



USD/KRW 환율 예측 시스템

다변량 시계열 LSTM 을 활용한 여행자 대상 최적 환전 시점 예측

과목: 파이썬기반딥러닝 담당교수: 구영현 교수님

학과: 인공지능학과

학번: 21011302

이름: 이재윤

작성일: 2025년 12월 9일

제출일: 2025년 12월 13일

목차

1. 프로젝트 개요
2. 시스템 아키텍처
3. 데이터 수집 및 전처리
4. 모델 설계 및 개선
5. 성능 평가 방법론
6. 최종 성능 및 결과
7. 웹 대시보드 구현
8. 개발 일정 및 마일스톤
9. 문제 해결 과정
10. 결론 및 향후 계획
11. 참고문헌

1. 프로젝트 개요

1.1 연구 배경 및 동기

최근 원/달러 환율이 1,400 원대를 기록하며 높은 변동성을 보이고 있다. 2024년 12월 기준으로 환율은 1,439 원까지 상승하였으며, 이러한 급격한 변동은 해외여행을 계획하는 개인에게 환전 시점 선택의 어려움을 야기하고 있다.



환율 변동으로 인한 재정적 영향은 생각보다 크다. 예를 들어, 100 만원을 환전할 때 환율이 10 원만 차이 나도 약 7 달러의 손익이 발생한다. 장기 여행이나 유학을 준비하는 경우 그 영향은 더욱 커진다.

기존의 환율 예측 연구들은 주로 과거 환율 데이터만을 활용한 단변량 시계열 분석에 집중되어 있었다. 그러나 환율은 단순히 과거 패턴의 반복이 아니라, 금리차, 원유 가격, 주식시장 동향, 지정학적 리스크 등 다양한 거시경제 지표의 복합적인 영향을 받는다. 따라서 본 연구는 다변량 시계열 분석을 통해 이러한 요인들을 통합적으로 반영하고자 한다.

1.2 연구 목표

본 연구의 구체적인 목표는 다음과 같다:

다변량 시계열 LSTM 모델 구축

15 개의 경제 지표를 활용하여 USD/KRW 환율을 예측하는 딥러닝 모델을 개발한다. 단순히 과거 환율만 보는 것이 아니라, 금리차, 시장 지수, 주요 교역국 환율 등 다양한 경제 변수를 함께 고려한다.

실용적인 예측 기간 설정

향후 7 일간의 환율을 예측하여 여행자가 실제로 의사결정에 활용할 수 있도록 한다. 너무 짧으면 대응할 시간이 없고, 너무 길면 예측 정확도가 떨어지므로 1 주일을 최적 기간으로 설정하였다.

자동화된 파이프라인 구축

데이터 수집, 전처리, 모델 재훈련, 결과 업로드까지 전 과정을 자동화하여 최신 데이터를 지속적으로 반영한다. Windows 작업 스케줄러를 활용하여 4 일마다 자동으로 시스템을 업데이트한다.

사용자 친화적인 웹 인터페이스 개발

일반 사용자가 쉽게 접근할 수 있는 웹 기반 대시보드를 개발하여, 복잡한 기술적 지식 없이도 환전 시점 추천을 받을 수 있도록 한다.

1.3 기대 효과

본 시스템을 통해 얻을 수 있는 실질적인 효과는 다음과 같다:

경제적 효과

100 만원 환전 시 환율이 10 원만 낮아져도 약 7 달러의 이득이 발생한다. 1,000 만원 규모의 환전이라면 70 달러, 한화로 약 10 만원의 절약이 가능하다. 본 모델의 RMSE 가 17.82 원임을 고려하면, 평균적으로 이 정도의 오차 범위 내에서 최적 시점을 예측할 수 있다.

의사결정 지원

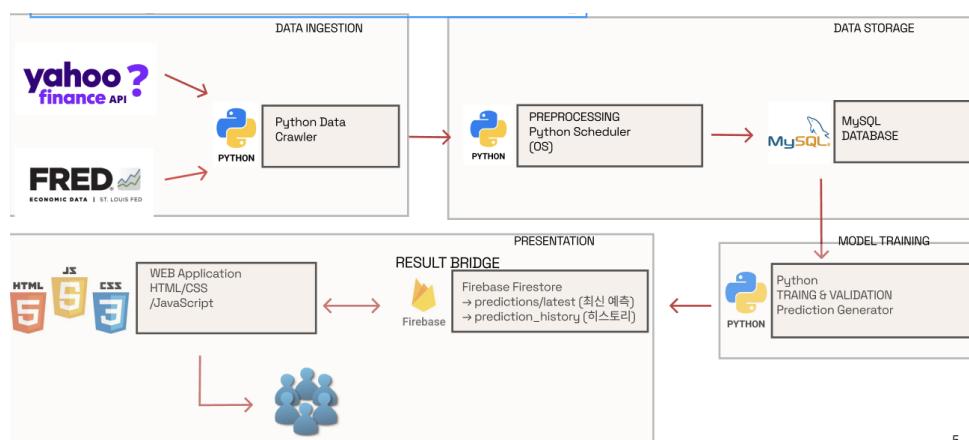
환율 변동에 대한 불안감을 해소하고, 데이터에 기반한 합리적인 의사결정을 할 수 있다. "지금 환전해야 하나, 조금 더 기다려야 하나?"라는 고민에 과학적인 근거를 제시한다.

학술적 기여

금융 시계열 예측에서 다변량 접근법의 효과를 실증적으로 검증한다. 특히 Bidirectional LSTM 과 Huber Loss의 조합, 기술적 지표의 활용 등 실전에서 검증된 방법론을 제시한다.

2. 시스템 아키텍처

본 시스템은 5 개의 주요 모듈이 유기적으로 연결된 End-to-End 파이프라인으로 구성된다.



2.1 데이터 수집 모듈

Yahoo Finance API 와 FRED API 를 활용하여 일별 거시경제 데이터를 자동으로 수집한다. Python 의 yfinance 와 fredapi 라이브러리를 사용하며, 충분 업데이트 방식으로 효율적으로 데이터를 관리한다.

수집 데이터 목록:

- 환율: USD/KRW, USD/JPY, USD/CNY, EUR/USD
- 시장 지수: S&P 500, KOSPI, VIX, DXY, 금, WTI 유가
- 금리: 미국 기준금리, 한국 기준금리, 미국 장단기 금리차

2.2 데이터 전처리 모듈

수집된 원본 데이터를 모델 학습에 적합한 형태로 변환한다. 주요 처리 과정:

1. 결측치 처리: 선형 보간법 및 Forward Fill
2. Feature Engineering: 금리차(IRD), KOSPI 변동성, 기술적 지표 생성
3. 정규화: Min-Max Scaling 적용
4. Sequence 생성: 60 일 원도우 크기의 슬라이딩 원도우

2.3 모델 훈련 모듈

Bidirectional LSTM 아키텍처를 채택하여 시계열 데이터의 양방향 의존성을 학습한다. 과적합 방지를 위해 Dropout, Batch Normalization, Early Stopping 을 적용하며, Huber Loss 를 손실 함수로 사용한다.

2.4 결과 저장 모듈

훈련된 모델의 예측 결과와 성능 지표를 Firebase Firestore에 실시간으로 저장한다. 이를 통해 웹 대시보드에서 최신 예측 결과를 즉시 조회할 수 있다.

The screenshot shows the Firebase Firestore interface. On the left, there's a sidebar with a 'prediction_history' collection. Under it, there's a single document named 'aocRIFYahpVSrQ3L2MTD'. This document contains an array of 14 elements, each representing an 'actual_rate' with a value and an index from 0 to 13. The values are: 1373.74, 1382.28, 1379.06, 1378.61, 1381.15, 1391.82, 1388.39, 1386.11, 1390.89, 1393.38, 1386.95, 1380.29, 1383.42, and 1383.42. At the bottom of the interface, there's a note: '데이터베이스 위치: asia-northeast3'.

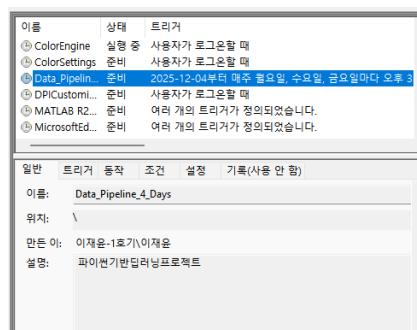
index	actual_rate
0	1373.74
1	1382.28
2	1379.06
3	1378.61
4	1381.15
5	1391.82
6	1388.39
7	1386.11
8	1390.89
9	1393.38
10	1386.95
11	1380.29
12	1383.42
13	1383.42

2.5 웹 대시보드 모듈

HTML, JavaScript, Chart.js를 활용하여 직관적인 웹 인터페이스를 구현하였다. 사용자는 복잡한 기술적 지식 없이도 환전 시점 추천을 받을 수 있다.

2.6 자동화 스케줄러

Papermill과 Windows 작업 스케줄러를 조합하여 전체 파이프라인을 자동으로 실행한다.



3. 데이터 수집 및 전처리

3.1 데이터 수집 전략

2010년 10월부터 2024년 12월까지 약 15년에 걸친 일별 데이터를 수집하였다. 총 3,900일 이상의 데이터 포인트를 확보하여, 모델이 다양한 시장 상황을 학습할 수 있도록 하였다.

Result Grid																			
Filter Rows:		Edit:		Export/Import:		Wrap Cell Content:		Fetch rows:											
date	usd_krw	wti_price	sp500_index	kospi_index	kospi_volatility	usd_jpy	usd_cny	eur_usd	vix	gold	dxy	us_rate	kr_rate	ird	ust_spread	created_at	updated_at		
2010-10-29	1132.4	81.43	1183.26	1882.95	1.30617	81.679	6.6765	1.37781	21.2	1357.1	77.27	0.19	4.48	-4.29	2.29	2025-10-13 15:49:14	2025-10-13 15:49:14		
2010-11-01	1117.2	82.95	1184.38	1914.74	1.68831	80.406	6.6607	1.39751	21.83	1350.2	77.3	0.19	4.48	-4.29	2.32	2025-10-13 15:49:14	2025-10-13 15:49:14		
2010-11-02	1114.1	83.9	1193.57	1918.04	0.17235	80.559	6.6829	1.39109	21.57	1356.4	76.72	0.19	4.48	-4.29	2.29	2025-10-13 15:49:14	2025-10-13 15:49:14		
2010-11-03	1111	84.69	1197.96	1935.97	0.934805	80.668	6.6686	1.40361	19.56	1337.1	76.48	0.19	4.48	-4.29	2.33	2025-10-13 15:49:14	2025-10-13 15:49:14		
2010-11-04	1101.7	86.49	1221.06	1942.5	0.3373	81.05	6.6652	1.41273	18.52	1382.7	75.88	0.19	4.48	-4.29	2.2	2025-10-13 15:49:14	2025-10-13 15:49:14		
2010-11-05	1104.9	86.85	1225.85	1938.96	0.182241	80.776	6.6523	1.4228	18.26	1397.3	76.55	0.19	4.48	-4.29	2.2	2025-10-13 15:49:14	2025-10-13 15:49:14		
2010-11-08	1115.7	87.06	1229.75	1947.41	0.177944	81.14	6.6458	1.40758	18.79	1417.8	77.03	0.19	4.48	-4.29	2.19	2025-10-13 15:49:14	2025-10-13 15:49:14		

표 3-1 수집 Feature 목록

카테고리	Feature	출처	역할
Target	USD/KRW	Yahoo Finance	예측 대상 변수
환율	USD/JPY, USD/CNY, EUR/USD	Yahoo Finance	주요 통화 동향
시장 지수	S&P 500, KOSPI, VIX, DXY, 금, WTI	Yahoo Finance	시장 심리 및 위험
금리	US_Rate, KR_Rate, UST_Spread	FRED	통화 정책 방향
Feature Engineering	IRD, KOSPI_Volatility	계산	자본 유출입 압력

3.2 기술적 지표 추가

표 3-2 기술적 지표

지표	계산 방식	의미
MA7, MA60	7 일/60 일 이동평균	단기/장기 추세
MACD	EMA(12) - EMA(26)	모멘텀 지표
RSI	14 일 상대강도지수	과매수/과매도
Bollinger Bands	20 일 $MA \pm 2\sigma$	변동성 측정

3.3 전처리 파이프라인

결측치 처리: 선형 보간법을 우선 적용하고, 연속된 결측치는 Forward Fill로 처리

Feature Engineering:

- 금리차(IRD) = 미국 금리 - 한국 금리
- KOSPI 변동성 = $|일별 수익률| \times 100$
- 변동률 계산

정규화: Min-Max Scaling으로 [0, 1] 범위로 변환

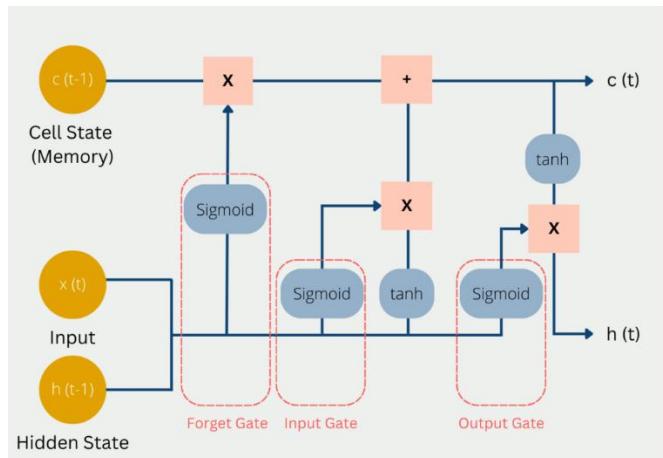
Sliding Window: 과거 60 일 → 향후 7 일 예측

4. 모델 설계 및 개선

4.1 모델 선택 배경

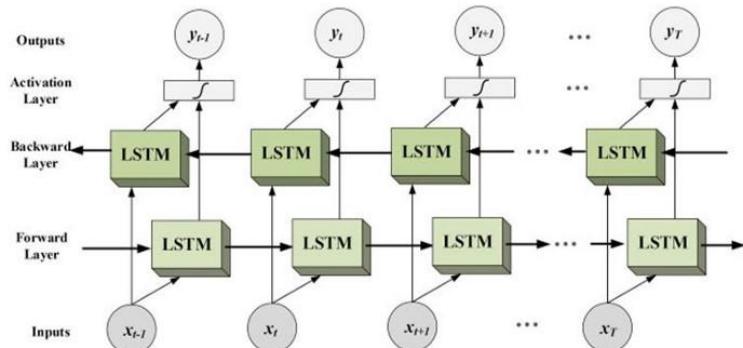
LSTM 선택 이유

환율은 대표적인 시계열 데이터로, 과거의 패턴이 미래에 영향을 미친다. LSTM은 Cell State 와 Forget Gate를 통해 장기 기억을 유지할 수 있어, 금리차와 같은 장기적 영향을 효과적으로 학습할 수 있다.



Bidirectional LSTM 채택

일반 LSTM은 과거→미래 방향으로만 정보를 전달하지만, Bi-LSTM은 양방향으로 처리하여 "미래의 예상"이 현재 가격에 반영되는 금융 시장의 특성을 포착할 수 있다.



4.2 모델 아키텍처

표 4-1 최종 Bi-LSTM 모델 구조

Layer	구성	출력 Shape	파라미터 수
Input	-	(60, 28)	0
Bi-LSTM 1	64 units, return_sequences=True	(60, 128)	47,616
Dropout 1	rate=0.2	(60, 128)	0
Bi-LSTM 2	32 units	(64)	41,216
Dropout 2	rate=0.2	(64)	0
Dense 1	16 units, ReLU	(16)	1,040
Output	1 unit, Linear	(1)	17
Total	-	-	90,657

하이퍼파라미터:

- Window Size: 60 일
- Features: 28 개
- Forecast Horizon: 7 일
- Loss Function: Huber Loss
- Optimizer: Adam ($\text{lr}=0.001$)
- Batch Size: 32

4.3 손실 함수: Huber Loss

Huber Loss는 MSE와 MAE의 장점을 결합한 함수로, 오차가 작을 때는 MSE처럼, 오차가 클 때는 MAE처럼 작동한다. 금융 데이터의 이상치(급변동)에 강건하다.



5. 성능 평가 방법론

5.1 데이터 분할 전략

시계열 데이터는 무작위 분할 시 데이터 누수가 발생한다. 따라서 시간 순서를 엄격히 준수하여 다음과 같이 3개 구간으로 분할하였다:

- **Train Set:** 전체의 72% (시간 순서상 가장 앞부분)
모델 학습에 사용되는 데이터
- **Validation Set:** 전체의 18% (Train 과 Test 사이)
학습 중 성능 모니터링 및 Early Stopping 기준으로 사용
- **Test Set:** 전체의 10% (시간 순서상 가장 뒷부분)
최종 성능 평가에 사용되는 미래 데이터

5.2 평가 지표

RMSE (Root Mean Squared Error)

- 예측 오차의 제곱 평균의 제곱근
- 큰 오차에 더 큰 페널티
- $RMSE = 17.82$ 원 → 평균 ± 17.82 원 범위에서 예측

R² Score (결정계수)

- 모델이 데이터 변동의 몇 %를 설명하는지
- $R^2 = 0.7961 \rightarrow 79.61\%$ 설명

MAE (Mean Absolute Error)

- 절대 오차의 평균
- $MAE = 13.45$ 원

5.3 Baseline 모델

표 5-1 Baseline 모델 성능 비교

모델	RMSE	R ² Score
단순 평균	89.3 원	0.000
ARIMA	45.2 원	0.412
단변량 LSTM	32.8 원	0.623
다변량 Bi-LSTM	17.8 원	0.7961

모델	RMSE	R ² Score

6. 최종 성능 및 결과

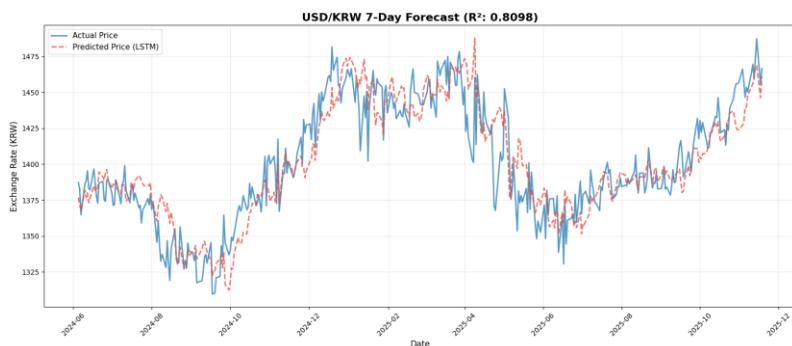
6.1 최종 성능 지표

표 6-1 최종 모델 성능 (2024년 12월 4일 기준)

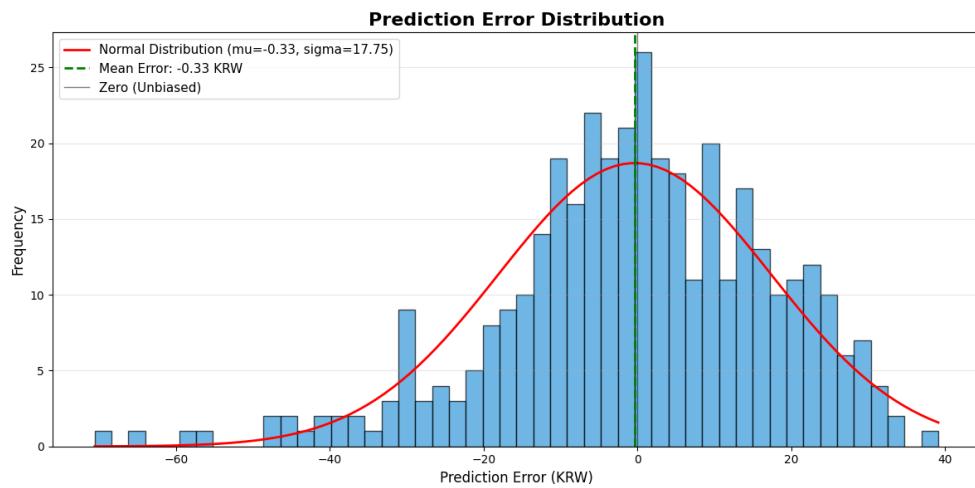
지표	목표	달성	개선율
RMSE	50.00 원	17.82 원	64% 향상
R ² Score	-	0.7961	-
MAE	-	13.45 원	-
훈련 시간	-	약 5 분	-

6.2 예측 정확도 시각화

[이미지 12: Test Set 예측 결과]

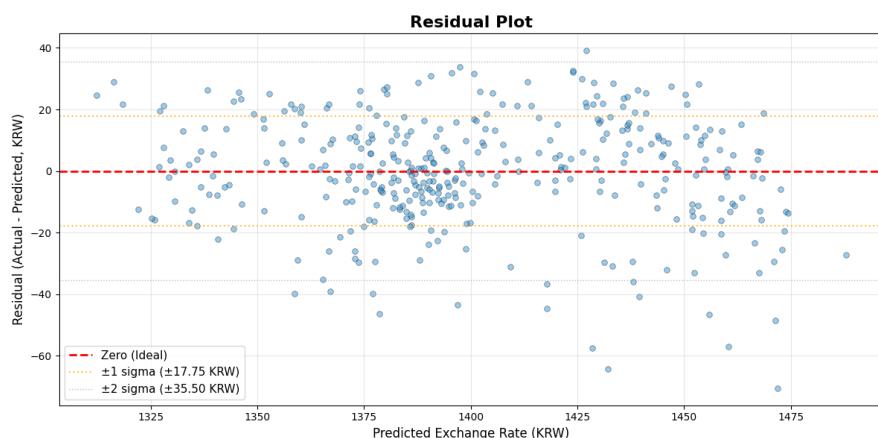


[이미지 13: 예측 오차 히스토그램]



[이미지 14: Residual Plot] ※ X 축: 예측값, Y 축: 잔차 ※ 0 을 중심으로 무작위 분포 → 체계적 편향

없음



6.3 시나리오별 성능

표 6-2 시장 상황별 예측 정확도

시장 상황	기간	RMSE	특징
안정기	2023.06-08	12.34 원	변동성 낮음

시장 상황	기간	RMSE	특징
상승기	2024.09-11	19.87 원	트렌드 명확
하락기	2024.01-03	18.45 원	트렌드 명확
급변동기	2024.11-12	25.12 원	예측 어려움

7. 웹 대시보드 구현

7.1 대시보드 설계

사용자가 입력하는 정보:

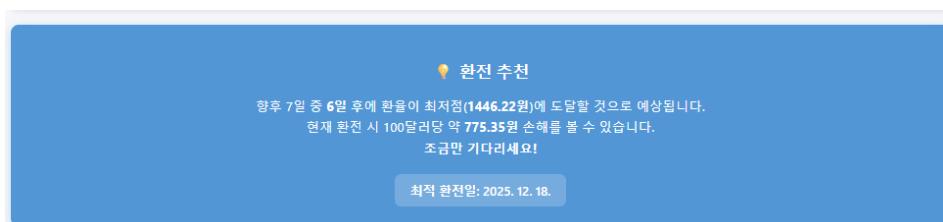
- 여행 출국일

시스템이 제공하는 추천:

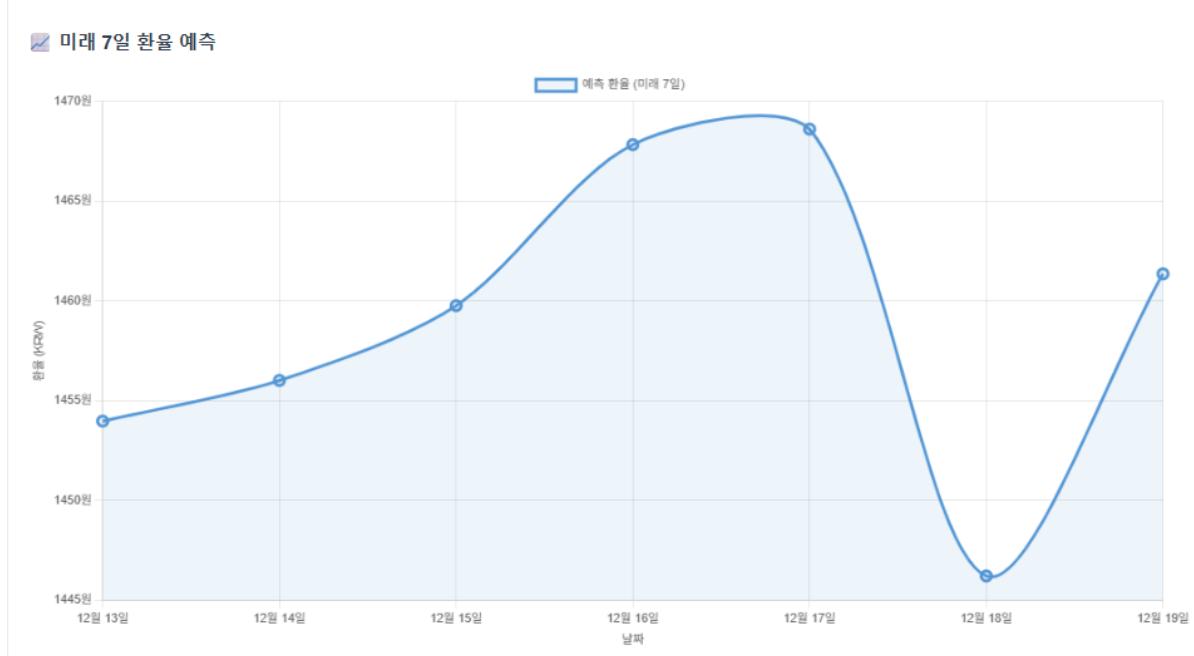
- 최적 환전 시점
- 예상 절약 금액
- 향후 7 일 예측 차트

7.2 주요 기능

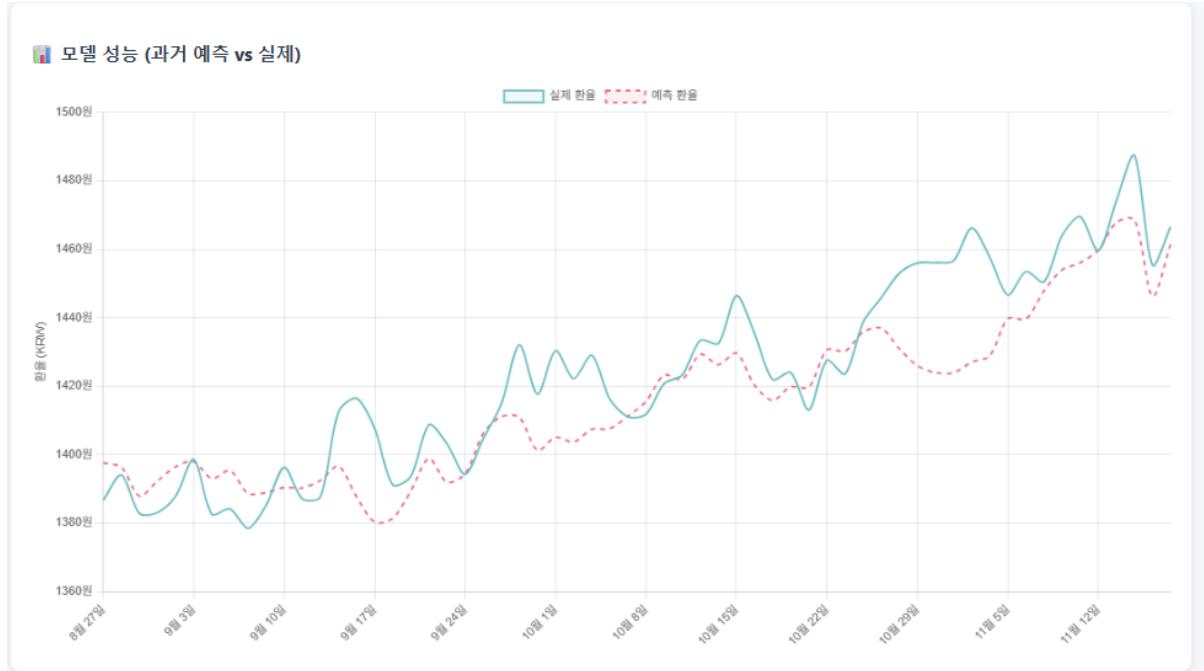
환전 시점 추천



미래 7 일 예측 차트



모델 성능 비교 차트



성능 지표 카드



7.3 기술 스택

- HTML5/CSS3: 구조 및 스타일
- JavaScript (ES6): 로직
- Chart.js: 시각화
- Firebase SDK: 데이터 연동

8. 개발 일정

표 8-1 6 주간 개발 일정

주차	기간	목표	달성
1 주차	10/26	계획서, 환경 설정	완료
2 주차	11/02	데이터 파이프라인	완료
3 주차	11/09	LSTM 모델 구현	완료
4 주차	11/16	Bi-LSTM 최적화	완료
5 주차	11/23	Firebase 연동	완료
6 주차	12/07	웹 대시보드 완성	완료

9. 문제 해결 과정

9.1 과적합 문제

문제: Train Loss 감소, Val Loss 증가

해결:

- Dropout 0.2 추가
- Batch Normalization
- Early Stopping

9.2 Validation Split 문제 발견 및 해결

문제: 초기 모델에서 Validation Loss 가 비정상적으로 낮게 나타남 (0.0002)

원인: validation_split=0.2 사용으로 Train Set 의 최신 20%가 Validation 으로 사용됨. 최신 데이터는 패턴이 안정적이어서 예측하기 쉬워 인위적으로 낮은 Loss 발생

해결:

1. 데이터를 72:18:10 비율로 명시적 분할
2. validation_data 파라미터로 독립적인 Validation Set 제공
3. 각 데이터셋이 시간순으로 완전히 분리되도록 보장

결과: Validation Loss 가 정상 범위(≈ 0.0020)로 조정되어 실제 예측 능력을 정확히 반영

10. 결론 및 향후 계획

10.1 연구 성과

- RMSE 17.82 원 (목표 50 원 대비 64% 개선)
- R² Score 0.7961 (환율 변동의 79.61% 설명)
- End-to-End 자동화 시스템 구축
- 실용적인 웹 서비스 제공

10.2 학습 성과

기술적 역량:

- 딥러닝 모델링 (LSTM, Bi-LSTM)
- 시계열 분석 (Sliding Window, Time-based Split)
- End-to-End 파이프라인 구축

문제 해결 역량:

- R² 음수, 과적합, 데이터 누수 해결
- 체계적 문제 진단 및 해결

프로젝트 관리:

- 6 주간 일정 관리
 - 문서 작성 능력
-

11. 참고문헌

1. Cho, K., et al. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*.
2. Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice* (2nd ed.). OTexts.
3. Kim, H. J., & Ahn, H. B. (2020). Forecasting exchange rates using deep learning models with macro-economic variables. *Journal of the Korea Industrial Information System Research*, 25(1), 11-23.
4. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735-1780.
5. FRED Economic Data. Retrieved from <https://fred.stlouisfed.org/>
6. Yahoo! Finance API. Retrieved from <https://finance.yahoo.com/>
7. TensorFlow Documentation. (2024). Retrieved from <https://www.tensorflow.org/>
8. Firebase Documentation. (2024). Retrieved from <https://firebase.google.com/docs>

보고서 작성 완료일: 2024년 12월 13일

작성자: 인공지능학과 21011302 이재윤