이터 가공, 그리고 ‘Plotly’, ‘Tensorflow’ 등의 모듈을 통해 데이터를 분석하고 가공하며 시각화를 진행한다.

본 시스템은 주식 입문자가 어려움을 겪을 수밖에 없는 환경에서 주식들의 데이터를 분석하고 가공하여 사용자에게 효율적으로 투자할 수 있도록 정보를 제공한다. 투자자들이 이 정보를 기반으로 더 안정적이고 효율적으로 투자하는 것을 기대할 수 있다.

더불어, 주식 분석 기법은 본 시스템에서 사용한 ‘추세 추종 기법’ 외에도 다양하게 존재한다. 다양한 기법을 활용하는 것은 효율적인 투자의 방법 중 하나이다. 따라서 다른 기법들을 추가함으로써 더욱 다양하고 신뢰성이 있는 정보를 전달할 수 있는 발전 가능성 또한 존재한다.

#### V. Reference

- [1] 김황후, “파이썬 증권 데이터 분석”, “한빛미디어”, 2020
- [2] 존 볼린저, “볼린저 밴드 투자기법”, “이레미디어”, 2010
- [3] 김환희, “텐서플로 2.0 프로그래밍”, “위키북스”, 2020
- [4] <https://dash.plotly.com/>
- [5] <https://plotly.com/python-api-reference/>