글로벌 이벤트와 감성 융합을 활용한 한국의 딥러닝 단기 환율 예측

Global Events and Sentiment Fusion for Deep Learning–Based Short-Term Exchange Rate Prediction in Korea

Jaeheung Parka, Kyungwon Kima,*∗*

*aSchool of International Trade and Business, Incheon National University, Incheon, 22012, Republic of Korea*

# 요약

# *본 연구는 원/달러 환율의 단기 예측 성능을 향상시키기 위해 거시·금융시장 지표와 더불어 뉴스 기반 이벤트(GDELT) 및 감성 요약 지표를 통합한 딥러닝 모델을 제안하였다. LSTM, GRU, CNN-LSTM, CNN-GRU 네 가지 아키텍처를 대상으로 다양한 입력 조합과 시차 구간(lookback=5~90일)을 실험한 결과, 이벤트·감성 통합 입력과 CNN-LSTM(lookback=20)의 조합이 최적 성능을 달성하였다(MSE=47.71, MAPE=0.39%). 변수 중요도 분석(Permutation Importance)을 통해 금·알루미늄·구리 등 원자재 가격, 미국 정책금리와 국채금리, KOSPI/KOSDAQ 지수 및 거래량 등이 환율 변동의 핵심 설명력을 가지는 것으로 확인되었으며, 뉴스 요약의 감성 변동성 역시 의미 있는 신호로 작용하였다. 시각화 결과, CNN-LSTM 기반 예측치는 실제 환율과 유사한 흐름을 포착하며 단기 방향성을 효과적으로 설명하였다. 본 연구는 사건·감성 정보의 활용이 환율 예측의 정확도를 높일 수 있음을 실증적으로 보여주며, 향후 금융시장 예측에 있어 텍스트 기반 신호의 중요성을 뒷받침한다.*

# *중심어 : 환율, 시계열 예측, 딥러닝, 글로벌 이벤트, 뉴스 감성 분석*

# Abstract

*This study proposes a deep learning framework for enhancing short-term forecasting of the KRW/USD exchange rate by integrating macro-financial indicators with news-based event (GDELT) and sentiment summary variables. Four architectures—LSTM, GRU, CNN-LSTM, and CNN-GRU—were evaluated across multiple input settings and lookback windows (5–90 days). The combination of event–sentiment integration and CNN-LSTM with a 20-day lookback achieved the best performance (MSE=47.71, MAPE=0.39%). Permutation importance analysis revealed that commodity prices (gold, aluminum, copper), U.S. policy and bond yields, and domestic equity indices (KOSPI/KOSDAQ) exerted the strongest explanatory power on exchange rate fluctuations, while sentiment volatility in news summaries also provided significant predictive signals. Visualization of the best-performing model demonstrated that CNN-LSTM predictions closely tracked actual exchange rate dynamics, effectively capturing short-term market direction. These findings provide empirical evidence that incorporating event and sentiment information can improve forecasting accuracy, underscoring the role of text-based signals in financial market prediction.*

*Keywords:* Exchange Rate, Time Series Forecasting, Deep Learning, Global Events, News Sentiment

*∗*Corresponding author: thekimk.kr@gmail.com

*ORCID(s):* 0000-0001-6530-8426 (K. Kim)

# Ⅰ. 서론

환율(Exchange rate)은 외국 통화 한 단위를 획득하기 위해 자국 통화를 얼마나 지불해야 하는지를 나타내는, 즉 자국 통화와 외국 통화 간의 교환 비율을 의미한다[1]. 이는 두 나라 통화의 상대적 가치를 보여주는 지표로, 오늘날 거의 모든 경제 활동이 국제 거래와 밀접하게 연계되어 있기 때문에 개인의 일상생활 뿐 아니라 기업 경영, 국가 경제 운영에 이르기까지 필수적인 정보로 기능한다[2]. 특히 원/달러 환율은 한국 대외경제에서 핵심적인 역할을 한다. 환율 변동은 수출입 가격, 외국인 투자, 물가 수준, 중앙은행의 통화정책 등 다양한 거시경제 변수와 밀접하게 연관되어 있으며, 환율의 향후 변화를 예측하는 것은 국가 경제의 안정성과 기업의 재무 전략 수립에 있어 중요한 과제로 인식된다. 그러나 환율은 경제 펀더멘털뿐만 아니라 지정학적 리스크, 글로벌 자본 흐름, 시장 심리 등 복합적인 요인의 영향을 받는다. 이로 인해 환율 변동은 종종 ‘랜덤워크(Random Walk)’로 설명되며, 이는 술에 취한 사람이 어디로 걸어갈지 예측하기 어렵듯 환율의 다음 시점을 정확히 전망하기 힘들다는 비유적 표현이다[3].

환율은 외환시장의 수요와 공급에 의해 결정되며, 장기적으로는 물가 수준이나 생산성 변화와 같은 경제 여건이 통화 가치에 영향을 미친다. 중기적으로는 대외거래와 거시경제정책이 주요 요인으로 작용하고, 단기적으로는 시장 참가자들의 기대, 주변국 환율 변동, 각종 뉴스에 의해 크게 좌우된다. 특히 시장 참가자들의 기대가 환율 상승 혹은 하락 쪽으로 쏠리면 자기실현적(Self-fulfilling) 거래를 통해 실제 환율 변동을 초래하기도 한다. 예컨대 다수가 환율 상승을 예상할 경우, 환율이 오르기 전에 외환을 선매입하려는 수요가 증가하여 실제로 환율 상승을 유발하게 된다. 또한 뉴스와 같은 정보 역시 시장 참가자들의 기대를 변화시켜 단기 환율 변동에 중요한 영향을 미친다. 실제로 2010년 5월 천안함 침몰조사 결과 발표로 지정학적 리스크가 부각되자 원/달러 환율이 단기간 급등한 사례가 있다[4].

이와 같은 환경에서 환율의 경로를 정밀하게 예측하는 것은 정책당국의 통화·재정정책 수립, 기업의 수출입 가격전략 조정, 금융기관의 리스크 관리에 필수적이다. 기존 국내 연구들은 이러한 필요성을 강조해왔다.

김우석·한규식(2021)은 팬데믹 이후 환율의 구조적 변화를 분석하며 변동성 확대에 따른 예측 필요성을 제시하였고[14], 오인정·김우주(2022)는 코로나19 구간을 대상으로 SARIMA와 ARDL 모형을 비교하여 국면별로 원/달러 예측모형의 성능이 달라진다는 점을 보여주었다[15].

그러나 이러한 연구들은 대부분 종가 기반의 시계열 모형 비교 또는 구조변화 탐색에 초점을 맞추었으며, 이벤트 데이터와 감성 분석 지표를 결합하여 환율을 예측한 사례는 드물다. 이에 본 연구는 기존 선행연구의 거시경제적·구조적 요인을 포괄하면서, 나아가 뉴스 감성 점수와 GDELT 기반 이벤트 지표를 통합한 융합적 접근을 시도한다. 이를 통해 환율 변동성의 구조적 요인과 심리적 요인을 동시에 반영하는 정교한 예측 모델을 제시하고자 한다.

이러한 맥락에서 시장 참여자들의 심리와 뉴스에 담긴 감정을 정량화해 반영하는 감성 분석(Sentiment Analysis)은 환율 예측의 새로운 단서를 제공할 수 있다. 따라서 본 연구는 감성 지표를 분석하여 이를 딥러닝 기반 시계열 예측 모델에 결합함으로써, 원/달러 환율 예측의 정확도를 향상시킬 수 있는 가능성을 탐색하고자 한다.

# Ⅱ. 문헌연구

환율 예측 연구들은 전통적으로 거시경제 변수와 금융시장 변수를 주요 독립변수로 설정하였다.  
Plakandaras et al.(2015)은 1999년부터 2011년까지의 데이터를 바탕으로 원자재(19개), 금속(10개), 주가지수(7개), 무역지수(3개), 금리(11개), 거시경제 지표(11개), 환율(8개) 등 광범위한 변수를 활용하였다. 다양한 모형을 비교한 결과, EEMD 기반 비선형 모형이 일간 및 월간 환율 예측에서 우수한 성능을 보였다[5].  
국내 연구로는 임현욱 외(2021)가 연합인포맥스와 Bloomberg 데이터를 사용하여 원/달러 환율을 예측하였다. 이들은 1년·10년 만기의 원화 IRS 금리, KTB 수익률, 본드 스왑 스프레드 등 금리 및 채권시장 변수를 중심으로 구성하였으며, ANN 모델이 평균 Hit Ratio 50.96%로 가장 우수한 성과를 보였다[6]. Cao et al.(2020)은 유가와 금값 같은 실물자산 변수 뿐 아니라 통화공급(M1, M2), CPI, PPI, 산업생산, 기준금리, 인플레이션율, 무역수지, 정책 불확실성 지수 등 거시경제 지표를 결합하였다. 이들은 제안한 DC-LSTM 모형이 CNN, LSTM, SVR, ARIMA보다 낮은 예측 오차(MAE 0.0145)를 기록했다고 보고하였다[7].  
또한 Qureshi(2025)는 USD/INR 환율을 대상으로 금리, 인플레이션율, GDP 성장률, 외환보유액 등 거시경제 펀더멘털 변수를 활용하였다. 다양한 머신러닝 모형을 비교한 결과, Random Forest가 RMSE 0.029로 가장 우수한 성과를 나타냈다[8]. 마지막으로 Wang et al.(2021)은 USD/CNY 환율 뿐만 아니라 나스닥, 다우존스, 상하이 종합지수, 항셍지수 등 글로벌 주가지수를 독립변수로 결합하였다. CNN, RNN, LSTM 등 여러 모델을 비교한 결과, CNN-TLSTM 모형이 MAPE 0.18945로 가장 높은 예측력을 보였다[9]. 종합하면, 기존 환율 예측 연구들은 공통적으로 금리, 물가, 무역수지, 주가지수, 원자재 가격 등 거시경제 및 금융 변수를 중심으로 환율 변동을 설명하고자 했음을 확인할 수 있다.

기존 선행연구 대비 본 연구의 가장 큰 차별성은 전통적 변수군에 외생적 사건을 반영하는 이벤트 기반 지표를 통합한 점에 있다. 이를 위해 GDELT(Global Database of Events, Language, and Tone) 데이터베이스를 활용하였으며, 이는 전 세계 언론(인쇄·방송·웹 기반)을 실시간으로 모니터링하여 사건 유형, 위치, 감정 톤, 인물 및 조직, 문서 수 등 다양한 속성으로 사건을 정량화하는 데이터베이스이다[16]. GDELT의 데이터는 전 세계 사회적 사건을 정량적 시계열 데이터로 전환할 수 있다는 점에서 기존 거시·금융 예측 변수의 한계를 보완할 수 있는 강력한 도구로 주목받고 있다.

경제 및 금융 분야에서도 GDELT 기반 이벤트 데이터를 예측 변수로 활용한 사례가 있다. 예컨대 ARIMA 모형에 GDELT 데이터를 추가하여 환율 예측 정확도를 약 1% 향상시키고, Gradient Boosting 기반 분류 모델에서는 정확도를 6% 상승시킨 연구[17]와, 유럽 채권시장 분석에서 GDELT 지표가 nowcasting 모델의 성능을 향상시킨 사례가 있다[18]. 또한 미국 주식시장 예측에서, GDELT 지표는 전통적 거시 경제 모형을 능가하는 설명력 및 예측 성능을 제공했다는 실증 결과도 보고되었다[19].

본 연구는 뉴스 기사와 온라인 댓글이라는 두 가지 비정형 텍스트 데이터를 대상으로 감성 분석을 수행하였다. 감성 분석은 자연어 처리(NLP)의 하위 분야로, 텍스트에 내재된 정서적 톤을 식별하고 분류하는 작업을 의미한다[20].

기존 연구에서는 규칙 기반 접근이나 전통적 머신러닝 기법이 주로 활용되었으나, 도메인 의존성, 언어 확장성, 은유·풍자와 같은 복잡한 뉘앙스 처리의 한계가 지속적으로 지적되었다[21]. 이러한 제약을 극복하기 위해 최근에는 Transformer 기반 딥러닝 모델이 적극 도입되었다. 특히 BERT 계열 모델은 양방향 문맥 정보를 활용하여 단어 의미를 정교하게 파악하고, 적은 데이터로도 파인튜닝이 가능하다는 장점이 있다.

금융 분야에서는 FinBERT와 같이 금융 특화 언어모델이 등장하여 기존 머신러닝 접근 대비 성능이 크게 향상되었으며[22], FiQA·Financial PhraseBank와 같은 벤치마크에서도 Transformer 기반 모델들이 전통적 방법을 압도하는 성능을 기록하였다[23]. 더 나아가 최근에는 GPT-4o와 같은 대규모 언어모델(LLM)이 프롬프트 튜닝이나 few-shot 학습만으로도 FinBERT와 동등하거나 그 이상의 성능을 보이는 것으로 보고되고 있다[24][25]. 이는 금융 뉴스·리포트 분석과 같은 복잡한 텍스트 처리에서 LLM이 높은 적응력과 확장성을 제공함을 보여준다.

이러한 흐름을 반영하여, 본 연구는 전통적인 규칙 기반·머신러닝 접근 대신 GPT-4o-mini 기반 LLM 감성 분석 절차를 설계하였다.

최근 금융시장 예측에서는 전통적인 거시·금융 변수 뿐 아니라 시장 참여자의 심리 변수를 반영하기 위해 감성 분석이 적극적으로 활용되고 있다.  
Mohan et al.(2019)은 Apple, Amazon, Microsoft 등 S&P500 주요 기업의 종가 정보와 함께 국제 뉴스 기사에서 추출한 긍·부정 감정 점수를 독립변수로 사용하였다. 분석 결과, 감정 변수를 포함한 RNN 모델이 MAPE 2.03으로 가장 우수한 성능을 보였다[10]. Jing et al.(2021)은 Shanghai Stock Exchange 데이터와 Eastmoney.com 투자자 게시글의 감정 점수를 CNN으로 분류하고, 이를 주가 기술적 지표와 함께 LSTM 모델에 입력하였다. 그 결과 CNN-LSTM 모델이 평균 MAPE 0.0449로 가장 뛰어난 예측력을 보여, 투자자 심리 반영의 효과성을 입증하였다[11]. 또한 정가연 외(2024)는 네이버 증권 뉴스, 한국은행 API, Naver DATALAB 데이터를 활용하여 코스피, S&P500, 유가, 금값, 원/엔 환율, 금리 등 거시·금융 변수와 함께 KLUE-BERT 기반 뉴스 감정 점수, 코스피 검색 빈도수 등 시장 심리 변수를 결합하였다. 실험 결과, CNN-GRU 모델이 MAPE 1.38%로 가장 낮아 감성 분석의 유용성을 보여주었다[12]. 이러한 연구들은 전통적인 금융 변수에 뉴스 및 게시글 감정 점수를 결합하는 융합적 독립변수 구성이 주가 예측 정확도를 높일 수 있음을 시사한다.

환율 예측에도 최근에는 감성 분석을 접목한 연구가 시도되고 있다. Ding et al.(2024)은 Investing.com과 ForexEmpire의 뉴스 및 댓글 데이터를 기반으로 감정 점수를 산출하였다. 이들은 해당 감성 지표를 USD/CAD, USD/MXN, USD/CNY, USD/JPY, USD/KRW 등 주요 교차환율, 원자재 가격(유가, 금, 구리, 옥수수 등), 글로벌 지수(S&P500, EuroStoxx50, VIX), 채권수익률(미국·유로존 10년물)과 같은 거시·금융 변수와 결합하였다. 다양한 모형을 비교한 결과, PSO-LSTM 모델이 RMSE 0.0958로 가장 높은 성능을 기록하였다[13]. 이러한 결과는 환율 예측에서도 단순한 거시경제 지표만으로 설명하기 어려운 단기 변동성을 뉴스 및 댓글에 담긴 시장 심리 변수가 효과적으로 보완할 수 있음을 보여준다.

<표 1>에 정리된 선행연구를 종합하면, 다음과 같은 시사점을 도출할 수 있다. 첫째, 환율 예측에서는 여전히 금리, 물가, 무역수지, 원자재 가격, 주가지수 등 거시경제 및 금융 변수가 핵심적 설명 변수로 기능하고 있다. 둘째, 최근 연구들은 뉴스, 투자자 게시글, 댓글 등 비정형 텍스트 데이터에서 추출한 감정 점수를 독립변수로 추가하여 예측 성능을 향상시키고 있으며, 이는 금융시장 예측에서 시장 심리 변수의 중요성을 뒷받침한다. 셋째, 감성 분석을 결합한 환율 예측 연구는 아직 초기 단계에 있으나, 뉴스·댓글 감성과 같은 심리 변수의 확장적 활용 가능성이 크다는 점에서 향후 연구 방향성을 제시한다. 따라서 본 연구는 이러한 시사점을 바탕으로, 거시경제 변수와 함께 GDELT 이벤트 톤, 뉴스 감성 분석을 통합한 딥러닝 기반 원/달러 환율 예측 모델을 제안한다. 이를 통해 환율 변동성을 설명할 수 있는 새로운 변수를 발굴하고, 기존 연구 대비 예측 정확도를 향상시키고자 한다.

**<표 1> 환율 예측 및 주가 예측과 감성 분석을 활용한 선행연구 분석 방향과 성능 요약**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 연구 목적 | 연구자(연도) | 데이터 소스 | 데이터 기간 | 주요 독립변수 | 방법론 | 성능 |
| **환율 예측** | Plakandaras et al.(2015)  [5] | 미기재 | 1999~2011 | 원자재(19), 금속(10), 주가지수(7), 금리(11), 무역·거시지표 등 | ARIMA, GARCH, AR-NN, MARS-NN, SVR, EEMD 결합 | 일간: EEMD-MARS-SVR, MARS-NN 우수 / 월간: EEMD-AR-SVR 우수 |
| 임현욱 외(2021)  [6] | 연합인포맥스, Bloomberg | 2008~2020 | IRS 금리, KTB 수익률, 본드 스왑 스프레드 | ANN, LR, DT | ANN 평균 Hit Ratio 50.96% |
| Cao et al.(2020)  [7] | 미기재 | 2016.6~2019.4 | 유가, 금값, CPI, PPI, 산업생산, 기준금리, 무역수지, 정책 불확실성 지수 등 | ARIMA, SVR, CNN, LSTM, DC-LSTM | DC-LSTM MAE 0.0145 |
| Qureshi(2025)  [8] | FRED, IMF, World Bank 등 | 2015~2023 | 금리, 인플레이션율, GDP 성장률, 외환보유액 | ARIMA, RW, RF, SVM | RF RMSE 0.029 |
| Wang et al.(2021)  [9] | Wind DB | 2006~2020 | USD/CNY, 나스닥, 다우, 상하이, 항셍 등 지수 | MLP, CNN, RNN, LSTM, CNN-LSTM, CNN-TLSTM | CNN-TLSTM MAPE 0.18945 |
| **주가 예측 + 감성분석** | Mohan et al.(2019)  [10] | 해외 뉴스 웹사이트 | 2013~2017 | S&P500 기업 종가, 뉴스 감정 | ARIMA, Prophet, RNN | RNN-pp MAPE 2.03 |
| Jing et al.(2021)  [11] | SSE, Eastmoney.com | 2017~2019 | 기술적 지표 , 게시글 감정 | CNN+LSTM | MAPE 0.0449 |
| 정가연 외(2024)  [12] | 네이버뉴스, KOSPI, DATALAB, 한은 API | 2021~2023 | 코스피·S&P500, 유가, 금값, 원/엔 환율, 금리, 검색 빈도, 감정 점수 | LSTM, GRU, CNN-LSTM, CNN-GRU | CNN-GRU + KLUE-BERT 감정 MAPE 1.38% |
| **환율 예측 + 감성분석** | Ding et al.(2025)  [13] | Investing.com, ForexEmpire | 2016~2024 | 뉴스·댓글 감정, 교차환율, 원자재, 글로벌 지수, 채권수익률 | PSO-LSTM, SVM, GRU, VAR, ARIMA/GARCH | PSO-LSTM RMSE 0.0958 |

# Ⅱ. 데이터 및 방법론

1. 종속변수 데이터

본 연구는 2020년 1월부터 2024년 12월까지의 일별 원/달러 환율 종가를 예측 대상으로 한다. 이 시기는 코로나19 팬데믹의 충격, 글로벌 통화정책의 급격한 전환, 지정학적 리스크의 확산 등 서로 다른 시장 국면을 모두 포함하고 있어 환율 변동성이 경제 전반에 미치는 파급효과를 분석하기에 적합하다. 실제로 2020년 이후 원/달러 환율은 팬데믹 초기의 급락, 2021년 저점 국면, 2022년 미국 연방준비제도의 초고속 금리 인상과 한·미 금리차 확대, 정권교체에 따른 국내 정치·정책 변화, 러시아·우크라이나 전쟁과 중동 지역 긴장 등 국내외 요인이 복합적으로 작용하면서 큰 폭의 상승세를 나타냈다.

<그림 1>에 제시된 환율 추이를 살펴보면, 2021년 초 약 1,070원 수준까지 하락한 이후, 2022년 5월에는 1,300원을 돌파하며 환율 변동성이 급격히 확대되었음을 확인할 수 있다. 특히 2021년 상반기의 일평균 환율 등락폭이 ±5원 수준에 불과했던 반면, 2022년 상반기에는 ±15원 수준으로 확대되었으며, 일간 표준편차 기준 환율 변동성 지표 또한 같은 기간 0.4%에서 1.2%로 세 배 가량 급등하였다. 이처럼 변동성이 커진 환율은 2022년 사상 최대의 무역수지 적자를 야기하였고, 원화 약세를 통해 수입물가 상승과 내수 위축, 소비심리 악화로 이어졌다. 반면 수출기업에게는 단기적 가격 경쟁력을 제공하였으나, 환율 급등에 따른 환 헤지 비용은 전년 대비 35% 증가하였고 중소기업을 중심으로 금융비용 부담이 가중되었다.

라인, 그래프, 스크린샷, 텍스트이(가) 표시된 사진

AI 생성 콘텐츠는 정확하지 않을 수 있습니다.

**<그림 1> 원/달러 환율 추이**

2. 독립변수 데이터

원/달러 환율을 예측하기 위해 본 연구에서 활용한 독립변수는 기존 환율 예측 연구에서 활용된 주요 요인을 기반으로 하되, 환율 변동의 구조적 요인과 심리적 요인을 동시에 포괄하도록 설계하였다.

선행연구에서는 주로 거시경제 및 금융시장 지표가 독립변수로 사용되었다. Plakandaras et al.(2015)은 원자재, 금속, 주가지수, 금리, 거시경제 변수 등 약 60개 지표를 활용하여 환율을 예측하였으며[5], 임현욱 외(2021)는 IRS 금리, KTB 수익률, 본드 스왑 스프레드 등 금리·채권시장 변수에 집중하였다[6]. Cao et al.(2020)은 유가와 금값 같은 실물자산 가격과 함께 M1·M2, CPI, PPI, 산업생산 등 거시지표를 포함하였고[7], Ali(2025)는 금리, 인플레이션율, GDP 성장률, 외환보유액을 핵심 변수로 설정하였다[8]. Wang et al.(2021)은 USD/CNY 환율뿐 아니라 나스닥, 다우존스, 상하이종합, 항셍 등 글로벌 주가지수를 결합하여 국제적 요인을 반영하였다[9].

최근에는 감성 분석을 통한 시장 심리 변수화 시도가 활발히 이루어졌다. Mohan et al.(2019)은 국제 뉴스 기사에서 추출한 감정 점수를 S&P500 주가와 결합하였으며[10], Jing et al.(2021)은 투자자 게시글의 감정을 CNN 기반 분류기로 산출해 주가 예측에 반영하였다[11]. 정가연 외(2024)는 KLUE-BERT 기반 뉴스 감정 점수를 거시 변수와 함께 투입하여 예측 정확도를 높였고[12], Ding et al.(2025)은 환율 관련 뉴스·댓글 감정을 교차환율, 원자재, 지수, 채권수익률과 결합하여 환율 예측 성능을 향상시켰다[13].

이러한 선행연구의 검토를 통해, 본 연구는 기존의 주요 독립변수들을 포괄하면서도 그 범위를 확장하였다. 우선 종속변수로는 원/달러 환율을 설정하였으며, 교차환율로는 USD/JPY와 USD/CNY를 포함하였다. 이는 원화 환율과 경쟁적 관계에 있는 주요 통화의 변동성을 반영하기 위함이다. 또한 국내외 주식시장 지표인 KOSPI, KOSDAQ, 다우존스, S&P500을 포함하여 자본 흐름과 위험 선호도의 영향을 고려하였다. 원자재 변수로는 유가(WTI), 금, 구리, 니켈, 알루미늄을 채택하여 글로벌 경기 사이클과 인플레이션 기대를 반영하였으며, 거시경제 변수로는 한국과 미국의 정책금리, 소비자물가지수, 생산자물가지수, 경상수지, 통화량(M1, M2)을 선정하였다. 이는 환율 결정에 직접적으로 작용하는 금리 차, 물가 수준, 대외거래, 유동성 요인을 반영하기 위함이다. 아울러 리스크 지표로는 VIX, OVX, KSVKOSPI를 포함하여 금융시장의 불확실성과 위험회피 성향을 계량화하였다.

3. 파생변수 데이터

(1) 이벤트 데이터

본 연구에서는 GDELT로부터 추출한 다섯 가지 핵심 이벤트 기반 변수를 활용하였다. 즉, (1) 전체 이벤트 수, (2) 평균 감성 톤, (3) 한국 관련 이벤트 톤, (4) 전체 문서 수, (5) 한국 관련 문서 수이다. 이벤트 지표는 기존의 거시적 구조적 요인(금리, 주가, 원자재 등) 이외에 '사건 기반 리스크 및 심리적 충격'을 반영할 수 있는 변수로서, 본 연구가 지닌 독창성을 강화해 준다.

(2) 감성분석 데이터

본 연구의 뉴스 감성 분석 데이터는 네이버 금융 뉴스의 ‘환율’ 섹션에서 2020년 1월 1일부터 2024년 12월 31일까지 총 35,633건의 기사를 수집한 것이다. 수집된 원문은 정규표현식을 활용해 대괄호·소괄호 내 불필요한 문구, 이메일 주소, 광고성 표현, 그리고 “=” 기호 이전의 문장 등을 제거하였다. 또한 제목과 본문이 동일한 중복 기사, 내용이 비어 있거나 공백만 포함된 기사, 분석 목적과 무관한 기사는 모두 제외하여 정보 노이즈를 최소화하였다.

정제된 뉴스는 GPT-4o-mini를 활용하여 원달러 환율과의 관련성을 기준으로 세부 분류하였다. 그 결과, 원달러 환율과 직접적으로 관련된 뉴스 22,756건, 간접적으로 관련된 뉴스 7,320건, 관련이 없는 뉴스 5,557건으로 구분되었다. 이후 감성 분석의 효율성과 정확성을 높이기 위해 ‘직접 및 간접 관련 뉴스’를 대상으로 추가적인 전처리를 수행하였다.

뉴스 본문은 길이가 다양하고 불필요한 수식어, 문맥적 중복, 장황한 서술을 포함하는 경우가 많아, 이를 그대로 감성 분석에 활용할 경우 모델의 입력 효율성과 분석 정확도에 제약이 발생할 수 있다. 따라서 본 연구에서는 감성 분석 전처리 과정의 일환으로 뉴스 본문 요약을 수행하였다. 요약 모델은 실험 기간 중 HuggingFace Hub에서 제공되는 한국어 요약 모델 중 다운로드 수가 높은 mT5-multilingual-XLSum, t5-base-korean-summarization, KoBART-summary-v3 세 가지를 비교하였다. 성능 평가는 AI Hub 한국어 문서 요약(신문기사) 데이터셋을 사용하였으며, 지표는 ROUGE-1, ROUGE-2, ROUGE-L을 적용하였다.

ROUGE는 “Recall-Oriented Understudy of Gisting Evaluation”의 약자로, 자동으로 생성된 텍스트를 평가하기 위한 지표 집합을 의미한다. 보통 텍스트 요약(TS, Text Summarization) 알고리즘이 생성한 요약의 품질을 평가하는 데 사용된다[22]. ROUGE-1은 단어 단위의 중복 정도를 나타내며, 요약이 원문 핵심 어휘를 얼마나 잘 포함하는지 평가한다. ROUGE-2는 연속된 2-gram 중복을 기준으로 하여, 문맥적 연결성과 유창성을 간접적으로 반영한다. ROUGE-L은 가장 긴 공통 부분수열(Longest Common Subsequence, LCS)을 활용하여, 요약이 원문의 전체 문장 구조와 맥락을 얼마나 잘 보존하는지를 평가한다[26].

그 결과 <표 2>에서 보듯 KoBART-summary-v3가 모든 지표에서 가장 높은 성능을 기록하여 최종 요약 모델로 채택하였다.

Dalal 외(2024)에 따르면 일반적으로 ROUGE-1 F1 점수가 0.3~0.4 정도이면 낮다고 볼 수 있고, 0.4~0.5 정도이면 보통이라고 볼 수 있으며, 0.5 이상이면 좋다고 볼 수 있다고 한다. ROUGE-2 F1 점수의 경우 0.1~0.2 정도이면 낮다고 볼 수 있고, 0.2~0.4 정도이면 보통이라고 볼 수 있으며, 0.4 이상은 좋다고 볼 수 있으며, ROUGE-L F1 점수는 0.3~0.4 정도이면 낮다고 볼 수 있고, 0.4~0.5 정도이면 보통이라고 볼 수 있으며, 0.5 이상은 일반적으로 좋다고 볼 수 있다고 한다. 다만 이런 기준은 언어, 도메인, 요약 길이, 참조 요약의 수 등에 따라 매우 달라지므로 절대적 기준이라기보다는 비교 맥락에서 참고하는 지침으로 활용된다.

**<표 2> 요약 모델 성능 표**

|  |  |  |  |
| --- | --- | --- | --- |
| 모델 | ROUGE-1 | ROUGE-2 | ROUGE-L |
| mT5\_XLSum | 0.2683 | 0.0762 | 0.2613 |
| T5-korean-summarization | 0.2356 | 0.0604 | 0.2322 |
| KoBART-summary-v3 | **0.4128** | **0.1632** | **0.3835** |

감성 분석은 GPT-4o-mini 기반의 LLM 프롬프트 분류 방식을 채택하였다. 감성 라벨 체계는 환율 변동의 금융적 맥락을 반영하여 정의되었다. 즉, 긍정은 원/달러 환율 하락(원화 강세), 위험선호 확대, 완화적 환경을 의미하고, 부정은 환율 상승(원화 약세), 위험회피 확대, 긴축적 환경을 의미하며, 중립은 방향성 판단이 어려운 경우를 뜻한다. GPT-4o-mini에는 few-shot 프롬프트 튜닝을 적용하여 별도의 파인튜닝 과정 없이 금융 문맥에 특화된 감성 분류를 수행하도록 하였다. 분류 결과는 긍정=+1, 중립=0, 부정=-1로 수치화하였으며, 이를 기반으로 날짜별 평균 감성 점수를 산출하여 일별 감정 지표(daily sentiment index)를 구축하였다.

또한 분석의 세분화를 위해 뉴스의 환율 관련성을 기준으로 세 가지 일별 변수 세트를 생성하였다. 첫째, 직접적인 환율 뉴스의 일별 평균 감성 점수 및 기사 수, 둘째, 간접적인 환율 뉴스의 일별 평균 감성 점수 및 기사 수, 셋째, 두 범주를 모두 포함한 전체 뉴스의 일별 감성 점수 및 기사 수를 각각 산출하였다. 이러한 세분화된 감성 지표는 후속 환율 예측 모델의 입력 변수로 활용되었다.

4. 데이터 전처리 및 통계량

결과적으로 본 연구는 거시·금융·원자재·리스크 변수와 더불어 이벤트 기반 변수 및 뉴스·댓글 감성 지표까지 포괄하여, 환율 변동의 구조적 요인과 단기 시장 심리 요인을 동시 고려하는 통합적 예측 프레임워크를 구축하고자 한다. <표 3>

**<표 3> 환율 예측에 사용한 데이터**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 대분류 | 출처 | 데이터명 | 기간 | 단위 | 추출 변수 | 활용 변수 |
| **목표 변수** | Yahoo Finance | KRW/USD | 2020.1–2024.12 | 일 | 종가 | Target (원/달러 환율) |
| **교차 환율** | Yahoo Finance | USD/JPY, USD/CNY | 2020.1–2024.12 | 일 | 종가 | USD/JPY 종가, USD/CNY 종가 |
| **시장 지표** | Investing.com | KOSPI, KOSDAQ, 다우존스, S&P500 | 2020.1–2024.12 | 일 | 종가, 거래량 | KOSPI 종가·거래량, KOSDAQ 종가·거래량, 다우존스 종가, S&P500 종가 |
| **원자재** | Yahoo Finance, Investing.com | WTI, 금(XAU), 구리, 니켈, 알루미늄 | 2020.1–2024.12 | 일 | 종가 | WTI 종가, 금, 구리, 알루미늄, 니켈 종가 |
| **거시 지표** | ECOS, FRED, 통계청, 한국은행 | 정책금리, CPI, PPI, 경상수지, 통화량(M1, M2) | 2020.1–2024.12 | 월 | 한국/미국 정책금리, 소비자물가지수, 생산자물가지수, 경상수지, M1, M2 | 동일 |
| **리스크 지표** | Yahoo Finance | VIX, MOVE, OVX, KSVKOSPI | 2020.1–2024.12 | 일 | 주식·채권·유가 변동성, 코스피 불안 지수 | 동일 |
| **이벤트 지표** | GDELT | events\_cnt, tone\_mean, tone\_kor\_mean, gkg\_doc\_cnt, gkg\_kr\_loc\_cnt | 2020.1–2024.12 | 일 | 이벤트 수, 평균 톤, 한국 관련 톤, 전체 GKG 문서 수, 한국 관련 문서 수 | 동일 |
| **감정 지표** | Naver News | 환율 관련 뉴스 | 2020.1–2024.12 | 일 | 뉴스 감정 점수 | 일일 감정 점수 평균, 뉴스 개수 |

데이터의 기술통계량은 <표 4>에 제시하였다. 표에는 각 변수별 관측치 수(Count), 평균(Mean), 최소값(Min), 제1사분위수(Q1), 중앙값(Median), 제3사분위수(Q3), 최대값(Max), 표준편차(Std)를 포함하였다.

**<표 4>** 변수들의 통계량 (N, Mean, Median, Std. Q1, Q3)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **변수명** | **Count** | **Mean** | **Min** | **25%** | **50%** | **75%** | **Max** | **Std** |
| **날짜** | 1186 | 2022-06-25 | 2020-01-02 | 2021-03-29 | 2022-06-25 | 2023-09-18 | 2024-12-30 | Nan |
| **원/달러 환율 종가** | 1186 | 1256.283 | 93.74322 | 1083.53 | 1181.822 | 1265.61 | 1333.02 | 1472.5 |
| **원/달러 환율 시가** | 1186 | 1256.633 | 93.7873 | 1083.87 | 1182.46 | 1266.325 | 1333.395 | 1471.1 |
| **코스피 지수** | 1186 | 2596.073 | 339.2823 | 1457.64 | 2399.31 | 2568.48 | 2758.637 | 3305.21 |
| **코스피 거래량** | 1186 | 7.13E+08 | 3.55E+08 | 824510 | 4.63E+08 | 6.14E+08 | 8.66E+08 | 3.46E+09 |
| **코스닥 지수** | 1186 | 841.1601 | 114.4289 | 428.35 | 762.185 | 847.505 | 919.4125 | 1060 |
| **코스닥 거래량** | 1186 | 1.31E+09 | 5.21E+08 | 6.2E+08 | 9.52E+08 | 1.17E+09 | 1.5E+09 | 5.07E+09 |
| **WTI 유가** | 1186 | 69.53344 | 23.29195 | 7.79 | 62.09 | 73.735 | 82.03 | 126.47 |
| **금(XAU) 가격** | 1186 | 1937.514 | 263.5906 | 1471 | 1778.307 | 1870.39 | 1985.455 | 2786.19 |
| **다우존스 지수** | 1186 | 33603.14 | 4657.789 | 18591.93 | 31022.09 | 33913.62 | 35627.82 | 45014.04 |
| **다우존스 거래량** | 1186 | 3.69E+08 | 1.13E+08 | 1.17E+08 | 3.02E+08 | 3.43E+08 | 4.02E+08 | 1.22E+09 |
| **S&P500 지수** | 1186 | 4250.18 | 764.0219 | 2237.4 | 3810.15 | 4200.01 | 4591.015 | 6090.27 |
| **통화량(M1)** | 1186 | 1219.395 | 103.6807 | 944.3064 | 1184.088 | 1221.11 | 1287.187 | 1371.921 |
| **통화량(M2)** | 1186 | 3616.388 | 347.3367 | 2935.64 | 3309.121 | 3718.333 | 3843.787 | 4183.535 |
| **한국 정책금리** | 1186 | 2.037789 | 1.303587 | 0.5 | 0.548387 | 1.75 | 3.5 | 3.5 |
| **미국 정책금리** | 1186 | 2.60763 | 2.31595 | 0.25 | 0.25 | 1.75 | 5.25 | 5.5 |
| **소비자물가지수(CPI)** | 1186 | 2.811298 | 1.658261 | -0.2 | 1.425 | 2.7 | 3.8 | 6.3 |
| **생산자물가지수(PPI)** | 1186 | 111.5014 | 7.401748 | 98.93 | 103.89 | 116.27 | 117.5 | 119.56 |
| **경상수지** | 1186 | 5333.791 | 4255.906 | -4205 | 1862 | 6365 | 7438 | 13096 |
| **코스피 변동성(KSVKOSPI)** | 1186 | 20.30508 | 6.604767 | 11.73 | 16.5625 | 18.38 | 22.415 | 69.24 |
| **달러/엔 환율** | 1186 | 127.8556 | 18.42469 | 102.678 | 109.3763 | 130.8025 | 145.341 | 161.607 |
| **달러/위안 환율** | 1186 | 6.867263 | 0.318528 | 6.3115 | 6.522125 | 6.94415 | 7.1455 | 7.3278 |
| **미국 10년물 국채금리** | 1186 | 0.267769 | 0.139671 | 0.0499 | 0.1415 | 0.2915 | 0.401 | 0.4988 |
| **구리 가격** | 1186 | 3.820936 | 0.631159 | 2.1195 | 3.553 | 3.8995 | 4.28875 | 5.119 |
| **알루미늄 가격** | 1186 | 2313.547 | 420.7198 | 1452 | 2126 | 2302.5 | 2534.125 | 3873 |
| **VIX(주식 변동성)** | 1186 | 21.43804 | 8.368224 | 11.86 | 15.855 | 19.63 | 25.0225 | 82.69 |
| **OVX(유가 변동성)** | 1186 | 44.72318 | 24.29387 | 23.73 | 34.28 | 39.005 | 47.485 | 325.15 |
| **이벤트 수(events\_cnt)** | 1186 | 133581.2 | 24788.61 | 0 | 117551 | 133423.5 | 150155 | 216375 |
| **이벤트 평균 톤(events\_tone\_mean)** | 1186 | -1.979816 | 0.258274 | -3.333169 | -2.133663 | -1.963814 | -1.832024 | 0 |
| **한국 관련 톤(events\_tone\_kor\_mean)** | 1186 | -0.832832 | 0.832322 | -3.924516 | -1.336452 | -0.710838 | -0.234973 | 1.06649 |
| **GKG 문서 수(gkg\_doc\_cnt)** | 1186 | 88501.87 | 24450.6 | 0 | 68474.75 | 85789 | 107725 | 157565 |
| **한국 관련 문서 수(gkg\_kr\_loc\_춧)** | 1186 | 617.3356 | 276.5282 | 0 | 435 | 547 | 750 | 3016 |
| **직접 뉴스 평균 감정 점수** | 1186 | -0.070743 | 0.429306 | -1 | -0.3636 | -0.07415 | 0.207875 | 1 |
| **직접 뉴스 기사 수** | 1186 | 12.69646 | 4.440581 | 0 | 10 | 13 | 16 | 20 |
| **간접 뉴스 평균 감정 점수** | 1186 | -0.040528 | 0.36739 | -1 | -0.1667 | 0 | 0 | 1 |
| **간접 뉴스 기사 수** | 1186 | 3.875211 | 3.31109 | 0 | 1 | 3 | 6 | 18 |
| **전체 뉴스 평균 감정 점수** | 1186 | -0.068784 | 0.351629 | -1 | -0.298525 | -0.0588 | 0.15 | 1 |
| **전체 뉴스 기사 수** | 1186 | 16.57167 | 3.709767 | 0 | 15 | 18 | 19 | 20 |

데이터 전처리 과정은 다음과 같은 절차를 거쳐 수행되었다. 우선, 전체 분석의 기준이 되는 날짜(Date) 변수와 원/달러 환율의 종가 및 시가 데이터를 구축하였다. 이때 주말 및 공휴일 등 비영업일은 제거하여 실제 금융시장의 거래일만을 반영하였다.

두 번째 단계에서는 날짜를 기준으로 일 단위 변수(예: 유가, 금 가격, 주요 주가지수, 변동성 지표 등)를 순차적으로 병합하였다. 또한, 정책금리, 소비자물가지수(CPI), 생산자물가지수(PPI), 경상수지, 통화량(M1, M2) 등과 같은 월 단위 거시경제 변수는 각 월의 대표값(해당 월 발표 기준)을 해당 월 전체 거래일에 동일하게 확장하여 병합하였다. 이를 통해 일 단위 종합 시계열 데이터프레임 내에서 월간 지표가 자연스럽게 반영되도록 구성하였다.

세 번째로, GDELT 이벤트 데이터 및 뉴스 감성 변수를 동일한 일 단위 주기로 맞추어 결합하였다. 이벤트 데이터는 글로벌 및 한국 관련 이벤트 수, 평균 톤(tone) 등을 포함하였고, 뉴스 감성 변수는 직접·간접 관련 뉴스의 일별 평균 감정 점수 및 기사 수로 구성하였다.

마지막으로, 결측치 처리는 전체 변수 중 뉴스 감성 데이터에서만 발생하였다. 2020–2024년 기간 동안 총 13일치의 뉴스 감정 값이 비어 있었으며, 해당 결측치는 분석의 일관성을 유지하기 위해 0으로 대체하였다(결측치=0). 그 외 변수에서는 결측이 존재하지 않아 별도의 보정 과정이 필요하지 않았다. 이와 같은 절차를 통해 모든 데이터는 일 단위로 정렬된 통합형 시계열 데이터프레임으로 구성되었으며, 최종적으로 1,186개의 관측치를 확보하였다.

데이터는 학습(80%), 테스트(20%)로 분할하였으며, 모든 수치형 변수는 RobustScaler로 정규화하였다. RobustScaler는 중앙값과 사분위 범위를 활용하여 극단값(outlier) 영향을 최소화하는 장점이 있다.

4. 분석 방법론: LSTM, GRU, CNN-LSTM, CNN-GRU

환율 예측 모델은 LSTM과 GRU 및 하이브리드 모델(CNN-LSTM, CNN\_GRU)을 대상으로 성능을 비교하였다.

최근 외환시장의 복잡성과 비정형적 거시경제 요인을 반영한 시계열 예측 분야에서는 통계 모형을 넘어 딥러닝 기반 아키텍처가 각광받고 있다. 그 중에서도 순환신경망(RNN) 계열의 LSTM(Long Short-Term Memory)과 GRU(Gated Recurrent Unit)는 장기 의존성 학습과 계산 효율성 면에서 주목할 만한 성능을 보인다.

LSTM은 Hochreiter와 Schmidhuber(1997)가 제안한 구조로, 메모리 셀과 게이트 메커니즘을 통해 vanishing gradient 문제를 완화하며 긴 시퀀스 학습에 강점을 가진다[28]. 반면 GRU는 update gate와 reset gate 두 가지 게이트만을 사용하여 구조가 단순하면서도 LSTM에 근접한 성능을 제공하며, 연산 효율성이 뛰어나다는 장점이 있다[29].

외환 예측에서도 이들 모델의 유용성이 다수 연구에서 입증되었다. 예를 들어 Qu와 Zhao(2019)는 EUR/USD 환율 예측에서 LSTM이 기존 RNN보다 RMSE와 MAE가 낮고, 예측 정확도 또한 우수함을 보고하였다[30]. Agustin과 De Melin(2024)은 ARS/USD 예측에서 GRU가 CNN보다 모든 지표에서 더 나은 성능을 보였으며, 이는 환율 데이터의 시계열적 패턴 포착에 순환 구조가 효과적임을 시사한다[31]. 또한 Islam & Hossain(2021)은 EUR/USD 등 주요 통화쌍을 대상으로 GRU와 LSTM을 결합한 하이브리드 모델을 제안하였고, 단일 모델 대비 예측 정확도가 향상됨을 확인하였다[32]. 종합하면, 선행연구들은 LSTM과 GRU가 외환 시장 예측에서 높은 잠재력을 지니며, 상황에 따라 하이브리드 접근이 예측 성능 개선에 유의미하게 작동할 수 있음을 보여준다.

딥러닝 기반 시계열 예측은 단순한 트렌드 분석을 넘어, 변수 간 복잡한 상호작용을 동시에 포착할 수 있는 하이브리드 구조로 발전하고 있다. 이 가운데 CNN-LSTM과 CNN-GRU는 Convolutional Neural Network(CNN)의 국소 특징 추출 능력과 순환신경망(RNN) 계열의 시계열 패턴 학습 능력을 결합한 대표적 모델이다. CNN-LSTM은 입력 시계열 데이터에 1차원 합성곱 계층(1D Convolution Layer)을 적용해 의미 있는 지역적 특징(feature)을 먼저 추출한 뒤, 이를 LSTM에 전달하여 장기적 의존성을 학습한다. CNN-GRU는 동일한 원리를 따르되, GRU의 단순화된 게이트 구조를 통해 연산 효율성과 학습 속도를 강화한 형태이다.

금융시장 예측에서도 이러한 하이브리드 모델은 단일 구조보다 우수한 성능을 보여왔다. 예컨대 Lu et al.(2020)은 주가 예측 실험에서 CNN-LSTM 모델이 가장 높은 예측 정확도를 기록했다고 보고하였으며[33], Jing et al.(2021) 또한 CNN-LSTM을 활용해 투자자 심리를 반영한 주가 예측에서 기존 모델을 능가하는 성과를 입증하였다[11].

따라서 본 연구는 뉴스 요약 및 댓글 감정 점수와 같은 비정형 텍스트 기반 심리 변수를 포함한 시계열 데이터를 CNN-LSTM과 CNN-GRU 모델에 적용하였다. 이로써 CNN이 시장 심리의 국소적 변동을 효과적으로 포착하고, RNN 계열이 이를 시간 축에서 장기적 흐름으로 연결함으로써, 원/달러 환율 예측의 정밀도를 한층 높일 수 있을 것으로 기대된다.

5. 예측 성능평가 지표

본 연구에서는 원/달러 환율 예측 모델의 성능을 다각도로 평가하기 위해 RMSE(Root Mean Squared Error), MSPE(Mean Squared Percentage Error), MAE(Mean Absolute Error), MAPE(Mean Absolute Percentage Error), MedAE(Median Absolute Error), MedAPE(Median Absolute Percentage Error) 총 여섯 가지 회귀 지표를 활용하였다.

첫째, RMSE는 예측 오차의 제곱 평균에 제곱근을 취한 값으로, 오차가 클수록 제곱에 의해 더 큰 영향을 미치게 된다. 따라서 모델이 큰 오차를 얼마나 효과적으로 억제하는지를 평가하는 데 적합하며, 데이터의 단위를 그대로 유지하기 때문에 해석이 용이하다. 둘째, MSPE는 예측 오차를 실제값으로 정규화한 후 제곱하여 평균한 값으로, 예측 오차의 상대적 크기를 강조한다. MSPE는 통계학 및 예측 모델링 분야에서 ‘Mean Squared Prediction Error’로도 불리며, 예측의 편향(bias)과 분산(variance)을 동시에 반영하는 지표로 활용된다. 셋째, MAE는 예측값과 실제값의 절대적 차이를 평균한 값으로, 모든 오차를 동일한 가중치로 평가하여 평균적인 예측 정확도를 파악하기에 적합하다. RMSE와 비교했을 때 이상치에 덜 민감하며, 데이터의 전반적인 오차 규모를 직관적으로 보여준다. 넷째, MAPE는 절대 오차를 실제값으로 나눈 뒤 백분율로 환산한 값으로, 모델의 오차를 직관적으로 이해할 수 있게 해준다. 그러나 실제값이 0 또는 매우 작은 경우 무한대 값이 발생할 수 있어 주의가 필요하며, 이로 인해 일부 연구에서는 MAPE의 대안 지표를 제안하기도 했다. 다섯째, MedAE는 절대 오차의 중앙값을 사용하여 이상치에 대한 영향을 최소화한 지표이다. 평균이 아닌 중앙값을 사용함으로써 데이터 분포의 왜곡에 덜 민감하며, 모델의 일관적 성능을 평가할 수 있다는 장점이 있다. 여섯째, MedAPE는 절대 백분율 오차의 중앙값으로, MAPE의 직관성을 유지하면서도 극단적인 오차의 영향을 완화한다. 특히 비정상 분포나 금융 시계열처럼 이상치가 자주 발생하는 데이터셋에서 안정적인 평가가 가능하다.

이처럼 다양한 지표를 함께 활용함으로써, 본 연구는 모델의 평균적 예측 정확도뿐만 아니라 이상치에 대한 강건성(robustness)과 실제 금융 예측 환경에서의 해석 가능성까지 종합적으로 검증하였다.

Ⅲ. 연구결과

3.1 예측 성능평가

본 연구는 원/달러 환율의 다음 영업일 종가를 예측하기 위해 LSTM, GRU, CNN-LSTM, CNN-GRU 네 가지 딥러닝 아키텍처를 비교하였다. 전체 데이터는 시계열 순서에 따라 80%를 학습용, 20%를 테스트용으로 분할하였으며, 입력 시퀀스의 길이(lookback window)는 5일, 10일, 20일, 30일, 60일, 90일로 설정하여 과거 기간의 길이에 따른 예측 성능 변화를 분석하였다. 모델의 불확실성을 최소화하기 위해 각각의 설정에서 시드(seed) 42, 55, 68을 적용해 세 차례 반복 학습을 수행하였고, 모든 결과는 이들의 평균값으로 평가하였다.

모든 실험에서 입력 데이터는 RobustScaler를 이용해 정규화하였다. 딥러닝 모델의 학습 성능은 입력 분포에 민감하게 반응하므로, 이상치의 영향을 완화하기 위해 평균과 표준편차 대신 사분위 범위를 사용하는 RobustScaler를 선택하였다.

입력 변수 조합은 네 가지 케이스로 구분하였다. (1) 거시·시장 변수만을 포함한 Baseline, (2) Baseline에 GDELT 이벤트 변수를 추가한 조합, (3) Baseline에 뉴스 요약 감성 지표를 추가한 조합, (4) Baseline에 이벤트 지표와 감성 지표를 모두 추가한 통합 조합이다.

모델 구조는 모두 코드 설계와 동일하게 구현되었다. LSTM과 GRU 모델은 64→32 유닛의 2층 순환 구조로 구성하였고, 각 층 사이에 Dropout(0.1)을 적용하였다. CNN-LSTM과 CNN-GRU는 1차원 합성곱(Conv1D, kernel\_size=3, causal padding, ReLU)을 통해 국소적 패턴을 추출한 뒤 Batch Normalization과 Dropout을 거쳐 각각 LSTM(64→32) 또는 GRU(64→32) 계층으로 시계열 의존성을 학습하였다. 모든 모델의 출력층은 Dense(1)이며, 손실 함수는 MSE로 고정하였다. <표 5>는 모델별 세부 하이퍼파라미터를 요약한 것이다.

**<표 5> 모델별 세부 하이퍼파라미터**

|  |  |  |
| --- | --- | --- |
| 알고리즘 | 주요 파라미터 | 값 |
| **LSTM** | Unit Numbers | 64 → 32 (2-스택) |
| Dropout Ratio | 0.1 |
| Dense Units | 1 |
| **GRU** | Unit Numbers | 64 → 32 (2-스택) |
| Dropout Ratio | 0.1 |
| Dense Units | 1 |
| **CNN-LSTM** | Filters | 32 |
| Kernel Size | 3 (causal padding) |
| Activation | ReLU |
| Batch Normalization | 적용 |
| LSTM Units | 64 → 32 |
| Dropout Ratio | 0.1 |
| Dense Units | 1 |
| **CNN-GRU** | Filters | 32 |
| Kernel Size | 3 (causal padding) |
| Activation | ReLU |
| Batch Normalization | 적용 |
| GRU Units | 64 → 32 |
| Dropout Ratio | 0.1 |
| Dense Units | 1 |

모델 성능 평가는 MSE, RMSE, MAE, MedAE, MAPE, MSPE, MedAPE의 6가지 지표를 사용하였다. MSE, RMSE, MAE, MedAE는 원 단위로, MAPE, MSPE, MedAPE는 백분율 단위로 산출되었다.

3.2 실험 결과

모델의 예측 성능을 검증하기 위해 Lookback을 5, 10, 20, 30, 60, 90일로 달리하며 LSTM, GRU, CNN-LSTM, CNN-GRU 네 가지 아키텍처를 동일한 조건에서 평가하였다. 각 실험은 Macro Only, Macro + Event, Macro + Sentiment, ALL의 네 가지 피처 조합으로 구성되었으며, 모든 경우에 대해 MSE, RMSE, MAE, MAPE, MedAE, MedAPE의 여섯 가지 지표를 산출하였다.

Lookback 5일 구간에서는 전반적으로 RMSE가 7.1~7.6 수준에서 형성되었으며,  
모델 간 성능 차이는 크지 않았으나, CNN-LSTM 모델의 ‘Macro + Event’ 조합이 RMSE 7.162로 가장 낮은 오차를 기록하였다.  
이는 짧은 기간 내 급변하는 환율의 단기 패턴을 합성곱 계층이 효과적으로 포착했기 때문으로 해석된다.

반면 GRU 단일 모델은 전반적으로 RMSE가 7.3 이상으로 높게 나타나, 단기 구간에서는 변동성이 큰 시계열을 안정적으로 추적하기 어려웠다. CNN-LSTM과 CNN-GRU 모델은 LSTM, GRU보다 낮은 MAPE를 보여, 구조적 복합성이 짧은 시계열에서도 의미 있는 패턴 학습에 기여했음을 알 수 있다.

<표 6> LookBack : 5

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Model** | **케이스** | **RMSE** | **MAE** | **MAPE** | **MedAE** | **MedAPE** |
| LSTM | Macro Only | 7.229 | 5.557 | 0.407 | 4.609 | 0.341 |
| LSTM | Macro + Event | 7.316 | 5.619 | 0.412 | 4.606 | 0.34 |
| LSTM | Macro + Sentiment | 7.183 | 5.521 | 0.405 | 4.623 | 0.343 |
| LSTM | ALL | 7.317 | 5.645 | 0.414 | 4.66 | 0.337 |
| GRU | Macro Only | 7.376 | 5.676 | 0.416 | 4.58 | 0.337 |
| GRU | Macro + Event | 7.552 | 5.816 | 0.426 | 4.749 | 0.346 |
| GRU | Macro + Sentiment | 7.566 | 5.852 | 0.428 | 4.755 | 0.35 |
| GRU | ALL | 7.67 | 5.929 | 0.434 | 4.782 | 0.352 |
| CNN\_LSTM | Macro Only | 7.403 | 5.625 | 0.413 | 4.609 | 0.339 |
| CNN\_LSTM | Macro + Event | 7.162 | 5.44 | 0.399 | 4.549 | 0.336 |
| CNN\_LSTM | Macro + Sentiment | 7.168 | 5.468 | 0.401 | 4.551 | 0.338 |
| CNN\_LSTM | ALL | 7.243 | 5.553 | 0.407 | 4.616 | 0.341 |
| CNN\_GRU | Macro Only | 7.41 | 5.607 | 0.411 | 4.546 | 0.332 |
| CNN\_GRU | Macro + Event | 7.169 | 5.428 | 0.398 | 4.448 | 0.328 |
| CNN\_GRU | Macro + Sentiment | 7.24 | 5.55 | 0.407 | 4.55 | 0.339 |
| CNN\_GRU | ALL | 7.288 | 5.593 | 0.41 | 4.61 | 0.338 |

Lookback 10일에서는 대부분의 모델이 5일보다 낮은 RMSE를 기록하였으며, 평균 RMSE가 7.28로 개선되었다. 특히 LSTM 계열 모델이 안정적인 수렴을 보였고, 그중 LSTM ‘Macro Only’의 RMSE가 7.089로 가장 낮았다. CNN-GRU 역시 7.116~7.19 수준의 결과를 보여, 짧은 구간에서도 CNN 기반 모델의 패턴 추출 능력이 유효함을 확인할 수 있었다. 그러나 GRU 모델은 여전히 7.5 이상으로, 시계열 길이를 늘려도 성능 향상 폭이 제한적이었다.

<표 7> LookBack : 10

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Model** | **케이스** | **RMSE** | **MAE** | **MAPE** | **MedAE** | **MedAPE** |
| LSTM | Macro Only | 7.089 | 5.468 | 0.401 | 4.665 | 0.343 |
| LSTM | Macro + Event | 7.11 | 5.479 | 0.401 | 4.569 | 0.332 |
| LSTM | Macro + Sentiment | 7.124 | 5.511 | 0.404 | 4.684 | 0.347 |
| LSTM | ALL | 7.103 | 5.497 | 0.403 | 4.628 | 0.344 |
| GRU | Macro Only | 7.739 | 6.055 | 0.442 | 4.802 | 0.352 |
| GRU | Macro + Event | 7.596 | 5.917 | 0.432 | 4.707 | 0.348 |
| GRU | Macro + Sentiment | 7.595 | 5.924 | 0.433 | 4.87 | 0.355 |
| GRU | ALL | 7.56 | 5.869 | 0.429 | 4.812 | 0.354 |
| CNN\_LSTM | Macro Only | 7.254 | 5.586 | 0.409 | 4.716 | 0.345 |
| CNN\_LSTM | Macro + Event | 7.4 | 5.65 | 0.414 | 4.716 | 0.345 |
| CNN\_LSTM | Macro + Sentiment | 7.177 | 5.502 | 0.403 | 4.496 | 0.334 |
| CNN\_LSTM | ALL | 7.165 | 5.523 | 0.404 | 4.787 | 0.35 |
| CNN\_GRU | Macro Only | 7.116 | 5.42 | 0.397 | 4.519 | 0.331 |
| CNN\_GRU | Macro + Event | 7.19 | 5.428 | 0.398 | 4.436 | 0.327 |
| CNN\_GRU | Macro + Sentiment | 7.158 | 5.508 | 0.403 | 4.776 | 0.354 |
| CNN\_GRU | ALL | 7.126 | 5.442 | 0.399 | 4.731 | 0.346 |

Lookback 20일 구간은 모든 모델에서 가장 우수한 성능을 나타낸 핵심 구간이었다. 전체 평균 RMSE는 7.132로 모든 실험 중 최저치를 기록했으며, MSE 역시 50.93으로 가장 낮았다. LSTM ‘Macro + Event’ 조합은 RMSE 6.946으로 전체 실험에서 최소 오차를 달성하였고, CNN-LSTM ‘ALL’ 조합 역시 RMSE 6.976으로 근소한 차이를 보였다. GRU와 CNN-GRU 계열 역시 대부분 이 구간에서 최적 성능을 보였으며, 각 모델별 최적 Lookback을 비교한 결과에서도 15개 중 10개 조합이 20일 구간에서 최소 RMSE를 기록하였다. 이는 약 1개월(20영업일) 단위의 중기 정보창이 환율 예측에서 가장 효율적인 시간 범위임을 의미한다.

<표 8> LookBack : 20

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Model** | **케이스** | **RMSE** | **MAE** | **MAPE** | **MedAE** | **MedAPE** |
| LSTM | Macro Only | 7.094 | 5.56 | 0.407 | 4.654 | 0.341 |
| LSTM | Macro + Event | 6.946 | 5.384 | 0.394 | 4.562 | 0.333 |
| LSTM | Macro + Sentiment | 7.053 | 5.533 | 0.405 | 4.663 | 0.341 |
| LSTM | ALL | 7.086 | 5.567 | 0.407 | 4.652 | 0.341 |
| GRU | Macro Only | 7 | 5.424 | 0.397 | 4.585 | 0.337 |
| GRU | Macro + Event | 7.303 | 5.692 | 0.416 | 4.819 | 0.354 |
| GRU | Macro + Sentiment | 7.375 | 5.767 | 0.421 | 4.761 | 0.346 |
| GRU | ALL | 7.376 | 5.717 | 0.418 | 4.799 | 0.351 |
| CNN\_LSTM | Macro Only | 7.101 | 5.483 | 0.401 | 4.484 | 0.329 |
| CNN\_LSTM | Macro + Event | 7.173 | 5.557 | 0.407 | 4.611 | 0.338 |
| CNN\_LSTM | Macro + Sentiment | 7.115 | 5.501 | 0.402 | 4.527 | 0.332 |
| CNN\_LSTM | ALL | 6.976 | 5.39 | 0.394 | 4.519 | 0.335 |
| CNN\_GRU | Macro Only | 7.128 | 5.481 | 0.401 | 4.519 | 0.33 |
| CNN\_GRU | Macro + Event | 6.996 | 5.383 | 0.394 | 4.44 | 0.326 |
| CNN\_GRU | Macro + Sentiment | 7.289 | 5.727 | 0.419 | 4.804 | 0.351 |
| CNN\_GRU | ALL | 7.094 | 5.5 | 0.402 | 4.629 | 0.34 |

Lookback 30일 구간에서는 RMSE가 다시 소폭 상승하여 평균 7.287을 기록하였다. LSTM과 CNN-LSTM 모델이 여전히 안정적인 수준을 유지한 반면, GRU 모델의 오차가 확대되며 변동성이 커졌다. 다만 LSTM ‘Macro + Event’ 조합은 RMSE 7.002로 비교적 낮은 오차를 유지하였고, 이는 30일 이상의 윈도우에서도 합성곱 계층이 불필요한 노이즈를 일부 억제하는 효과를 발휘한 것으로 해석된다.

<표 9> LookBack : 30

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Model** | **케이스** | **RMSE** | **MAE** | **MAPE** | **MedAE** | **MedAPE** |
| LSTM | Macro Only | 7.122 | 5.523 | 0.404 | 4.536 | 0.333 |
| LSTM | Macro + Event | 7.002 | 5.363 | 0.392 | 4.441 | 0.321 |
| LSTM | Macro + Sentiment | 7.146 | 5.593 | 0.409 | 4.66 | 0.341 |
| LSTM | ALL | 7.116 | 5.565 | 0.406 | 4.666 | 0.344 |
| GRU | Macro Only | 7.598 | 5.985 | 0.436 | 4.801 | 0.352 |
| GRU | Macro + Event | 7.483 | 5.841 | 0.426 | 4.809 | 0.352 |
| GRU | Macro + Sentiment | 7.642 | 6.047 | 0.441 | 4.991 | 0.363 |
| GRU | ALL | 7.569 | 5.979 | 0.436 | 4.833 | 0.354 |
| CNN\_LSTM | Macro Only | 7.052 | 5.426 | 0.397 | 4.533 | 0.331 |
| CNN\_LSTM | Macro + Event | 7.098 | 5.411 | 0.396 | 4.366 | 0.317 |
| CNN\_LSTM | Macro + Sentiment | 7.205 | 5.605 | 0.409 | 4.562 | 0.336 |
| CNN\_LSTM | ALL | 7.155 | 5.539 | 0.405 | 4.688 | 0.343 |
| CNN\_GRU | Macro Only | 7.543 | 5.761 | 0.421 | 4.718 | 0.347 |
| CNN\_GRU | Macro + Event | 7.353 | 5.593 | 0.409 | 4.676 | 0.341 |
| CNN\_GRU | Macro + Sentiment | 7.178 | 5.533 | 0.404 | 4.488 | 0.325 |
| CNN\_GRU | ALL | 7.336 | 5.7 | 0.416 | 4.693 | 0.34 |

Lookback 60일과 90일 구간에서는 모든 모델에서 RMSE가 다시 증가하였다. 평균 RMSE는 각각 7.516과 7.458로, 짧은 구간보다 오차가 확대되는 경향을 보였다. 이는 학습 윈도우가 지나치게 길어짐에 따라 과거의 패턴이 현재 환율 변동과의 연관성을 약화시키는 ‘정보 희석 효과’가 발생했기 때문으로 보인다. LSTM 모델은 60일 이후에도 상대적으로 안정적인 결과를 보였으나, GRU 모델은 RMSE가 7.9 이상으로 급등하여 장기 구간에서의 일반화 성능 저하가 뚜렷했다.

<표 10> LookBack : 60

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Model** | **케이스** | **RMSE** | **MAE** | **MAPE** | **MedAE** | **MedAPE** |
| LSTM | Macro Only | 7.342 | 5.623 | 0.41 | 4.476 | 0.326 |
| LSTM | Macro + Event | 7.339 | 5.574 | 0.406 | 4.404 | 0.317 |
| LSTM | Macro + Sentiment | 7.367 | 5.673 | 0.413 | 4.58 | 0.334 |
| LSTM | ALL | 7.244 | 5.555 | 0.405 | 4.491 | 0.325 |
| GRU | Macro Only | 7.927 | 6.097 | 0.444 | 5.039 | 0.368 |
| GRU | Macro + Event | 7.849 | 6.05 | 0.44 | 5.01 | 0.366 |
| GRU | Macro + Sentiment | 7.581 | 5.9 | 0.429 | 4.756 | 0.343 |
| GRU | ALL | 8.172 | 6.38 | 0.464 | 5.3 | 0.387 |
| CNN\_LSTM | Macro Only | 7.276 | 5.579 | 0.406 | 4.468 | 0.327 |
| CNN\_LSTM | Macro + Event | 7.429 | 5.64 | 0.411 | 4.743 | 0.343 |
| CNN\_LSTM | Macro + Sentiment | 7.474 | 5.757 | 0.419 | 4.7 | 0.343 |
| CNN\_LSTM | ALL | 7.395 | 5.695 | 0.415 | 4.609 | 0.337 |
| CNN\_GRU | Macro Only | 7.414 | 5.632 | 0.41 | 4.806 | 0.344 |
| CNN\_GRU | Macro + Event | 7.421 | 5.66 | 0.412 | 4.601 | 0.331 |
| CNN\_GRU | Macro + Sentiment | 7.527 | 5.835 | 0.424 | 4.731 | 0.343 |
| CNN\_GRU | ALL | 7.493 | 5.729 | 0.417 | 4.595 | 0.337 |

<표 11> LookBack : 90

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Model** | **케이스** | **RMSE** | **MAE** | **MAPE** | **MedAE** | **MedAPE** |
| LSTM | Macro Only | 7.261 | 5.535 | 0.403 | 4.522 | 0.329 |
| LSTM | Macro + Event | 7.281 | 5.544 | 0.404 | 4.358 | 0.316 |
| LSTM | Macro + Sentiment | 7.294 | 5.592 | 0.407 | 4.59 | 0.333 |
| LSTM | ALL | 7.238 | 5.548 | 0.404 | 4.451 | 0.326 |
| GRU | Macro Only | 7.665 | 5.868 | 0.426 | 4.768 | 0.345 |
| GRU | Macro + Event | 7.803 | 5.996 | 0.436 | 4.956 | 0.364 |
| GRU | Macro + Sentiment | 7.789 | 6.052 | 0.439 | 4.905 | 0.359 |
| GRU | ALL | 7.709 | 5.951 | 0.433 | 4.921 | 0.356 |
| CNN\_LSTM | Macro Only | 7.382 | 5.644 | 0.411 | 4.638 | 0.338 |
| CNN\_LSTM | Macro + Event | 7.392 | 5.64 | 0.41 | 4.427 | 0.323 |
| CNN\_LSTM | Macro + Sentiment | 7.403 | 5.677 | 0.413 | 4.601 | 0.337 |
| CNN\_LSTM | ALL | 7.458 | 5.739 | 0.417 | 4.612 | 0.339 |
| CNN\_GRU | Macro Only | 7.273 | 5.511 | 0.401 | 4.694 | 0.338 |
| CNN\_GRU | Macro + Event | 7.35 | 5.61 | 0.408 | 4.549 | 0.329 |
| CNN\_GRU | Macro + Sentiment | 7.513 | 5.777 | 0.42 | 4.758 | 0.353 |
| CNN\_GRU | ALL | 7.517 | 5.744 | 0.417 | 4.712 | 0.346 |

Lookback별 평균 성능을 종합하면, 전체 모델과 케이스의 평균 RMSE는 20일 구간에서 7.132로 가장 낮았고, 그 다음이 10일(7.281), 30일(7.287) 순이었다. 반면 60일(7.516)과 90일(7.458)은 가장 높은 값을 나타냈다. 이는 지나치게 짧거나 긴 구간보다는 중간 길이의 시계열 창이 예측 안정성과 변동 포착의 균형을 가장 잘 달성한다는 점을 시사한다.

<표 12> lookback별 평균 지표값

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **lookback** | **RMSE** | **MAE** | **MAPE** | **MedAE** | **MedAPE** | **MSE** | **MSPE** |
| 20 | 7.132 | 5.542 | 0.405 | 4.627 | 0.339 | 50.93 | 0.003 |
| 10 | 7.281 | 5.611 | 0.411 | 4.682 | 0.344 | 53.084 | 0.003 |
| 30 | 7.287 | 5.654 | 0.413 | 4.654 | 0.34 | 53.21 | 0.003 |
| 5 | 7.331 | 5.617 | 0.412 | 4.615 | 0.34 | 53.788 | 0.003 |
| 90 | 7.458 | 5.714 | 0.416 | 4.654 | 0.339 | 55.716 | 0.003 |
| 60 | 7.516 | 5.774 | 0.42 | 4.707 | 0.342 | 56.623 | 0.003 |

모델별·케이스별 최적 조합을 살펴보면, LSTM 계열은 주로 20일 구간에서 최소 RMSE를 기록하였고, GRU 계열 또한 대부분의 케이스가 동일 구간에서 최적점을 보였다. CNN-LSTM은 ‘Macro Only’ 및 ‘Macro + Event’ 조합에서 30일, 나머지 케이스에서 20일이 최적이었으며, CNN-GRU는 단기(10일) 및 중기(20일) 구간에서 우수한 결과를 보였다. 특히 CNN-LSTM의 ‘ALL’ 조합은 RMSE 6.976, MAE 5.390, MAPE 0.394%로 전체 실험 중 가장 균형 잡힌 성능을 나타냈으며, 이는 이벤트 및 감성 정보가 거시변수의 한계를 보완하며 환율의 단기·중기 변동을 효과적으로 설명함을 보여준다.

<표 13> 모델별 케이스별 최적 조합

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Model** | **케이스** | **Best Lookback** | **RMSE** | **MAE** | **MAPE** | **MedAE** | **MedAPE** |
| LSTM | Macro + Event | 20 | 6.946 | 5.384 | 0.394 | 4.562 | 0.333 |
| CNN\_LSTM | ALL | 20 | 6.976 | 5.39 | 0.394 | 4.519 | 0.335 |
| CNN\_GRU | Macro + Event | 20 | 6.996 | 5.383 | 0.394 | 4.44 | 0.326 |
| GRU | Macro Only | 20 | 7 | 5.424 | 0.397 | 4.585 | 0.337 |
| CNN\_LSTM | Macro Only | 30 | 7.052 | 5.426 | 0.397 | 4.533 | 0.331 |
| LSTM | Macro + Sentiment | 20 | 7.053 | 5.533 | 0.405 | 4.663 | 0.341 |
| LSTM | ALL | 20 | 7.086 | 5.567 | 0.407 | 4.652 | 0.341 |

종합적으로 볼 때, Lookback 20일 구간에서 모델의 예측력이 가장 높았으며, CNN-LSTM과 CNN-GRU 등 합성곱 기반 순환 모델이 전통적 RNN 구조보다 전반적으로 우수한 결과를 보였다. 이러한 결과는 CNN 계층이 단기적 급변 패턴을 포착하고, LSTM/GRU 계층이 시간적 연속성을 학습함으로써 서로 보완적인 역할을 수행했기 때문으로 해석된다. 또한 이벤트(GDELT)와 뉴스 감성 변수를 함께 포함한 ‘ALL’ 조합의 성능이 모든 Lookback 구간에서 안정적으로 우수하게 나타나, 거시적 요인에 더해 정성적 요인을 통합적으로 반영하는 접근법이 환율 예측에 효과적임을 실증적으로 확인할 수 있었다.

3.2 변수 중요도 분석

모델이 어떤 정보를 통해 예측력을 확보했는지를 정량적으로 파악하기 위해 Permutation Importance 분석을 수행하였다. 이 기법은 개별 피처의 값을 무작위로 섞은 후, 그로 인한 모델 예측 성능(MSE) 저하 정도를 계산함으로써 해당 피처가 모델 성능에 기여하는 상대적 중요도를 측정한다. 따라서 회귀계수에 직접 의존하지 않고, 비선형적 구조와 변수 간 상호작용을 고려할 수 있다는 점에서 딥러닝 기반 시계열 모델의 해석 가능성을 보완하는 데 유용하다.

분석 결과, S&P500 종가(0.397)가 전체 변수 중 가장 높은 중요도를 보였으며, 그 뒤를 경상수지(0.348), 다우존스 지수 (0.281)가 차지하였다. 이는 미국 증시 및 대외수지와 같은 글로벌 금융·거시 변수들이 원/달러 환율 변동에 가장 직접적인 영향을 미친다는 점을 보여준다. 특히 S&P500과 다우존스 지수는 투자심리와 위험자산 선호도를 반영하는 대표적 글로벌 지표로서, 환율의 단기·중기 변동에 큰 설명력을 제공하는 것으로 나타났다.

흥미로운 점은 events\_tone\_mean(0.229)이 네 번째로 높은 중요도를 기록했다는 것이다. 이는 GDELT 이벤트 데이터에서 추출된 사건 톤(감성)의 평균값으로, 글로벌 뉴스 내 긍·부정 정서의 편향이 환율 움직임과 유의한 상관관계를 가진다는 사실을 의미한다. 특히 이벤트 감성 변수의 영향력이 CPI(0.121)나 WTI(0.081) 등 전통적 거시 변수보다도 높게 나타났다는 점에서, 이벤트 정보가 경제 펀더멘털 변수 이상으로 시장 심리에 작용함을 보여준다.

그 외에도 CPI, WTI Price, KOSDAQ 종가, M2, Copper 가격 등이 주요 설명 변수로 확인되었다. 이는 물가·유동성·원자재 가격이 복합적으로 작용해 환율에 영향을 미친다는 점을 뒷받침한다. 반면 한국 정책금리, M1, usd/jpy, us10y 등의 변수는 상대적으로 낮은 중요도를 보였는데, 이는 단기 환율 변동 예측에서는 국내 단독 요인보다 글로벌 및 심리적 요인의 설명력이 더 크다는 점을 시사한다.

텍스트, 스크린샷, 소프트웨어, 번호이(가) 표시된 사진

AI 생성 콘텐츠는 정확하지 않을 수 있습니다.

**<그림 2> Permutation Importance (Top 20)**

3.3 시각화

본 절에서는 최적 구성(Macro+Event, LSTM, lookback=20)을 이용하여 2020년부터 2024년까지의 원/달러 환율 예측 결과를 시각화하였다<그림 3>. 그래프는 실제 값(Actual)과 모델의 예측 값(Predicted)을 전체 구간에 걸쳐 비교하였으며, 학습·검증에 사용된 구간과 테스트 구간은 파란색 점선으로 구분하였다.

전체 시계열 비교 결과, 모델은 환율의 장기적 추세뿐만 아니라 중단기적 변동 패턴을 비교적 안정적으로 추적하는 모습을 보였다. 특히 코로나19 이후 글로벌 유동성 확대 국면(2020–2021), 미 연준의 급격한 긴축과 환율 상승(2022), 그리고 이후 조정 국면(2023–2024)에 걸쳐 방향성과 국면 전환을 대체로 잘 포착하였다. 이는 이벤트 정보가 단순 시계열 신호를 넘어, 국면 전환기에서 유의미한 보조 지표로 작용했음을 시사한다.

또한 <그림 4>는 테스트 구간을 확대하여 실제 값과 예측 값의 오차를 상세히 비교한 것이다. 전반적으로 실제 값의 국지적 급등락을 예측하기에는 한계가 있었으나, 평균 수준과 변동성의 큰 방향성은 안정적으로 재현하였다.

결론적으로, 제안된 최적 모델은 환율의 구조적 추세와 단기적 변동성을 모두 일정 수준 포착하며, 이벤트·뉴스 감성 정보를 추가할 경우 단순 거시변수만을 사용할 때보다 전반적 예측력이 향상됨을 시각적으로 확인할 수 있었다.

텍스트, 그래프, 스크린샷, 라인이(가) 표시된 사진

AI 생성 콘텐츠는 정확하지 않을 수 있습니다.

**<그림 3> 원/달러 환율 예측 결과**

그래프, 스크린샷, 라인, 텍스트이(가) 표시된 사진

AI 생성 콘텐츠는 정확하지 않을 수 있습니다.

**<그림 4> Test 구간 예측 결과**

Ⅳ. 결론

본 연구는 원/달러 환율 예측의 정확도 향상을 위해 거시·시장 변수에 이벤트(GDELT) 및 뉴스 감성 요약 지표를 결합하고, LSTM, GRU, CNN-LSTM, CNN-GRU 네 가지 딥러닝 아키텍처를 비교·분석하였다. 2020년부터 2024년까지의 일별 시계열 데이터를 기반으로, 세 가지 시드(42, 55, 68)를 적용한 반복 실험을 통해 평균 성능을 산출하였다. 모든 모델은 동일한 학습 조건(epoch=80, batch=64, Adam optimizer, MSE 손실함수) 하에 학습되었으며, 입력 시퀀스 길이(lookback)를 5, 10, 20, 30, 60, 90일로 변화시켜 시간창 길이에 따른 예측력 변화를 검증하였다.

실험 결과, LSTM 모델의 ‘Macro + Event’ 조합과 lookback=20일 설정이 전체 실험 중 가장 우수한 성능(RMSE=6.946, MAE=5.384, MAPE=0.394%)을 기록하였다. 이는 약 1개월 규모의 시계열 윈도우가 환율의 정보 효율성이 가장 높은 구간임을 시사하며, 과도하게 짧거나 긴 윈도우(5일 이하, 60일 이상)는 오히려 노이즈 축적으로 인해 성능이 저하됨을 보여준다. LSTM 모델은 순환 구조를 통해 시계열의 장기 의존성을 안정적으로 학습하면서도, 이벤트(GDELT) 변수의 시점별 변화를 효과적으로 반영하여 단기 변동성까지 포착한 것으로 해석된다.

한편, CNN-LSTM과 CNN-GRU 모델은 합성곱 계층을 통한 단기 패턴 인식에 강점을 보여 전체적으로 낮은 RMSE를 유지했으나, LSTM 단일 구조의 단순성과 안정성이 중기 예측 구간(20일)에서는 오히려 더 높은 예측 효율을 보였다. GRU 모델은 전반적으로 LSTM 대비 오차가 크고 분산이 높아, 시계열 길이가 짧을수록 예측 변동성이 확대되는 경향을 보였다.

변수 중요도 분석(Permutation Importance) 결과에서는 S&P500 종가(0.397), 경상수지(0.348), 다우존스 종가(0.281)가 최상위에 위치하여, 글로벌 주식시장과 대외수지 요인이 환율 변동에 가장 강한 영향을 미치는 것으로 나타났다. 또한 events\_tone\_mean(0.229)이 네 번째로 높은 중요도를 기록하며, 뉴스 기반 사건의 정서적 톤이 전통적 거시 변수(CPI, WTI 등)보다 높은 설명력을 보였다. 이는 시장의 ‘심리적 방향성’이 환율의 단기 변동을 결정하는 주요 신호로 작용할 수 있음을 의미한다. 반면, 개별 이벤트(GDELT) 변수들은 일부 영향력을 보였으나 그룹 단위의 평균 중요도는 낮게 나타나, 데이터의 노이즈 및 국면 의존성이 작용했을 가능성이 제기된다.

시각화 결과에서도 제안된 CNN-LSTM 모델이 2020–2024년 전 기간 동안 환율의 구조적 추세와 단기 변동성을 안정적으로 추적하는 것으로 확인되었다. 특히 코로나19 팬데믹, 미 연준의 금리 인상, 글로벌 공급망 충격 등 주요 거시 이벤트 구간에서 환율의 방향성과 변동 폭을 실질적으로 반영하였다. 이는 사건 및 감성 정보가 단기 방향성 및 시장 불확실성 예측에 있어 보조지표로서 유용함을 실증적으로 보여준다.

종합하면, 본 연구는 (1) 거시·시장 변수에 사건·감성 요인을 결합한 복합 입력 구조의 유효성, (2) 약 20영업일 규모의 최적 시간창, (3) CNN-LSTM 기반 하이브리드 구조의 예측 안정성과 해석 가능성을 실증적으로 제시하였다. 이러한 결과는 환율 예측을 위한 실무적·정책적 활용 측면에서 의미가 있다. 금융기관은 사건·감성 지표를 보조 피처로 활용함으로써 환위험 관리 및 단기 환헤지 전략의 정밀도를 높일 수 있으며, 정책 당국은 시장 심리 변동을 조기 감지하는 데이터 기반 정책 보조지표로 활용할 수 있을 것이다.

다만, GDELT 이벤트 데이터의 노이즈와 중복성, 그리고 비정상 구간(예: 지정학적 충격)에서의 예측 불안정성은 본 연구의 한계로 남는다. 향후 연구에서는 다양한 뉴스·SNS 기반 감성 데이터의 통합, 시점별 중요도 동적 가중화, Transformer 기반 시계열 모델과의 비교 등을 통해 사건·감성 정보의 활용 가능성을 더욱 확장할 수 있을 것으로 기대된다.

부록

*A. 변수 상관관계 히트맵*

<그림 5>는 변수 간 상관관계 히트맵을 제시한 것이다. 결과를 보면 환율과 주가, 금리, 원자재 가격 등 주요 거시·시장 지표들은 서로 강한 양(+)의 상관 혹은 음(–)의 상관을 보이며, 전형적인 금융 변수들 간 공분산 구조가 나타난다. 뉴스 감성 지표들 역시 서로 밀접한 상관관계를 형성하고 있음을 확인할 수 있다.

패턴, 스크린샷, 사각형, 텍스트이(가) 표시된 사진

AI 생성 콘텐츠는 정확하지 않을 수 있습니다.

**<그림 5> 변수 상관관계 히트맵**

# Data Availability

The data can be provided on request.

# Declaration of Competing Interest

The authors declare no conflict of interest.

# CRediT authorship contribution statement

**Jaeheung Park:** Conceptualization, Software, Validation, Data Curation, Writing Original Draft, Preparation.

**Kyungwon Kim:** Methodology, Preparation, Software, Writing Review and Editing, Visualization, Supervision.

# References

[1] 한국은행. *환율의 정의*. 한국은행. https://www.bok.or.kr/portal/main/contents.do?menuNo=200407

[2] 김인준, & 이영섭. (2019). *국제경제론*. 박영사.

[3] 한국경제. (2023, 10월 22일). [다산칼럼] 환율 예측이 늘 빗나가는 까닭. *한국경제신문*. https://www.hankyung.com/article/2023102265501

[4] 한국은행. *환율의 변동요인*. 한국은행. https://www.bok.or.kr/portal/main/contents.do?menuNo=200407

[5] Plakandaras, V., Papadimitriou, T., & Gogas, P. (2015). Forecasting daily and monthly exchange rates with machine learning techniques. *Journal of Forecasting*, *34*(7), 560-573.

[6] 임현욱, 정승환, 이희수, & 오경주. (2021). 국고채, 금리 스왑 그리고 통화 스왑 가격에 기반한 외환시장 환율예측 연구: 인공지능 활용의 실증적 증거. *지식경영연구*, *22*(4), 71-85.

[7] Cao, W., Zhu, W., Wang, W., Demazeau, Y., & Zhang, C. (2020). A deep coupled LSTM approach for USD/CNY exchange rate forecasting. *IEEE Intelligent Systems*, *35*(2), 43-53.

[8] Qureshi, A. M. (2025). ML FORECASTING OF EXCHANGE RATES MACHINE LEARNING PREDICTION OF EXCHANGE RATES USING MACROECONOMIC INDICATORS Analysis and Use of Various ML Techniques (SVM, Random Forest). *Available at SSRN 5287273*.

[9] Wang, J., Wang, X., Li, J., & Wang, H. (2021). A prediction model of CNN-TLSTM for USD/CNY exchange rate prediction. *Ieee Access*, *9*, 73346-73354.

[10] Mohan, S., Mullapudi, S., Sammeta, S., Vijayvergia, P., & Anastasiu, D. C. (2019). Stock price prediction using news sentiment analysis. 2019 IEEE fifth international conference on big data computing service and applications (BigDataService)

[11] Jing, N., Wu, Z., & Wang, H. (2021). A hybrid model integrating deep learning with investor sentiment analysis for stock price prediction. *Expert Systems with Applications*, *178*, 115019.

[12] 정가연, 이혁제, 이준영, & 이제혁. (2024). 금융 특화 감정분석 모델과 딥러닝 시계열 예측 모델을 활용한 코스피 지수 예측. *대한산업공학회지*, *50*(4), 240-250.

[13] Ding, H., Shi, X., Deng, R., Faroog, S., Dewi, D. A., Abdullah, S. N., & Malek, B. A. (2024). Eur/usd exchange rate forecasting incorporating text mining based on pre-trained language models and deep learning methods. *arXiv preprint arXiv:2411.07560*.

[14] 김우석, & 한규식. (2021). COVID-19 가 원달러환율에 미친 영향. *금융지식연구*, *19*(1), 33-58.

[15] 오인정, & 김우주. (2022). SARIMA 와 ARDL 모형을 활용한 COVID-19 구간별 원/달러 환율 예측. *지능정보연구*, *28*(4), 191-209.

[16] GDELT Project. The GDELT Project. https://www.gdeltproject.org

[17] Schroeder, C., Winterbottom, S., Sitter, R., & Perdigones, J. L. (2013). Big Data Analysis of Human Societal Events: Indications for Forecasting Currency Exchange Rates.

[18] Consoli, S., Tiozzo Pezzoli, L., & Tosetti, E. (2020). Information extraction from the GDELT database to analyse EU sovereign bond markets. Workshop on Mining Data for Financial Applications,

[19] Blanqué, P., Ben Slimane, M., Cherief, A., Le Guenedal, T., Sekine, T., & Stagnol, L. (2022). Monitoring Narratives: an Application to the Equity Market. *Amina and Le Guenedal, Théo and Sekine, Takaya and Stagnol, Lauren, Monitoring Narratives: An Application to the Equity Market (April 7, 2022)*.

[20] Gupta, S., Ranjan, R., & Singh, S. N. (2024). Comprehensive study on sentiment analysis: From rule-based to modern llm based system. *arXiv preprint arXiv:2409.09989*.

[21] Cambria, E., Poria, S., Gelbukh, A., & Thelwall, M. (2018). Sentiment analysis is a big suitcase. *IEEE Intelligent Systems*, *32*(6), 74-80.

[22] Araci, D. (2019). Finbert: Financial sentiment analysis with pre-trained language models. *arXiv preprint arXiv:1908.10063*.

[23] Nasiopoulos, D. K., Roumeliotis, K. I., Sakas, D. P., Toudas, K., & Reklitis, P. (2025). Financial Sentiment Analysis and Classification: A Comparative Study of Fine-Tuned Deep Learning Models. *International Journal of Financial Studies*, *13*(2), 75.

[24] Shen, Y., & Zhang, P. K. (2024). Financial sentiment analysis on news and reports using large language models and finbert. 2024 IEEE 6th International Conference on Power, Intelligent Computing and Systems (ICPICS),

[25] Kang, J.-W., & Choi, S.-Y. (2025). Comparative investigation of gpt and finbert’s sentiment analysis performance in news across different sectors. *Electronics*, *14*(6), 1090.

[26] Barbella, M., & Tortora, G. (2022). Rouge metric evaluation for text summarization techniques. *Available at SSRN 4120317*.

[27] Zhang, M., Li, C., Wan, M., Zhang, X., & Zhao, Q. (2024). ROUGE-SEM: Better evaluation of summarization using ROUGE combined with semantics. *Expert Systems with Applications*, *237*, 121364.

[28] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735-1780.

[29] Zheng, W., & Chen, G. (2021). An accurate GRU-based power time-series prediction approach with selective state updating and stochastic optimization. *IEEE Transactions on Cybernetics*, *52*(12), 13902-13914.

[30] Qu, Y., & Zhao, X. (2019). Application of LSTM neural network in forecasting foreign exchange price. Journal of Physics: Conference Series,

[31] Agustin, F., & De Melin, P. (2024). Comparison of GRU and CNN Methods for Predicting the Exchange Rate of Argentine Peso (ARS) against US Dollar (USD). *International Journal Artificial Intelligent and Informatics*, *2*(1), 9-16.

[32] Islam, M. S., & Hossain, E. (2021). Foreign exchange currency rate prediction using a GRU-LSTM hybrid network. *Soft Computing Letters*, *3*, 100009.

[33] Lu, W., Li, J., Li, Y., Sun, A., & Wang, J. (2020). A CNN‐LSTM‐based model to forecast stock prices. *Complexity*, *2020*(1), 6622927.