

## [금융데이터 기반 최종 프로젝트 결과 보고서]

훈련과정명	[iM 뱅크] iM DiGital Banker Academy (데이터 분석 전문가 양성 과정)
진행 기간	2025. 07. 04 – 2025. 08. 01
팀 명	이건 다섯 번째 레슨, 좋은 건 너만 보기
팀원	김인호, 박승욱, 이나경, 이우태, 전은진
주제	iM 환율적 참견시점: 수출입 법인을 위한 외환 전략 부제: 현업 금융 마케팅을 위한 시계열 추이 패턴 분석 및 인사이트 도출

### 프로젝트 주제 선정 배경

1997년 12월 16일 우리나라는 원/달러 환율의 일일 변동폭 10%를 완전히 폐지함으로써 환율이 시장기능에 따라 움직이는 자유변동환율제도를 채택하였다. 2007년 충북대 경제학과 교수 이연호는 정부에서 선물환 시장을 적극 육성하여 기업들의 선물환 시장에 대한 접근을 용이하게 하고, 결제통화의 다변화와 선물환 시장을 적극 활용해 환율변동에 대한 방어 능력을 키우는 것이 필요하다고 제언하였다. 하지만, 2025년 2월 중소기업중앙회의 조사에 따르면, 중소기업 2곳 중 1곳은 환리스크에 대응하지 않고 있다는 결과가 확인되었다.

한국은행에서 발표한 중소기업 환율 리스크 분석 연구(2024)에 따르면, 환율 변동성이 증가하면 영업이익률이 감소하여 이 손실이 중소기업 부도로까지 이어지는 사실이 확인되었다. 수출입을 기반으로 하는 국내 기업들, 그중에서도 중소·중견기업은 환율의 급격한 변동에 직접적으로 노출되어 있는 것이다. 기업은 통상적으로 대금 결제가 특정 시점에 집중되는 구조를 가지고 있어, 환율이 불리하게 변동할 경우 큰 손실을 감수해야 하는 구조적 취약성이 있다. 그러나 중소기업의 경우, 환리스크에 대응할 수 있는 인력이 부족하여 환 리스크에 대한 전략이 부재하고, 환율 변동에 대한 대응은 여전히 사후적으로 환리스크에 대응하고 있다.

최근 글로벌 외환시장은 그 어느 때보다 복잡한 양상을 보이고 있다. 미국의 고금리 기조 장기화, 일본의 통화정책 전환 가능성, 중국의 경기 둔화 우려 등 주요국 경제정책이 상호 충돌하면서 전 세계 환율의 방향성과 변동폭에 큰 영향을 미치고 있다. 이에 따라 환율 예측 가능성은 점점 낮아지고 있으며, 기업들이 체감하는 환리스크 또한 빠르게 확대되고 있는 실정이다.

현재로서는 기존 시장에서 제공 중인 환율 정보 제공 서비스는 달러 위주로만 한정되어 있거나, 단순히 과거 데이터를 추세선으로 제공하는 수준에 머물러 있다. 자동화 시스템으로 기업의 실제 의사결정에 충분한 도움을 줄 수 있는 서비스가 부재하다. 수출입 기업이 보다 선제적으로 환리스크에 대응할 수 있도록 돋는 AI 기반의 환율 예측 및 전략 제안 시스템 구축이 시급한 과제로 떠오르고 있다.

본 프로젝트는 이러한 산업적·정책적 니즈에 주목하여, 단순한 예측을 넘어 기업이 실질적으로 활용할 수 있는 의사결정 지원 플랫폼을 제작하고자 한다. 이를 통해 수출입 기업이 환율 변동의 불확실성 속에서도 보다 안정적이고 전략적인 의사결정을 내릴 수 있도록 지원하는 것이 본 과제의 출발점이자 목표이다.

## 프로젝트 주제 문제 정의

환율 변동성 확대는 수출입 기업의 수익성과 재무 안정성에 직접적인 영향을 미치는 리스크 요인이다. 환율이 유리하게 작용할 경우 수익이 늘어나지만, 반대로 급등락이 발생할 경우에는 계약 이행조차 어려워질 정도로 피해가 확대될 수 있다. 특히 수출입 계약이 장기화되거나, 결제 통화가 복수 통화로 구성된 경우, 환율 변동은 예측 불가능한 손실로 이어지기 쉽다.

실제로 최근 환율 급등기에는 다수의 기업이 통화선도 계약으로 오히려 손실을 입는 사례가 나타났다. 이는 단순히 해지 수단을 활용하는 것만으로는 충분하지 않으며, 정확한 환율 예측과 그에 따른 전략적 의사결정이 병행되어야 함을 시사한다. 하지만 현재 기업이 활용할 수 있는 예측 기반 시스템은 제한적이며, 주로 전문가의 직관이나 과거 경험에 의존하는 방식에 머물러 있다.

또한, 환율 변동에 영향을 미치는 요인은 단순한 가격 흐름 외에도 금리 차, 정책 리스크, 글로벌 유동성, 경제 불확실성 지수(EPU), 원자재 가격, 뉴스 이슈 등 복합적인 외부 요인에 의해 결정된다. 이를 종합적으로 고려하지 않고 단순 시계열만으로 예측할 경우, 실제 기업이 필요로 하는 시점에 적절한 인사이트를 제공하기 어렵다.

따라서 본 프로젝트는 다음과 같은 문제를 정의한다:

- 주요 환율(USD/KRW, JPY/KRW, CNY/KRW 등)의 정밀한 예측을 위한 모델 설계가 필요하다.
- 수출입 구조 및 보유 통화, 결제 시점 등에 따라 다른 전략이 필요한 기업 상황을 반영해야 한다.
- 예측 결과를 단순 출력이 아닌 손익 시뮬레이션, 전략 제안, 정책 요약 등 실질적 의사결정에 활용 가능한 형태로 전달해야 한다.
- 경제 이벤트, 뉴스 토픽, EPU 지수 등 외부 요인을 반영하여 종합적 판단이 가능한 예측 시스템으로 확장해야 한다.

이러한 정의를 바탕으로 본 프로젝트는 환율 예측 모델 구축에 그치지 않고, 기업의 환리스크 대응 전략을 자동으로 제안하는 통합형 웹 서비스로 구현하는 것을 최종 목표로 설정하였다.

## 현황

2025년 기준, 글로벌 환율 환경은 점점 더 불안정해지고 있다. 원/달러 환율은 단기 급등락을 반복하며 예측 가능성이 낮아졌고, 미국의 재정건전성 악화, 일본 금리 기조 변화, 중국 경기 둔화 등 외생 변수들이 환율에 복합적으로 작용하고 있다. 특히 달러 인덱스(DXY)와 국제 유가, 비트코인 가격 변동성까지 환율에 영향을 주면서, 기업 입장에서는 환리스크에 대한 예측이 더욱 어려워진 상황이다.

여기에 더해, 2025년 8월부터 시행 예정인 미국의 대(對)한국 전 제품 25% 관세 부과 예고는 글로벌 외환 시장에 심각한 정책 리스크로 작용하고 있다. 실제 발표 직후 원/달러 환율은 1% 이상 급등했으며, 이는 정치·경제적 이벤트가 환율에 미치는 영향을 극명하게 보여주는 사례이다. 이러한 환경은 기업의 사전 대응 필요성을 더욱 강하게 뒷받침한다.

하지만 국내 기업들의 대응 역량은 구조적으로 미흡한 수준이다. 한국은행 실증 분석에 따르면, 환율 변동성이 커질수록 중소기업의 영업이익률은 하락하고 차입금 비율은 증가하며, 환위험이 실질적인 재무위기로 이어지는 경우도 많았다. 그럼에도 불구하고 많은 중소기업은 외환 파생상품이나 자연 해지 전략을 활용하지 못하고 있으며, 기존 정부·은행의 외환 지원도 사후적, 일률적 지원에 그치는 경우가 많다.

최근에는 경기도의 '옵션형 환변동보험' 전액 지원, IBK 기업은행의 환율 챌린지 이벤트, K-SURE의 환변동보험 확대 등 정책적 지원이 확대되고 있지만, 이는 환율 변동 이후에 대응하는 방식이 대부분이다. 기업 입장에서 환율 변동 전에 미리 예측하고 대응할 수 있는 선제적 시스템은 여전히 부족하다.

더욱이 기존 시중은행에서 제공하는 환율 전망 서비스는 달러 중심이고, 모델 정확도나 설명력이 부족하여, 다양한 통화를 거래하는 기업의 실제 판단에 도움이 되기 어렵다. 이는 실질적인 환리스크 대응의 공백으로 작용하고 있으며, 특히 인프라가 부족한 중소기업일수록 그 영향은 더 크다.

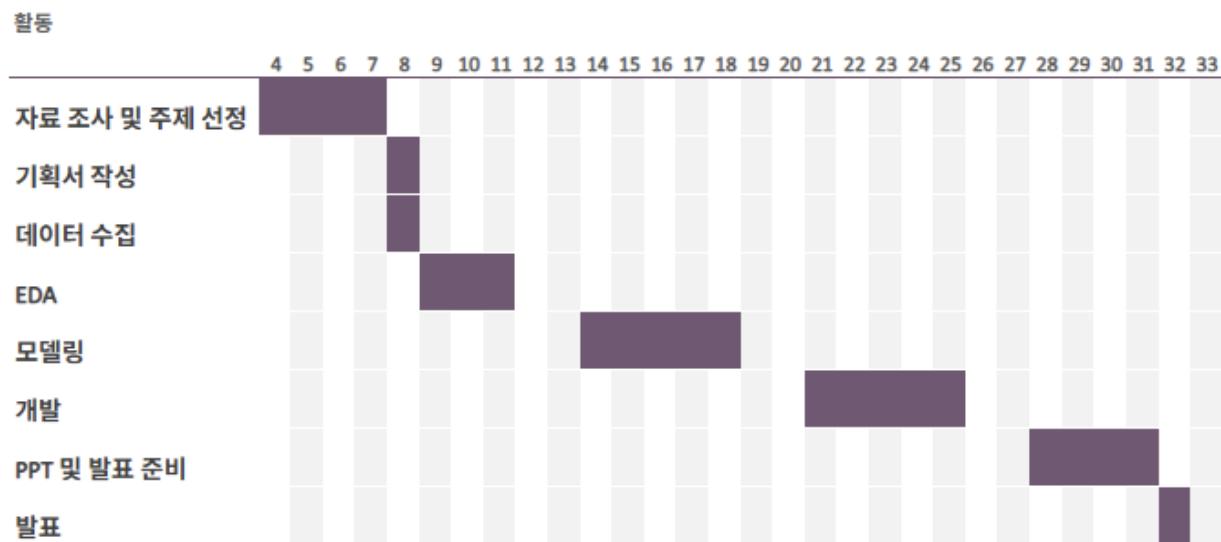
결과적으로, 국내 다수의 수출입 기업은 높은 환리스크 노출에도 불구하고, 정교한 예측과 전략 제안을 동시에 제공받을 수 있는 시스템이 부족하다. 본 프로젝트는 이와 같은 현실적인 공백을 해소하고, 기술과 데이터를 기반으로 실질적 환리스크 대응이 가능한 시스템을 구현함으로써 기업의 의사결정 품질과 재무 안정성 향상에 기여하고자 한다.

#### 팀 구성 및 역할

이름	역할
김인호	시각화/대시보드 설계, 기획, 발표
박승욱	리서치, EDA & 모델링
이나경	리서치, 기록자, 개발자
이우태	개발자, EDA & 모델링
전은진	팀장, 개발자, EDA & 모델링

[표-1] 팀 구성 및 역할

## 진행 일정



[표-2] 진행 일정

## 프로젝트 목표

본 프로젝트는 환율 변동성에 직접적으로 노출되는 수출입 기업을 대상으로, 정밀한 환율 예측 모델을 기반으로 한 환리스크 대응 솔루션을 제공하는 것을 목표로 한다. 이를 통해 궁극적으로 ① 고도화된 환율 예측 모델의 구축과 ② 예측 결과를 활용한 전략 제안형 웹 시스템 구현이라는 두 가지 핵심 목표를 달성하고자 한다.

첫째, 정밀한 환율 예측 모델을 구축하고자 한다. LSTM 과 같은 시계열 딥러닝 모델 뿐 아니라 XGBoost, RandomForest 등의 머신러닝 기반 모델을 함께 활용하여, 주요 통화(USD/KRW, JPY/KRW 등)의 향후 단기 및 중기 환율 변동을 예측한다. 예측 정확도를 높이기 위해 환율 자체의 기술적 지표 뿐 아니라, 거시경제, 금융시장, 무역지표, 원자재 가격, 시장심리 등 다양한 변수군을 포함한다. 주요 변수에는 환율 종가, 이동평균선, 모멘텀 수익률 등 기초 시계열 정보 뿐 아니라, 기준금리, CDS 프리미엄, 외국인 투자 흐름, 무역수지, 제조업 PMI, 소비자심리지수, 유가, 금, DXY 등 경제 환경을 다면적으로 반영한 지표들이 포함된다. 또한 RSI, MACD, 볼린저 밴드 등 시장 기술 지표 역시 예측 모델의 정교화에 활용된다.

둘째, 실시간 환율 예측 결과를 기반으로 전략을 제공하는 통합형 웹 서비스를 구현한다. 이 시스템은 수출입 기업이 환율 리스크에 효과적으로 대응할 수 있도록 다음과 같은 기능을 제공한다.

### ● 기능 1: 날짜별 주요 이슈 및 경제지표 흐름 제공

사용자가 원하는 날짜를 선택하면, 해당일의 환율에 영향을 미칠 수 있는 주요 경제 뉴스 키워드를 확인할 수 있도록 지원한다. 주요 뉴스와 함께 EPU 지수, 환율 추세 그래프 등을 종합적으로 시각화 하여 직관적인 정보 탐색이 가능하도록 구성하였다.

### ● 기능 2: 환율 예측기 및 전략 제안

사용자가 예측 대상 통화, 예측 시점, 보유 금액, 기업 유형(수입/수출)을 입력하면, 환율 예측 결과를 바탕으로 손익 시뮬레이션을 수행하고 맞춤형 환리스크 대응 전략을 제안한다. 전략 제안은 기업의 수출입 특성에 따라 자동으로 조정된다.

### ● 기능 3: 최신 통화정책 방향 요약 제공

한국은행의 통화정책 발표 내용을 간략히 요약해 제공함으로써, 사용자가 최근의 정책 기조를 쉽게 파악할 수 있도록 지원한다.

이러한 기능을 통해 본 프로젝트는 기업별 맞춤형 환 위험 전략 수립을 돕고, 환율 예측 기반의 실용적 인사이트를 제공함으로써 실제 비즈니스에 활용 가능한 데이터 분석 시스템으로 발전시키는 것을 지향한다.

### 데이터

#### [USD]

테이블 정의서		프로젝트명	환율적 참견 시점 : 수출입 법인을 위한 외환 전략		
		작성자	이나경, 김인호, 이우태		
		Table 명	달러		
No	컬럼ID	컬럼명	Type	내용	비고
1	KRW=X	환율	float	미국 달러 대비 원화 환율 증가	예측 대상 변수
2	CL=F	WTI유	float	서부 텍사스산 중질유(WTI) 증가	입력 변수
3	GC=F	금	float	금 선물 증가	입력 변수
4	^GSPC	S&P500	float	미국 S&P500 주가지수 증가	입력 변수
5	^DJI	다우존스	float	미국 다우존스 산업평균지수 증가	입력 변수
6	CNY=X	위안	float	위안화 대비 미국 달러 환율 증가	입력 변수
7	^N225	닛케이	float	일본 닛케이225 지수 증가	입력 변수
8	^KS11	코스피	float	한국 코스피 지수 증가	입력 변수
9	^SIXC	나스닥	float	미국 나스닥 종합지수 증가	입력 변수
10	VIX	VIX	float	미국 시장 변동성 지수(Volatility Index) 증가	기준 대시
11	MA20	20일 이동평균	float	20일간 환율의 단순 이동평균	기술적 지표
12	STD20	20일 표준편차	float	20일간 환율의 표준편차	기술적 지표
13	Upper	볼린저 상단선	float	MA20 + 2 × STD20 값, 볼린저밴드 상단	기술적 지표
14	Lower	볼린저 하단선	float	MA20 - 2 × STD20 값, 볼린저밴드 하단	기술적 지표
15	EMA12	EMA12	float	12일 지수이동평균 (Exponential Moving Average)	기술적 지표
16	EMA26	EMA26	float	26일 지수이동평균	기술적 지표
17	MACD	MACD	float	EMA12 - EMA26, 이동평균 수렴 확산지수	기술적 지표
18	Signal	시그널선	float	MACD의 9일 지수이동평균	기술적 지표
19	RSI	RSI	float	14일 상대강도지수 (Relative Strength Index)	기술적 지표
20	Target	환율 방향 클래스	int	다음날 환율의 변화 방향 (0: 보합, 1: 상승, 2: 하락, 기준: 11월 차이)	분류 타겟 변수
21	alpha1~101	알파변수	float	시계열예측을 위한 파생변수	기술적 지표
22	중국환율_시가	중국환율 시가	float	중국 위안화 대비 환율 관련 데이터	입력 변수
23	중국환율_고가	중국환율 고가	float	중국 위안화 대비 환율 관련 데이터	입력 변수
24	중국환율_저가	중국환율 저가	float	중국 위안화 대비 환율 관련 데이터	입력 변수
25	중국환율_종가	중국환율 종가	float	중국 위안화 대비 환율 관련 데이터	예측 대상 변수
26	중국환율_변동%	중국환율 변동률	float	중국 위안화 대비 환율 관련 데이터	입력 변수
27	WTI유_종가	WTI유 종가	float	WTI 원유 가격 정보	입력 변수
28	WTI유_시가	WTI유 시가	float	WTI 원유 가격 정보	입력 변수
29	WTI유_고가	WTI유 고가	float	WTI 원유 가격 정보	입력 변수
30	WTI유_저가	WTI유 저가	float	WTI 원유 가격 정보	입력 변수
31	WTI유_거래량	WTI유 거래량	float	WTI 원유 가격 정보	입력 변수
32	WTI유_변동%	WTI유 변동률	float	WTI 원유 가격 정보	입력 변수
33	금_종가	금 종가	float	금의 시세	입력 변수
34	금_시가	금 시가	float	금의 시세	입력 변수
35	금_고가	금 고가	float	금의 시세	입력 변수
36	금_저가	금 저가	float	금의 시세	입력 변수
37	금_변동%	금 변동률	float	금의 시세	입력 변수
38	S&P500_종가	S&P500 종가	float	미국 S&P500 지수	입력 변수
39	S&P500_시가	S&P500 시가	float	미국 S&P500 지수	입력 변수
40	S&P500_고가	S&P500 고가	float	미국 S&P500 지수	입력 변수
41	S&P500_저가	S&P500 저가	float	미국 S&P500 지수	입력 변수
42	S&P500_변동%	S&P500 변동률	float	미국 S&P500 지수 (단위: %)	입력 변수
43	다우존스_종가	다우존스 종가	float	미국 다우존스 산업평균지수	입력 변수
44	다우존스_시가	다우존스 시가	float	미국 다우존스 산업평균지수	입력 변수
45	다우존스_고가	다우존스 고가	float	미국 다우존스 산업평균지수	입력 변수
46	다우존스_저가	다우존스 저가	float	미국 다우존스 산업평균지수	입력 변수
47	다우존스_거래량	다우존스 거래량	float	미국 다우존스 산업평균지수	입력 변수
48	다우존스_변동%	다우존스 변동률	float	미국 다우존스 산업평균지수 (단위: %)	입력 변수
49	상해종합_고가	상해종합 고가	float	중국 상하이 증권거래소에 상장된 기업을 대상으로 구성된 종합 주가 지수	입력 변수
50	상해종합_저가	상해종합 저가	float	중국 상하이 증권거래소에 상장된 기업을 대상으로 구성된 종합 주가 지수	입력 변수
51	상해종합_거래량	상해종합 거래량	float	중국 상하이 증권거래소에 상장된 기업을 대상으로 구성된 종합 주가 지수	입력 변수
52	상해종합_변동%	상해종합 변동률	float	중국 상하이 증권거래소에 상장된 기업을 대상으로 구성된 종합 주가 지수 (단위: %)	입력 변수
53	닛케이_종가	닛케이 종가	float	일본 닛케이225 주가 지수	입력 변수
54	닛케이_시가	닛케이 시가	float	일본 닛케이225 주가 지수	입력 변수
55	닛케이_고가	닛케이 고가	float	일본 닛케이225 주가 지수	입력 변수
56	닛케이_저가	닛케이 저가	float	일본 닛케이225 주가 지수	입력 변수
57	닛케이_변동%	닛케이 변동률	float	일본 닛케이225 주가 지수 (단위: %)	입력 변수
58	코스피_종가	코스피 종가	float	한국 코스피 지수	입력 변수
59	코스피_고가	코스피 고가	float	한국 코스피 지수	입력 변수
60	코스피_저가	코스피 저가	float	한국 코스피 지수	입력 변수
61	코스피_거래량	코스피 거래량	float	한국 코스피 지수	입력 변수
62	코스피_변동%	코스피 변동률	float	한국 코스피 지수 (단위: %)	입력 변수
63	나스닥_종가	나스닥 종가	float	미국 기술주 중심의 나스닥 지수	입력 변수
64	나스닥_시가	나스닥 시가	float	미국 기술주 중심의 나스닥 지수	입력 변수
65	나스닥_저가	나스닥 저가	float	미국 기술주 중심의 나스닥 지수	입력 변수
66	나스닥_거래량	나스닥 거래량	float	미국 기술주 중심의 나스닥 지수	입력 변수
67	나스닥_변동%	나스닥 변동률	float	미국 기술주 중심의 나스닥 지수 (단위: %)	입력 변수
68	VIX_종가	VIX 종가	float	S&P500 옵션시장에서 파생된 변동성 지수	입력 변수
69	VIX_시가	VIX 시가	float	S&P500 옵션시장에서 파생된 변동성 지수	입력 변수

70	VIX_고가	VIX 고가	float	S&P500 옵션시장에서 파생된 변동성 지수	입력 변수
71	VIX_저가	VIX 저가	float	S&P500 옵션시장에서 파생된 변동성 지수	입력 변수
72	VIX_거래량	VIX 거래량	float	S&P500 옵션시장에서 파생된 변동성 지수	입력 변수
73	VIX_변동%	VIX 변동률	float	S&P500 옵션시장에서 파생된 변동성 지수 (단위: %)	입력 변수
74	시가총액_전체	시가총액 전체	float	시가총액	입력 변수
75	시가총액_외국인보유	시가총액 외국인보유	float	외국인 보유 시가총액	입력 변수
76	시가총액_비율	시가총액 비율	float	시가총액 비율 (단위: %)	입력 변수
77	주식수_전체	전체 주식수	float	전체 주식수	입력 변수
78	주식수_외국인보유	외국인보유 주식수	float	외국인 보유 주식수	입력 변수
79	주식수_비율	주식수 비율	float	주식수 비율 (단위: %)	입력 변수
80	CD금리(91일)	CD금리(91일)	float	CD 금리	입력 변수
81	미국정책금리	미국정책금리	float	미국정책금리	입력 변수
82	한국(M1)변동%	한국(M1)변동률	float	한국 M1 통화량 (단위: %)	입력 변수
83	한국(M2)조원	한국(광의통화)조원	float	한국 M2 통화량 (단위: 조원)	입력 변수
84	한국(M2)변동%	한국(M2)변동률	float	한국 M2 통화량 변동 (단위: %)	입력 변수
85	미국(M1)십 억달러	미국(광의통화)십 억달러	float	미국 M1 통화량 (단위: 십 억달러)	입력 변수
86	미국(M2)십 억달러	미국(광의통화)십 억달러	float	미국 M2 통화량 (단위: 십 억달러)	입력 변수
87	소비자심리지수	소비자 심리지수	float	국내 소비자심리지수	입력 변수
88	생산자물가지수	생산자 물가지수	float	국내 생산자심리지수	입력 변수
89	산업생산지수	산업 생산지수	float	국내 산업생산지수	입력 변수
90	외환보유액(억달러)	외환 보유액(억달러)	float	외환 보유액 (단위: 억달러)	입력 변수
91	경상수지	경상수지	float	경상수지	입력 변수
92	미국소비자물가지수	미국소비자물가지수	float	미국소비자물가지수	입력 변수
93	비트코인_검색량	비트코인 검색량	int	구글 검색 트렌드 데이터를 기반으로 한 비트코인 검색량	입력 변수
94	일본엔_검색량	일본엔 검색량	int	구글 검색 트렌드 데이터를 기반으로 한 일본엔 검색량	입력 변수
95	유로(EUR)_검색량	유로(EUR) 검색량	int	구글 검색 트렌드 데이터를 기반으로 한 유로(EUR) 검색량	입력 변수
96	S&P 500_검색량	S&P 500 검색량	int	구글 검색 트렌드 데이터를 기반으로 한 S&P 500 검색량	입력 변수
97	코스피_검색량	코스피 검색량	int	구글 검색 트렌드 데이터를 기반으로 한 코스피 검색량	입력 변수
98	중국EPU	중국 경제 불확실성 지수	int	중국 EPU지수	입력 변수

[표-3] USD/KRW 예측 모델 변수 정의서

## [CNY]

테이블 정의서		프로젝트명	환율적 참견 시점 : 수출입 법인을 위한 외환 전략		
		작성자	이나경, 김인호, 이우태		
		Table 명	위안		
No	컬럼ID	컬럼명	Type	내용	비고
1	Date	날짜	datetime	2007-01-14부터 2025-07-28까지	시계열 데이터 기준 날짜
2	Shanghai Composite_Open	상해종합주가지수 시가	float	중국 상하이 증권거래소에 상장된 기업을 대상으로 구성된 종합 주가 지수	입력 변수
3	Shanghai Composite_High	상해종합주가지수 고가	float		입력 변수
4	Shanghai Composite_Low	상해종합주가지수 저가	float		입력 변수
5	Shanghai Composite_Close	상해종합주가지수 종가	float		입력 변수
6	Shenzhen Component_Open	선전 종합 지수 시가	float	선전 중시에 상장된 모든 기업들의 주가를 시가총액으로 가중 평균하여 계산하는	입력 변수
7	Shenzhen Component_High	선전 종합 지수 고가	float		입력 변수
8	Shenzhen Component_Low	선전 종합 지수 저가	float		입력 변수
9	Shenzhen Component_Close	선전 종합 지수 종가	float		입력 변수
10	Hang Seng_Open	항셍지수 시가	float	홍콩증권거래소에 상장된 모든 기업의 시총을 대표하는 홍콩증시의 대표 주가지	입력 변수
11	Hang Seng_High	항셍지수 고가	float		입력 변수
12	Hang Seng_Low	항셍지수 저가	float		입력 변수
13	Hang Seng_Close	항셍지수 종가	float		입력 변수
14	Hang Seng China Enterprises_Open	홍콩H지수 시가	float	HSCEI	입력 변수
15	Hang Seng China Enterprises_High	홍콩H지수 고가	float	홍콩증권거래소에 상장된 중국 국영 기업들 중 우량 기업들을 모아 만든 지수	입력 변수
16	Hang Seng China Enterprises_Low	홍콩H지수 저가	float		입력 변수
17	Hang Seng China Enterprises_Close	홍콩H지수 종가	float		입력 변수
18	Close	위안 환율 종가	float	위안 환율	예측 대상 변수
19	Open	위안 환율 시가	float		입력 변수
20	High	위안 환율 고가	float		입력 변수
21	Low	위안 환율 저가	float		입력 변수
22	Change	위안 환율 변화율	float		입력 변수

[표-4] CNY/KRW 예측 모델 변수 정의서

## [JPY]

테이블 정의서		프로젝트명	환율적 참견 시점 : 수출입 법인을 위한 외환 전력		
No	컬럼ID	컬럼명	Type	내용	비고
1	Date	날짜	datetime	2013-03-04부터 2021-10-01까지	시계열 데이터 기준 날짜
2	Nikkei225_Open	닛케이225 시가	float	일본 닛케이 평균주가	입력 변수
3	Nikkei225_High	닛케이225 고가	float		입력 변수
4	Nikkei225_Low	닛케이225 저가	float		입력 변수
5	Nikkei225_Close	닛케이225 종가	float		입력 변수
6	TOPIX_Open	TOPIX 시가	float	TOPIX Growth Market 250 Index	입력 변수
7	TOPIX_High	TOPIX 고가	float		입력 변수
8	TOPIX_Low	TOPIX 저가	float		입력 변수
9	TOPIX_Close	TOPIX 종가	float		입력 변수
10	Mothers_Open	MTHR 시가	float	Tokyo Stock Exchange Growth Market 250 Index	입력 변수
11	Mothers_High	MTHR 고가	float		입력 변수
12	Mothers_Low	MTHR 저가	float		입력 변수
13	Mothers_Close	MTHR 종가	float		입력 변수
14	Close	엔 환율 종가	float	엔 환율	예측 대상 변수
15	Open	엔 환율 시가	float		입력 변수
16	High	엔 환율 고가	float		입력 변수
17	Low	엔 환율 저가	float		입력 변수
18	Change	엔 환율 변화율	float		입력 변수

[표-5] JPY/KRW 예측 모델 변수 정의서

## [EUR]

테이블 정의서		프로젝트명	환율적 참견 시점 : 수출입 법인을 위한 외환 전력		
No	컬럼ID	컬럼명	Type	내용	비고
1	Date	날짜	datetime	2007-03-30부터 2025-07-28까지	시계열 데이터 기준 날짜
2	DAX_Open	DAX 시가	float	독일 프랑크푸르트 증권거래소에 상장된 40개 주요 기업의 주가를 기준으로 산출되는 독일 대표 주가지수	입력 변수
3	DAX_High	DAX 고가	float		입력 변수
4	DAX_Low	DAX 저가	float		입력 변수
5	DAX_Close	DAX 종가	float		입력 변수
6	EUROSTOXX50_Open	EUROSTOXX50 시가	float	유로존(유로화를 사용하는 유럽 국가들의 경제 연합)의 주요 기업 50개를 대표하는 주가지수	입력 변수
7	EUROSTOXX50_High	EUROSTOXX50 고가	float		입력 변수
8	EUROSTOXX50_Low	EUROSTOXX50 저가	float		입력 변수
9	EUROSTOXX50_Close	EUROSTOXX50 종가	float		입력 변수
10	CAC_Open	CAC 시가	float	유로넥스트 파리에 상장된 40개의 주요 기업 주식 가격을 바탕으로 산출되는 프랑스의 대표적인 주가지수	입력 변수
11	CAC_High	CAC 고가	float		입력 변수
12	CAC_Low	CAC 저가	float		입력 변수
13	CAC_Close	CAC 종가	float		입력 변수
14	Close	유로 환율 종가	float	유로 환율	예측 대상 변수
15	Open	유로 환율 시가	float		입력 변수
16	High	유로 환율 고가	float		입력 변수
17	Low	유로 환율 저가	float		입력 변수
18	Change	유로 환율 변화율	float		입력 변수

[표-6] EUR/KRW 예측 모델 변수 정의서

## [감성분석]

최종적으로는 폐기된 모델에 사용된 테이블 정의서이다.

테이블 정의서		프로젝트명	환율적 참견 시점 : 수출입 법인을 위한 외환 전력	
No	컬럼ID	컬럼명	Type	내용
1	뉴스_식별자	뉴스_식별자	object	뉴스_식별자
2	일자	날짜	datetime	뉴스_발행_일자
3	키워드	키워드	object	뉴스_키워드_주제
4	특성추출(가중치순 상위 50개)	특성추출	object	뉴스_키워드_가중치순_상위_50개
5	본문	본문	object	뉴스_본문
6	sentiment	감성_지표	float	감성_지표

[표-7] 감성분석 변수 정의서

## [환율변동성지수]

테이블 정의서		프로젝트명	환율적 참견 시점 : 수출입 법인을 위한 외환 전력	
No	컬럼ID	컬럼명	Type	내용
1	뉴스_식별자	Date	datetime64	일자
2	일자	일자	datetime64	EPU가 해당하는 월로 연-월-01로 구성
3	미환율_증가	미환율_증가	float64	원달러_증가
4	한국EPU	한국EPU	float64	공식 제공되고 있는 한국 EPU
5	one_Korea_EPU	one_Korea_EPU	float64	한국 EPU 개발 논문을 기반으로 국내 12개 주요 신문을 활용해 산출한 EPU 지수
6	two_Korea_EPU	two_Korea_EPU	float64	논문에 수록된 환율, 재정 등 정책 관련 단어를 EPU 범주에 재분류한 확장형 지수
7	three_Korea_EPU	three_Korea_EPU	float64	KDI(한국개발연구원)가 공식 제공하는 EPU 관련 단어집 기반 지수

[표-8] 환율변동성지수 변수 정의서

## [통화정책방향 브리핑 ]

테이블 정의서		프로젝트명	환율적 참견 시점 : 수출입 법인을 위한 외환 전력	
No	컬럼ID	컬럼명	Type	내용
1	Date	Date	object	통화정책방향 브리핑 개최일자
2	Content	Content	object	통화정책방향 브리핑 자막
3	default_summary	default_summary	object	기본 요약 내용
4	sola_summary	sola_summary	object	COSTAR 프롬프트 적용한 요약 내용

테이블 정의서		프로젝트명	환율적 참견 시점 : 수출입 법인을 위한 외환 전력	
No	컬럼ID	컬럼명	Type	내용
1	Date	Date	object	통화정책방향 브리핑 개최일자
2	Content	Content	object	통화정책방향 브리핑 자막
3	default_summary	default_summary	object	기본 요약 내용
4	sola_summary	costar_summary	object	COSTAR 프롬프트 적용한 요약 내용

[표-9] 통화정책방향 브리핑 변수 정의서

## [뉴스]

환율변동성지수 산출과 월별 주요이슈 토픽모델링에 사용한 테이블 정의서이다.

테이블 정의서		프로젝트명	환율적 참견 시점 : 수출입 법인을 위한 외환 전력	
No	컬럼ID	컬럼명	Type	내용
1	일자	일자	int64	연_월_일
2	키워드	키워드	object	뉴스_키워드
3	특성추출(가중치순 상위 50개)	특성추출(가중치순 상위 50개)	object	가중치순_상위_50개
4	본문	본문	object	뉴스_본문 내용
5	제목	제목	object	뉴스_제목

[표-10] 뉴스 토픽모델링 변수 정의서

## 진행 내용

### 1. 활용 도구 및 기법

**분석 도구:** Python, Pandas, NumPy, Matplotlib, Seaborn

**모델링:** LSTM, XGBoost, RandomForest, 감성 분석(BERT 기반)

**시각화/UI:** Streamlit

**기타 기법:** EPU 지수 활용, 뉴스 토픽 모델링, COSTAR 프롬프트 프레임워크

### 2. WBS 일정표

이건 다섯 번째 레슨 좋은 건 너만 보기 WBS							
과제명	iM 환율적 참견시점: 수출입 법인을 위한 외환 전략					일정관리담당자	전운진
팀장명	전운진	과제기간	2025.07.04 ~ 2025.07.30			최종작성일	기간경과율
WBS코드	작업이름	담당자	기간	시작날짜	예상 완료 날짜	실제 완료날짜	계획율
1.1	Sprint 1 - 프로젝트 준비						
1.1.1	자료 조사 및 주제 선정		4	2025. 7. 4	2025. 7. 7	2025. 7. 7	100%
1.1.2	기획서 작성		1	2025. 7. 8	2025. 7. 8	2025. 7. 8	100%
1.2	Sprint2- 데이터 수집 및 전처리						
1.2.1	법인 데이터 전처리		1	2025. 7. 8	2025. 7. 8	2025. 7. 8	100%
1.2.2	외부 데이터 수집		1	2025. 7. 8	2025. 7. 8	필요시 처리	
1.3	Sprint3 - EDA 및 피처 생성						
1.3.1	EDA 수행 및 주요 인사이트 도출		5	2025. 7. 9	2025. 7. 13	2025. 7. 13	50%
1.4	Sprint4 - 모델링 및 평가						
1.4.1	원달러 환율 예측 모델 생성	이우태, 김인호	7	2025. 7. 14	2025. 7. 20	2025.7.29	200%
1.4.2	일별 환율변동성 지표 제작	전운진	7	2025. 7. 14	2025. 7. 20	2025.7.26	200%
1.4.3	뉴스기반 감성분석	이나경, 박승욱	7	2025. 7. 14	2025. 7. 20	2025.7.20	100%
1.4.4	프롬프트 요약 모델 생성	전운진	7	2025. 7. 14	2025. 7. 20	2025.7.20	100%
1.5	Sprint5 - 시스템 구축						
1.5.1	API 연결		7	2025. 7. 21	2025. 7. 27	2025.7.29	133%
1.5.2	Streamlit 구현	이나경, 전운진	7	2025. 7. 21	2025. 7. 27	2025.7.29	133%
1.6	Sprint6 - 발표						
1.6.1	최종 보고서 작성		3	2025. 7. 28	2025. 7. 30	2025.7.30	100%
1.6.2	발표자료 제작		3	2025. 7. 28	2025. 7. 30	2025.7.30	100%
1.6.3	발표		1	2025. 7. 31	2025. 7. 31	2025.7.30	100%

[표-11] WBS 일정표

### 3. 초기 프로젝트 기획 내용

#### [감성분석 환율 예측 모델 폐기]

초기에 일별 뉴스 감성분석을 단독으로 환율 예측, 일별 환율 상승/하락 예측을 시도했으나, 성능의 한계로 감성분석은 폐기하였다. 참고한 논문에서도 감성분석을 활용한 논문 성능이 accuracy 0.5~0.6 이 최대였으나, 한 번 시도해보자는 의의로 감성분석 기반의 환율 예측 모델을 개발하였다. 하지만, 논문과 동일하게 하락을 0.6 정도로 예측하고 상승은 0.3 정도 예측하는 등 0.7 이상을 넘지 못했다. 실무에서의 활용성이 부족하다고 생각되어 감성분석 환율 예측 모델을 폐기하게 되었다.

#### [최종 조합 적용 결과]

```
change_cut: 0.03, alpha: 0.7, threshold: 0.5000, class_weight_1: 12.0, macro f1-score: 0.4959
```

#### 분류 리포트:

	precision	recall	f1-score	support
0	0.72	0.65	0.68	261
1	0.28	0.35	0.31	103
accuracy			0.56	364
macro avg	0.50	0.50	0.50	364
weighted avg	0.59	0.56	0.58	364

#### 혼동 행렬:

```
[[169  92]
 [ 67  36]]
```

[표-12] 감성분석 환율 예측 모델 classification\_report

#### [환율 예측 분류 모델의 어려움]

주요하게 참고한 논문 '딥러닝을 활용한 원화 환율 예측: 시장 및 웹데이터와 거시경제 지표의 활용(황미라, 2020)'에서 환율을 분류로 예측한 모델 성능은 accuray 71.42%에 달했다. 자체 구축한 환율 예측 모델의 경우 f1-score 가 최대 상승을 0.34, 하락을 0.7 정도 예측할 수 있었다.

표 3.6: 데이터별 환율 방향성 예측력 비교

다음 표는 각 데이터별로 환율 방향성을 예측한 결과이다. 모든 데이터의 조합을 사용하였을 때 예측력이 가장 높음을 확인할 수 있다.

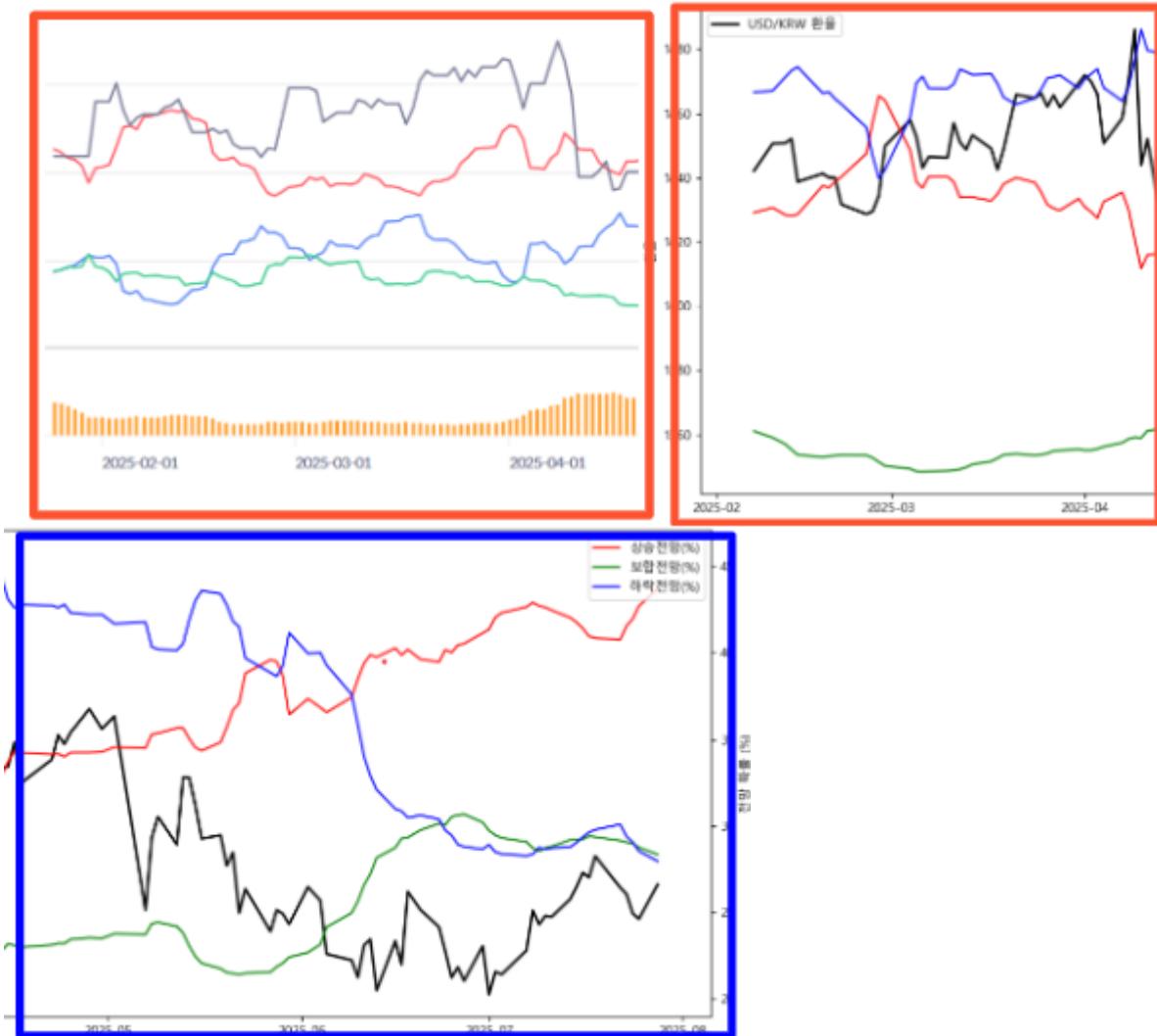
Model	Market	Macro	Web	ACC
LSTM	o	o	o	71.242%
	o	x	x	68.537%
	x	o	x	65.638%
	x	x	o	67.556%
	o	x	o	69.339%
	o	o	x	68.637%
	x	o	o	68.136%

#### 분류 리포트:

	precision	recall	f1-score	support
0	0.75	0.67	0.71	390
1	0.30	0.39	0.34	145

[표-13] 레퍼런스 성능과 개발한 모델 성능 비교

아래의 왼쪽 사진은 KB 국민은행에서 제공중인 일변 환율 상승/하락/보합 확률이고 오른쪽은 자체 구축한 환율 분류 모델의 상승/하락 확률이다. KB의 경우에는 상승할 확률이 제일 컸지만, 자체 구축한 모델의 경우 강력하게 다음 시점이 하락할 것으로 예측하였다. 대폭 하락 전 하락을 예측하는 등 분류 모델의 성능이 좋았고 실제로 다음 시점에서 급격히 환율이 하락하기 시작하였다.



[표-14] USD/KRW 기준 환율 예측 그래프와 직접 개발한 환율 예측 그래프 비교

논문에서 accuracy 70%가 최대였던 점과 KB의 상승/하락/보합 분류 모델과 비교하였을 때, 분류 모델의 성능이 논문만큼 구현한 것으로 추정되었다. 하지만, 성능 지표적으로 상승, 하락의 예측이 f1-score 를 0.7 을 넘기지 못하여 활용성 측면에서 보다 많은 기능을 소비자에게 제공할 수 있는 플랫폼을 기획하고자 환율 예측을 회귀 모델로서 진행하였다.

## 4. 최종 프로젝트 기획 내용

감성분석을 활용한 환율 예측 모델을 폐기하는 대신, 한국은행 총재가 발표하는 통화정책방향 브리핑 요약 서비스, 주식 시장의 공포 탐욕 지수를 벤치마킹하여 환율 변동성 지수를 개발하고자 하였다.

### [새로운 서비스 제공 - 환율변동성지수&뉴스 기반 월별 주요 이슈]

원달러와 상관계수가 상위 3 번째이면서도 뉴스 기반으로 만들어져 정성적인 활용까지 가능한 한국 EPU를 지수를 환율변동성지수로 사용하였다.

#### - 환율변동성지수

**첫번째. EPU 지수 개발자 Barker 가 제공하는 공식 한국 EPU 지수 데이터**

**두번째. 한국 EPU 개별 논문을 기반으로 국내 12 개 주요 신문을 활용해 자체 산출**

**세번째. 논문에 수록된 환율, 재정 등 정책 관련 단어를 E·P·U 범주에 재분류한 확장형 지수**

**네번째. KDI(한국개발연구원)가 공식 제공하는 EPU 관련 단어집 기반**

총 4 가지 방법으로 만들어진 한국 EPU 지수를 비교하였고 최종적으로 성능이 좋았던 세번째 확장형 EPU 지수를 환율변동성지수로 최종 결정지었다.

#### - 뉴스 기반 월별 주요 이슈

EPU에 사용한 데이터와 동일한 뉴스 데이터를 사용했으며, 토픽 모델링 특성상, 시간이 오래 걸려 수집한 뉴스데이터 중 가장 많은 기사를 발간한 상위 3 개 신문사만 선정하여 2023년부터 진행하였다. 누락값 제거 등을 하여 문장화 처리를 해주고, 월별로 데이터를 그룹화하여 문장을 벡터화 하였다. 그 뒤에 LDA 모델을 학습시켜 월마다 3 개의 토픽이 추출되도록 하여, 시계열 흐름 속 주요 이슈 키워드 파악이 가능하도록 월간 주요 뉴스를 토픽 모델링하였다.

### [새로운 서비스 제공 - 통화정책방향 브리핑 요약]

통화정책방향 브리핑 요약은 초기에 완전한 자동화 시스템을 구축하기 위해 웹스크래핑이 아닌 유튜브에서 한국은행이 업로드하는 통화정책방향 재생목록에 있는 브리핑 자막 내용을 가져오고자 했다. 하지만, 동영상의 키를 특정할 수 없고, 재생목록에서 영상을 가져오는 데 어려움이 있어 결국 한국은행 공식사이트에 업로드 된 자막 내용을 웹스크래핑하여 데이터를 적재하였다. 따라서 한국은행 통화정책방향 브리핑 요약은 반자동화 시스템으로 구축할 수 있었다.

### [환율 예측 회귀 모델 - 4 개 통화]

본 연구의 목적은 주요 외환 종목(USD, CNY, JPY, EUR)에 대해 각종 경제 및 금융 지표를 기반으로 실시간 예측이 가능한 환율 예측 모델을 구축하는 데에 있다. 이를 위해 딥러닝 기반의 환율 예측 모델과 기존 레퍼런스를 참고하였으나, 거시경제 지표는 그 특성상 발표 주기의 한계로 인해 실시간성 반영이 어려운 점이 존재하였다. 따라서 초기 모델은 거시 지표를 제외한 상태에서 구축하고, 이후 성능 개선의 일환으로 제한적으로 활용하는 방식으로 접근하였다.

수집된 각 통화별 데이터는 다음과 같다.

USD: 2003년 12월 1일부터 수집되었으며, 총 4,316 개의 인스턴스를 확보

CNY: 2007년 1월 1일부터 수집되었으며, 총 4,848 개의 인스턴스를 확보

JPY: 2013년 3월 4일부터 2021년 10월 1일까지 수집되었으며, 총 1,976 개의 인스턴스를 확보

EUR: 2007년 3월 30일부터 수집되었으며, 총 4,574 개의 인스턴스를 확보 각 환율의 입력 변수로는 해당 통화와 연관된 주요 주가 지수를 활용하였다. USD 모델의 경우 S&P 500, 다우존스 산업평균지수, 나스닥 지수를 포함하였으며, CNY 모델은 상하이 종합지수, 선전성분지수, 항셍지수, 항셍 중국기업지수를 포함하였다. JPY 모델은 닛케이 225, 토픽스, 마더스 지수를 EUR 모델은 DAX, EUROSTOXX50, CAC 지수를 각각 입력 변수로 사용하였다.

## 분석 결과

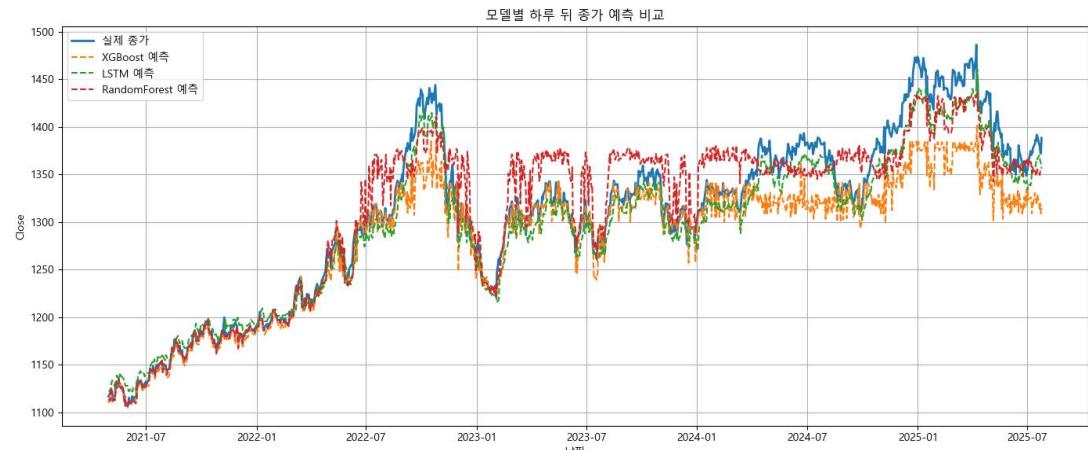
### 1. 환율 예측 회귀 모델

이번 프로젝트에서는 다양한 시계열 예측 모델을 활용하여 환율 변동을 예측하고자 하였다. 그 결과, 다음과 같은 성능 지표를 도출하였다:

- 예측 모델 별 성능 지표 (LSTM, XGBoost, RandomForest 등)
- 평가 지표: RMSE, MAE, R<sup>2</sup>, F1-score 등
- 모델 성능 비교 표/그래프
- 예측 결과 시각화 (ex. 실제 vs 예측 환율 그래프)
- 감성분석 결과나 뉴스 분석 성능도 포함 가능

환율 예측 회귀 분석 결과, JPY를 제외한 대부분의 통화에서 예측 기간이 길어질수록 모델 성능이 저하되는 경향이 나타났으며, 이는 시계열의 불확실성이 누적됨에 따라 장기 예측에서의 신뢰도 하락으로 해석할 수 있다. 다만, LSTM 모델은 R<sup>2</sup> 지표상 성능이 낮게 나타난 경우에도 추세선 측면에서는 비교적 정확한 방향성을 제시하는 경향이 관찰되었다.

USD 환율 예측 그래프는 다음과 같다.



[표-15] USD 하루 뒤 종가 예측 그래프



[표-16] USD 세달 뒤 종가 예측 그래프

하루와 세 달을 비교하였을 때, 예측 기간이 길어짐에 따라 모델의 성능과 추세선이 부정확해지는 것을 확인할 수 있었다. LSTM의 경우 성능이 다른 모델과 마찬가지로 성능이 저하하였으나, 그럼에도 추세선은 상당히 유사하게 예측하는 것으로 나타났다.

## 2. 환율 변동성 지수의 개발

원달러 환율 상승하락 분류 모델에 한국 EPU 지수를 변경해가며 모델 성능을 비교해보았을 때, 자체 산출한 확장형 EPU 지수가 성능이 가장 좋았다.

### 분류 리포트:

	precision	recall	f1-score	support
0	0.75	0.77	0.76	390
1	0.33	0.31	0.32	145

[표-17] 자체 개발 EPU 성능

모델링 결과로써 기존 제공되던 논문보다 좋은 성능을 보일 수 있는 한국 EPU 지수라는 사실을 도출하였으며, 기존 사용되던 단어 대비 다음 단어를 사용하는 것이 더 좋은 EPU 지수를 산출할 수 있다고 결론지었다.

종류	키워드
E	<p>"경제", "경기"</p> <p># — 기본 정책 키워드 —</p> <p>"정부", "청와대", "국무회의", "국회", "의회", "당국",      "한국은행", "한은", "중앙은행",      "기획재정부", "기재부", "금융위원회", "금융위",      "정책", "재정", "입법", "법안", "법률", "예산",      "세금", "규제", "규정", "적자", "부채", "채무",      "연방준비제도", "연준", "Fed", "FRB",      "구조개혁", "구조조정",      # — 통화·금융 정책 —</p> <p>"통화정책", "금융완화", "추가완화", "양적완화", "질적완화",      "통화긴축", "유동성긴축", "マイ너스 금리", "통화할인율", "재할인율",      "통화 운용", "공개시장조작", "물가안정목표", "물가목표",      "금융통화위원회", "금통위", "금융통화운영위원회", "금융통의",      "기준금리", "정책금리", "금리정책", "통화당국", "환금리",      # — 재정 정책 —</p> <p>"재정정책", "정부 예산", "추가경정예산", "추경", "일반 회계", "특별 회계",      "재정적자", "재정수지", "정부지출", "재정지출",      "사회보장성 지출", "국민연금 보험료", "국민건강 보험료",      "의료비 지출", "간병비 지출", "의료보수수가", "의료수가",      "공무원급여", "공적개발원조", "ODA",      "국방비", "군비",      "국채 발행 잔여", "공공부문 부채", "재정부채", "국채", "정부부채", "지방      채",      "경기부양", "경기부양책",      # — 불확실성 키워드 —</p> <p>"불확실", "불확실성", "리스크", "불투명",      "불안", "우려", "걱정"</p>
P	
U	

[표-18] EPU 기반 키워드 도출

### 3. 뉴스 기반 월별 주요 이슈

월별로 토픽 3 개가 도출되고, 토픽별로 상위 키워드 10 개가 산출되도록 모델링 작업을 진행했다.

#### 선택한 날짜: 2025년 05월 01일

- 토픽 1: 후보, 대선, 국민, 대통령, 이재명, 이날, 선거, 김문수, 사람, 민주당
- 토픽 2: 사업, 시장, 대비, 증가, 기준, 가격, 매출, 서비스, 지역, 확대
- 토픽 3: 미국, ai, 기업, 중국, 한국, 투자, 트럼프, 시장, 관세, 글로벌

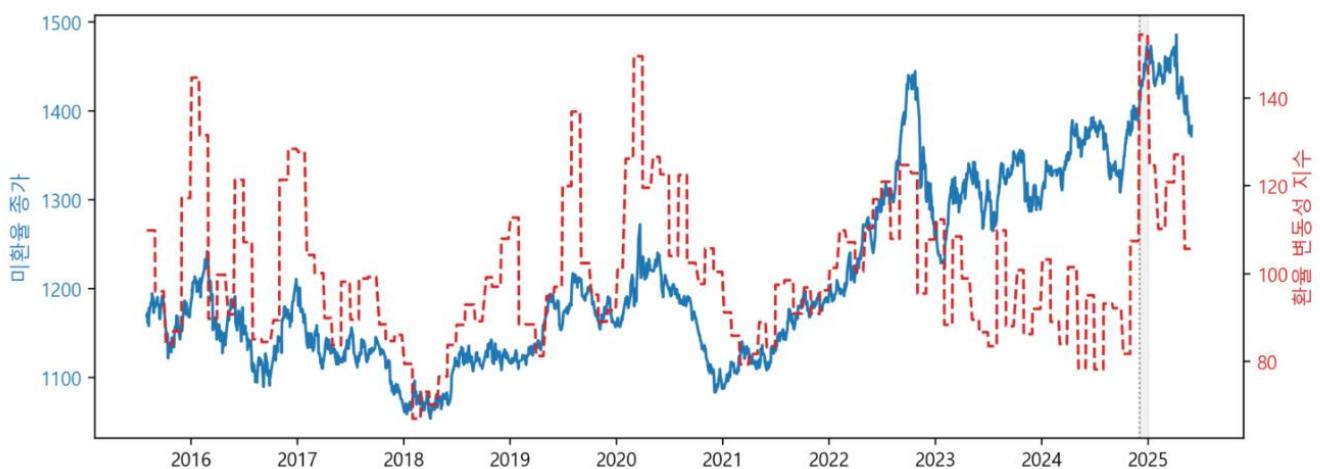
[표-19] 월별 토픽 상위 10 개

특히 환율이 급등하거나, 급락하였을 때, 정성적으로 환율 변동의 급변한 이유를 찾는 데 도움이 되었다. 환율이 급등하였던 2024년 12월에 한국에 비상계엄이 발생한 것이 확인되었다. 발생한 비상계엄에 따라 탄핵, 국민 같은 키워드가 함께 노출되었다.

**선택한 날짜: 2024년 12월 01일 ↗**

- 토픽 1: 대통령, 탄핵, 국회, 미국, 대표, 국민, 의원, 이날, 비상계엄, 윤석열
- 토픽 2: 기업, ai, 사업, 지원, 시장, 서비스, 지역, 투자, 계획, 내년
- 토픋 3: 공개, 방송, 서울, 배우, 사람, 한국, 작품, 생각, 게임, 시작

**미환율 및 환율 변동성 지수 추이**



[표-20] 미환율 및 환율 변동성 지수 추이

#### 4. 통화정책방향 브리핑 요약

정성적으로 살폈을 때는 SOLAR Pro2 를 API 로 연결하여 모델링을 했을 때, 가장 좋은 요약 결과를 도출했다. 또한, 2023년 11월 개최된 프롬프트 엔지니어링 대회의 우승자가 사용한 COSTAR 프롬프트를 사용하였을 때, n-gram 기반의 Rouge 평가에서 가장 좋은 성능을 보였다. 하나하나 숫자가 일치하는지 확인하였을 때도 일치하는 결과를 보였다.

금융 정책을 요약하는 모델이기 때문에 단어의 일치도가 가장 높은 SOLAR Pro2 모델에 COSTAR 프롬프트 엔지니어링을 적용한 모델을 최종 선택하였다.

모델	단어	문맥	문장구조
GPT4.0	0.05	0.05	0.05
Solar Pro2	0.29	0.20	0.24
GPT4.0 + COSTAR	0.13	0.09	0.12
<b>Solar Pro2 + COSTAR</b>	<b>0.33</b>	<b>0.23</b>	<b>0.26</b>

[표-21] 통화정책방향 브리핑 요약 모델 단어 일치도

## 인사이트 도출 및 해석

본 프로젝트는 단순한 환율 예측 모델 개발을 넘어서, 실제 외환 리스크 관리 현장에서 유의미한 의사결정을 지원할 수 있는 통합 예측·해석·전략 제안 시스템을 설계하고 구현하였다는 점에서 실용적 의의가 있다.

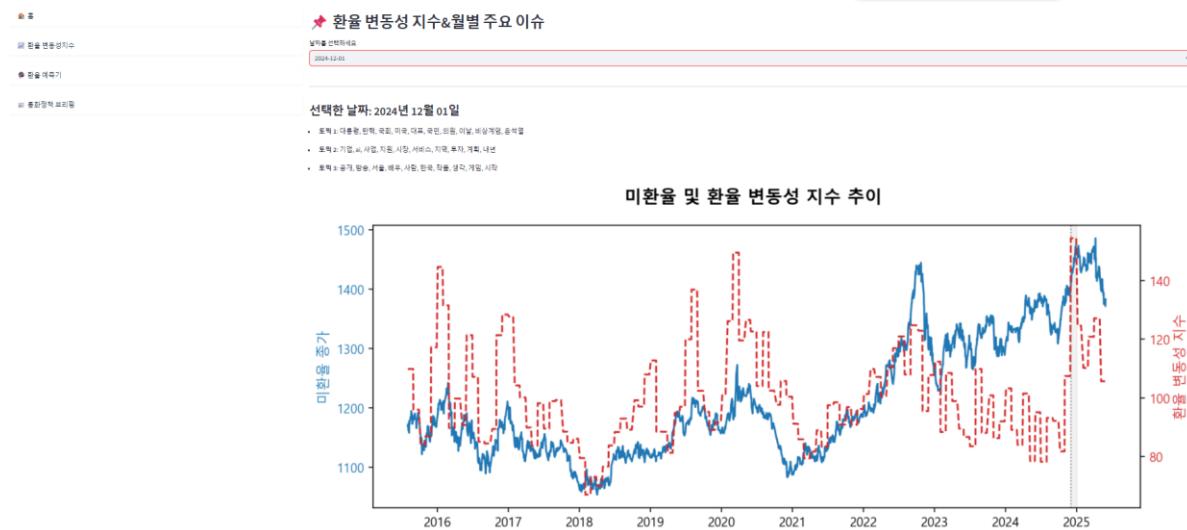
특히 본 시스템은 기업의 환리스크 대응을 자동화하고, 복잡한 거시·금융 정보를 직관적으로 해석 가능하도록 시각화 하는 다양한 기능들을 통합하였다. 주요 결과 및 기능별 인사이트는 다음과 같다.

### 1. 환율 변동성 기반 위험성 판단과 월별 주요 이슈 파악

사용자가 원하는 날짜를 선택하면, 해당 시점의 ▲환율 급등락 여부 ▲관련 뉴스 키워드 ▲정책 불확실성 지수(EPU) ▲환율 시계열 흐름이 한눈에 시각화된다.

이를 통해 사용자는 단순히 과거 환율 흐름을 확인하는 것을 넘어서, 특정한 시점에 환율이 급등 또는 급락했다면, 어떤 거시경제 이벤트나 정책 환경이 환율에 영향을 미쳤는지 맥락을 이해할 수 있다.

※ 그림 1. 환율 급등 시기 주요 뉴스 및 EPU 흐름 시각화 예시



[표-22] 환율 급등 시기 주요 뉴스 및 EPU 흐름 시각화 예시

### 2. 4 개의 통화 환율 예측기 기능

수출입 기업의 경우, 본 시스템을 통해 실제 거래일 기준으로 환율을 사전에 점검하고, 선물환 매도, 옵션 매수, 계약 시점 조정 등 다양한 전략을 시뮬레이션해볼 수 있다. 사용자가 예측하고자 하는 통화 4 종 유로, 엔, 달러, 위안을 선택할 수 있다. 통화를 선택함에 따라 예측하고자 하는 시점, 수출/수입기업 선택, 환전하고자 하는 외환을 입력할 수 있고, 이에 따라 콜/풋옵션, 선물을 통해 예상되는 환리스크에 따라 대응할 수 있는 시스템을 마련했다. 소정의 비용을 지불하고 해당 시점에 권리를 행사하거나 행사하지 않아도 되도록 콜/풋옵션을 선택하여 예측 결과의 리스크를 줄일 수 있다.

또한 한국은행 통화정책방향 브리핑과 환율변동성지수를 통한 환율 흐름을 보조함으로써 배경 이슈 해석이 동시에 제공되기 때문에, 단순 수치 기반 판단이 아닌 행동 유도형 의사결정 시스템으로 사용할 수 있을 것이다.

- 수출 기업: 상승 예측 시, 수출 단가 재조정 또는 환 헤지 실행 판단
- 수입 기업: 하락 예측 시, 환전 시점 조정 또는 대금 결제 분할
- 개인 투자자/무역 중개업자: 실시간 전략 가이던스를 통해 중단기 투자 판단
- 예상 손익 (환율 시뮬레이션 기반)
- 콜/풋옵션 활용 여부 및 전략 추천
- 권장 환전 시점 및 환 헤지 여부 판단

이전 종가	예측 종가	예측 결과
0	1602.99	1604.7469 상승

환전 계산 결과	
현재 환전: 약 8,015 원	
하루 후 환전: 약 8,024 원	
→ 약 9 원 이득	
현재 환율은 1,602.99원이며, 하루 후 1,604.75원으로 상승이 예상됩니다.	

**추천 전략**

수출기업에게 추천되는 전략은 환율 상승 시 수익 증대 효과, 추가 대금 불필요입니다.

[표-23] 환율 예측 결과 및 권고 전략 시각화 예시

### 3. 통화정책 브리핑 요약 기능

한국은행의 기준금리 발표, 통화정책 방향, 금융통화위원회 회의록 등은 환율에 영향을 미치는 핵심 외생 변수이다. 현실적으로 1 시간이나 되는 한국은행의 통화정책방향 브리핑을 전부 시청하는 것은 어려움이 있었다. 소비자는 뉴스 기사로 주관적으로 요약된 기사로만 한국은행의 통화정책방향 브리핑을 확인할 수 있었으나, 본 시스템은 주요 통화정책 브리핑을 자동으로 요약하여 제공함으로써, 환율 예측 결과로 설명되지 않는 정책 기반 환율 변동을 해석 가능하게 한다.

예측된 환율 상승이 실제로는 기준금리 동결 등의 외생 변수로 인해 상쇄되었는지 여부를 판단하는 데 도움을 주며, 사용자는 보다 전략적이고 배경 맥락에 기반한 의사결정을 수행할 수 있다.

**최신 통화정책방향 브리핑**

날짜 선택:

- 2025-07
- 2025-05
- 2025-04
- 2025-02
- 2025-01

**요약 내용 (2025-07)**

**2025년 7월 한국은행 통화정책방향 요약**

- 1. 기준금리 결정**
  - 현행 2.5% 유지 (금융통화위원회 전원 일치)
    - 물가 안정세(2% 내외) 지속 예상
    - 성장세둔화 우려 및 무역협상 불확실성 증대
    - 금융안정 리스크(수도권 주택가격 상승, 기계부채 증가)에 대한 우려
- 2. 대외 경제 환경**
  - 세계경제: 고관세 영향으로 성장둔화 전망
    - 미국: 관세 선수요 효과로 2분기 성장을 일시 상승, 하반기 둔화 전망
    - 유로·중국: 재정 확대에도 통상 갈등으로 성장세 완만
  - 물가:
    - 미국: 소비자물가 하락세 지속이나 관세 전가 영향으로 불확실성 존재
    - 유로: 수요 압력 원화로 2% 미만 안정세 예상
  - 국제금융시장:
    - 미·중 무역협상 진전 및 중동 긴장 원화로 위험자산 선호
    - 미 장기금리 소폭 하락, 달러화 약세 지속

[표-24] 통화정책 브리핑 별 요약 결과

#### 4. 실무 적용 예시 및 전략적 활용 가능성

본 프로젝트는 단순한 환율 예측 모델 개발에 그치지 않고, 예측 결과를 실제 비즈니스 의사결정에 연계할 수 있도록 설계되었다는 점에서 실용적 의의를 갖는다. 특히 수출입 기업이 환율 변동성에 효과적으로 대응할 수 있도록 다양한 기능을 통합한 시스템을 구현하였으며, 각 기능별로 다음과 같은 활용성과 인사이트가 도출된다.

우선, 사용자가 특정 날짜를 선택하면 해당 일자의 주요 환율 관련 뉴스 키워드와 경제 이슈, 정책 불확실성(EPU 지수) 및 환율 추세를 함께 시각화하여 제공하는 기능을 통해, 환율 급등락의 배경을 다각도로 파악할 수 있다. 이는 단순히 과거 흐름을 보여주는 것이 아니라, 유사한 정책 환경이나 뉴스 흐름이 재현될 경우의 환율 변동 가능성을 유추할 수 있도록 하여, 사전 대응 전략 수립에 도움이 된다.

다음으로, 환율 예측기 기능은 사용자가 예측 대상 통화, 금액, 시점, 기업 유형(수입/수출) 등을 입력하면, LSTM 기반 예측 모델을 통해 향후 환율을 예측하고, 이에 따른 손익 시뮬레이션과 전략 제안을 함께 제공한다. 이를 통해 사용자는 환매수·환매도 권고, 콜옵션 또는 풋옵션 활용 등 상황에 따른 맞춤형 환리스크 대응 전략을 자동으로 받아볼 수 있으며, 예측 결과에 대한 해석과 행동 지침을 동시에 확보할 수 있다.

또한, 통화정책 브리핑 기능은 한국은행의 기준금리 결정, 통화정책 방향 발표 등 주요 정책 정보를 요약하여 제공함으로써, 환율 예측 모델로 설명되지 않는 외생적 변동 요인을 정책적 맥락에서 보완적으로 해석할 수 있도록 돕는다. 이를 통해 사용자는 예측 결과를 단순 수치로 받아들이는 것이 아니라, 정책 배경까지 고려한 전략적 판단을 수행할 수 있다.

이러한 기능들을 통해 본 시스템은 단순 예측 도구를 넘어, 환율 예측 → 이슈 해석 → 전략 제안이라는 일련의 의사결정 지원 흐름을 사용자에게 제공한다. 특히 수출 기업의 경우 환율 상승 예측 시 선물환 매도나 수출 단가 조정 등을 고려할 수 있으며, 수입 기업은 환율 하락 예측을 바탕으로 환전 시점 조정이나 옵션 매수 전략을 활용할 수 있다. 나아가 개인 투자자나 무역 중개업자 또한 해당 시스템을 통해 실시간 예측 결과와 정책 해석, 전략 제안을 종합적으로 활용함으로써 보다 정교한 투자 및 비즈니스 판단이 가능해진다.

본 프로젝트는 예측 정확도 향상 뿐만 아니라, 사용자의 정책적 리스크 감내 수준에 맞춘 유연한 전략 설계가 가능하다는 점에서 실무적 활용 가치가 높다고 생각된다. 향후에는 실제 기업의 외화 수지 흐름과 연동하여, 기업 맞춤형 환리스크 조기경보 시스템으로 확장될 수 있을 것이다.

## 주요 알고리즘 요약

### 1. 환율 예측 모델

- 사용 모델: LSTM, XGBoost, RandomForest
- 분석 기법: 감성 분석(BERT 등), 시계열 입력 구조(seq\_len)
- 설명: 각 알고리즘의 목적 및 적용 방식 (ex. LSTM 은 환율 시계열 예측용)
- 보조 기법: 하이퍼파라미터 튜닝, EarlyStopping, feature 중요도 등

### 2. 환율변동성지수 - EPU 산출 수식

- 언론사 n 개 정의
- 경제, 불확실성, 정책 카테고리에 각각 하나의 용어를 공동으로 포함하는 기사를 수집
- 기간에 맞게 뉴스 기사에 대한 데이터를 수집
- 신문사, 월별로 관련 기사 수를 집계 후, 동일 신문의 총 기사 수로 집계하여 상대 빈도 수를 얻음
- 각 신문의 상대 EPU 수를 전체 표본 기간의 단위 표준 편차로 표준화
- 신문의 표준화 시리즈를 월별로 평균하여 전체 월간 한국 EPU 지수
- EPU 디수를 표본 기간 중 평균 100 으로 정규화

### 3. 자체 개발 환율변동성지수 테스트 원달러 분류 모델

- EPU(경제정책불확실성) 지수를 입력변수로 사용하여 환율의 상승 가능성(의일 상승폭  $\geq 5$  원)을 예측
- 예측 모델: LSTM 기반 이진 분류 모델 + Focal Loss + 하이퍼파라미터 튜닝
- 모델 구조 : 입력 → LSTM(64) → LSTM(32) → Dropout(0.2) → Dense(1, sigmoid)

## 4. 뉴스 기반 월별 주요 이슈 토픽 모델링

- 사용 알고리즘 : LDA
- 연, 월 기준으로 데이터를 그룹핑
- 각 월의 키워드문장을 문서로 간주하고 LDA 수행
- 단어 벡터화 : CountVectorizer 사용
- LDA 3 개 토픽 추출
- 각 토픽에서 상위 키워드 10 개 추출

## 5. 통화정책방향 브리핑 요약

- SOLAR Pro2 API 연결
- COSTAR 프롬프트 엔지니어링 진행
- C(맥락) : 한국은행 총재가 발표하는 통화정책방향
- O(목적) : 독자가 장시간 발표된 내용을 요약본을 통해 편하게 인지
- S(스타일) : 전문적인 글쓰기 스타일
- T(톤) : 격식 있고 신뢰감을 주는 톤
- A(대상) : 한국은행 총재의 발표내용을 이해하고 요약할 수 있는 능력을 갖춤
- R(응답 형식) : 예시 형식 제공 및 여러 지시문 입력(자세한 내용은 코드에서 확인 가능)

### 핵심코드

#### 1. 환율예측 – 분류

##### [USD/KRW]

2025년 5월까지의 실제 및 보간 데이터를 기반으로 '환율 예측용 통합 데이터셋'을 생성한다.

- 시계열 기반 예측을 위한 다변수 입력 구성 : 시장지표 + 거시지표 + 웹데이터 + 파생지표
- 예측 대상 : USD/KRW 환율 종가
- 일별/월별 외부 데이터 통합

```
# 1. 일별 데이터 병합 (df1 ~ df10) + (df18, df19)
daily_dfs = [df1, df2, df3, df4, df5, df6, df7, df8, df9, df10, df18, df19]
for df_ in daily_dfs:
    df_base = pd.merge(df_base, df_, on='Date', how='left')

# 2. 월별 데이터 병합을 위한 '기준년월' 컬럼 생성
df_base['기준년월'] = df_base['Date'].dt.to_period('M').astype(str)
df_base['기준년월'] = pd.to_datetime(df_base['기준년월'], format='%Y-%m')

# 3. 월별 데이터 병합 (df11 ~ df21)
monthly_dfs = [df11, df12, df13, df14, df15, df16, df17, df20, df21]
for df_ in monthly_dfs:
    df_base = pd.merge(df_base, df_, on='기준년월', how='left')

# 4. 월데이터 병합 (df22, df23)
df_base = pd.merge(df_base, df22, on='기준년월', how='left')
df_base = pd.merge(df_base, df23, on='기준년월', how='left')
```

### ● 기본 전처리

```
# 1. object 타입 컬럼 중 숫자형으로 변환 가능한 것 선택
obj_cols = df_base.select_dtypes(include='object').columns

# 2. 수치형으로 변환 시도 (콤마, % 제거 포함)
for col in obj_cols:
    df_base[col] = (
        df_base[col]
        .astype(str)
        .str.replace(',', '', regex=False) # 쉼표 제거
        .str.replace('%', '', regex=False) # % 기호 제거
        .replace('-', '') # 음수 아닌 '-' 빈값 처리
    )

# 숫자로 변환 (에러 시 NaN)
df_base[col] = pd.to_numeric(df_base[col], errors='coerce')

# → 무한대 결측으로 처리
df_base.replace([np.inf, -np.inf], np.nan, inplace=True)
```

### ● 이진 분류 준비

```
df = df_base.copy()
df['next_day_close'] = df['미환율_종가'].shift(-1)
df['change'] = df['next_day_close'] - df['미환율_종가']
df['target'] = (df['change'] >= change_cut).astype(int)
df = df.dropna().reset_index(drop=True)
```

### ● 클래스 불균형 완화 및 노이즈 보강

```
def add_jitter(X, sigma_ratio=0.05):
    X_jittered = X.copy()
    for i in range(X.shape[1]):
        feature_std = np.std(X[:, i])
        noise = np.random.normal(loc=0.0, scale=sigma_ratio * feature_std, size=X[:, i].shape)
        X_jittered[:, i] += noise
    return X_jittered
```

### ● Threshold 튜닝을 통한 예측 민감도 최적화

```
# 모델 예측 확률값 (1차원 배열로 평탄화)
y_proba = model.predict(X_test_seq).flatten()

# 최적 임계값 기준으로 미진 분류 결과 생성
y_pred = (y_proba > best_threshold).astype(int)
```

### ● 시계열 입력 생성 함수

30 일간의 데이터를 기반으로 시퀀스를 생성하는 함수다. LSTM 입력에 사용된다.

```

def create_sequences(data, targets, seq_len=30):
    X, y = [], []
    for i in range(len(data) - seq_len):
        X.append(data[i:i+seq_len])
        y.append(targets[i+seq_len])
    return np.array(X), np.array(y)

# ♦ Bidirectional LSTM 모델 정의 + Focal Loss 적용
# 시계열 특성을 양방향으로 학습하여, 클래스 불균형 보정을 위해 Focal Loss를 커스터마이징
# 하여 사용합니다.
def focal_loss(gamma=2., alpha=0.25):
    def loss(y_true, y_pred):
        eps = tf.keras.backend.epsilon()
        y_pred = tf.clip_by_value(y_pred, eps, 1. - eps)
        return tf.reduce_mean(alpha * tf.pow(1 - y_pred, gamma) * -y_true * tf.math.log(y_pred))
    return loss

model = Sequential([
    Bidirectional(LSTM(72)),
    Dropout(0.2),
    Dense(48, activation='relu'),
    Dense(3, activation='softmax')
])
model.compile(loss=focal_loss(), optimizer=Adam(0.0007), metrics=['accuracy'])

    ♦ Temperature Scaling 적용
Softmax 확률값을 부드럽게 조정하여, 과도한 확신(overconfidence)을 완화합니다.
def softmax_temperature(logits, T=0.7):
    exp_logits = np.exp(logits / T)
    return exp_logits / np.sum(exp_logits, axis=1, keepdims=True)

```

### ● 요약 및 특징

- 예측 대상 : 다음날 환율 방향 ( $\uparrow/\rightarrow/\downarrow$ )
- 입력 변수 : WTI 유가, 금값, S&P500, VIX 등 주요 경제지표 + 기술적 지표 (RSI, MACD, Bollinger Band 등)
- 모델 구조 : Bidirectional LSTM + Dropout + Dense(Softmax)
- 핵심 기법 :
  - Focal Loss (클래스 불균형 보정)
  - Temperature Scaling (확률 민감도 보정)
  -

## [CNY/KRW]

```
from tensorflow.keras.layers import Input, LSTM, Dropout, Dense
from tensorflow.keras.models import Model

inputs = Input(shape=(10, N))
x = LSTM(64, return_sequences=True)(inputs)
x = LSTM(32)(x)
x = Dropout(0.2)(x)
outputs = Dense(1, activation='sigmoid')(x)

model = Model(inputs, outputs)
```

## 2. 환율예측 - 회귀

실시간 환율 예측 모델 구축

환율 예측 모델에서의 데이터 수집은 다음과 같은 기준으로 진행되었다. USD, JPY, EUR 환율은 야후파이낸스(Yahoo Finance) API를 통해 수집하였으며, CNY 환율은 해당 플랫폼에서 과거 데이터의 제공이 불가능한 관계로 인베스팅닷컴(Investing.com)에서 수동 수집한 후, 웹 크롤링을 통해 실시간성 구현을 시도하였다.

참고로, 데이터 수집 시 야후파이낸스의 period='25y' 옵션을 활용하였으나, 이후 경제 지표와 결합하는 과정에서 일부 시계열 간 수집 기간 차이가 발생하였다. JPY의 경우, 실시간 구현에는 한계가 있었으나 데이터의 양은 충분하다고 판단되어 모델링을 진행하였다.

```
import yfinance as yf
import pandas as pd

def dollar_indicator():
    tickers = {
```

```

    "S&P500": "^GSPC",
    "DowJones": "^DJI",
    "NASDAQ": "^IXIC"
}

data_list = []

for name, symbol in tickers.items():
    ticker = yf.Ticker(symbol)
    df = ticker.history(period="25y", interval="1d")
    df = df.reset_index()

    df = df[["Date", "Open", "High", "Low", "Close"]]
    df = df.rename(columns={
        "Open": f"{name}_Open",
        "High": f"{name}_High",
        "Low": f"{name}_Low",
        "Close": f"{name}_Close"
    })
    df["Date"] = pd.to_datetime(df["Date"]).dt.tz_localize(None)
    data_list.append(df)

df_merged = data_list[0]
for df in data_list[1:]:
    df['Date']=pd.to_datetime(df['Date'])
    df_merged = pd.merge(df_merged, df, on="Date", how="inner")

df_merged['Date'] = pd.to_datetime(df_merged['Date']).dt.strftime('%Y-%m-%d')
df_merged['Date'] = pd.to_datetime(df_merged['Date'])
return df_merged

import cloudpickle
with open("dollar_indicator.pkl", "wb") as f:
    cloudpickle.dump(dollar_indicator, f)

dollar_indi_df=dollar_indicator()

def real_times(symbol):
    ticker = yf.Ticker(symbol)
    df = ticker.history(period=f"25y", interval="1d")

```

```

df = df.copy()
df = df[["Open", "High", "Low", "Close"]]
df.columns = [f"{col}" for col in df.columns]
df["Date"] = df.index.date
df.reset_index(drop=True, inplace=True)
# 변동량 계산
df["Change"] = df["Close"].diff()

df['Date']=pd.to_datetime(df['Date'],format='%Y-%m-%d')

df=df.reindex(columns=['Date','Close','Open','High','Low','Change'])
return df

import cloudpickle
with open("real_times.pkl", "wb") as f:
    cloudpickle.dump(real_times, f)

dollar_df=real_times('USDKRW=X')

df=pd.merge(dollar_indi_df,dollar_df,on='Date',how='inner')

```

모델은 예측 기간을 하루, 일주일, 한 달, 세 달로 설정하고, 각 구간마다 XGBoost, LSTM, Random Forest 모델을 비교하여 가장 우수한 성능을 보인 모델을 선택하였다. 성능 평가지표로는 평균절대오차(MAE), 평균제곱근오차(RMSE), 결정계수( $R^2$ )를 활용하였다.

```

# 데이터 정리
df = df.sort_values('Date')
df['Date'] = pd.to_datetime(df['Date'])
df['target'] = df['Close'].shift(-1)
df = df.dropna().reset_index(drop=True)

# 피처, 타겟 설정
drop_cols = ['Date', 'target']
X = df.drop(columns=drop_cols, errors='ignore')
y = df['target'].values

# 학습/테스트 분할
split_idx = int(len(df) * 0.8)
X_train_raw, X_test_raw = X.iloc[:split_idx], X.iloc[split_idx:]

```

```

y_train_raw, y_test_raw = y[:split_idx], y[split_idx:]

# 정규화
scaler_X = MinMaxScaler().fit(X_train_raw)
X_train_scaled = scaler_X.transform(X_train_raw)
X_test_scaled = scaler_X.transform(X_test_raw)

scaler_y = MinMaxScaler().fit(y_train_raw.reshape(-1, 1))
y_train_scaled = scaler_y.transform(y_train_raw.reshape(-1, 1))
y_test_scaled = scaler_y.transform(y_test_raw.reshape(-1, 1))

# 시퀀스 생성 함수
def make_sequence(X, y, seq_len):
    X_seq, y_seq = [], []
    for i in range(len(X) - seq_len):
        X_seq.append(X[i:i+seq_len])
        y_seq.append(y[i+seq_len])
    return np.array(X_seq), np.array(y_seq)

seq_len = 20
X_train_seq, y_train_seq = make_sequence(X_train_scaled, y_train_scaled, seq_len)
X_test_seq, y_test_seq = make_sequence(X_test_scaled, y_test_scaled, seq_len)

X_train_flat = X_train_scaled[seq_len:][:len(X_train_seq)]
X_test_flat = X_test_scaled[seq_len:][:len(X_test_seq)]
y_train_flat = y_train_scaled[seq_len:][:len(X_train_seq)]
y_test_flat = y_test_scaled[seq_len:][:len(X_test_seq)]

# 모델 훈련 및 예측
## XGBoost
model_xgb = XGBRegressor(n_estimators=200, learning_rate=0.05, random_state=42)
model_xgb.fit(X_train_flat, y_train_flat.ravel())
y_pred_xgb = scaler_y.inverse_transform(model_xgb.predict(X_test_flat).reshape(-1, 1)).re

## LSTM
model_lstm = Sequential()
model_lstm.add(LSTM(128, input_shape=(seq_len, X_train_seq.shape[2])))
model_lstm.add(Dense(1))
model_lstm.compile(optimizer='adam', loss='mse')
early_stop = EarlyStopping(patience=15, restore_best_weights=True)
model_lstm.fit(X_train_seq, y_train_seq, epochs=100, batch_size=32, validation_split=0.2,
               callbacks=[early_stop], verbose=0)

```

```

y_pred_lstm = scaler_y.inverse_transform(model_lstm.predict(X_test_seq)).reshape(-1)

## RandomForest
model_rf = RandomForestRegressor(n_estimators=200, random_state=42)
model_rf.fit(X_train_flat, y_train_flat.ravel())
y_pred_rf = scaler_y.inverse_transform(model_rf.predict(X_test_flat)).reshape(-1, 1).resh

# 실제값 복원
y_true = scaler_y.inverse_transform(y_test_seq).reshape(-1)

# 성능 출력
for name, pred in zip(['XGBoost', 'LSTM', 'RandomForest'],
                      [y_pred_xgb, y_pred_lstm, y_pred_rf]):
    rmse = np.sqrt(mean_squared_error(y_true, pred))
    mae = mean_absolute_error(y_true, pred)
    r2 = r2_score(y_true, pred)
    print(f"{name}")
    print(f"MAE: {mae:.4f}")
    print(f"RMSE: {rmse:.4f}")
    print(f"R^2: {r2:.4f}")

```

데이터 전처리 과정에서는 수집된 데이터의 날짜를 기준으로 통합한 후, 종가(Close) 기준으로 다음 날 환율을 예측하는 형태의 타겟 변수(target)를 설정하였다. 전체 데이터는 80%의 학습 데이터와 20%의 테스트 데이터로 분할하여 사용하였으며, 정규화를 위해 MinMaxScaler를 적용하였다. 특히, 시계열 특성을 반영하기 위해 입력 데이터는 시퀀스 길이 20으로 구성된 시퀀스 데이터를 생성하여 LSTM 모델의 입력 형태에 맞추었다. 모델 학습은 XGBoost, RandomForest, LSTM 세 가지 알고리즘을 대상으로 수행하였다. 각각의 모델은 다음과 같은 설정으로 학습되었다.

- XGBoost: 학습률 0.05, 추정기 200 개
- RandomForest: 추정기 200 개
- LSTM: 은닉 유닛 128, 에포크 100, 조기 종료 조건(patience=15) 설정

모든 모델은 테스트 데이터셋에 대해 예측을 수행한 후, 성능 지표로 평균 절대 오차(MAE), 평균 제곱근 오차(RMSE), 결정계수( $R^2$ )를 산출하여 비교하였다. 실험 결과, LSTM 모델은 하루 뒤 환율 예측에서 가장 높은 결정계수( $R^2=0.9555$ )를 보이며 우수한 성능을 나타냈다. 또한 시각화 결과를 통해서도, LSTM 모델이 실제 환율 추세를 가장 안정적으로 추종하는 것으로 나타났다. 이상의 결과는 시계열적 특성을 반영하는 딥러닝 기반 모델이 단기 예측에 있어 유의미한 예측력을 가짐을 시사한다. 특히 고빈도 경제지표와 함께 사용할 경우, 정책적 판단이나 외환 리스크 대응 전략 수립에 활용 가능성이 높을 것으로 기대된다.

위 모델의 성능을 추가적으로 파악하기 위하여 그래프를 통하여 추세선을 확인하였다.

- #시각화

```
date_test = df['Date'].iloc[seq_len + split_idx: ].reset_index(drop=True)
n = min(len(date_test), len(y_true), len(y_pred_xgb), len(y_pred_lstm), len(y_pred_rf))

plt.figure(figsize=(14, 6))
plt.plot(date_test[:n], y_true[:n], label='실제 종가', linewidth=2)
plt.plot(date_test[:n], y_pred_xgb[:n], label='XGBoost 예측', linestyle='--')
plt.plot(date_test[:n], y_pred_lstm[:n], label='LSTM 예측', linestyle='--')
plt.plot(date_test[:n], y_pred_rf[:n], label='RandomForest 예측', linestyle='--')
plt.title("모델별 하루 뒤 종가 예측 비교")
plt.xlabel("날짜")
plt.ylabel("Close")
plt.legend()
plt.grid(True)
plt.tight_layout()
plt.show()
```

다음으로 해당 모델을 서비스 UI 와 연동하기 위하여 피클로 저장하는 과정을 수행하였다.

```
import joblib
joblib.dump(model_lstm, 'dollar_하루.pkl')
joblib.dump(scaler_X, 'dollar_scaler_X_하루.pkl')
joblib.dump(scaler_y, 'dollar_scaler_y_하루.pkl')
```

- 최종적으로, 환율별 모델 성능은 다음과 같다.

### USD

하루: LSTM ( $R^2 = 0.9555$ )

일주일: LSTM ( $R^2 = 0.8287$ )

한 달: LSTM ( $R^2 = -0.1602$ )

세 달: LSTM ( $R^2 = -0.3518$ )

### CNY

하루: LSTM ( $R^2 = 0.9106$ )

일주일: LSTM ( $R^2 = 0.4090$ )

한 달: Random Forest ( $R^2 = -0.6879$ )

세 달: Random Forest ( $R^2 = -3.4918$ )

### JPY

하루: Random Forest ( $R^2 = 0.9802$ )

일주일: XGBoost ( $R^2 = 0.8240$ )

한 달: LSTM ( $R^2 = 0.5062$ )

세 달: LSTM ( $R^2 = 0.6423$ )

**EUR**

하루: XGBoost ( $R^2 = 0.9852$ )

일주일: LSTM ( $R^2 = 0.8981$ )

한 달: LSTM ( $R^2 = 0.8069$ )

세 달: LSTM ( $R^2 = 0.2789$ )

### 3. 환율변동성 지수

```

import re
import pandas as pd

# ★ 주요 키워드 정의
e = 경제를 정의하는 단어 입력
p = 정책을 정의하는 단어 입력
u = 불확실성을 정의하는 단어 입력

# ★ EPU 관련 기사 추출
cols = ['키워드', '특성추출(가중치순 상위 50개)', '본문', '제목']
joined = df[cols].fillna('').agg(' '.join, axis=1)

cond1 = joined.str.contains('|'.join(map(re.escape, e)))
cond2 = joined.str.contains('|'.join(map(re.escape, p)))
cond3 = joined.str.contains('|'.join(map(re.escape, u)))

df_epu = df[cond1 & cond2 & cond3]

# 일자 정리
df_epu['일자'] = pd.to_datetime(df_epu['일자'], format='%Y%m%d').dt.to_period('M').dt.to_
    # 신문사별 EPU 기사수 및 전체 기사수 집계
epu_count = df_epu.groupby(['일자', '언론사']).size().reset_index(name='epu기사수')
total_count = df.groupby(['일자', '언론사']).size().reset_index(name='전체기사수')
total_count['일자'] = pd.to_datetime(total_count['일자'], format='%Y%m%d').dt.to_period('M')

# 상대빈도 및 표준화
merged = total_count.merge(epu_count, on=['일자', '언론사'], how='left').fillna(0)
merged['상대빈도수'] = merged['epu기사수'] / merged['전체기사수']
merged['표준화빈도'] = merged['상대빈도수'] / merged.groupby('언론사')['상대빈도수'].tran

# 월별 평균 → EPU 지수 생성
monthly_epu = merged.groupby('일자')['표준화빈도'].mean()
korea_epu = (monthly_epu / monthly_epu.mean() * 100).rename('Korea_EPU')

# 저장
korea_epu.reset_index().to_csv('one_2009_to_2011.csv', index=False)

```

#### 4. 뉴스 기반 월별 주요 이슈 토픽 모델링

```
# ✓ 키워드 전처리 및 문장화
상위언론사데이터['키워드'] = 상위언론사데이터['키워드'].apply(
    lambda x: x if isinstance(x, list) else str(x).replace('...', '').split(',')
)
상위언론사데이터['키워드문장'] = 상위언론사데이터['키워드'].apply(lambda x: ' '.join(x))

# ✓ LDA 토픽 추출 함수
def lda_by_month(df, n_topics=5, top_n=10):
    result = []
    for (year, month), group in df.groupby(['연', '월']):
        texts = group['키워드문장'].tolist()
        if len(texts) < 2:
            continue
        vectorizer = CountVectorizer(max_df=0.95, min_df=1, stop_words='english')
        try:
            X = vectorizer.fit_transform(texts)
        except ValueError:
            continue
        if X.shape[1] == 0:
            continue
        lda = LatentDirichletAllocation(n_components=n_topics, random_state=42)
        lda.fit(X)
        words = vectorizer.get_feature_names_out()
        for idx, topic in enumerate(lda.components_):
            top_keywords = [words[i] for i in topic.argsort()[:-top_n - 1:-1]]
            result.append({
                '연': year, '월': month,
                '토픽번호': idx + 1,
                '상위키워드': ', '.join(top_keywords)
            })
    return pd.DataFrame(result)

# ✓ 실행
monthly_topics = lda_by_month(상위언론사데이터, n_topics=3, top_n=10)
print(monthly_topics)
```

## 5. 통화정책방향 브리핑 요약

최종 선택하게 된 SOLA 모델에 COSTAR 프롬프트 엔지니어링을 한 핵심코드만 작성하였다.

```
load_dotenv()
sola_api_key = os.getenv('sola_api_key')
df = pd.read_csv('기자간담회.csv')[['Date', 'Content']]
```

```
system_prompt = """
```

당신은 글쓰기 전문가입니다. 아래 조건을 바탕으로 글을 작성해주세요:

**1. Context:**

- 주제는 한국은행 총재가 발표하는 통화정책방향입니다.

**2. Objective:**

- 독자가 장시간 발표된 내용을 요약본을 통해 편하게 인지할 수 있도록 합니다.

**3. Style:**

- 전문적인 글쓰기 스타일.

**4. Tone:**

- 격식 있고 신뢰감을 주는 톤.

**5. Audience:**

- 은행 직원으로 한국은행 총재의 발표 내용을 이해하고 요약할 수 있는 능력을 갖춘 사람들입니다.

**6. Response:**

- 가능한 한 심층 요약 글로 작성해주세요.
- 발표 내용을 주제별로 구분하여 항목별로 정리된 형태로 요약해주세요.
- 항목 제목은 발표 내용에 맞게 자유롭게 생성하되, 일관된 서식(숫자 또는 기호)을 사용해주세요.
- 발표에서 다룬 주제가 많을 경우 항목 수를 늘려도 좋습니다.

예시 형식:

1. oo 관련 발표 내용 요약
2. △△ 배경 및 한국은행의 입장
3. 향후 대응 방향

```
"""
```

## 참고 자료 항목

### [기획배경]

- 송영철 외. (2004). 중소기업 환율 리스크 분석 연구: 재무 및 경제적 영향을 중심으로. n.p.: 중소벤처기업연구원.
- 새 정부 기조 따라...'포용 금융' 박차 가하는 은행권 . (2025).

<https://www.sentv.co.kr/article/view/sentv202507220147#rs%EC%9E%90%EC%B2%B4>.

- 「중소기업 환율 리스크 분석 연구」(자율연구과제), 송영철, 2024

- 2025년 경제 최대 리스크 '환율', 중소기업부터 덮친다 | 출처 : 아시아경제 |

<https://www.asiae.co.kr/article/2024123110314818396>.

- '외환시장 거래시간 연장' 1년...하루 거래량 16.3% 늘었다 . (2025).

<https://www.korea.kr/news/policyNewsView.do?>

- '외환시장 거래시간 연장' 1년...하루 거래량 16.3% 늘었다 . (2025).

<https://www.korea.kr/news/policyNewsView.do?>

### [환율 - 데이터 수집처]

- 프로젝트에 사용된 데이터 출처
- 구글 검색량 : Google trends
- 경제 불확실성 지수 : Economic Policy Uncertainty
- 시장 데이터, 거시경제 지표 : Investing.com

### [환율 - 참고 논문 및 보고서]

- 딥러닝을 활용한 원화 환율 예측: 시장 및 웹데이터와 거시경제 지표의 활용 = Deep Learning Approach for USD/KRW Exchange Rate Forecasting / 황미라 한국과학기술원 / 2020
- GARCH 모형을 이용한 한국 원화의 환율변동성 추정 / 모수원, 이광배 / 한국무역통상학회 / 2016
- 데이터 마이닝 기법을 활용한 환율예측 : GARCH 와 결합된 랜덤 포레스트 모형(서종덕, 2016)

### [환율변동성지수 개발 - 데이터 수집처]

- 빅카인즈

### [환율변동성지수 개발 - 논문 및 보고서]

- Macroeconomic effects of uncertainty shocks: Evidence from Korea(조두연, 김후상, 2023)
- KOSTAT-CIS 를 이용한 경기도 경기선행지수의 작성 및 활용(김정훈, 황상연, 2014)
- 경제정책 불확실성과 원/달러 환율 변동성(김경민, 2025)
- New economic policy uncertainty indexes for South Korea(이경희, 최주희 외, 2020)
- 환율 변동성의 예측과 장기기억(김창범, 2013)
- 확률적 변동성 모형을 이용한 원화 환율 변동성 추정(천도현 외, 2017)
- KDI 한국개발연구원 EPU 지수