

외생변수와 전이학습을 통한 오일변동성 예측  
- 뉴스데이터, S&P500 활용 -

김주환, 김정현, 유동연

지도 교수 : 송 재 욱 (인)

심사위원 교수 : (인)

한양대학교 산업공학과

## 국문 초록

본 연구는 뉴스기사와 S&P500, 그리고 전이학습을 활용하여 Brent원유선물 변동성 예측을 진행하였다. 뉴스기사 제목을 감정분석 또는 임베딩벡터를 그대로 활용하여 고려하였고, 금융위기 기간 예측성능 향상을 위해 S&P500도 같이 고려하였다. 타 연구들과 비교하여 트랜스포머의 문장 임베딩 벡터를 그대로 활용하는 시도를 하였다는 점과 오일변동성 예측에 전이학습 유효성을 확인한다는 점에서 차별점을 가진다. 연구에 사용된 오일데이터는 WTI원유현물 데이터와 Brent원유선물 데이터를 사용하였고, 각각 1986년 부터 2024년까지, 2002년부터 2024년까지 데이터가 수집되었다. 뉴스기사와 S&P500도 해당기간에 맞춰 수집하여 고려하였다. 이 때, 딥러닝 모델로는 BiLSTM을 사용하였으며, 길이가 가장 긴 WTI원유현물 데이터를 통해 베이스 모델을 만들고 그보다 길이가 짧은 Brent원유선물 변동성 예측에 전이학습을 적용하였다. 분석결과, 전이학습은 더 많은 시계열 데이터 확보와 함께 예측성능 향상을 이루었고, 뉴스데이터 감정분석과 S&P500 활용은 이전 변동성만 고려할 때보다 더 높은 예측성능을 보여주었다. 문장임베딩 벡터를 그대로 사용하는 것은 예측 정확도에 긍정적인 영향을 미치지 않았다. 본 연구는 다음과 같은 결과를 바탕으로 적절한 외생변수의 활용과 전이학습은 시계열변동성 예측 정확도 향상에 도움이 됨을 주장한다.

주 용어: 임베딩벡터, 감정분석, S&P500, 전이학습, 오일변동성예측

## 목차

### 국문초록

1. 서론 .....	6
2. 문헌 고찰 .....	7
3. 데이터 및 방법론 .....	8
3.1 데이터 처리 및 분석 .....	8
3.1.1 뉴스 데이터 처리 .....	8
3.1.3 데이터 분석 .....	9
3.2 방법론 .....	12
3.2.1 BiLSTM .....	12
3.2.2 전이 학습 .....	14
3.2.3 자연어 처리 .....	14
3.3 Performance Evaluation .....	15
4. 결과 .....	16
4.1 자연어 처리 결과 .....	16
4.2 Total-Period 결과 .....	19
4.3 Sub-Period 결과 .....	20
4.4 예측모델별 DM-Statistic 비교 .....	23
5. 결론 및 향후 연구방안 .....	23
6. 참고문헌 .....	25
Abstract .....	27

## 표 목차

표 1	.....	9
표 2	.....	12
표 3	.....	16
표 4	.....	19
표 5	.....	20
표 6	.....	23

## 그림 목차

그림 1	10
그림 2	11
그림 3	11
그림 4	12
그림 5	13
그림 6	13
그림 7	14
그림 8	17
그림 9	17
그림 10	17
그림 11	18
그림 12	18
그림 13	18
그림 14	19
그림 15	19
그림 16	20
그림 17	20
그림 18	21
그림 19	22
그림 20	22

## 1. 서론

원유는 국가 경제 발전을 위한 전략적 상품으로 원유가격 변동성은 세계 경제 전반에 걸쳐 상당한 영향을 미친다. 최근에 일어난 COVID-19 사태와 러시아-우크라이나 전쟁 등은 원유의 가격에 큰 영향을 주어 금융 시장의 변동성을 극대화시켰고, 이는 공급망 교란, 에너지 수급 불균형, 국제 무역 긴장으로 이어지며 경제적 불확실성을 더욱 심하게 하였다.

이러한 맥락에서 원유 변동성을 예측하는 것은 정책 입안자, 투자자 및 기업들에게 더욱 중요한 과제가 되었다. 금융 변동성은 자산 가격의 변동성을 측정하는 중요한 지표로 리스크 관리 및 투자 전략 수립에 필수적인 요소이다. 특히 석유와 같은 원자재 시장의 금융 변동성은 글로벌 경제에 직접적인 영향을 미치기 때문에, 이를 정확하게 예측하는 것은 경제적 안정성을 유지하는데 있어 중요한 과제이다.

본 논문은 유가 변동성을 예측하는데 S&P500, 유가 관련 뉴스 기사, 그리고 딥러닝 접근법 중 하나인 전이학습을 활용하여 정확성을 높이하고자 한다. 인공지능을 활용하여 금융 변동성을 예측한 기존의 연구에서는 시계열 데이터의 길이가 짧아 학습이 원활하게 되지 않는 문제가 발생하였는데 이를 해결하고자 길이가 긴 WTI원유현물 데이터를 학습하여 베이스 모델로 만들고 길이가 짧은 다른 유가 선물 데이터를 예측하는데 전이학습을 활용하였다. 또한, S&P500과 원유 간 변동성 파급 효과의 존재를 확인하고, 유가 관련 뉴스 기사를 예측에 활용함으로써 가격 데이터로 반영할 수 없는 시장 심리를 효과적으로 반영하고자 한다. 이를 통해 투자자들은 신뢰할 수 있는 유가 예측정보를 바탕으로 시장의 리스크를 관리하고 적절한 대응 전략을 수립할 수 있을 것이다.

## 2. 문헌 고찰

최근까지도 원유가격 및 변동성 예측에 머신러닝을 활용한 연구는 활발하게 진행되고 있다. Yu et al.(2008)은 원유 현물 가격 예측을 위해 EMD 기반 신경망 학습을 제안하였으며, 이는 금융 시계열 데이터의 비선형 패턴을 잘 포착할 수 있음을 보여주었다. He et al.(2012)는 신호를 다양한 주파수 대역으로 분해하는 웨이블릿 변환과 여러 개별 모델을 결합한 앙상블 모델을 통해 원유가격을 예측하였다. 이는 기존의 단일 모델보다 성능이 뛰어나며, 다양한 시나리오에 높은 적응력을 보여주었다. He et al.(2019)은 기존의 딥러닝 모델이 짧은 시계열 데이터에 대해 성능이 저하되는 문제를 해결하기 위해 전이 학습을 사용하여 금융 시계열 데이터 예측 성능을 올리는 방법론을 제시하였다. 저자는 첫 번째 데이터로 모델을 사전 학습한 후, 두 번째 데이터로 모델의 일부 레이어를 고정하고 나머지 레이어를 학습시켰다. 이후 타겟 데이터셋을 사용하여 전체 모델을 파인튜닝하여 성능을 올렸다. 그 결과, 전이 학습 전략을 사용한 모델이 그렇지 않은 모델보다 우수한 성능을 보였다. Li et al.(2021)은 뉴스 감정분석을 통해 원유 선물 시장의 수익률과 변동성을 예측하는 딥러닝 모델을 제안하였다. 연구는 뉴스 감정지수가 시장 참가자의 심리를 반영하여 원유 선물 시장 움직임에 영향을 미칠 수 있음을 보여주었다. Wu et al.(2021)은 COVID-19 팬데믹 동안 소셜 미디어 정보를 바탕으로 미국 원유 시장을 예측하는 연구를 진행하였다. 이 논문은 팬데믹 기간 동안 소셜 미디어에서 발생한 감성 변화가 원유 변동성에 미치는 영향을 연구하였고, 이를 활용하여 더 정확한 예측을 할 수 있음을 확인하였다. Zhang et al.(2023)은 거시 변수를 사용하여 유가 변동성을 예측성을 향상시키는 연구를 제안했다. FRED-MD 데이터 베이스에 있는 127개 거시변수 중 상대적으로 유용한 변수를 선택하여 머신러닝 모델의 입력 데이터로 사용함으로써 예측 성능을 향상 시켰다. Abdollahi(2023)은 최신 자연어 처리 기술을 사용하여 유가 변동성을 예측하는 접근법을 제안하였다. 저자는 유가 관련 뉴스 및 트위터 피드를 통해 감정지수를 추출하고, 이를 GARCH 모델과 BiLSTM 모델의 혼합 구조로 통합하였다. 평가 지표로 RMSE(Root Mean Square Error)와 MAE(Mean Absolute Error)를 사용했으며, 평가 모델로는 7개의 모델을 사용하고 예측 기간은 16주, 22주, 28주로 설정하였다. 모든 예측 기간에서 뉴스 감성 지수를 사용한 모델이 가장 낮은 RMSE와 MAE값을 보였으며, 특히 News-GARCH-BiLSTM 모델이 가장 낮은 예측 오차를 기록하였다. 본 논문은 뉴스 감성 지수가 유가 변동성 예측에서 트위터 감성 지수보다 더 높은 예측 성능을 보여준다는 것을 입증하였으며, GARCH와 BiLSTM을 결합한 하이브리드 모델이 단일 모델보다 더 나은 예측 정확도를 보여주었다. 그러나 뉴스와 트위터에서 추출한 감정지수의 경우 특정 이벤트나 상황에 따라 감성의 편향이 발생할 수 있으며, 데이터를 감정지수로 변환하는 과정에서 정보가 손실될 수 있다는 한계를 지니고 있다.

자연어 처리 과정에서의 이러한 정보 손실을 줄이기 위한 노력은 꾸준히 이어져 오고 있다. 현재 가장 널리 쓰이는 방법은 자연어를 벡터로 변환하는 기법이다. Mikolov et al.(2013)은 Word2Vec의 Skip-gram과 Continuous Bag of Words (CBOW) 모델을 통해 단어 벡터를 효율적으로 학습하는 방법을 제안하였다. Word2Vec 기법은 자연어를 임베딩 벡터라고 불리는 밀집 벡터로 표현하는 것을 가능하게 하였다. 이후 Reimers and Gurevych(2019)가 제안한 Sentence-BERT는 단어를 넘어 문장 전체의 임베딩 성능을 크게 향상시켰다. 이러한 문장 임베딩 방법들은 지속적으로 연구되고 있다.

앞서 소개한 선행 연구들은 뉴스와 트위터의 기사의 감정분석 결과를 외생 변수로 사용하여 변동성 예측의 성능을 높이려 했다. 그러나 짧은 시계열 데이터에는 활용하기 어려우며, 뉴스 기사 개수에 따른 편향 및 정보 손실이 발생할 수 있다는 문제가 있었다. 이러한 문제를 해결하기 위해 두 가지 방법을 제안한다. 첫 번째 방법은 길이가 긴 원유 현물 데이터를 바탕으로 베이스모델을 학습하고, 학습된 가중치를 이용해 길이가 짧은 원유 선물 시계열 변동성을 예측하는 전이 학습 방식을 사용하는 것이다. 두 번째 방법은 자연어 처리 과정에서 정보 손실과 감성 편향 문제를 해결하기 위해 문장을 임베딩 벡터 등으로 변환하는 것이다.

위 방법들을 통해 원유 데이터에서의 전이 학습의 유효성을 연구하고, 기존 선행 연구가 자연어를 처리한 방식의 유효성을 검증 및 개선안을 찾고자 한다.

### 3. 데이터 및 방법론

#### 3.1 데이터 처리 및 분석

##### 3.1.1 뉴스 데이터 처리

변동성 예측에 있어서 외생 변수로 활용할 자연어는 뉴스의 제목만을 사용했다. 뉴스의 제목 데이터는 Google News 에서 크롤링을 통해 약 50,000건을 수집했다. Crude, Brent 등의 키워드들로 검색하여 우선 수집한 뒤, 결측치가 발생한 주에 한해서 키워드를 oil로 수정 후 기사 수집을 진행했다. 이후 주간 데이터로 활용하기 위해 내용의 중요성 등을 고려하여 불필요한 뉴스를 제거했다. 이후 한 주에 하나의 뉴스 기사만을 남겨두었으며, 맞춤법 교정, 특수문자 제거 등 전처리 과정을 거쳐 데이터를 정제하였다.

##### 3.1.2 오일가격데이터 처리

1986년 1월부터 2023년 12월까지의 WTI원유현물 데이터는 Alphavantage의 데이터를 활용하였고, 2000년 2월부터 2023년 12월까지의 Brent 원유선물가격은 Yahoo Finance의 데이터를 활용하였다. 오일가격은 일별데이터로 수집되었으며 daily log return을 구한 뒤 20주를 Window-size로 하여 Zhong et al.(2023)이 정의한 실현변동성을 계산하였다. 수식은 다음과 같다.

$$RV_t = \sqrt{RVar_t} = \sqrt{\sum_{j=1}^M r_{t,j}^2}, \quad t = 1, \dots, T. \quad (1)$$

- $RV_t$ : 시점  $t$ 의 실현 변동성
- $\sqrt{RVar_t}$ : 시점  $t$ 의 실현 분산의 제곱근
- $\sum_{j=1}^M r_{t,j}^2$ : 시점  $t$ 의  $M$ 개의 하위 간격에서 수익률  $r_{t,j}$ 의 제곱 합

현실 오일변동성 예측은 일간 단위가 아닌 주간 단위로 필요하기 때문에 일간 빈도로 있는 실현변동성은 Haugom et al.(2014)이 정의한 주간 변동성을 통해 주간 빈도로 변경하였다. 수식은 다음과 같다.

$$RV_t^{(w)} = \frac{1}{5}(RV_t^{(d)} + RV_{t-1d}^{(d)} + \dots + RV_{t-4d}^{(d)}) \quad (2)$$

- $RV_t^{(w)}$ : 시점  $t$ 의 주간 실현 변동성
- $RV_t^{(d)}$ : 시점  $t$ 의 일간 실현 변동성



본 연구에서는 전이학습의 유효성 실험을 위해 변동성 데이터를 베이스모델용 데이터와 재학습용 데이터로 나누었다. 베이스모델용 데이터는 전이학습을 위해 베이스 모델을 학습시키는 훈련 데이터를 의미한다. 재학습용 데이터는 실제 예측이 이루어지는 기간으로 미리 훈련된 베이스모델 가중치를 그대로 사용하면서 출력레이어만 재학습하고 한 포인트를 예측하게 된다.

현실에선 원유 변동성의 분포가 계속 변할 것이므로 변동성 예측모델의 주기적인 재학습이 필요하다. 때문에 재학습은 한 포인트 예측 이후 매번 진행되며, 2000년 2월부터 예측이 끝난 시점까지 Brent원유선물 변동성 데이터로 재학습이 진행된다.

Table 1. Number of Volatility

Weekly Data	Full Sample	Number of Observations	Backbone-Sample Estimation	Number of Observations	Retrain-Sample Predictions	Number of Observations
WTI Volatility	1986-01-31 ~2024-01-05	1976	1986-01-31 ~2004-12-31	988	2005-01-07 ~2024-01-05	988(사용X)
Brent Volatility	2000-02-04 ~2024-01-05	1247	2000-02-04 ~2004-12-31	257	2005-01-07 ~2024-01-05	990

### 3.1.3 데이터 분석

본 연구에서는 2005년 1월부터 2023년 12월까지의 Brent원유선물 변동성을 예측한다. 이 기간은 글로벌 금융위기(2008년-2009년), 국제유가 폭락(2014년), Covid-19(2020년), 러시아-우크라이나 전쟁(2022년)을 포함한 기간으로 <Figure 1>은 예측기간동안 WTI원유현물의 수익률과 변동성을 나타내는 것이며, <Figure 2>는 예측기간동안 Brent 원유선물의 수익률과 변동성을 나타낸다. 수익률은 일간 로그수익률로 계산되었고, 변동성은 실험에 실제로 사용된 주간 실현변동성을 의미한다. 원유가격은 주요 이벤트가 발생하는 기간 동안 급격한 변동을 보이며, 이는 원유가격 변동이 에너지 시장과 관련된 불확실성에 영향을 받을 수 있음을 의미한다.(Zhang et al., 2024). <Table 2>는 예측기간 동안 WTI원유현물 변동성과 Brent원유선물 변동성의 통계량을 나타낸다. WTI 원유현물 변동성이 Brent원유선물 변동성보다 평균적으로 더 큰 변동폭을 가지고

있고, 두 시계열 모두 정상성을 띄지만 데이터분포는 정규분포를 따르지 않는다. <Figure 1>과 <Figure 2>를 통해 두 변동성 시계열은 매우 유사한 움직임을 가지며 <Figure 3>에선 예측기간 동안 WTI원유현물 변동성과 Brent원유선물 변동성의 실제 상관관계수가 0.82로 매우 높은 양의 상관관계를 가짐을 확인한다. 이를 근거로 WTI원유현물을 통해 Brent원유선물 변동성 전이학습을 고려하였다.

또한, Brent원유선물 변동성 예측에 S&P500를 고려하였다. Bagcilar et al.(2021)에 따르면 S&P500, 원유, 금 자산 간에는 양방향 수익률과 변동성 파급 효과가 존재하는 것으로 나타났다. 그들은 Diebold와 Yilmaz(2012)의 스피로버 지수를 사용하여 S&P 500지수, 원유, 금 간의 수익률 및 변동성 파급 효과를 분석하였다. 1986년 1월부터 2018년 8월까지의 월별 실현 변동성 및 수익률 데이터를 활용하여 수익률과 변동성의 파급 효과를 조사하였고, 연구결과에 따르면 전체 표본의 경험적 증거는 석유가 정보 전달 메커니즘에서 중심적인 역할을 수행한다는 구조와 일치하였다.

이를 근거로 S&P500을 외생변수로 고려하였으며, <Figure 3>은 예측기간동안 S&P500의 일간 로그수익률을 나타낸다. 그림에서 볼 수 있듯이 원유 변동성 그래프에 큰 변화를 주었던 기간에 S&P500의 수익률에도 비슷한 변화가 나타남을 확인하였다.

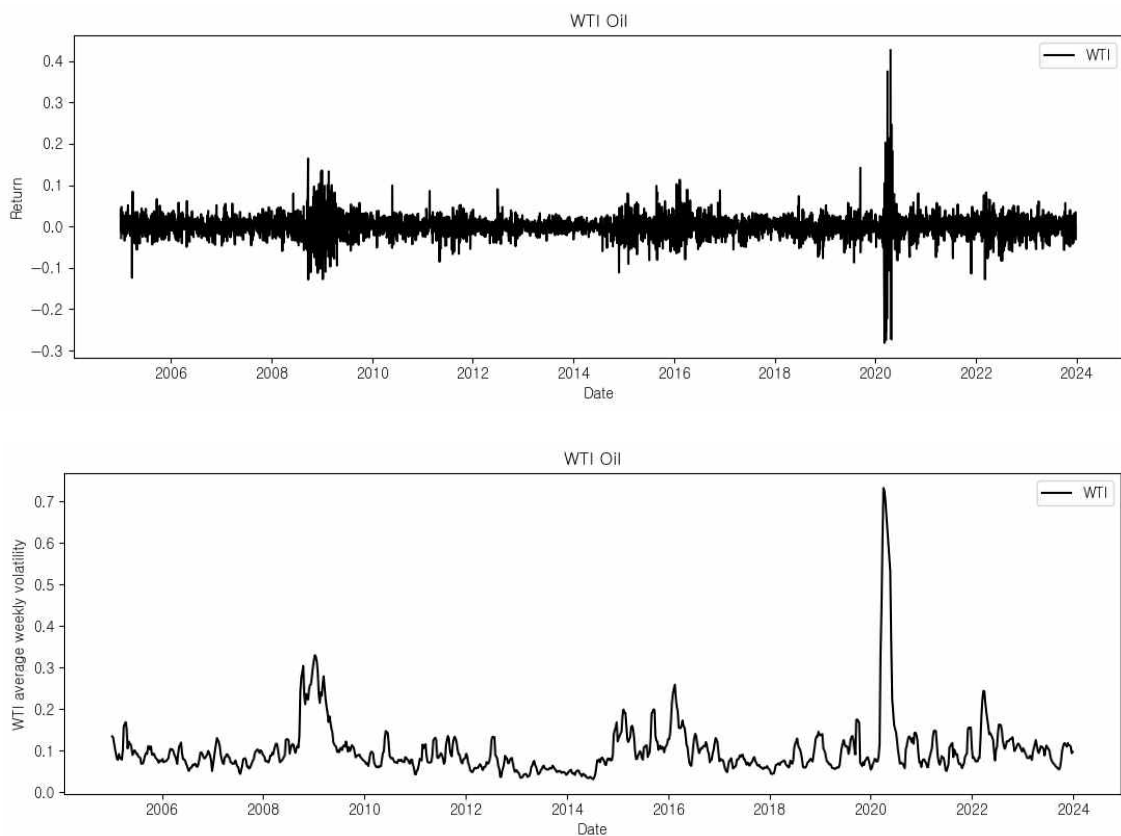


Figure 1. WTI daily log return & average weekly volatility

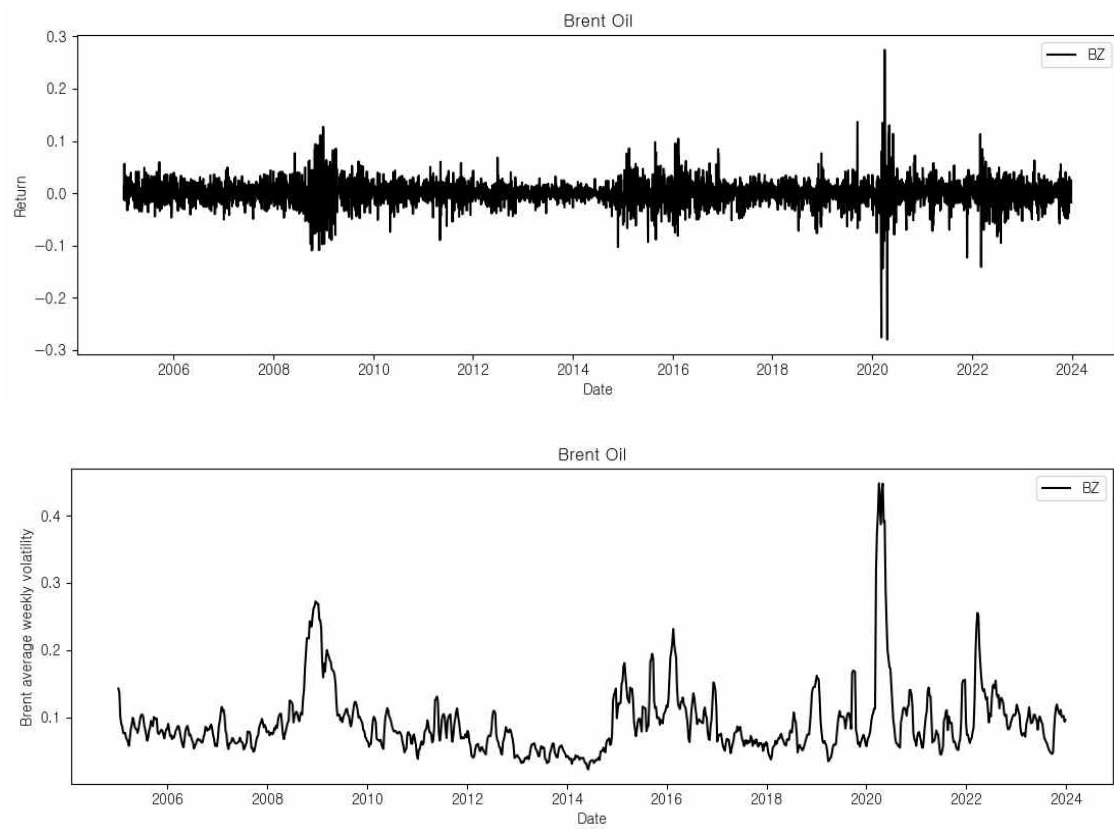


Figure 2. Brent daily log return & average weekly volatility

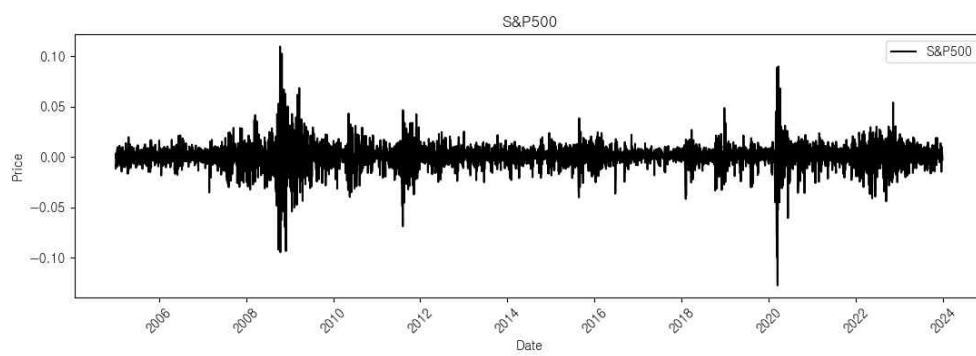


Figure 3. S&P500 daily log return

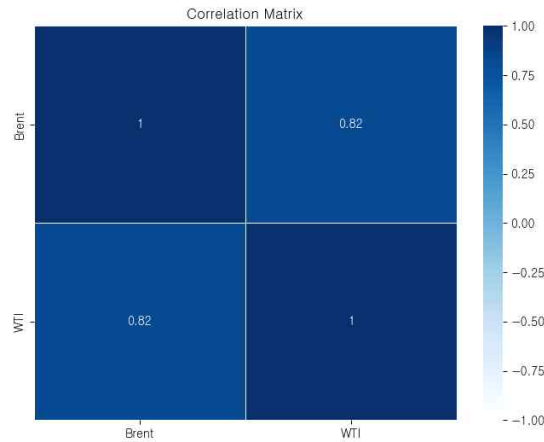


Figure 4. Correlation Matrix between WTI volatility and Brent volatility

Table 2. Summary Statistics of Volatility

Total	mean	std	min	25%	50%	75%	max	kurto sis	skew ness	JB	ADF
WTI	0.1010 1	0.0583 74	0.0307 18	0.0683 07	0.0877 47	0.1169 5	0.7325 4	27.526 1	3.8558	67,590 ***	-6.720 2***
Brent	0.0937 22	0.0480 48	0.0226 41	0.0652 28	0.0833 24	0.1067 7	0.4482 2	14.360 3	2.9392	12,603 ***	-4.556 7***

참고: 이 표는 WTI원유현물, Brent원유현물 변동성에 대한 설명적 통계를 나타낸다. 이 연구에서는 평균(mean), 표준편차(Std), 왜도(Skewness), 첨도(Kurtosis), 자크-베라 통계(JB), 증강 디키-풀러 테스트(ADF) 등이 사용된다. 별표 \*\*\*, \*\* 및 \*는 1%, 5% 및 10% 수준에서 귀무가설을 거부함을 의미한다.

## 3.2 방법론

### 3.2.1 BiLSTM

본 논문은 시계열 데이터 분석에서 자주 사용되는 LSTM(Long Short-Term Memory) 네트워크의 확장 버전인 BiLSTM 모델을 사용하였다(Hochreiter and Schmidhuber, 1997). LSTM은 순환 신경망(RNN)의 일종으로 장기 의존성 문제를 해결하기 위해 설계되었다.

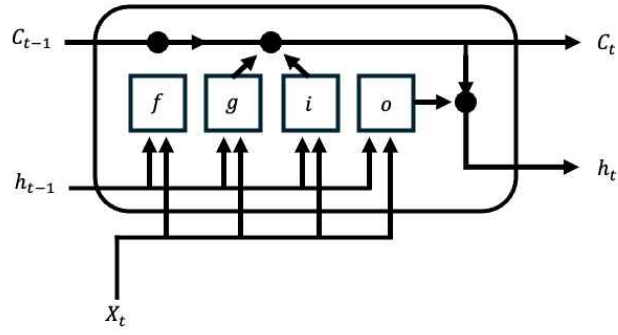


Figure 5. LSTM cell

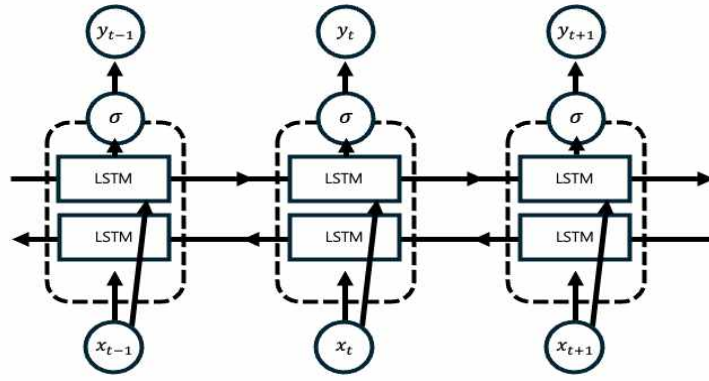


Figure 6. Bi-LSTM

LSTM의 수식은 다음과 같다.

$$g_t = \sigma(b_g + U_g x_t + V_g h_{t-1}) \quad (3)$$

$$i_t = \sigma(b_i + U_i x_t + V_i h_{t-1}) \quad (4)$$

$$o_t = \sigma(b_o + U_o x_t + V_o h_{t-1}) \quad (5)$$

$$f_t = \tanh(b_f + U_f x_t + V_f h_{t-1}) \quad (6)$$

$$c_t = g_t \odot c_{t-1} + i_t \odot f_t \quad (7)$$

$$h_t = o_t \odot \tanh(c_t) \quad (8)$$

LSTM 네트워크는 입력 게이트, 망각 게이트, 출력 게이트를 통해 정보의 흐름을 조절한다. 입력 게이트  $i_t$ 는 새로 들어오는 정보의 중요성을 판단하고, 망각 게이트  $g_t$ 는 이전 상태를 얼마나 유지할지 결정한다. 입력 및 망각 게이트의 결과를 사용하여 셀 상태를 갱신하고 (5), 출력 게이트  $o_t$ 는 현재 상태를 기반으로 활성화 값을 사용하여 최종 은닉 상태( $h_t$ )를 계산하여 이를 다음 타임스텝으로 전달한다.  $\sigma$ 는 활성화 함수를 나타내고,  $b$ 는 편향 항이다.  $U_i$ ,  $U_g$ ,  $U_o$ 는 각각 입력 게이트, 망각 게이트, 출력 게이트의 가중치 행

렬을 나타내고,  $V_i, V_g, V_o$ 는 각각의 입력 게이트, 망각 게이트, 출력 게이트의 순환 가중치 행렬을 나타낸다. BiLSTM(양방향 LSTM)은 두개의 LSTM 네트워크를 사용하여 입력 시퀀스를 순방향과 역방향 모두 처리함으로써, 긴 시퀀스 데이터나 복잡한 문맥을 이해하는데 효과적이며 단방향 LSTM보다 향상된 성능을 제공한다. <Figure 6>에서  $X_t$ 는 입력변수이다. 입력변수는 본 실험에서 20주간 유가 상품 변동성과 이와 관련된 신문 기사에서 자연어 처리를 통해 추출한 벡터로 구성했다. 입력된 변수를 순방향 LSTM 셀과 역방향 LSTM 셀을 순차적으로 거쳐 시퀀스 데이터를 처리하고 두 LSTM에서 각각 출력된 정보를 결합한 뒤 예측값을 도출했다.

### 3.2.2 전이 학습

전이학습(Transfer learning)은 한 도메인에서 학습된 모델을 다른 관련 도메인에 적용하는 머신러닝 기법이다. Gururangan et al.(2020)의 연구에서는 자원이 부족한 상황에서 전이 학습은 성능을 효율적으로 향상시킬 수 있는 방법임을 보여줬다. 딥러닝 모델은 학습하는데 많은 데이터가 필요한데, 금융 데이터는 길이가 짧은 한계를 지니고 있다. 본 연구는 전이학습을 도입하여 한계를 극복하고 성능을 향상시키고자 했다. 전이 학습 구현을 위해 Python의 Pytorch 라이브러리를 사용하였다. 사전학습 단계에서는 오일 관련 인덱스 중 길이가 가장 긴 WTI원유현물 데이터를 사용해서 모델을 학습시켰다. 파인튜닝 단계에서는 사전 학습된 모델 레이어의 가중치를 고정 시키고, 마지막 출력 레이어의 가중치만 WTI원유현물 데이터에 비해 길이가 짧은 Brent원유현물 데이터를 학습시켜 예측 성능을 향상시켰다. <Figure 7>에서 하늘색은 재학습 시 가중치를 업데이트하지 않는 레이어를 의미하고, 빨간색은 재학습시 가중치가 업데이트되는 레이어이다. 본 연구에서는 마지막 출력레이어만을 재학습하여 전이학습을 진행한다.

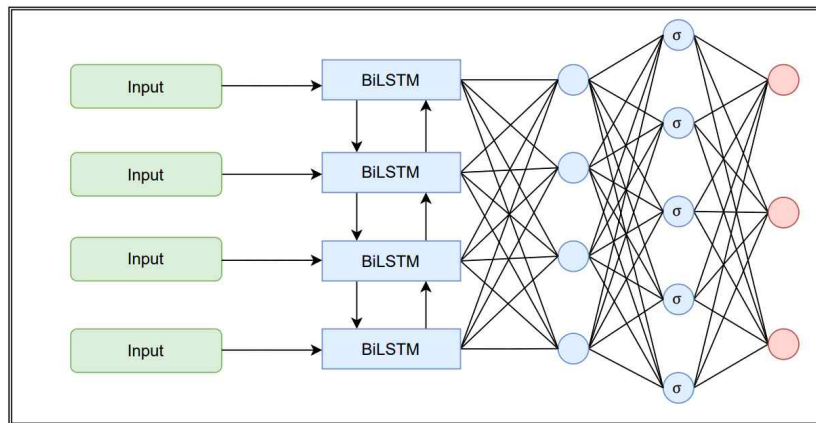


Figure 7. Transfer Learning

### 3.2.3 자연어 처리

본 연구에서는 뉴스 제목 데이터를 외생변수로 활용하기 위해 다음과 같은 방법들을 시도하였다. 먼저, 감성분석을 위해 BERT의 파생 모델인 RoBERTa에 증류 기법을 적용한 DistilRoBERTa 모델을 사용하였다. 특히 금융 뉴스의 감성 분류를 목적으로 파인튜닝된 오픈 소스 모델인“mrm8488/distilroberta-finetuned

-financial-news-sentiment-analysis“를 활용하였다. 증류 기법을 사용한 이 모델은 기본 BERT와 RoBERTa 모델에 비해 높은 효율을 보여 감정분석에 적합하였다. 다음으로, 임베딩 벡터화 방법을 사용하였다. 센텐스임베딩 벡터는 텍스트 데이터를 고차원의 벡터로 변환하여 문장의 의미를 최대한 많이 담을 수 있다는 장점이 있다. 본 연구에서는 “sentence-transformers/ all-MiniLM-L6-v2“ 모델을 사용하여 뉴스 데이터를 384차원의 벡터로 임베딩하였다.

Zhang, G, et al.(2024)에 따르면 주성분분석(PCA)과 같은 간단한 방법으로 문장 임베딩의 차원을 거의 50%까지 축소하여도 다운스트림에서 성능저하가 거의 발생하지 않았다고 주장한다. 이를 근거로 이 연구에서는 384차원의 임베딩벡터와 UMAP기법을 활용해 BiLSTM 에 적절한 차원(75, 100)으로 축소하여 성능을 비교하였다. 또한, 텍스트 데이터를 다룰 때 고전적 방법론 중 잘 알려진 정수 인코딩을 사용하여 벡터화를 진행했다. 따라서 본 연구는 선행 연구에서 사용된 감정 분석이 외생 변수로서 최선의 방법인지 검증하고, 더 나은 방법이 있는지를 연구하였다. 이를 통해 뉴스 데이터를 활용한 예측 성능을 최대화하고자 하였다.

### 3.3 Performance Evaluation

예측 성능을 평가하기 위해서 MAE(Mean Absolute Error)와 RMSE(Root Mean Square Error)를 사용하였다. MAE는 예측 값과 실제 값 사이의 절대 오차들의 평균을 나타낸다. 오차의 절대 값을 사용하기에 모든 오차를 동일하게 처리하며, 이상치에 덜 민감하다. RMSE는 예측 값과 실제 값 사이의 제곱 오차들의 평균을 구한 후, 그 값의 제곱근을 취한 값을 나타낸다. 모델의 예측이 실제 값과 얼마나 가까운지를 측정하지만, 오차의 제곱을 사용하여 큰 오차에 더 큰 패널티를 부여하기 때문에 MAE보다 이상치에 더 민감하다. 두 평가 지표의 수식은 아래와 같다.

$$MAE = \frac{1}{n} \sum_{i=1}^n | \sigma_i - \hat{\sigma}_i | \quad (9)$$

- $n$ : 관측치의수
- $\sigma_i$ : 실제 값
- $\hat{\sigma}_i$ : 예측 값

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\sigma_i - \hat{\sigma}_i)^2} \quad (10)$$

- $n$ : 관측치의수
- $\sigma_i$ : 실제 값
- $\hat{\sigma}_i$ : 예측 값

예측 모델간의 성능을 비교하는데 Diebold-Mariano (DM) test를 사용하였다. DM 통계량을 계산하는 수식은 아래와 같다.

$$DM = \frac{\bar{d}}{\sqrt{\hat{\sigma}_d^2 / T}} \quad (11)$$

- $\bar{d}$ : 예측 오차 차이의 평균
- $\hat{\sigma}_d^2$ : 예측 오차 차이의 분산
- $T$ : 예측길이의길이

DM test의 가설검정은 다음과 같다

- 귀무가설( $H_0$ ) : 두 모델의 예측 성능에 차이가 없다.
- 귀무가설( $H_1$ ) : 두 모델의 예측 성능에 차이가 있다.

## 4. 결과

### 4.1 자연어 처리 결과

Table 3. Total-Period 오차측정지표

Total	Vol	Vol+Embedding	Vol+UMAP_100	Vol+UMAP_75	Vol+IE	Vol+Sentiment
MAE	0.01263	0.03257	0.01731	0.01654	0.01444	0.01160
RMSE	0.01835	0.04228	0.02259	0.02171	0.01953	0.01672

위 표는 외생변수로 사용한 임베딩 벡터의 추출 방식별 예측 오차와 정수인코딩, 감정지수 결과를 비교했다. 384차원의 임베딩 벡터를 그대로 활용한 예측그래프 <Figure 9>를 보면 정답과 매우 동떨어진 예측을 수행함을 확인하였다. 이는 어텐션 메커니즘이 없는 LSTM 특성상 차원수가 클수록 성능이 떨어지는 것으로 해석된다. 반면에 100차원으로 축소된 예측그래프 <Figure 10>과, 75차원으로 축소된 예측그래프 <Figure 11>을 보면 임베딩 벡터를 그대로 하는 것보다 훨씬 더 정답과 비슷한 예측을 수행하는 것을 확인할 수 있고, 100차원과 75차원의 성능지표 차이가 크지 않을 것을 보아 적당한 차원축소는 문장임베딩벡터의 본래 정보를 유지하는 것으로 보인다.(Zhang, G, et al., 2024) 자기 자신만을 고려한 예측그래프 <Figure 8>을 보면 후행성이 극명하게 발생하는 것을 확인할 수 있다. 그러나 감정지수를 활용한 예측그래프<Figure 13>을 보면 오차지표에서도 성능이 개선됐을 뿐만 아니라, 후행성도 완화되었음을 가시적으로 확인하였다.



본 연구에서는 위와 같은 결과를 근거로 뉴스데이터를 감정지수로 활용하는 방식을 채택했다. 임베딩 벡터를 그대로 활용하는 방식은 긍정적인 영향을 주지 못했지만, 그것은 이 연구의 모델이 임베딩 벡터가 담고 있는 정보들을 온전히 해석하지 못했기 때문으로 해석된다. 때문에 임베딩 벡터를 활용하려면 좀 더 고도화된 모델이 필요할 것이다.

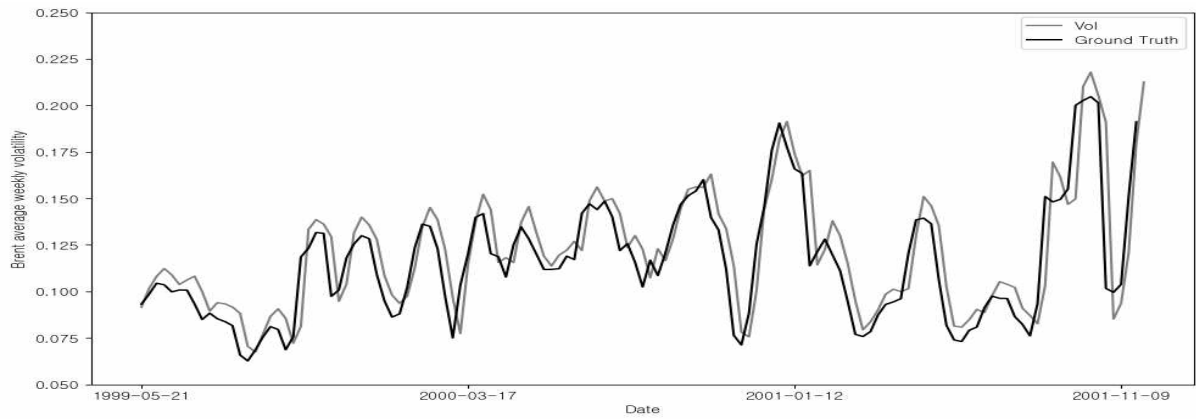


Figure 8. Prediction with only volatility

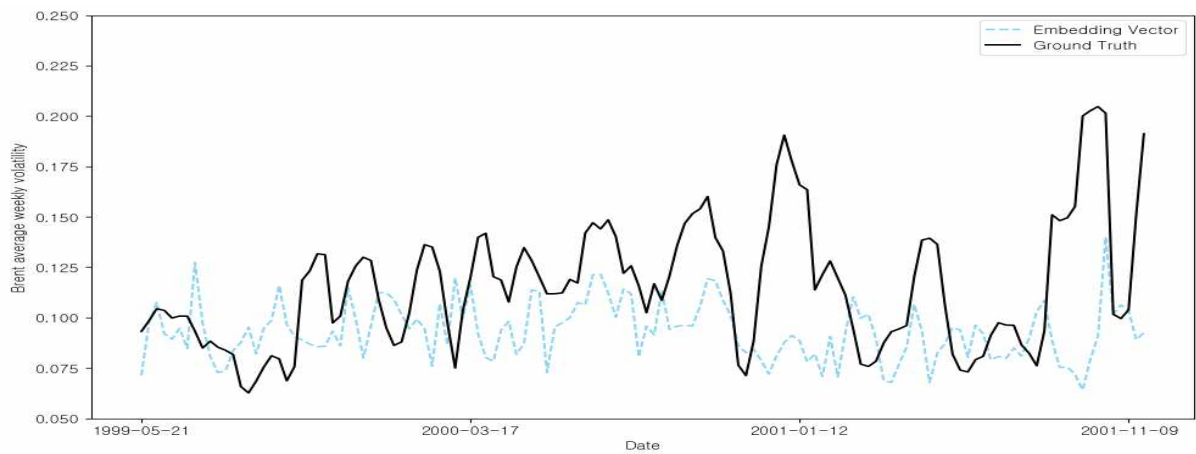


Figure 9. Prediction with volatility and embedding vector

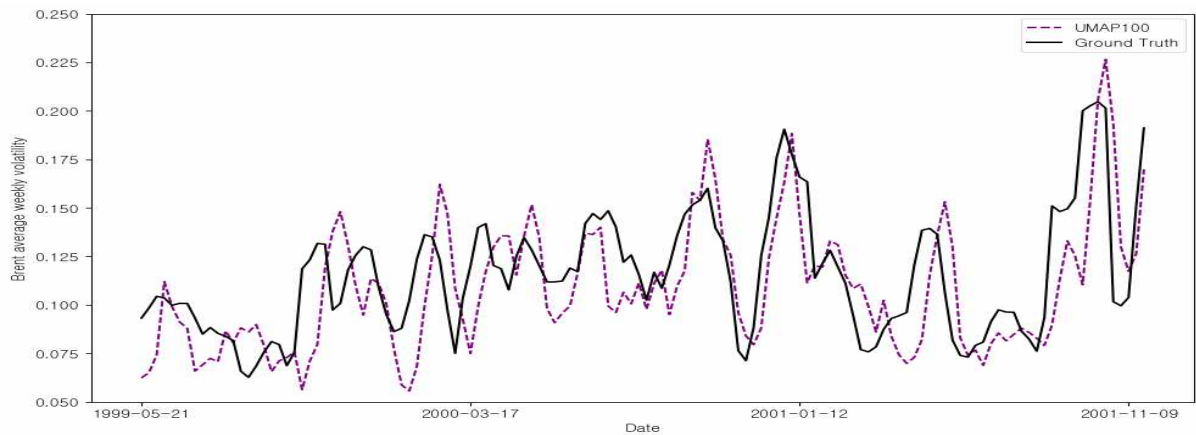


Figure 10. Prediction with volatility and UMAP100

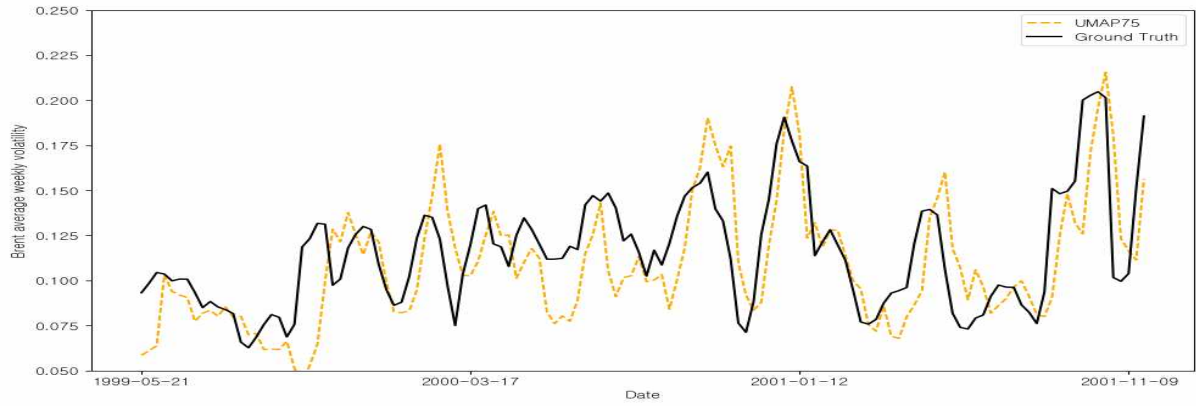


Figure 11. Prediction with volatility and UMAP75

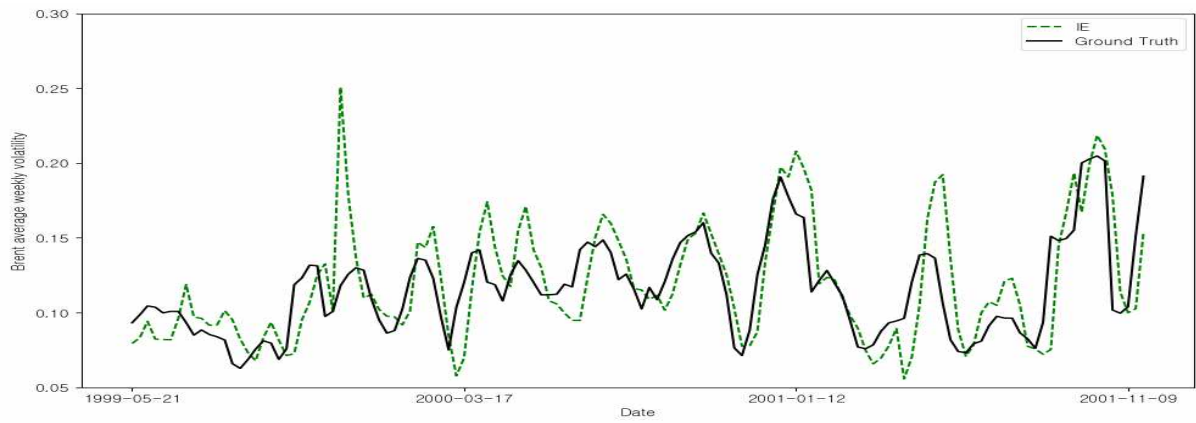


Figure 12. Prediction with volatility and IE

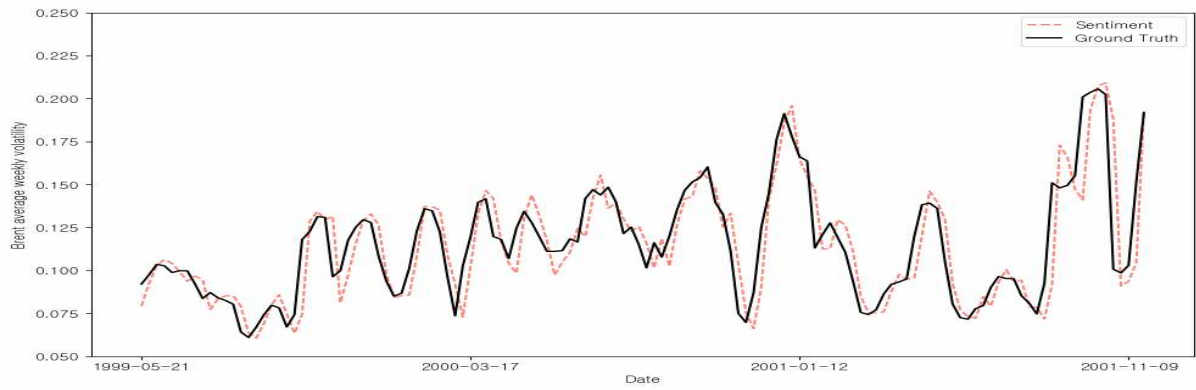


Figure 13. Prediction with volatility and sentiment

## 4.2 Total-Period 결과

Table 4. Total-Period 오차측정지표

Total	BZ-BZ	WTI-BZ	BZ-BZ-E	WTI-BZ-E	BZ-BZ-S	WTI-BZ-S	BZ-BZ-ES	WTI-BZ-ES
MAE	0.00956	0.00910	0.00936	0.00903	0.00913	0.00854	0.009	0.00842
RMSE	0.01893	0.01755	0.01601	0.01503	0.01671	0.0147	0.01689	0.01462

<Table 4>를 통해 Total-Period (2005.05 ~ 2024.01)에 대한 각 모델들의 오차측정지표를 확인할 수 있다. WTI-BZ-ES 모델이 RMSE와 MAE에서 각각 0.01462, 0.00842로 가장 성능이 뛰어났다. 이는 뉴스테이터를 활용한 감정지수와 S&P500이 전체적으로 긍정적인 상호작용을 하여 원유 변동성예측에 도움이 됨을 보여 준다. 또한 전이학습을 한 모델들이 하지 않은 모든 모델들 보다 전체적으로 낮은 오차를 보여줌으로, 본 연구는 원유 변동성 예측에 전이 학습을 활용하는 것이 타당함을 주장한다. <Figure 14 ~ 17>은 전체기간에 대하여 전이학습을 수행한 모델들 각각의 그래프를 나타낸다.

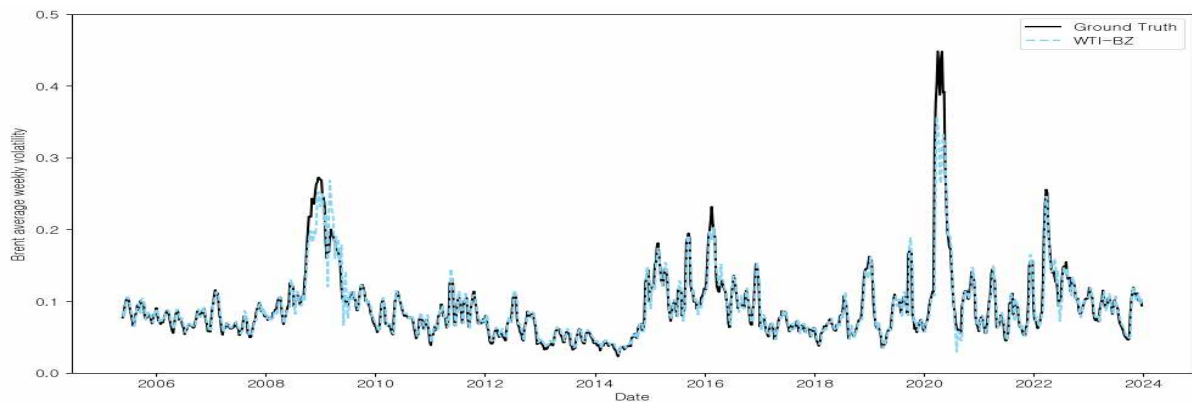


Figure 14. WTI-BZ prediction

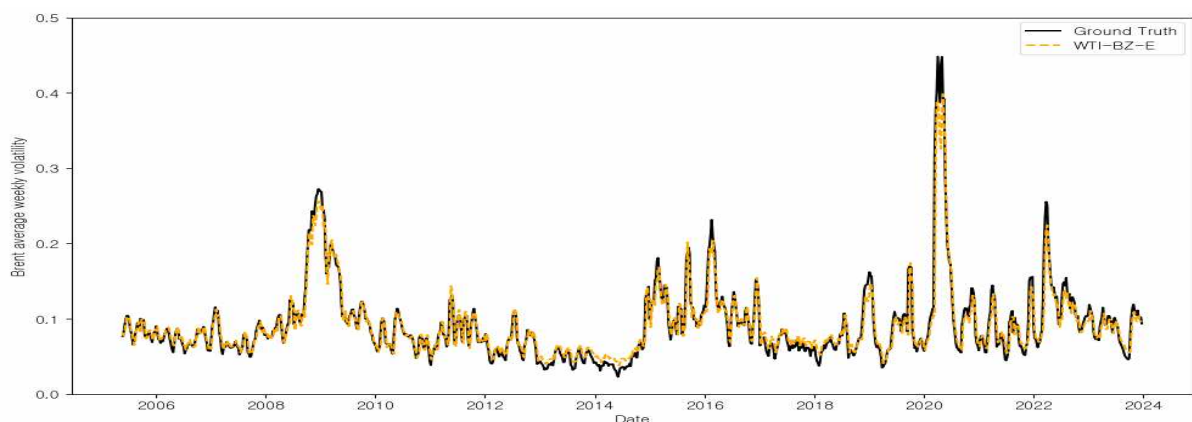


Figure 15. WTI-BZ-E prediction

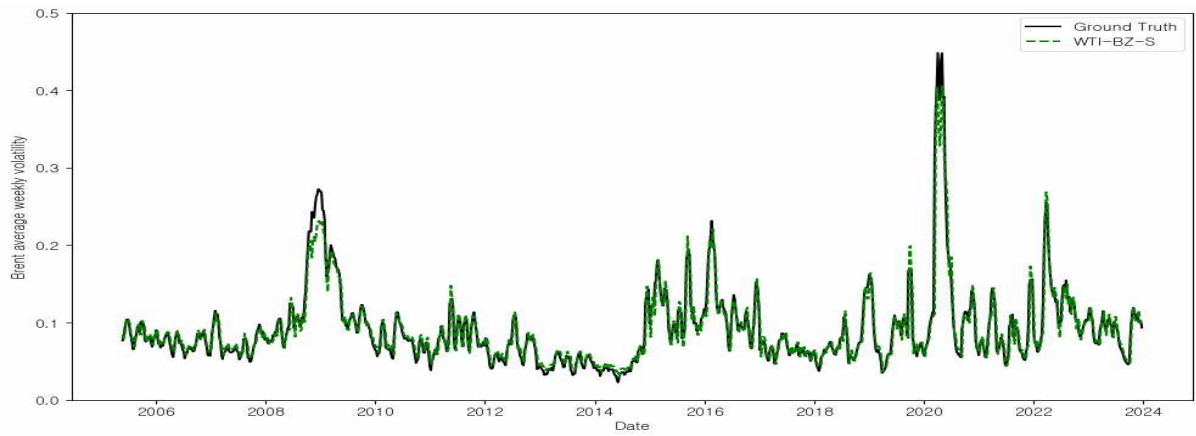


Figure 16. WTI-BZ-S prediction

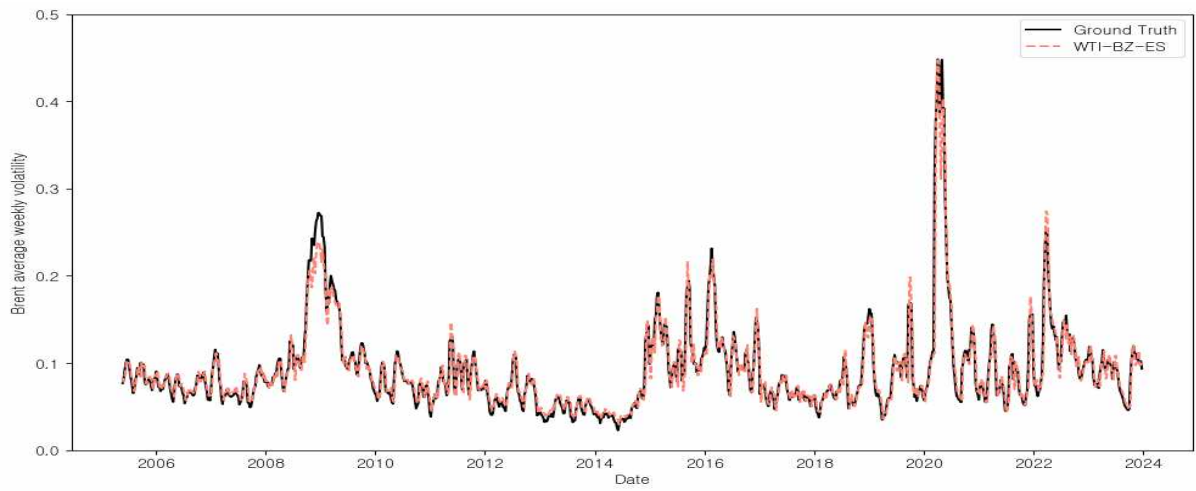


Figure 17. WTI-BZ-ES prediction

#### 4.3 Sub-Period 결과

Table 5. Sub-Period 오차측정지표

	BZ-BZ	WTI-BZ	BZ-BZ-E	WTI-BZ-E	BZ-BZ-S	WTI-BZ-S	BZ-BZ-ES	WTI-BZ-ES
~2008-09								
MAE	0.00511	0.00488	0.00551	0.00512	0.00526	0.00530	0.00516	0.00532
RMSE	0.00681	0.00652	0.00710	0.00688	0.00696	0.00703	0.00692	0.00701
2008-09~2009-06								
MAE	0.02815	0.03045	0.02095	0.01359	0.02212	0.02095	0.02428	0.02045
RMSE	0.03411	0.03564	0.02531	0.01756	0.02615	0.02595	0.02862	0.02481
2019-12~2022-09								
MAE	0.01922	0.01756	0.01654	0.01594	0.01702	0.01452	0.01733	0.01369
RMSE	0.03932	0.03502	0.03081	0.02851	0.03337	0.02711	0.03369	0.02614

<Table 5>는 Sub-Period에 대한 분석 결과이며 원유 변동성 예측에 대한 결과를 세 개의 sub-period로 나누어 분석하였다. 2005년 5월부터 2008년 9월까지 기간은 금융위기 발생 전 기간으로 변동성이 작은 구간을 대표한다. 2008년 9월부터 2009년 8월까지 기간은 금융위기가 발생한 구간으로 변동성이 큰 구간을 의미한다. 2019년 12월부터 2022년 9월까지 기간은 코로나 기간으로 최근 기간에 대한 예측결과를 분석한다. 변동성이 작은 구간에 대해서는 S&P500과 감정분석을 모두 고려하여 전이학습을 수행한 것이 변동성만 사용하여 전이학습을 수행하지 않은 것보다 근소하게 성능이 떨어지는 모습을 확인할 수 있다. 이는 변동성이 작은 구간에서 시계열의 자기상관성이 높음을 알 수 있다. 하지만 변동성이 작은 구간의 예측그래프인 <Figure 18>를 살펴보면 모델간의 예측동향이 거의 비슷함을 확인할 수 있다.

그러나 변동성이 급격하게 커지는 2008년 금융위기 구간에서는 S&P500, 감정분석, 그리고 전이학습을 고려하는 것이 낮은 예측 오차를 보여주었다. 특히 S&P500만 고려하여 예측을 진행한 경우가 다른 경우들보다 RMSE를 매우 크게 감소시켰다. 이는 Bagcilar et al.(2021)가 주장하는 S&P 500, 원유, 금 자산 간의 양방향 수익률과 변동성 파급 효과가 금융위기 구간에서 더 두드러짐을 알 수 있었다. 또한, 감정분석과 S&P500을 모두 고려한 경우도 변동성만 고려할 때보다 낮은 RMSE를 보여주어 유효성을 확인할 수 있었고, 금융위기 구간 예측 그래프인 <Figure 19>를 보면 외생변수를 고려하지 않은 WTI-BZ 모델은 불안정한 예측결과를 보여준다. 그러나 S&P500과 감정지수를 고려하면 훨씬 안정적인 예측동향을 나타낸다.

코로나 바이러스로 세계 경제가 크게 흔들렸던 구간인 2019년 12월부터 2022년 9월 기간도 마찬가지로 외생변수들을 고려하고 전이학습을 수행했을 때가 더 우수한 성능을 보였다. 또한, 이 구간에선 감정지수만을 활용하여 전이학습을 수행했을 때가 S&P500만을 활용하여 전이학습을 수행했을 때보다 우수한 성능을 보였다. 가장 최근에 기간에 대해서는 감정지수와 S&P500, 전이학습을 모두 활용한 모델이 오차가 가장 낮다. 감정분석의 유효성은 최근 기간에서 더 우수하였다. 또한, 코로나기간 예측그래프를 보여주는 <Figure 20> 을 보면 감정지수를 고려한 예측 그래프만 후행성이 발생하지 않음을 확인할 수 있다. 뉴스데이터가 일부구간에서 원유 변동성에 대한 선행적 정보를 담고 있음을 확인하였다.

이처럼 변동성이 커지는 위기 구간에서는 단순 시계열 데이터로 예측할 때보다 뉴스 데이터의 감정지수와 S&P500을 사용했을 때가 훨씬 좋은 예측 성능을 보여줄 수 있었다. 이를 통해 경제 위기 등으로 변동성이 커지는 구간에서 예측을 위해 뉴스 데이터 분석 결과와 S&P500 등 경제지표가 유용함을 알 수 있었다.

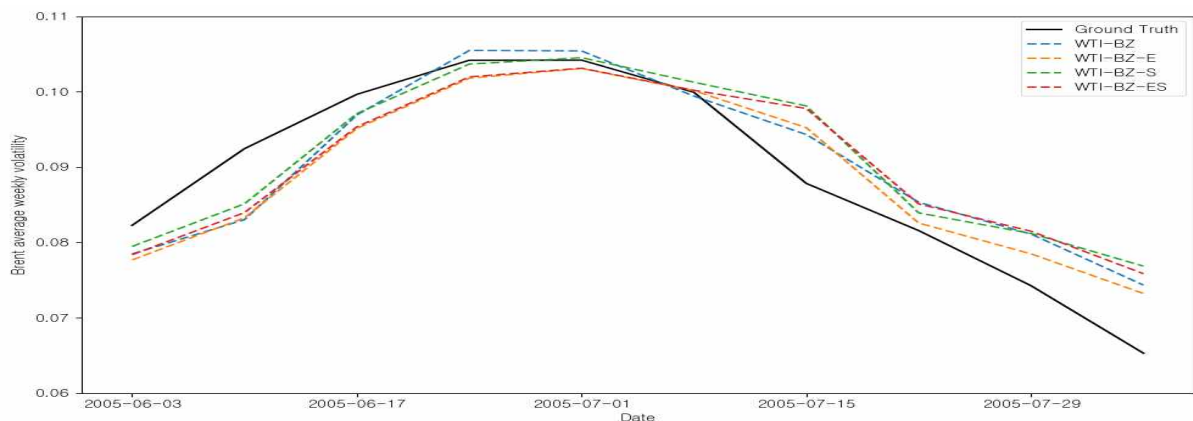


Figure 18. 2005 sub-period

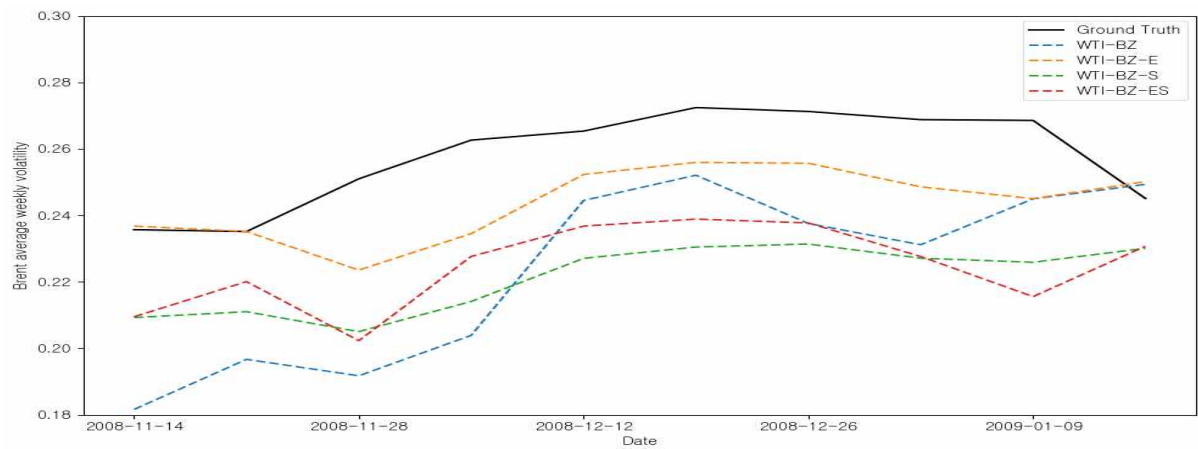


Figure 19. 2008 sub-period

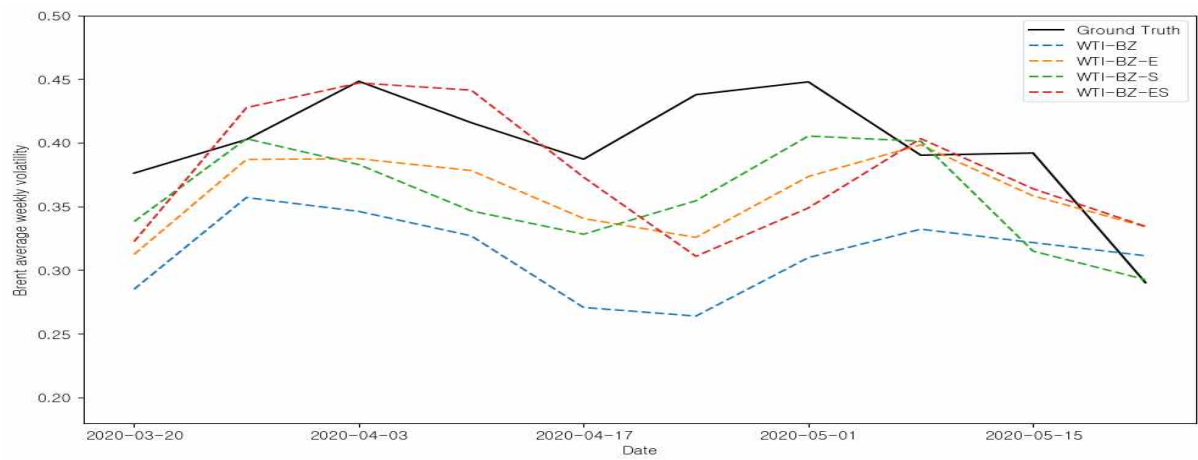


Figure 20. Covid-19 sub-period

#### 4.4 예측모델별 DM-Statistic 비교

기본 모델인 BZ-BZ와 다른 모델간의 성능을 비교해보기 위해 Diebold-Mariano (DM) test를 해보았다. 모든 DM-statistic이 p-value 5% 유의 수준을 넘지 못하였으므로, 거시적인 관점에서 각 모델 간의 유의미한 성능차이가 없다는 귀무가설을 채택한다. 그러나 전이학습을 하고, 외생변수들을 고려했을 때 p-value가 낮아지는 것을 보아 미시적으로 성능이 개선됨을 확인할 수 있다.

Table 6. BZ-BZ와 다른 모델간의 성능 비교

MODEL 1	MODEL2	DM-statistic	p-value
BZ-BZ	BZ-BZ-E	1.0924	0.2746
	BZ-BZ-S	1.2067	0.2275
	BZ-BZ-ES	1.1866	0.2354
	WTI-BZ	1.0236	0.3060
	WTI-BZ-E	1.1026	0.2702
	WTI-BZ-S	1.2091	0.2266
	WTI-BZ-ES	1.2071	0.2274

## 5. 결론 및 향후 연구방안

본 연구에서는 변동성이 크고 데이터의 길이가 짧은 유가 선물 데이터의 가격 예측의 정확성을 높이 고자 했다. Abdollahi(2023)의 연구에서는 WTI원유현물을 예측하는 데에 있어서 자연어, 그 중 트위터와 뉴스 제목 데이터를 외생변수로 활용했다. 하지만 실제 유가 선물 데이터는 WTI원유현물보다 길이가 훨씬 짧은 경우가 많았다. 본 연구에서는 위와 같은 상황에서도 예측의 정확성을 높이고, 추가적인 외생변수를 사용하여 예측 모델의 성능을 향상시키고자 했다.

이 과정에서 유가 데이터 중 길이가 가장 긴 WTI원유현물로 베이스모델을 만들어 선물 데이터에 전이학습을 적용해 유가 선물 데이터의 길이가 짧아 예측이 어렵다는 한계점을 극복하고자 했다. 또한, S&P500을 추가적인 외생변수로 활용하여 예측의 정확성을 높이 고자 했다. 뉴스 기사 데이터의 처리에 있어서는 임베딩 벡터, 정수 인코딩의 활용 성능을 특정 구간에서 감정분석과 비교하여, 가장 우수한 방법을 채택하고자 했다. 그 결과 감정분석이 가장 좋은 성능을 보였으며, 실제 예측 모델의 외생 변수로 활용하게 되었다.

예측 모델로 BiLSTM을 사용하여 전이학습과 외생변수를 적용한 결과, WTI원유현물로 베이스모델을 만들고 뉴스 기사 제목의 감정지수와 S&P500지수를 외생변수로 활용하여 Brent원유선물을 예측했을 때 가 평가 지표인 MAE와 RMSE 측면에서 가장 뛰어난 성능을 보였다. 특히 Brent원유선물로만 예측을 진



행했을 때보다 훨씬 높은 정확도를 보였다. 이를 통해 시계열 데이터에서의 전이학습의 유효성과 적절한 외생 변수의 선택이 예측 모델의 성능 향상에 크게 기여함을 확인할 수 있었다.

본 연구에서는 자연어 처리와 전이학습을 활용한 측면에서 뉴스 제목 데이터의 유효성에 한계가 있음을 확인하였다. 이전 연구에서도 원유와 관련된 뉴스 데이터를 활용한 이후에도 후행성이 완전히 제거되지 않았다. 감정지수가 후행성을 완화하는 모습을 보여주었지만, 이 연구에서도 마찬가지로 후행성이 여전히 존재함을 확인하였다. 후행성이 없어지지 않는 원인은 많은 뉴스들이 원유의 가격 변동 소식만을 다루고 있어 이미 정보가 선반영되어 있기 때문으로 판단된다.

향후 연구 방안으로는 다음과 같다. 첫째, 시계열 데이터 예측의 모델과 관련하여 다양한 시도를 해볼 수 있다. 예를 들면, 시계열 데이터 예측에 있어서 범용적인 모델인 BiLSTM 외에 NBEATXs와 같이 최신의 모델들을 사용해볼 수 있다. 또한, BiLSTM 모델의 복잡성을 더욱 높여서 자연어와 경제지표를 포함해 더 많은 외생변수를 사용하여 성능을 비교해볼 수 있다. 둘째, 본 연구의 한계점인 예측의 후행성을 제거하기 위해서 뉴스 제목 데이터의 유효성을 더욱 철저하게 검증해볼 필요가 있다. 또한, 오일과 관련된 뉴스만이 아닌, 다른 영역에서의 뉴스들도 활용하는 방안을 생각해볼 필요가 있다.



## 6. 참고문헌

- Yu, L., Wang, S., & Lai, K. K. (2008). Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm. *Energy economics*, 30(5), 2623-2635.
- He, K., Yu, L., and Lai, K.K. (2012). Crude oil price analysis and forecasting using wavelet decomposed ensemble model. *Energy*, 46(1), 564-574.
- He, Q. Q., Pang, P. C. I., & Si, Y. W. (2019). Transfer learning for financial time series forecasting. In *PRICAI 2019: Trends in Artificial Intelligence: 16th Pacific Rim International Conference on Artificial Intelligence, Cuvu, Yanuca Island, Fiji, August 26-30, 2019, Proceedings, Part II 16* (pp. 24-36). Springer International Publishing.
- Li, Y., Jiang, S., Li, X., & Wang, S. (2021). The role of news sentiment in oil futures returns and volatility forecasting: Data-decomposition based deep learning approach. *Energy Economics*, 95, 105140.
- Wu, B., Wang, L., Wang, S., & Zeng, Y. R. (2021). Forecasting the US oil markets based on social media information during the COVID-19 pandemic. *Energy*, 226, 120403.
- Zhang, Y., Wahab, M. I. M., & Wang, Y. (2023). Forecasting crude oil market volatility using variable selection and common factor. *International Journal of Forecasting*, 39(1), 486-502.
- Abdollahi, H. (2023). Oil price volatility and new evidence from news and Twitter. *Energy Economics*, 122, 106711.
- Mikolov, T., Chen, K., Corrado, G.S., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. International Conference on Learning Representations.
- Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. *Conference on Empirical Methods in Natural Language Processing*.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735-1780.
- Zhong, J., Cao, W., & Tang, Y. (2023). Tail risk of international equity market and oil volatility. *Finance Research Letters*, 58, 104365.
- Haugom, E., Langeland, H., Molnár, P., & Westgaard, S. (2014). Forecasting volatility of the US oil market. *Journal of Banking & Finance*, 47, 1-14.
- Balcilar, M., Ozdemir, Z. A., & Ozdemir, H. (2021). Dynamic return and volatility spillovers among S&P

500, crude oil, and gold. *International Journal of Finance & Economics*, 26(1), 153–170.

Suchin Gururangan, Ana Marasović, Swabha Swayamdipta, Kyle Lo, Iz Beltagy, Doug Downey, and Noah A. Smith. (2020). Don’ t Stop Pretraining: Adapt Language Models to Domains and Tasks. *58th Annual Meeting of the Association for Computational Linguistics*, pages 8342–8360.

Zhang, X., & Guo, Q. (2024). How useful are energy-related uncertainty for oil price volatility forecasting?. *Finance Research Letters*, 60, 104953.

Zhang, G., Zhou, Y., & Bollegala, D. (2024). Evaluating Unsupervised Dimensionality Reduction Methods for Pretrained Sentence Embeddings. *International Conference on Language Resources and Evaluation*.

## Abstract

This study employs a combination of news articles, the S&P500 index, and transfer learning to predict the volatility of Brent crude oil futures. The titles of news articles were analyzed using sentiment analysis or embedding vectors, and the S&P500 index was also considered to enhance prediction performance during the financial crisis. This study differs from other studies in that it attempts to utilize the sentence embedding vector of the transformer in its original form and confirms the effectiveness of transfer learning in predicting oil volatility. The oil data utilized in the study consisted of the spot price of WTI crude oil and the futures price of Brent crude oil. The data was collected from 1986 to 2024 and 2002 to 2024, respectively. Additionally, news articles and the S&P500 index were collected and considered for the period. In this case, a bidirectional long short-term memory (BiLSTM) deep learning model was employed, with the longest WTI spot price data serving as the base model. Transfer learning was applied to the shorter Brent futures volatility prediction. The analysis demonstrated that transfer learning enhanced prediction performance with more time series data. Furthermore, the incorporation of news data sentiment analysis and the S&P500 index yielded higher prediction performance than considering solely historical volatility. However, the use of sentence embedding vectors as is did not yield a positive impact on prediction accuracy. These results indicate that the incorporation of appropriate exogenous variables and transfer learning can enhance the accuracy of time series volatility forecasting.

Keywords: embedding vectors, sentiment analysis, S&P500, transfer learning, oil volatility forecasting