

A Machine Learning-Based Analysis of Short-Term Surges and Listing Premiums for Newly Listed Altcoins on Korean Exchanges Using Offshore Prices and Social Signals

Eun Hong Park*, Yeong-In Lee*, Ha Young Kim**

*Student, Graduate School of Information, Yonsei University, Seoul, Korea

**Associate Professor, Graduate School of Information, Yonsei University, Seoul, Korea

[Abstract]

This study investigates the predictability of post-listing price surges and cross-exchange premiums on the Upbit market. This Study is to demonstrate the empirical validity of an integrated analytical framework that combines offshore market signals with online social signals. By integrating pre-listing Binance data with quantified social sentiment from online communities, we evaluated the ex-ante identification of market anomalies, specifically short-term surges and the formation of the "Kimchi Premium". Empirical results indicate that while the MLP model is superior for detecting price surges, a Logistic Regression model augmented with social factors is most effective for identifying compound events of surges and premiums simultaneously. These findings suggest that combining global price dynamics and investor sentiment, provides a robust basis for screening structural inefficiencies in segmented cryptocurrency markets.

▶ **Key words:** Machine Learning, Cryptocurrency, New Listing, Kimchi Premium, Social Sentiment

[요 약]

최근 가상자산 시장에서 신규 상장은 유동성 공급과 가격 재평가라는 순기능과 함께 단기 과열 및 거래소 간 가격 괴리 확대라는 위험을 동시에 수반한다. 국내 원화 시장은 규제 및 자본 이동 제약 등으로 인해 해외 거래소 대비 가격 괴리가 잦아, 상장 직후 급등·프리미엄 현상을 사전에 식별하는 것이 투자자 보호와 시장 효율성 제고 측면에서 중요한 과제로 대두되고 있다. 본 연구는 최근 5년간 업비트 원화 시장에 신규 상장된 종목 중 바이낸스에 선상장된 코인을 대상으로, 해외 거래소의 선행 시세 정보와 온라인 커뮤니티 기반 소셜 신호를 입력 변수로 구성하고, 상장 직후 단기 급등 및 프리미엄 발생 여부를 사전 식별 가능성을 실증적으로 평가하였다. 상장 후 단기 급등 여부와 프리미엄 발생 여부를 이진 분류 문제로 정의하고, 세 가지 분류 모형의 성능을 비교·분석하였다. 분석 결과, 단기 급등 예측에서는 MLP 모형이 가장 높은 AUC와 정밀도를 보였으며 단기 급등과 프리미엄이 동반되는 예측에서는 소셜 신호를 결합한 로지스틱 회귀 모형이 가장 우수한 성능을 보였다. 이러한 결과는 단기 급등 가능성과 프리미엄 위험을 조기 스크리닝하는 데 활용될 수 있는 실증적 근거를 제공하며, 상장 이벤트 기반 시장 모니터링 체계 설계에 시사점을 제공한다.

▶ **주제어:** 기계학습, 암호화폐, 신규 상장, 김치 프리미엄, 소셜 센티먼트

- First Author: Eun Hong Park, Yeong-In Lee, Corresponding Author: Ha Young Kim
- *Eun Hong Park (pehhh03@yonsei.ac.kr), Graduate School of Information, Yonsei University
- *Yeong-In Lee (zeroin@yonsei.ac.kr), Graduate School of Information, Yonsei University
- **Ha Young Kim (hayoung.kim@yonsei.ac.kr), Graduate School of Information, Yonsei University
- Received: 2025. 12. 31, Revised: 2026. 01. 14, Accepted: 2026. 01. 23.

I. Introduction

최근 암호화폐 시장이 빠르게 성장하면서 신규 자산의 거래소 상장은 투자자들의 이목이 집중되는 핵심 이벤트로 자리 잡았다. 전통 금융시장의 IPO와 달리 암호화폐 시장은 절차가 상대적으로 유연하고 국가별 규제 환경과 거래소 정책에 따라 상장 시점 및 종목 구성이 상이하다는 특징을 갖는다. 해외 대형 거래소에서 이미 거래되고 있는 자산이 국내 원화(KRW) 시장에 추가 상장될 때 새로운 투자 수요가 유입되며, 상장일 전후로 단기간에 가격이 크게 상승하는 이른바 '상장 뱀(Listing Pump)' 현상이 여러 종목에서 관찰되어 왔다[1,2]. 이러한 현상은 신규 시장 진입에 따른 유동성 프리미엄이 가격에 반영되는 일반적인 과정일 수 있으나, 정보 우위에 있는 소수와 추격 매수에 나서는 개인 투자자 간의 비대칭이 심화될 경우 단기 과열과 급락이 반복되는 가격 왜곡을 초래할 수 있다[3,4].

기존의 암호화폐 가격 예측 연구들은 주로 비트코인(BTC)나 이더리움(ETH)과 같은 시가총액 상위 자산을 대상으로 LSTM, GRU 등의 딥러닝 모델을 적용하여 향후 가격 수준 또는 수익률 방향을 예측하는 데 초점을 맞추어 왔다[5-7]. 이러한 접근은 일반적인 시장에서의 추세 및 변동성 예측에는 유용할 수 있지만, 개별 알트코인의 신규 상장과 같은 특수한 상황에서 발생하는 초과수익률을 설명하거나 상장 직후의 급등 여부를 사전에 탐지하는 데에는 한계가 있다. 더욱이 한국 시장은 자본 이동의 제약과 규제 환경으로 인해 글로벌 시세와 괴리가 발생하는 '김치 프리미엄(Kimchi Premium)'이라는 독특한 구조적 특징을 가진다[8-10]. 해외 거래소에서의 선행 가격 흐름과 국내 진입 시점의 가격 괴리율은 상장 초기 수익률에 영향을 미칠 수 있는 잠재적 요인임에도 불구하고, 이를 체계적으로 측정하고 예측 모형에 반영하여 상장 직후 급등 여부를 사전에 탐지하는 연구는 아직 미흡한 실정이다[11].

본 연구의 핵심은 해외 거래소의 선행 시세 데이터와 온라인 커뮤니티 기반 소셜 신호를 결합한 통합적 분석 프레임워크가, 국내 암호화폐 시장의 특수성인 상장 직후 단기 급등 및 거래소 간 프리미엄 현상을 사전에 식별할 수 있는지를 실증적으로 검증하는 데 있다. 본 연구의 차별점은 다음과 같다.

첫째, 단일 시장의 시세만을 분석하는 기존 연구의 한계를 넘어 해외 선행 거래소 가격과 국내 상장 시점 기준 가격을 비교함으로써, 상장 시점의 프리미엄을 핵심 변수로 도출하였다.

둘째, 정량적 시세 정보에 온라인 커뮤니티의 티커 언급

량 및 증가율 등을 결합한 데이터셋을 구축하고, 시장 신호만 사용한 경우 대비 시장과 소셜 신호를 결합한 경우가 상장 직후의 수급 불균형 및 가격 왜곡 현상을 더 높은 예측 성능으로 식별할 수 있음을 비교 실험을 통해 확인하였다.

셋째, 본 연구는 기계학습 모델을 활용하여 상장 이벤트 관련 이상 급등 및 프리미엄 발생 가능성을 조기 스크리닝할 수 있는 실증적 근거를 제공한다.

II. Related work

2.1. Structural Characteristics of the Domestic Cryptocurrency Market

국내 가상자산 시장은 업비트(Upbit), 빗썸(Bithumb) 등 원화(KRW) 시장 중심의 거래소를 기반으로 생태계가 형성되어 있다. 최근 기사에 따르면, 한국 암호화폐 시장의 현물 거래대금은 글로벌 전체의 약 9% 수준으로 단일 국가 기준으로 미국·일본과 함께 상위권을 차지하고 있다[12]. 국내 시장에서는 국내 주요 거래소 현물 거래대금 가운데 업비트가 약 70% 내외의 점유율을 유지하고 있다[13]. 2021년 기준 바이낸스(Binance)와 같은 대형 거래소는 500개 이상 수준의 토큰을 지원하는 반면, 국내 주요 거래소는 규제와 엄격한 심사 기준으로 인해 약 200개 수준의 상대적으로 제한된 수의 종목만을 지원한다[14]. 상장 정책의 비대칭성은 다수의 알트코인이 해외 거래소에서 먼저 가격을 형성한 후, 일정 시차를 두고 국내 원화 시장에 신규 상장되는 구조적 환경을 조성한다.

주식 시장의 IPO와 유사하게, 가상자산 시장에서도 신규 상장 시점 전후로 양(+)의 수익률이 나타난다는 연구들이 존재한다[15]. 코인이 대형 거래소에 신규 상장되거나 타 거래소에 교차 상장될 때 유동성 확대와 투자자들의 투기적 수요가 결합되어 단기간에 가격이 급등하는 경우가 빈번하다. 국내 시장의 경우, 이러한 상장 효과가 해외 거래소와의 가격 괴리와 맞물릴 경우 상장 직후 단기 가격 변동성이 더욱 확대될 가능성이 높다.

한국 시장은 자본 이동의 제약과 독자적인 규제 환경으로 인해 글로벌 시세와 가격이 괴리되는 김치 프리미엄 현상이 지속적으로 발생한다. 기존 연구들은 이를 시장 비효율성이라기보다, 차익거래 비용과 투자자 심리에 의해 결정되는 구조적 요인이라고 설명하였다[8-10]. 양철원(2019)은 비트코인의 국내외 가격 차이를 분석한 연구에서 2015년 7월부터 2018년 12월까지의 비트코인 가격을 바탕으로 한국 거래소와 해외 거래소 간 평균 약 223달러,

최대 8,358달러에 달하는 김치 프리미엄이 지속적으로 존재했음을 증명하였다[16]. 이후 연구들 또한 원화 시장과 달러화 시장 간 가격 불균형 및 차익거래 수익성을 분석하며, 한국 시장이 글로벌 가격과 구조적으로 괴리될 수 있는 환경에 놓여 있음을 입증하였다[17].

이러한 김치 프리미엄은 비트코인뿐만 아니라 일부 알트코인에서도 국지적으로 관측되며, 국내 거래소 신규 상장 직후에는 투기적 수요가 급증하는 반면 제한된 유동성 공급으로 인해 단기적인 가격 괴리가 더욱 확대되는 경우도 있다. 그러나 기존 연구들은 주로 거시적인 시장 전체의 프리미엄 추이에 집중해 왔으며, 개별 종목의 상장 프리미엄이 상장 직후의 단기 수익률 및 급등 여부와 어떤 관계를 갖는지에 대한 분석은 미흡하다.

따라서 본 연구는 선행 연구의 개념을 확장하여, 상장 시점 개별 프리미엄을 핵심 설명 변수로 설정한다. 이를 통해 가격 괴리가 상장 이벤트 이후의 급등 현상에 미치는 영향을 실증적으로 검증하고자 한다.

2.2. Cryptocurrency Price Prediction

암호화폐의 높은 변동성과 비선형성을 예측하기 위해 초기 연구에서는 ARIMA, GARCH 계열과 같은 계량경제학적 시계열 모형이 주로 활용되었다[18]. 해당 모형들은 과거 수익률의 자기상관 구조와 조건부 이분산성을 이용해 비트코인의 변동성을 추정하고, 추가 지수·변동성 지수(VIX) 등 외생 변수와의 관계를 분석하는 데 유용하게 사용되었다[19]. 그러나 ARIMA, GARCH 계열 모형은 선형 구조가 전제되므로 복잡한 비선형 패턴과 급격한 구조 변화가 빈번하게 발생하는 암호화폐 시계열의 특성을 충분히 반영하기 어렵다.

이러한 한계를 보완하기 위해 최근에는 LSTM(Long Short-Term Memory), GRU(Gated Recurrent Unit)와 같은 인공신경망 기반 딥러닝 모델이 암호화폐 가격 예측에 활발히 활용되고 있다[20]. 유기섭(2025)은 비트코인 시장을 대상으로 GRU, LSTM, Random Forest 세 가지 모형을 비교한 연구에서 GRU 모형이 단기 수익률 예측에서 가장 낮은 예측 오차를 기록하여 상대적으로 우수한 성능을 보였다고 설명하였다[21]. 김태욱·손정우(2025) 역시 비트코인 가격을 예측하는 딥러닝 모형에 한국 채권지수 등 안전자산 관련 변수를 추가하여 Granger 인과성 검정을 통해 선택한 변수들을 LSTM 구조에 반영함으로써 기존 단일 시계열 기반 모형보다 예측 성능이 개선됨을 보였다[22]. 이들 연구는 LSTM·GRU와 같은 딥러닝 모델이 시계열 데이터의 장기 의존성을 학습하는 데 효과적이고, 비

트코인·이더리움과 같이 거래 이력이 충분히 축적된 자산의 가격 추세를 예측하는 데 적합하다는 것을 증명하였다.

그러나 시가총액이 낮고 거래량과 유동성이 제한적인 알트코인의 경우, 딥러닝 모형이 학습 데이터의 노이즈에 과적합되거나 학습된 결과를 해석하기 어렵다는 한계가 존재한다. 김선웅(2023)은 주요 알트코인의 가격과 기술 지표를 활용해 Random Forest, XGBoost, LSTM 등의 성능을 비교한 결과, 변동성이 큰 구간이나 관측치 수가 상대적으로 적은 종목에서는 앙상블 모형이 딥러닝 모형보다 안정적인 예측 성능을 보이는 경우가 많았으며, 특히 변수 중요도를 통해 예측에 기여한 요인을 직관적으로 해석할 수 있다고 언급하였다[23].

가격 정보만으로는 시장 참여자의 기대와 공포, 투기적 심리와 같은 비정형 정보를 충분히 포착하기 어렵다는 한계를 극복하기 위하여 뉴스 기사나 검색량, 소셜 미디어 등의 텍스트 데이터를 결합한 연구도 진행되었다[24]. 원종관(2021)은 국내와 해외 비트코인 시장을 비교하면서 가격 및 거래량에 뉴스 텍스트에서 추출한 용어와 감성 정보를 결합한 separated RNN 구조를 제안하였고, 텍스트 정보를 포함한 모형이 가격 정보만을 사용한 모형보다 예측 성능이 개선됨을 보였다[25]. 류진환(2025)은 가격, 기술 지표, 온라인 검색량 등을 동시에 입력하는 멀티모달 딥러닝 구조를 통해 일별 가격 등락(up/down) 방향을 예측하는 연구를 수행하였으며, 텍스트·검색 기반 지표가 등락 예측에 유의미한 설명력을 제공함을 증명하였다[26].

이와 같은 선행 연구들은 가상자산 가격 예측에 있어 머신러닝 모형 및 딥러닝 모형이 전통적인 선형 시계열 모형보다 우수한 성능을 보일 수 있다는 점과 뉴스나 커뮤니티, 검색량 등 텍스트 기반 커뮤니티 신호가 가격 변동의 선행 지표로 사용될 수 있다는 점을 시사한다. 신규 상장과 같이 투자자 기대 심리가 가격 형성에 큰 비중을 차지하는 상황에서는, 가격·거래량과 더불어 커뮤니티 언급 신호를 함께 고려하는 멀티모달 접근이 단기 급등 여부를 예측하는 데 유효한 방법이 될 수 있다.

III. Dataset Construction

본 연구에서 사용한 데이터 구축 절차는 Fig 1과 같다. 해외 거래소와 국내 거래소에서 1분 단위 OHLCV(Open, High, Low, Close, Volume) 시계열 데이터를 수집하여 코인 단위 티커 데이터셋을 구성하고, 동시에 온라인 커뮤니티로부터 소셜 텍스트 데이터를 크롤링하여 티커별 언

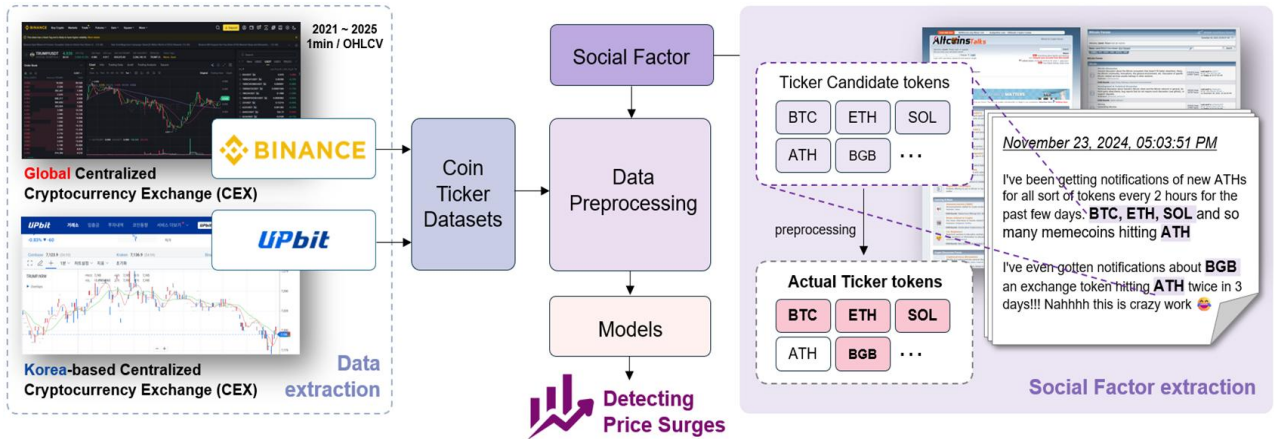


Fig. 1. Overall data pipeline from exchanges and social forums to prediction models

급량을 추출하였다.

위의 두 가지 정보는 전처리 단계를 거쳐 가격과 거래량 기반의 특성과 연도별 소셜 팩터로 변환되며, 이후 4장에서 설명하는 분류 모형의 입력 변수로 사용된다.

3.1. Sampling Procedure

본 연구는 2021년 1월부터 2025년 12월까지 업비트 (Upbit)의 원화(KRW) 시장에 신규 상장된 종목을 모집단으로 한다. 이 중 글로벌 시세 비교를 위해 바이낸스 (Binance)에 이미 상장되어 1분봉 시세 데이터가 존재하는 종목을 1차적으로 선별하였다. 수집 결과 업비트 원화 시장 상장 종목은 총 144개였으며, 이 가운데 바이낸스 테더(USDT) 시장 데이터를 확보할 수 있는 교집합 종목 107개를 최종 분석 대상으로 확정하였다.

상장 시점 t_0 는 업비트 API 상에서 해당 코인의 첫 거래가 관측되는 분 단위 타임스탬프로 정의하였으며, 이를 기준으로 업비트 데이터 $[t_0, t_0+7d]$ 와 바이낸스 데이터 $[t_0-7d, t_0+7d]$ 구간의 1분봉 OHLCV를 확보하였다.

3.2. FX Conversion and Currency Unification

가격 및 거래량 데이터는 각 거래소의 공식 API를 통해 1분 단위의 OHLCV 데이터를 수집하였다. 서로 다른 통화 단위를 통일하기 위해 야후 파이낸스에서 제공하는 2021년부터 2025년까지의 원/달러(USD/KRW) 환율 데이터를 활용하였다. 바이낸스의 USDT 가격은 달러(USD)와 1:1로 페깅된다고 가정하고, 각 분봉 시점 t 에 대응하는 일별 환율 FX_t 를 적용하여 원화 환산 가격 ($P^{Bin, KRW}$)을 산출하였다.

$$P^{Bin, KRW}(t) = P^{Bin, USD}(t) \times FX(d(t))$$

3.3. Price Premium

환율 변환이 완료된 바이낸스 가격과 업비트 가격 간의 차이를 정량화하기 위하여 거래소 간 가격 괴리율(Gap)을 다음과 같이 정의하였다.

$$Gap(t) = \frac{P^{Up}(t) - P^{Bin, KRW}(t)}{P^{Bin, KRW}(t)}$$

시장 시점의 괴리는 두 거래소 간의 유동성 차이, 정보 반영의 시차, 그리고 국내 시장 특유의 일시적 수급 불균형이 반영된 지표로 해석될 수 있다. 본 연구에서는 상장 직전 및 직후에 형성되는 이 가격 괴리율을 주요한 파생 변수로 추출하여, 해당 변수가 상장 이후의 가격 급등 현상과 어떠한 상관관계를 갖는지 분석한다.

3.4. Social Factor

정량적 시장 데이터 외에 투자자들의 관심도를 반영하기 위해 암호화폐 전문 포럼인 Bitcointalk와 Altcointalk에서 2021년부터 2025년까지의 게시글 텍스트 데이터를 크롤링하였다. 두 포럼에서 신규 코인-알트코인 관련 논의가 집중되는 게시판만을 대상으로 게시글 제목과 본문, 게시 시각 정보를 수집하였다. 수집된 텍스트에서 영문 대문자 2-10자로 이루어진 토큰을 후보 티커로 추출한 뒤, 일반 단어나 약어 등을 제거하고 실제 거래소에 상장된 코인 목록과 매칭하여 최종 티커별 언급 횟수를 산출하였다. 동일 월-동일 티커에 대해 게시판 간 언급 빈도는 합산하였으며, 게시글 내 등장 여부를 기준으로 1회 이상 언급 시 1회로 카운트하였다.

분석 대상인 107개 티커에 대한 월별 언급량 $d_{i,m}$ 을 추출한 후, 월별 전체 언급량 변동을 보정하기 위해 정규화된 소셜 팩터를 구성하였다. 구체적으로 월 m 의 전체 언급량과 티커 i 의 언급 비중 $S_{i,m}$ 를 다음과 같이 정의한다.

$$S_{i,m} = \frac{d_{i,m}}{D_m}, D_m = \sum_i d_{i,m}$$

이후 월별 분포에서 티커 간 상대적 관심도를 비교하기 위해 $S_{i,m}$ 에 대하여 z-score 표준화를 적용하여 소셜 팩터 $z_{i,m}$ 을 산출하였다.

최종적으로 각 코인의 상장월을 t_0 라 할 때, 소셜 팩터에는 상장 시점 이후 정보가 포함되지 않도록 t_0 의 전월부터 직전 6개월($t_0-6 \sim t_0-1$)의 $z_{i,m}$ 만을 사용하였다. 또한 최근 관측치일수록 더 큰 가중치를 부여하기 위해 EWMA (Exponentially Weighted Moving Average)로 요약하여 소셜 팩터 $SF_i(t_0)$ 를 구성하였다.

$$SF_i(t_0) = \sum_{k=1}^6 w_k z_{i,t_0-k}$$

여기서 w_k 는 t_0 에 가까운 관측치일수록 더 큰 값을 갖도록 정의되는 EWMA 가중치로 EWMA는 최근 관측치일수록 더 큰 비중을 부여하되 가중치가 지수적으로 감소하는 이동평균/평활화 방법이다. 상장 직전 단기적인 관심 급증과 지속적 관심 추이를 동시에 반영할 수 있다. 예를 들어 상장월이 2025년 5월인 경우, 2024년 11월부터 2025년 4월까지의 월별 z-score를 EWMA로 요약하여 소셜 팩터를 구성한다.

3.5 Data Splitting

분류 모델의 일반화 성능을 평가하기 위해 107개 상장 종목을 티커 단위로 분할하였다. train = 74개, val = 16개, test = 17개로 티커를 할당하였다. 각 티커는 하나의 상장 이벤트로 간주되며, 하나의 티커가 여러 세트에 동시 등장하지 않도록 분리하여 데이터 누수를 방지하였다. 분할 시 세 라벨의 비율이 극단적으로 치우치지 않도록, 라벨 분포를 고려한 계층적 분할에 준하는 방식을 차용하였다.

IV. Experiments

4.1. Experimental setup

본 연구에서는 국내 거래소 상장 이벤트를 대상으로 (1) 상장 직후 단기 급등 여부와 (2) 상장 시점 프리미엄을 예측하는 두 가지 이진 분류를 정의하였다. 3.1장에서 정의한 바와 같이, 업비트 상장 기준 시점을 t_0 라 두고, 상장 후 H 일 이내 구간 $[t_0, t_0+H]$ 에서의 분봉 종가가 $P^{Up}(t)$ 일 때, 상장 후 최대 수익률은 다음과 같다.

$$R_{\max}(H) = \max_{t \in [t_0, t_0+H]} \frac{P^{Up}(t) - P^{Up}(t_0)}{P^{Up}(t_0)}$$

임계값 θ 를 기준으로, 단기 급등 여부 라벨 m_{up} 은 다음과 같다.

$$y_{up} = \begin{cases} 1, & R_{\max}(H) \geq \theta \\ 0, & otherwise \end{cases}$$

상장 후 H 일 이내 최대 수익률이 θ 이상인 경우를 1, 그렇지 않은 경우를 0으로 라벨링한다. 예를 들어 $H = 3$ 일, $\theta = 0.3$ 으로 설정하면 상장 후 3일 이내 최대 수익률이 30% 이상인 경우를 1로 간주한다.

거래소 간 가격 괴리율 $Gap(t)$ 은 3.3장에서와 동일하다. 상장 시점 이후 H_{gap} 일 이내 구간에서의 괴리율 최대값을 다음과 같이 정의한다.

$$G_{\max}(H_{gap}) = \max_{t \in [t_0, t_0+H_{gap}]} Gap(t)$$

상장 시점 프리미엄 라벨 y_{gap} 은 급등 여부 라벨과 동일하다.

$$m_{gap} = \begin{cases} 1, & G_{\max}(H_{gap}) \geq \delta \\ 0, & otherwise \end{cases}$$

여기서 δ 는 상장 직후 국내 - 해외 가격 괴리가 어느 수준 이상일 때 프리미엄으로 볼 것인가를 결정하는 임계값이다. 마지막으로 단기 급등과 프리미엄이 모두 충족되는 경우를 포착하기 위해 결합 라벨을 사용하여 세 가지 이진 라벨을 구성한다. 각 라벨에 대해 동일한 입력 변수와 분류 모델을 적용하여 예측 성능을 비교 및 분석하며 변수에 대한 자세한 설명은 아래의 Table 1을 참고하면 된다.

4.2. Model Selection

4.2.1 Logistic Regression (Logistic Regression)

Logistic Regression는 이진 분류 문제에서 가장 기초적이면서도 널리 사용되는 선형 모델이다. 입력 특성들의 선형 결합을 시그모이드(Sigmoid) 함수에 통과시켜, 특정 클래스에 속할 확률을 0과 1 사이의 값으로 출력한다. 본 연구에서는 비선형 모델들과의 성능을 비교하기 위한 선형 기준 모델로 Logistic Regression를 포함하였다.

4.2.2 LightGBM (LGBM)

LightGBM은 트리 기반 앙상블 모델로, 대용량 데이터에서도 빠른 학습 속도와 우수한 예측 성능을 보인다. 특히 잎 중심(Leaf-wise) 트리 분할 방식을 사용하여 깊이가 깊은 트리 구조를 통해 변수 간의 복잡한 비선형 상호작용을 효과적으로 포착한다. 본 연구에서는 정형 데이터 분류

에 강점이 있는 LightGBM을 활용하여 단기 급등 여부와 프리미엄 여부를 예측하였다.

4.2.3 Multi-Layer Perceptron (MLP)

MLP는 입력층(Input Layer), 은닉층(Hidden Layer), 출력층(Output Layer)으로 구성된 가장 기본적인 형태의 인공신경망이다. 각 뉴런의 연결 강도(Weight)와 비선형 활성화 함수(Activation Function)를 통해 입력 변수들 사이의 복잡하고 추상적인 관계를 학습할 수 있다. 본 연구에서는 상장 전·후 구간의 요약 통계량과 프리미엄 지표, 소셜 팩터 등을 고정된 길이의 벡터로 변환하여 입력하였으며, 이를 통해 전통적인 트리 기반 모델 대비 신경망 기반 모델의 유효성을 검증하고자 하였다.

4.3. Evaluation Metrics

본 연구에서는 상장 종목의 단기 급등 및 프리미엄 여부를 예측하는 분류 모형의 성능을 평가하기 위해 다음의 지표들을 사용하였다. 이상의 지표들은 세 가지 라벨에 대하여 동일하게 산출하였으며, 모형 간 성능 비교와 변수 구성에 따른 성능 차이 분석에 활용하였다.

4.3.1 Area Under the ROC Curve (AUC)

AUC는 임계값을 여러 수준으로 바꾸어가면서, 1을 얼마나 잘 잡아내는지 0을 얼마나 잘 걸러내는지를 동시에 고려한 지표이다. 값이 0.5면 무작위 추정과 유사한 수준, 1에 가까울수록 급등과 비급등을 전반적으로 잘 구분하는 평가 지표이다. 검증 데이터셋의 AUC를 기준으로 모형과 하이퍼파라미터를 비교하고, 최종 성능은 테스트 데이터셋의 AUC 값을 사용하였다.

4.3.2 F1-Score

실제 투자 환경에서는 하나의 임계값을 기준으로 이진 분류를 수행해야 하므로, 정밀도(Precision)와 재현율(Recall)의 조화 평균인 F1-score를 사용하였다. 검증 데이터셋에서 여러 후보 임계값 중 F1-score가 최대가 되도록 임계값을 선택하고, 선택된 임계값을 테스트 데이터셋에 적용하여 최종 F1-score를 계산하였다.

Table 1. System Environment

Variable	Description
Target Variable(Labels) *Within H days after listing	
y_up	Indicator for a price surge: 1 if the Upbit price increases by at least 10% relative to T_0 ; 0 otherwise
y_gap	Indicator for a premium event: 1 if the Upbit-Binance price gap reaches at least 10% at any time; 0 otherwise
y_both	Indicator for a joint event: 1 if both $y_{up} = 1$ and $y_{gap} = 1$; 0 otherwise.
Outcome Metrics *Within H days after listing	
up_maxret	Maximum cumulative return on Upbit
gap_maxret	Maximum Upbit-Binance premium
Pre-listing Price and Volatility *Pre-listing window	
pre_n	Number of minute-level observations in the pre-listing window
pre_logmean	Mean log return
pre_logstd	Standard deviation of log returns
pre_abs_p95	95% of absolute log returns
Social Factors	
social_share	Monthly ticker mention share
social_norm	EWMA of z-scored monthly mention share over the 6 months before listing ($t_0 - 6$ to $t_0 - 1$)

4.3.3 Accuracy

정확도는 전체 샘플 중에서 모델이 올바르게 분류한 비율로, 본 연구의 데이터는 1 비율이 상대적으로 낮은 불균형 구조를 가지므로, 정확도는 AUC 및 F1-score를 보완하는 참고 지표로 활용하였다.

4.3.4 Precision

정밀도는 예측된 급등 확률을 기준으로 상위 10%에 해당하는 종목만을 선택했을 때, 이들 가운데 실제 급등한 종목의 비율을 의미한다. 모델이 가장 급등했다고 평가한 상위 소수 종목만 고른 경우의 성공률을 측정하기 위한 지표이다.

4.3.5 Lift

향상도는 모델이 선택한 상위 구간의 정밀도가 무작위 선택 대비 얼마나 개선되었는지를 나타낸다. 예측 확률 기준 상위 10% 종목의 정밀도를 전체 비율로 나눈 값으로 정의한다. 무작위로 종목을 선택했을 때보다 높은 성공 확률로 급등 종목을 포착한다는 의미이다.

V. Results

5.1. Experimental Results

본 장에서는 세 가지 분류 모형에 대하여 시장 정보만을 사용한 경우(Market-only)와 시장 정보에 소셜 팩터를 추가한 경우(Market + Social)의 성능을 비교한다. 모든 평가는 테스트 데이터셋을 기준으로 수행하였으며, 불균형 라벨 구조를 고려하여 AUC, F1-score, 정확도, 상위 10% 구간 정밀도, 향상도를 동시에 살펴보았다.

Table 2는 상장 후 1일 내 5% 이상 급등 여부를 변수로 설정했을 때의 성능을 비교 분석하였다. 전반적으로 MLP가 가장 우수한 성능을 보였다. 특히 Market-only 설정에서 AUC 0.83, F1 0.63, ACC 0.65, Precision 0.50, Lift 1.70 수준을 기록하며, 다른 모형 대비 높은 분류 성능과 투자 효율성을 동시에 달성하였다. 소셜 팩터를 추가한 경우에도 MLP는 AUC 0.84로 소폭 상승하며 ACC는 0.71 개선되었으며, F1은 유사한 수준을 유지하였다. 이는 MLP가 상장 전후 구간에서 계산된 요약 통계량 및 변동성 지표 등 비선형 패턴을 효과적으로 학습하며, 소셜 팩터가 추가되더라도 핵심 구분 능력이 견고하게 유지됨을 시사한다. Logistic Regression은 Market-only에서 AUC 0.72, F1 0.47, ACC 0.47, Precision 0.40, Lift 1.36 수준으로 MLP 대비 성능이 낮았으나, 소셜 팩터를 추가했을 때 AUC가 0.80으로 크게 상승하였다. AUC와 ACC의 개선 폭이 뚜렷하게 나타나 소셜 신호가 선형 모형의 구분 성능을 보완하는 역할을 할 수 있음을 보여준다. 반면 LightGBM은 소셜 팩터를 추가한 후에는 AUC가 0.57로 감소하였다. 이는 소셜 팩터가 트리 기반 모형에서는 일관된 성능 개선으로 이어지지 않음을 시사한다.

Table 3은 1일 내 3%/5%/7% 급등 여부와 프리미엄 여부를 함께 고려하여 여러 임계값의 결과를 평균한 성능을 제시한다. 전반적으로 Table 2 대비 AUC가 다소 높게 나타났다. 이는 상장 직후 단기 수익률 자체보다 급등 또는 프리미엄이 동반되는 구조적 상황이 상장 전 해외 거래소 시세 정보와 소셜 관심도로 비교적 안정적으로 설명될 수 있음을 시사한다.

모형별로는 Logistic Regression이 가장 우수한 성능을 보였다. market-only에서 이미 AUC 0.85, F1 0.62, ACC 0.67, Precision 0.37, Lift 1.49로 높은 성능을 기록하였고, 소셜 팩터를 추가한 Market + Social 설정에서는 AUC가 0.89로 추가 개선되었다. Precision 역시 0.43으로 상승하여(0.37→0.43) 상위 10% 구간에서의 선별 정확도가 개선되었으며, ACC는 0.68로 유사한 수준을 유지하

였다. MLP는 Table 3에서도 안정적인 성능을 유지하였다. Market-only에서 AUC 0.79, F1 0.51, ACC 0.57, Precision 0.33, Lift 1.36을 기록하였고, 소셜 팩터를 추가한 경우 AUC는 0.79로 동일하게 유지되었다. 이는 소셜 팩터가 추가되면서 MLP가 복합 이벤트 종목을 상위 후보군에서 더 높은 정밀도로 선별할 수 있음을 의미한다.

종합하면, Table 2에서는 MLP가 가장 높은 성능을 보였고, Logistic Regression은 소셜 팩터 추가 시 AUC 및 ACC에서 큰 폭의 개선이 관찰되었다. Table 3에서는 Logistic Regression이 가장 높은 AUC를 기록하며 전체적인 구분 성능이 가장 우수했고, MLP 역시 소셜 신호 추가 시 정밀도와 향상도가 개선되어 상위 10% 전략 관점에서 효율성이 강화되었다. 따라서 소셜 팩터는 급등과 프리미엄을 동시에 식별하는 경우 정밀도 및 상위 후보군 선별 성능을 보완하는 방향으로 기여할 수 있음을 확인하였다.

Table 2. Performance comparison for surges $\geq 5\%$ within 1 day

Method		AUC	F1	ACC	Precision	Lift
LightGBM	Market	0.67	0.50	0.53	0.30	1.02
	+ Social	<u>0.57</u>	0.51	0.53	0.30	1.02
LogReg	Market	<u>0.72</u>	0.47	<u>0.47</u>	0.40	1.36
	+ Social	0.80	0.51	0.82	0.40	1.36
MLP	Market	<u>0.83</u>	0.63	<u>0.65</u>	0.50	1.70
	+ Social	0.84	<u>0.62</u>	0.71	0.50	1.70

Table 3. Average performance for compound events of surges (3%/5%/7%) and Kimchi Premium

Method		AUC	F1	ACC	Precision	Lift
LightGBM	Market	0.78	<u>0.45</u>	0.65	<u>0.33</u>	<u>1.36</u>
	+ Social	0.78	0.55	<u>0.66</u>	0.40	1.61
LogReg	Market	<u>0.85</u>	0.62	0.67	<u>0.37</u>	1.49
	+ Social	0.89	<u>0.57</u>	0.68	0.40	<u>1.37</u>
MLP	Market	0.79	<u>0.51</u>	<u>0.57</u>	<u>0.33</u>	<u>1.36</u>
	+ Social	0.79	0.57	0.59	0.41	1.51

Table 4. Summary of top 3 assets by composite score

Ticker	MINA	OG	SEI
T_0	2023/10/24	2025/09/22	2023/08/16
R_max	116.6	78	153.8
G_max	280.3	264.0	130.9
S(=RxG)	3.27	2.06	2.01
pre_log_std	0.01	0.04	0.02
pre_vol_mean	205,125	533,724	4,929,458

5.2. Result Analysis

상장 펌핑 강도를 종합적으로 평가하기 위하여 상장 후 단기 최대 수익률 과 최대 가격 괴리율을 결합한 점수를 사용해 상위 3개 종목의 그래프를 원화 가격 기준으로 살펴보았다. Fig 2의 시계열 패턴 분석 결과, 상위 랭크된 종목들은 상장 직후 가격 급등 구간에서 김치 프리미엄이 동반 상승하는 현상을 보였다.

Table 4는 복합 점수 상위 3개 종목의 상장 후 수익률·괴리율과 상장 전 변동성·거래량 특성을 요약한 것이다. 예를 들어 가장 높은 점수를 기록한 MINA의 경우, 상장 후 구간에서 단기 최대 수익률은 116.6%인 반면, 동시간대 최대 괴리율은 280.3%에 달한다. 이는 해당 종목의 급등이 펀더멘털 개선보다는, 국내 시장의 제한된 유동성과 차익거래 제약에 기인한 일시적 수급 불균형에 의한 현상임을 시사한다.

반면 SEI의 경우 최대 수익률 153%이 최대 괴리율 130%을 상회하며, 프리미엄 확대보다는 고평가로 매수세가 더 크게 기여한 사례로 해석할 수 있다. 이는 수익률과 괴리율을 분리해 해석할 필요성을 보여준다. 상위 종목들의 상장 전 바이낸스 거래 데이터를 분석한 결과, 공통적으로 상장 직전 구간에서 높은 거래량과 변동성이 관찰되었다. 특히 MINA와 SEI는 평균 거래량과 로그수익률 표준편차가 전체 표본 평균을 유의하게 상회하였다. 이는 해외 시장에서 이미 형성된 높은 관심도와 투기적 수요가 국내 상장 시점의 초기 자금 유입을 가속화시키고, 여기에 김치 프리미엄이 결합되면서 급등과 프리미엄으로 귀결되었다. 이를 통해 본 논문이 제안하는 예측 모형이 최대 수익률과 괴리율 간의 비선형적 관계를 학습하여 신규 상장 이벤트의 급등 여부를 선별할 수 있다는 점을 실증적으로 확인할 수 있다.

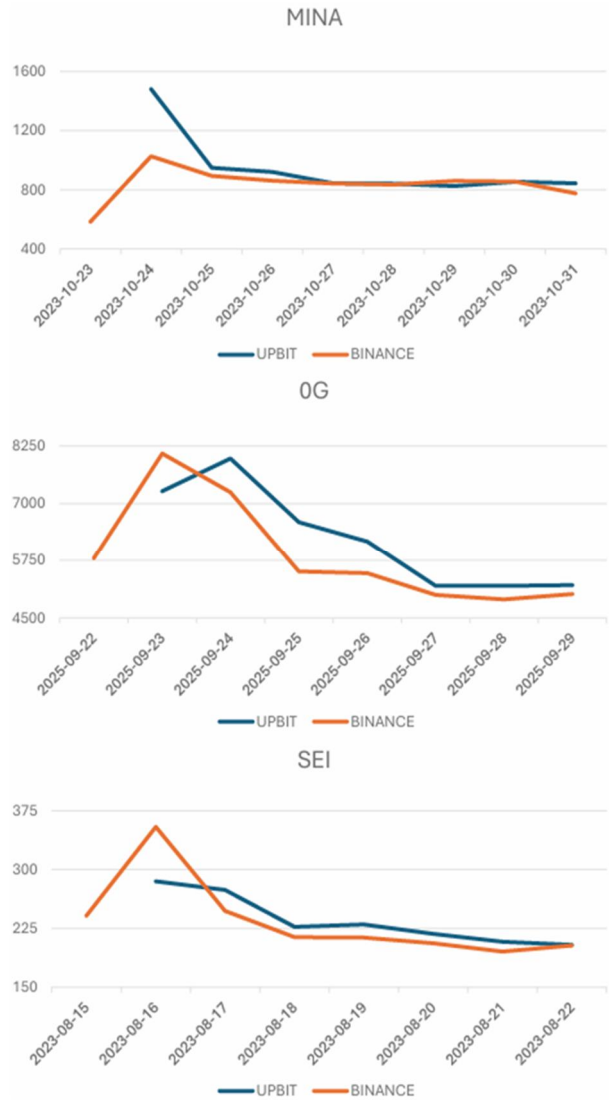


Fig. 2. Daily price trends for the top 3 assets

VI. Conclusion

본 연구는 2021년 1월부터 2025년 12월까지 업비트 원화 마켓에서 신규 상장된 가상자산을 대상으로, 상장 직후 단기 급등과 거래소 간 프리미엄 발생 가능성을 사전에 식별할 수 있는지를 실증적으로 검증하였다. 바이낸스의 선행 시세 정보와 온라인 커뮤니티 기반 소셜 신호를 결합한 분석 프레임워크가 상장 직후 단기 급등과 프리미엄 발생 여부를 얼마나 효과적으로 식별하는지를 체계적으로 비교 평가하였다. 이를 위해 바이낸스 분 단위 시계열과 환율 정보를 결합한 시장 변수, 그리고 온라인 커뮤니티 텍스트로부터 구축한 소셜 팩터를 입력 특성으로 사용하였으며, Logistic Regression, LightGBM, MLP 세 가지 분류 모

형과 Market-only과 Market + Social을 체계적으로 비교하였다.

실험 결과, 단기 급등 여부 예측 과제에서는 MLP 모형이 전반적으로 가장 우수한 성능을 보였다. 반면 급등과 프리미엄을 함께 고려한 Table 3에서는 모형 전반에서 평균 AUC가 Table 2보다 높게 나타나는 경향이 관찰되었다. 또한 소셜 팩터는 단일 급등 예측에서는 모델에 따라 효과가 제한적이거나 상이하게 나타났으나, 복합 이벤트 식별에서는 정밀도와 향상도의 개선을 통해 조기 스크리닝 관점의 활용 가능성을 보여주었다.

이러한 결과는 다음의 시사점을 제공한다. 첫째로 국내 가상자산 상장 이벤트와 같이 표본이 제한되고 시장 구조적 요인이 강하게 작동하는 문제에서는 MLP와 같은 적절한 복잡도의 신경망이나 Logistic Regression와 같이 단순하거나 해석 가능한 모형이 실무적으로 유용한 수준의 예측 성능과 안정성을 제공할 수 있다는 점이다. 둘째로 소셜 팩터는 단일 급등 예측에서는 제한적이거나 모델 의존적인 효과를 보였으나, 프리미엄을 포함한 복합 이벤트 식별에서 정밀도 및 향상도 개선을 통해 조기 스크리닝 관점의 가치를 보여주었다.

한편 본 연구는 티커 단위 테스트 표본 수가 제한적이라 성능 지표의 분산이 크고 모델 간 차이에 대한 통계적 강건성을 단정하기 어렵다는 한계를 가진다. 향후 연구에서는 상장 이벤트 표본을 확장하고 연도 기준의 시간 분할 및 반복 검증을 병행하여 미래 예측 시나리오에 가까운 방식으로 일반화 성능을 재검증할 필요가 있다. 이러한 후속 연구는 국내 코인 시장에서 반복적으로 나타나는 상장 펌핑과 프리미엄 현상을 보다 정교하게 이해하고, 투자자 보호 및 시장 안정성 제고를 위한 실무적 모니터링 체계와 정책적 대응 방안 마련에 기여할 것으로 기대된다.

ACKNOWLEDGEMENT

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2024S1A5C3A03046579)

REFERENCES

- [1] I. Makarov and A. Schoar, "Trading and arbitrage in cryptocurrency markets," *Journal of Financial Economics*, vol. 135, no. 2, pp. 293-319, 2020. <https://doi.org/10.1016/j.jfineco.2019.07.001>
- [2] H. Lee, "A study on cross-border price deviations of virtual assets," *Korean Finance Association*, vol. 37, no. 2, pp. 79-111, 2024. <https://doi.org/10.37197/ARFR.2024.37.2.3>
- [3] J. Son and J. Kim, "A study in bitcoin volatility through economic factors," *Journal of Society for e-Business Studies*, vol. 24, no. 4, pp. 109, 2020. <https://doi.org/10.7838/jsebs.2019.24.4.109>
- [4] H. Benedetti and E. Nikbakht, "Returns and network growth of digital tokens after cross-listings," *Journal of Corporate Finance*, vol. 66, Art. no. 101853, 2021. <https://doi.org/10.1016/j.jcorpfin.2020.101853>
- [5] S. W. Kim, "Prediction performance of LSTM-based economic policy uncertainty index on asset prices: Traditional asset vs. digital asset," *Journal of Digital Contents Society*, vol. 23, no. 6, pp. 1105-1113, 2022. <https://doi.org/10.9728/dcs.2022.23.6.1105>
- [6] P. K. Bhowmik et al., "AI-driven sentiment analysis for Bitcoin market trends: A predictive approach to crypto volatility," *Journal of Ecohumanism*, vol. 4, no. 4, pp. 266-288, 2025. <https://doi.org/10.62754/JOE.V4I4.6729>
- [7] G. Yu, "Bitcoin market prediction using the GRU analysis model: A comparative study with LSTM and random forest models," *Journal of Information Technology Applications and Management*, vol. 32, no. 1, pp. 1-13, 2025. <https://doi.org/10.21219/jitam.2025.32.1.001>
- [8] J. H. Oh, "The determining factors of Kimchi premium in the cryptocurrency market," *Journal of e-Business Studies*, vol. 20, no. 2, pp. 215-228, 2019. <https://doi.org/10.20462/TeBS.2019.4.20.2.215>
- [9] K. Lee et al., "The determinants of Bitcoin prices in Korea," *Korean Journal of Financial Studies*, vol. 48, no. 4, pp. 393-415, 2019. <https://doi.org/10.26845/KJFS.2019.08.48.4.393>
- [10] G. Cho et al., "Statistical arbitrage transactions of cryptocurrency using Kimchi premium and exchange-rate fluctuation prediction," *Journal of Korea Academia-Industrial Cooperation Society*, vol. 24, no. 10, pp. 354-363, 2023. <https://doi.org/10.5762/KAIS.2023.24.10.354>
- [11] H. Kim et al., "Exploring cryptocurrency influence factors using a feature selection algorithm," *The Journal of Internet Electronic Commerce Research*, vol. 19, no. 5, pp. 185-197, 2019. <https://doi.org/10.37272/JIECR.2019.10.19.5.185>
- [12] Maeil Business Newspaper, <https://stock.mk.co.kr/news/view/710059>
- [13] Hankyungbusiness, <https://magazine.hankyung.com/business/article/202410271745b>

- [14] Xangle, South Korea's Cryptocurrency Market: ① Exchanges, <https://xangle.io/en/research/detail/737>
- [15] L. Ante, "Market reaction to exchange listings of cryptocurrencies," Blockchain Research Lab Working Paper Series, no. 3, 2019. <https://doi.org/10.2139/ssrn.3450301>
- [16] C. W. Yang, "A study on arbitrage trading using the price difference of Bitcoin," *Asset Management Review*, vol. 7, no. 2, pp. 1-20, 2019. <https://doi.org/10.23007/amr.2019.7.2.1>
- [17] J. Kim, "A survey of cryptocurrencies based on blockchain," *Journal of The Korea Society of Computer and Information*, vol. 24, no. 2, pp. 67-74, 2019. <https://doi.org/10.9708/jksci.2019.24.02.067>
- [18] P. Katsiampa, "Volatility estimation for Bitcoin: A comparison of GARCH models," *Economics Letters*, vol. 158, pp. 3-6, 2017. <https://doi.org/10.1016/j.econlet.2017.06.023>
- [19] U. Lee, "A literature study on predicting AI-based cryptocurrency value trends," in *Proceedings of the Korean Institute of Information Scientists and Engineers*, pp. 2010-2012, 2021.
- [20] T. Hong et al., "Predicting cryptocurrency prices using explainable artificial intelligence based on deep learning," *Journal of Intelligence and Information Systems*, vol. 29, no. 2, pp. 129-148, 2023. <https://doi.org/10.13088/jiis.2023.29.2.129>
- [21] G. Sim, "A comparative study on smoothing techniques for performance improvement of an LSTM learning model," *Journal of The Korea Society of Computer and Information*, vol. 28, no. 1, pp. 17-26, 2023. <https://doi.org/10.9708/jksci.2023.28.01.017>
- [22] T. Kim and J. Sohn, "A deep learning-based predictive modeling of Bitcoin prices: Analysis with a focus on investor rationality and risk management," *Journal of The Korea Society of Computer and Information*, vol. 30, no. 9, pp. 11-19, 2025. <https://doi.org/10.9708/jksci.2025.30.09.011>
- [23] S. W. Kim, "Prediction performance of altcoin prices using machine learning algorithms," *Journal of Digital Contents Society*, vol. 24, no. 1, pp. 141-151, 2023. <https://doi.org/10.9728/dcs.2023.24.1.141>
- [24] S. Lee and J. Park, "A comparative study of stock and cryptocurrency market prediction using internet search volume," *Asset Management Review*, vol. 9, no. 2, pp. 41-61, 2021. <https://doi.org/10.23007/amr.2022.9.2.41>
- [25] J. Won and T. Hong, "Cryptocurrency prediction using text mining and deep learning techniques: Comparison of Korean and U.S. markets," *Knowledge Management Research*, vol. 22, no. 2, pp. 1-17, 2021. <https://doi.org/10.15813/kmr.2021.22.2.001>
- [26] T. Hong, "Predicting cryptocurrency price movement based on multimodal learning and deep learning," *Korea Intelligent Information Systems Society*, vol. 31, no. 1, pp. 325-337, 2025. <https://doi.org/10.13088/jiis.2025.31.1.325>

Authors



Eun Hong Park received the bachelor's degree in the in the Department of Business Administration from Soongsil University in 2021. She is a master student at the Yonsei University Graduate School of Information

since 2024. Her current research interests include deep learning and financial engineering.



Yeong-In Lee received the B.S. in the Department of Business Administration and Visual Media Design from Sookmyung Women's University, Seoul, Korea, in 2018. She is currently pursuing an integrated M.S.

-Ph.D. degree at the Graduate School of Information, Yonsei University, Seoul, Korea since 2023. Her current research interests include robotics, computer vision and deep learning.



Ha Young Kim is an Associate Professor at Graduate School of Information, Yonsei University, Korea. She received her Ph.D. degree at department of Mathematics, Purdue University, USA.

From 2011 to 2016, she was a research staff member in Samsung Advanced Institute of Technology (SAIT) of Samsung Electronics, Korea, working on various recognition systems with deep learning. Her primary research areas are deep learning and computational Finance. She has published in leading journals, including *Information Fusion*, *Applied Soft Computing*, *Expert Systems with Applications*, *Stochastic and Dynamics*, *Computers in Biology and Medicine*, *PLoS ONE*, *automation in construction*, *Journal of Computing in Civil Engineering* and *Annals of Finance*. She is the the inventor of 8 patents and 13 patent applications.