

## 『제1회 KRX 금융 빅데이터 활용 아이디어 경진대회』 제안서

1 아이디어 명	분 단위 데이터분석을 통한 종목별 당일주가 변동 예측 -수급 및 정보비대칭성을 중심으로-	2 구분	개인 <input type="checkbox"/>	팀 <input checked="" type="checkbox"/>
3 팀 명	Micro-A.N.T	4 팀 구분 (2 '팀' 체크 시)	일반 <input checked="" type="checkbox"/>	법인 <input type="checkbox"/>
5 법인 명 (4 '법인' 체크 시)		6 법인등록번호 (4 '법인' 체크 시)		

### 아이디어 개요

- 1) KRX의 일중 데이터(오전: 09:00 or 09:11 ~ 11:00)를 활용하여 KOSPI 구성 종목들의 당일 주가 변동 예측  
- 일중 장 초반의 수급 데이터를 활용하여 KOSPI 구성 종목들의 당일 증가 상승 여부를 예측
- 2) 다양한 머신러닝 및 딥러닝 모델에 기반한 예측 모형  
- 일중데이터를 가공한 양질의 투자정보를 이용하여 개별종목들의 당일 주가 변동을 예측하는데 있어서 다양한 머신러닝과 딥러닝 모델을 활용
- 3) 거래소의 미시데이터를 활용한 투자 정보를 제공하고, 시장 참여자 사이에 정보 비대칭 해소 기대  
- 투자 의사결정에 있어서 예측 모형의 결과를 유용한 정보로 활용함에 따라 시장 참여자 사이에 정보 비대칭 해소 기대

### 아이디어 제안 배경 및 목적

#### 1) Market Microstructure analysis

시장 미시구조(Market Microstructure)는 시장에서 자본 거래 및 재화 교환이 어떻게 발생하는지에 대해 미시적인 관점으로 연구하는 금융의 한 분야로 거래 메커니즘과 가격 형성 과정에 대한 연구를 주로 시행함. 컴퓨팅 파워의 업그레이드는 금융 시장 내 데이터 기반의 의사결정과 알고리즘 적용의 보편화로 이어져, 시장 미시적 데이터에 대한 수요는 대폭 증가하는 상황이며, 이러한 상황 속에서 한국거래소가 보유한 미시적 데이터를 활용하여 특정 거래 메커니즘 및 가격 형성 및 가격 발견, 거래 타이밍 등에 대한 정보를 시장 참여자에게 제공 가능.

#### 2) 시장 참여자 간 정보 비대칭 완화

KOSPI 지수 및 개별종목 등락에 영향력이 큰 외국인 투자자의 수급 데이터를 분석함으로써 시장 참여자들에게 유의미한 추가적 정보 제공이 가능. 정보 제공을 통해 한국거래소는 시장 참여자 사이에 정보 비대칭 문제를 완화할 수 있을 뿐만 아니라, 금융시장의 효율성 저하와 소비자 보호 관련해서 발생할 수 있는 문제들을 사전 예방할 수 있을 것이라 기대됨. 이는 시장참여자에 대한 서비스 혁신을 일궈낼 수 있으며, 선진형 증권시장 구축에 있어 성공적인 시도가 될 것임.



## 아이디어 제안 배경 및 목적

### 3) 거래당일 거시 경제상황 영향 고려

금융 자산의 가격에 이미 시장에 공개된 정보는 반영되었기 때문에 공개된 정보로는 초과수익을 기록할 수 없다는 준강형 효율적 시장가설에 입각했을 때, 장 초반에 형성된 지수 및 개별 종목의 주가는 등락에 큰 영향을 미치는 미 증시, 중국 증시, 환율 등 거시 경제상황에 대한 정보가 충분히 반영되었다고 보고 장 초반의 데이터를 사용함으로써 거래당일 거시 경제 상황의 영향을 고려. 장 초반에 형성되는 영향을 시간대별로 09시에 전일 미국 증시, 10시 30분 중국 증시 개장을 비롯하여 당일 거시경제 움직임을 반영하기 위한 11시까지의 데이터를 분석함.

### 4) 당일 투자자 의사 반영

다수 연구에 따르면, 포트폴리오 재조정자와 정보를 가진 투자자 간의 상호작용으로 인해 주식시장 초반에 거래량과 유동성 및 가격 변동성 등이 크게 기록되는 패턴을 띰. 이러한 사유로 시장 참여자들은 장 초반의 시장 상황을 고려하여 당일 투자 의사결정을 행하게 되는데, 이처럼 시장 참여자들의 의사가 반영되어 양질의 정보를 지닌 장 초반의 수급 데이터에 주목하고, 장 후반까지의 흐름을 예측하고자 함.

## 활용 데이터

### 1) 주식 - [유가증권]일별 시세정보

- 시가총액

### 2) 주식 - [유가증권]일중 시세정보

- (파생)거래금액 / 시가총액(%)
- 종목별 거래량 기준 매수매도불균형
- 주문수량기준깊이
- (파생)HS 역선택비용 / 호가단위
- (파생)일중강도 / 거래량
- (파생)평균호가스프레드/호가단위

### 3) 주식 - [유가증권]투자자별 시세정보

- 투자자 코드(9000)
- (파생)외국인 순매수(도)금액
- (파생)외국인 체결강도(%)

## 제품·서비스·콘텐츠 세부내용

1) 1분 데이터(09:00~11:00)를 활용하여 종목별 당일 시가 대비 종가 변동예측

- 전체 컬럼 (총 8개)

- 수급 데이터 컬럼 (6개)

- 거래대금/시가총액(%)

- 해당 종목의 거래금액을 당일 시가총액으로 나눈 값

- 주식의 시가총액에 비하여 얼마나 거래가 많이 이루어졌는지 확인할 수 있는 지표

- 외국인 순매수(도)금액

- 해당 종목에 대해 외국인 투자자가 거래한 매수대금에서 매도대금을 뺀 값

- 외국인 체결강도(%)

- 투자자별 거래실적 데이터에서 외국인 투자자의 매수거래량 / 매도거래량 x 100으로 계산한 값

- 해당 종목에 대한 외국인 투자자의 매수세/매도세의 강도를 나타냄

- 종목별 거래량 기준 매수매도불균형

- 종목별 매수주도거래량을 매도주도거래량으로 나눈 값

- 매수(도)자 호가가 제출되어 거래가 성사된 것일 경우 매수(도)자주도거래로 구분

- 일종강도 / 거래량

-  $(2 \times \text{종가} - \text{고가} - \text{저가}) / (\text{고가} - \text{저가})$

- 가격의 절대적인 크기와 상관없이 종가가 고가와 저가의 평균에 비해 어느 정도 위치에 있는지를 나타내는 지표

- 주문수량기준깊이

- 최우선 매도호가와 최우선 매수호가의 잔량 합의 평균



## 제품·서비스·콘텐츠 세부내용

### - 정보비대칭성 컬럼 (2개)

#### • HS역선택 비용/호가단위

- 평균 유효스프레드에서 평균 실현스프레드를 차감한 값으로 매수자와 매도자 사이의 정보 비대칭 정도를 나타냄
- 기업 별 형성된 주가마다 호가를 할 수 있는 최소단위의 가격이 다르기 때문에 HS역선택 비용을 호가 단위로 나눠줌

#### • 평균호가 스프레드 /호가단위

- 호가 스프레드(최우선 매도&매수호가의 차이)의 평균
- 기업 별 형성된 주가마다 호가를 할 수 있는 최소단위의 가격이 다르기 때문에 평균호가스프레드를 호가 단위로 나눠줌

### 2)변수별 검정통계량 확인

- 웰치의 t 검정(Welch's t test)

### 3)변수별 상관성 확인

### 4) 예측 모델에서는 머신러닝, 딥러닝 모델 활용(train:test는 7:3,8:2로 검증)한 변동 예측

- 머신러닝 모델 : 로지스틱회귀분석, Decision Tree, RandomForest, Xgboost, Catboost
- 딥러닝 모델 : DNN

### 5) 평가지표

- Accuracy을 통한 모델 성능 평가

## 기존 상품·서비스·콘텐츠와의 차별성

### 1) 정보 적시성

시장 정보는 거래 즉시 신속하게 확인할 수 있어야 하며 투자판단에 있어서 오류가 생기지 않도록 최신의 정보가 제공되어야 하는데, 개인 투자자 같이 수급과 관련하여 지연된 정보를 활용하고 있던 일부 시장 참여자들에게 한국거래소의 분별 수급 데이터를 분석한 후 도출된 정보를 즉각 제공함으로써 적시성이 고려된 정보에 기반한 투자의사 결정이 가능.

### 2) 정보 비대칭 지표 고려

정보 비대칭 현상에서 정보 우위를 가지는 시장 참여자가 발생시키는 신호가 비정보 거래자의 매수·매도 거래량에 영향을 미친다는 신호이론의 관점에서 볼 때, 정보 비대칭 자료는 개별 종목들의 당일 주가 변동에 중요 지표로 작용함. 따라서 정보의 비대칭 정도를 대용할 수 있는 한국거래소의 평균 호가스프레드, HS 역선택 비용 데이터를 변수로 예측 모델에 활용하고자 함.

### 3) 일중 매매 관련 투자정보 전달

한국거래소의 일중 데이터를 활용하여 시가 대비 종가의 상승 여부를 예측함으로써, 일중 거래를 통한 초과수익 기록 및 변동성 관리에 유용한 정보 전달이 가능.

### 시장성 및 사업화 가능성

- 1) 웹사이트 또는 앱으로 시장 미시 데이터와 분석 내용을 즉각 제공하여 정보 지연율은 낮추고, 시장 참여자들의 정보 접근성과 용이성은 높임으로써 시장 건전성 제고
- 2) 장 초반 데이터 분석을 통해 개별종목의 주가 변동을 예측하여, 시장 참여자에게 추후 매매시점 및 매매전략, 헷징 전략 수립 등에 있어서 유용한 투자 정보 제공 가능
- 3) 기존에 쉽게 구할 수 있는 일별데이터보다는 한국거래소에서 제공하는 미시 데이터(일중데이터)를 활용하여 투자자들에게 어필 및 일중데이터 사용 유도
- 4) 최신 인공지능 기술을 활용한 시장수익률 대비 높은 수익 창출을 통한 사업 다각화

### 기타(아이디어 강점 등 자유 기재)

- 1) 일중 데이터를 이용하여 당일 주가 변동을 예측함으로써, 한국거래소가 보유한 시장 미시 데이터에 대한 수요 증가를 기대할 수 있으며 이는 정보 사업 확대로도 이어질 수 있음
- 2) 한국 거래소의 유용한 투자 참고 정보를 제공을 통해, 시장 참여자 간의 정보 비대칭을 해소하여 시장의 공정한 가격 형성과 거래의 안정성 및 효율성을 도모하고자 하는 거래소의 역할 제고 가능
- 3) 제공된 데이터 외에 한국거래소가 보유한 더 많은 시장 미시 데이터를 이용한다면, 머신러닝 및 딥러닝 예측 모델의 성능 개선 시도가 가능하리라 판단됨