|  |
| --- |
| **인공지능 기반 퀀트 알고리즘 개발**  **요 약**  현재 머신러닝은 각종 분석을 위해 사용되고, 이 분석한 정보들을 바탕으로 여러 서비스를 제공한다. 머신러닝을 적용하면 좋은 상황 중 하나가 빅데이터인데 사람이 미처 파악하지 못할 데이터간의 관계를 파악할 수 있기 때문이다. 특히나 주식은 수십 년의 주가 지수와 여러 종목별 주가가 있는 빅데이터다. 이 데이터들을 이용하여 수익률을 예측할 수 있다면 투자자가 더 안정적이고 효율적인 투자를 할 수 있게 될 것이다. 기존의 사람이 정의한 주가 예측 공식들이 아닌 컴퓨터가 데이터를 통해 예측하는 인공지능을 개발한다. |

**1. 서론**

**1.1. 연구배경**

돈을 많이 벌고자 하는 욕망은 대부분의 인간이 가지고 있는 욕구이다. 하지만 노동을 통해 얻을 수 있는 소득은 정해져 있다. 따라서 사람들은 투자를 하기 시작했고 주식은 현재 가장 많은 사람들이 이용하는 투자 수단이 되었다. 하지만 단순히 기업 이미지, 기업 대표, 인기 등을 보고 주식을 구매하는 것은 투자가 아니라 투기에 가깝다. 그래서 합리적이고 효율적인 투자를 위한 여러 분석 기법들이 탄생했다. 실제로 이러한 분석 기법을 이용한 투자는 수익률을 봤을 때 통계적으로 유의미한 결과를 보여줬다.

최근 머신러닝 분야가 빠르게 성장함에 따라 여러 분야에 머신러닝들이 적용되고 있다. 사람이 해결하기 어려웠던 문제나 정확도가 떨어졌던 문제들이 머신러닝들로 해결되고, 분석하기 어려운 데이터들에 대한 분석도 머신러닝으로 이루어진다. 나아가 머신러닝은 단순히 문제 해결에만 사용되는 것에 그치지 않고 이를 적용한 서비스를 제공하거나, 최근에는 사업 아이템 자체가 머신러닝이 메인인 스타트업도 생겨나고 있다. 특히 머신러닝이 사람보다 더 좋은 성능을 발휘하는 분야 중 하나가 빅데이터다.

주식시장은 공식적인 데이터들이 아주 많이 쌓여있는 빅데이터라고 볼 수 있다. 위에서 언급한 분석 기법들이 아직도 유용하게 사용되고 있지만, 주식시장이 빅데이터라면 머신러닝을 적용해보지 않을 이유가 없다. 현재도 머신러닝을 이용해서 주가 예측을 위한 시도는 계속되고 있다.

**1.2. 연구목표**

충분한 기간의 시간단위 주가 데이터와 주가 지수 데이터를 딥러닝 모델로 학습시키면, 시계열 예측이 가능하며, 이를 통해 유의미한 수준의 정확도를 가진 모델을 설계한다. 해당 모델을 통해 근미래의 주가 예측 정보를 제공하여 투자할 종목 선택에 도움을 주도록 한다.

**2. 관련연구**

**2.1. 엘리어트 파동 이론**

엘리어트 파동 이론이란, 1939년에 미국 회계사인 엘리어트가 발표한 주가의 변동은 상승 5파와 하락 3파로 움직이고, 끝없이 순환하면서 시장의 추세를 이어간다는 이론이다. 인간의 모든 행동은 패턴, 비율, 시간의 세 가지로 구성되고 이것이 주식시장에도 반영된다는 배경 하에 진행되는 분석 기법이다. 여러 차트 분석 기법들 중에 가장 많이 사용되는 이론이지만 타당성은 현재까지도 토론 대상이다.

**2.2. 재무제표를 이용한 RIM 계산**

RIM은 잔여이익모델법의 약자로써 기업의 자기자본에 미래 잔여이익의 현재가치를 더한 값을 가지고 기업가치를 구하는 방법이다. 이 방법은 주가를 예측하는 용도가 아니라 자신이 생각하는 매수가, 매도가를 결정하는 용도로 사용한다.

**2.3. 딥러닝 모델 종류**

근 몇년간, AI는 많은 관심을 받아왔고 많은 분야에서 활용되고 있다. 머신러닝, 더 깊게는 딥러닝까지 많은 연구가 이루어지고 있는데, 이 과정에서 ANN, DNN, CNN, RNN과 같은 모델들이 만들어졌다.

**2.3.1 K-NN**

K-최근접 이웃(K-Nearest Neighbor, KNN)은 지도 학습 알고리즘 중 하나이다. 어떤 데이터가 주어지면 그 주변의 데이터를 살펴본 뒤 더 많은 데이터가 포함되어 있는 범주로 분류하는 방식이다. K-NN의 특징은 훈련이 따로 필요 없다는 것이다. 훈련 데이터를 기반으로 모델을 만들고 테스트 데이터로 테스트를 하는 방식이다.

**2.3.2 SVM**

서포트 벡터 머신(support vector machine SVM)은 기계학습의 분야 중 하나로 패턴 인식, 자료분석을 위한 지도 학습 모델이며, 주로 분류와 회귀 분석을 위해 사용한다. 회귀를 위한 SVM은 SVR(support vector regression)이라고 불린다. SVR에 의해 생성된 모델은 손실 함수가 모델 예측에 가까운 모든 학습 데이터를 무시하기 때문에 학습 데이터의 하위 집합에만 의존한다.

**2.3.3 LSTM**

RNN의 경우, 데이터간의 거리가 멀어질수록 데이터의 상관관계가 불명확해지는데 이를 개선한 모델이 바로 LSTM(Long Short Term Memory Network)이다. 가장 기본적인 RNN 모델의 경우, 학습이 거듭될수록 초기의 weight값이 유지되지 않아 학습률이 저하되는 Long-Term Dependency 현상이 생기는데 LSTM에서는 Cell State라는 layer를 추가해서 weight를 계속 기억할 것인지 결정하여 이 문제를 개선했다. 이를 통해 직전의 데이터만 참고하는 Vanilla RNN과 달리 LSTM은 cell state에서 다룬 weight를 참고할 수 있게 되었다.

**3. 프로젝트 내용**

**3.1. 모델 학습에 필요한 데이터**

**3.1.1. 주가 및 지수**

시계열 데이터로 구성된 주가 데이터가 필요하다. 주가 데이터는 상한가, 하한가, 시가, 종가, 거래량 등으로 구성되어져 있고, 지수 데이터는 KOSPI, S&P500, NASDAQ 등이 있다. 시계열 데이터로 구성되어 있으니 순환신경망 모델에 학습시킬 수 있고, 주변 값과 비교하는 방식인 K-NN과 SVM에도 사용할 수 있다.

**3.1.2. 재무제표**

재무제표는 해당 기업의 주가에서 보여지지 않는 부분을 알 수 있다. 수익이나 부채 등 단순히 수치상으로 알 수 있는 정보들부터 여러 수치들 사이의 상관관계를 통해 기업의 현재 상태를 알 수 있다. 예를 들면, 매출은 증가하지 않았는데 매출채권은 급증했다면 실적부진을 만회하기 위해서 신용도가 떨어지는 신규 거래처에 외상으로 제품을 팔았을 가능성이 있다.

**3.2. 머신러닝 모델**

주가 및 지수, 재무제표를 통해 사용할 feature를 선택한다. 해당 feature 데이터들을 같은 시간에 측정한 데이터끼리 맞춰서 전처리한다. 그리고 이를 머신러닝 모델에 학습시켜서 근미래의 주가를 예측한다.

**4. 진행 일정**

|  |  |
| --- | --- |
| 진행 주차 | 내용 |
| 3/29 ~ 4/4 | 재무제표에서 주가에 영향을 미치는 지표 조사 |
| 4/5 ~ 4/11 | 증권사 API 사용방법 공부 및 숙지 |
| 4/12 ~ 4/18 | 증권사 API 테스트 |
| 4/19 ~ 4/25 | 인공지능을 사용하여 적용할 퀀트 알고리즘 연구 및 테스트 |
| 4/26 ~ 5/2 | 중간 보고서 작성 |
| 5/3 ~ 5/9 | 인공지능을 사용하여 적용할 퀀트 알고리즘 연구 및 테스트 |
| 5/10 ~ 5/16 | 인공지능을 사용하여 적용할 퀀트 알고리즘 연구 밀 테스트 |
| 5/17 ~ 5/23 | 성능이 잘 나오는 알고리즘 선별 및 최종 코드 작성 시작 |
| 5/24 ~ 5/30 | 개발 마무리 |
| 5/31 ~ 6/6 | 최종 보고서 작성 |

**5. 결론 및 기대효과**

재무제표 기반으로 한 투자가 가장 합리적이고 좋은 방법이지만 모두가 이를 공부하는 것을 원하진 않을 것이다. 실제로 비전문가인 일반 투자자들은 재무제표를 읽고 분석하는 것이 어렵고 오래걸린다고 생각한다. 그래서 차트 분석 기법 같은 훨씬 쉽지만 타당성이 많이 떨어지는 분석 기술을 사용하는 투자자들이 많다. 하지만 이 마저도 하지 않는 투자자들이 대부분이다. 이 연구에서 개발한 주가 예측 모델은 차트 분석 기법보다 타당성이 높고, 효율적인 주식 선별을 제공할 수 있다.

**6. 참고문헌**

[1] Tensorflow, [**https://www.tensorflow.org/**](https://www.tensorflow.org/)

[2] 송유정, 이종우. (2017). 텐서플로우를 이용한 주가 변동 예측 딥러닝 모델 설계 및 개발. 한국정보과학회 학술발표논문집, (), 799-801.

[3] 이지훈. (2017). 딥러닝을 이용한 주가 예측 모델. 국내석사학위논문 숭실대학교 대학원, 서울