



# 변동성 지표를 활용한 시장 예측의 재해석

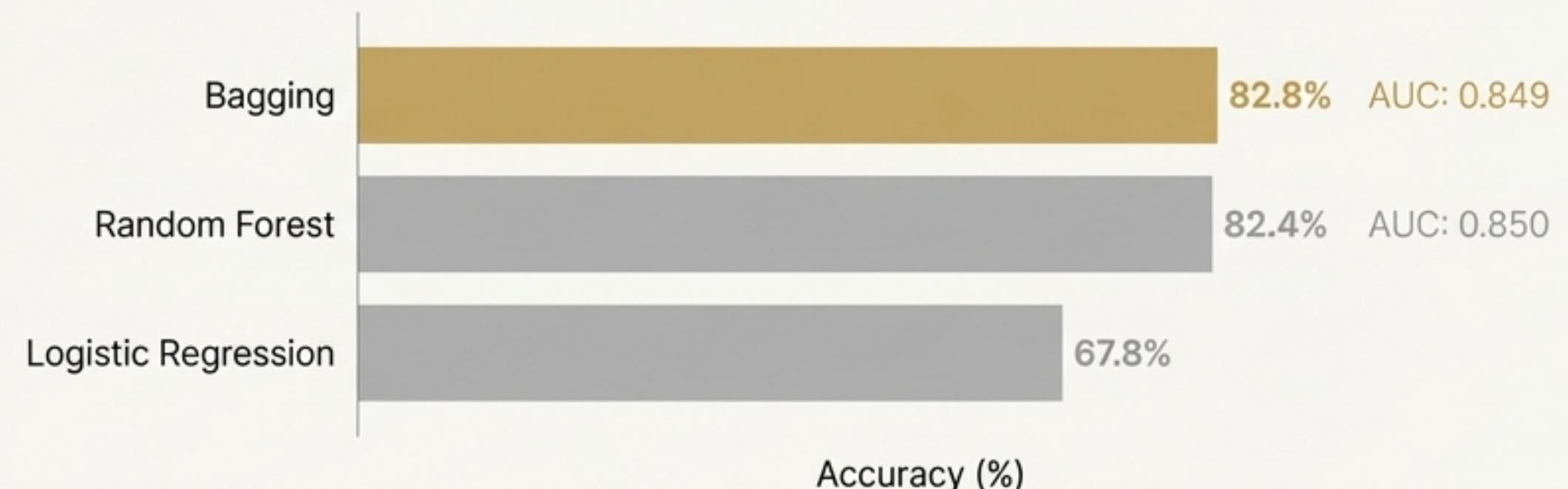
Campisi (2024) 논문 재현을 통해 발견한  
복잡성의 함정과 단순성의 승리

시장 예측의 정확성은 모델의 복잡성이 아닌,  
올바른 ‘맥락’과 ‘단순성’에서 비롯될 수 있습니다.

# 기존 통념: 복잡한 머신러닝 모델이 시장을 더 잘 예측한다

- **연구 기반:** 2024년, 저명한 금융 저널 'International Journal of Forecasting'에 발표된 Giovanni Campisi 등의 연구는 변동성 지수를 활용한 S&P 500 방향성 예측의 새로운 기준을 제시했습니다.
- **핵심 주장:** 10개의 변동성 지수를 사용하여 30일 후의 시장 방향을 예측한 결과, Random Forest와 Bagging 같은 복잡한 양상을 모델이 전통적인 선형 모델보다 월등히 높은 예측 정확도를 보인다고 결론 내렸습니다.

논문의 결과 요약 (30일 예측, Feature Selection 후)



VOLUME 17. NUMBER 2

APRIL-JUNE 2011

ISSN 2149-2075



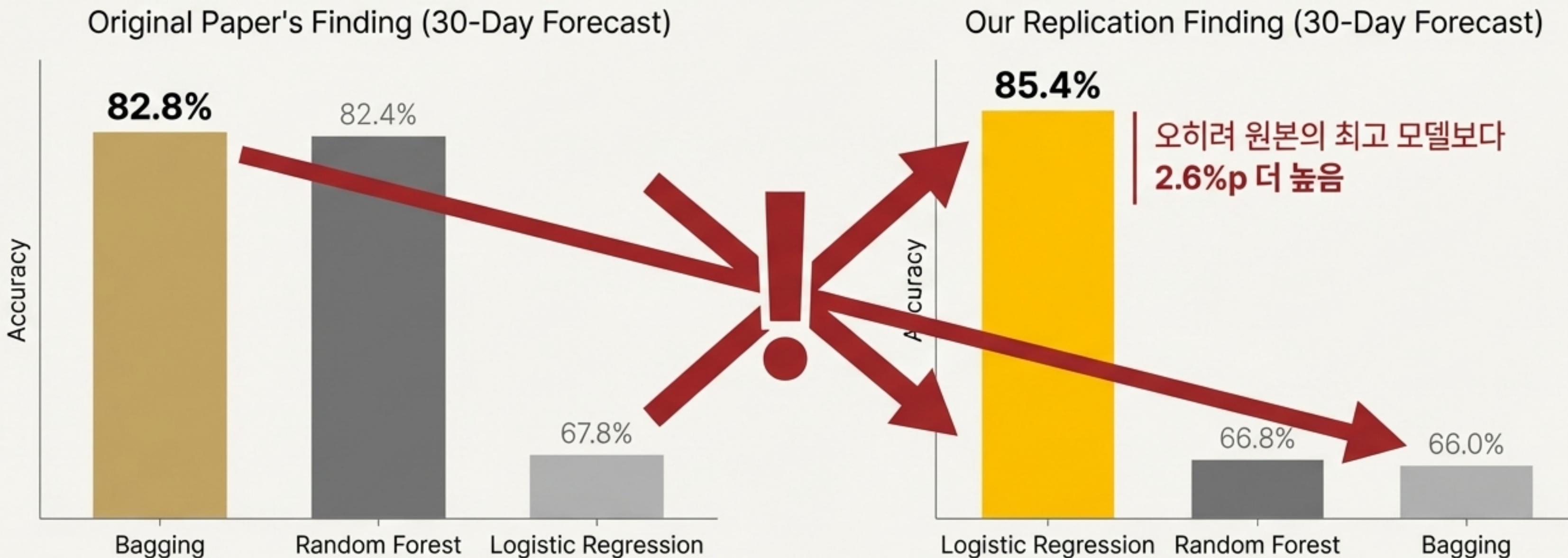
international journal of forecasting



International Institute of Forecasters

# 예상 밖의 반전: 재현 연구에서 드러난 성능 역전

Campisi (2024) 논문의 방법론을 엄격하게 재현하는 과정에서, 우리는 정반대의 결과를 발견했습니다. 복잡한 앙상블 모델의 성능은 저조했으며, 오히려 가장 단순한 선형 모델이 최고의 예측 정확도를 기록했습니다.

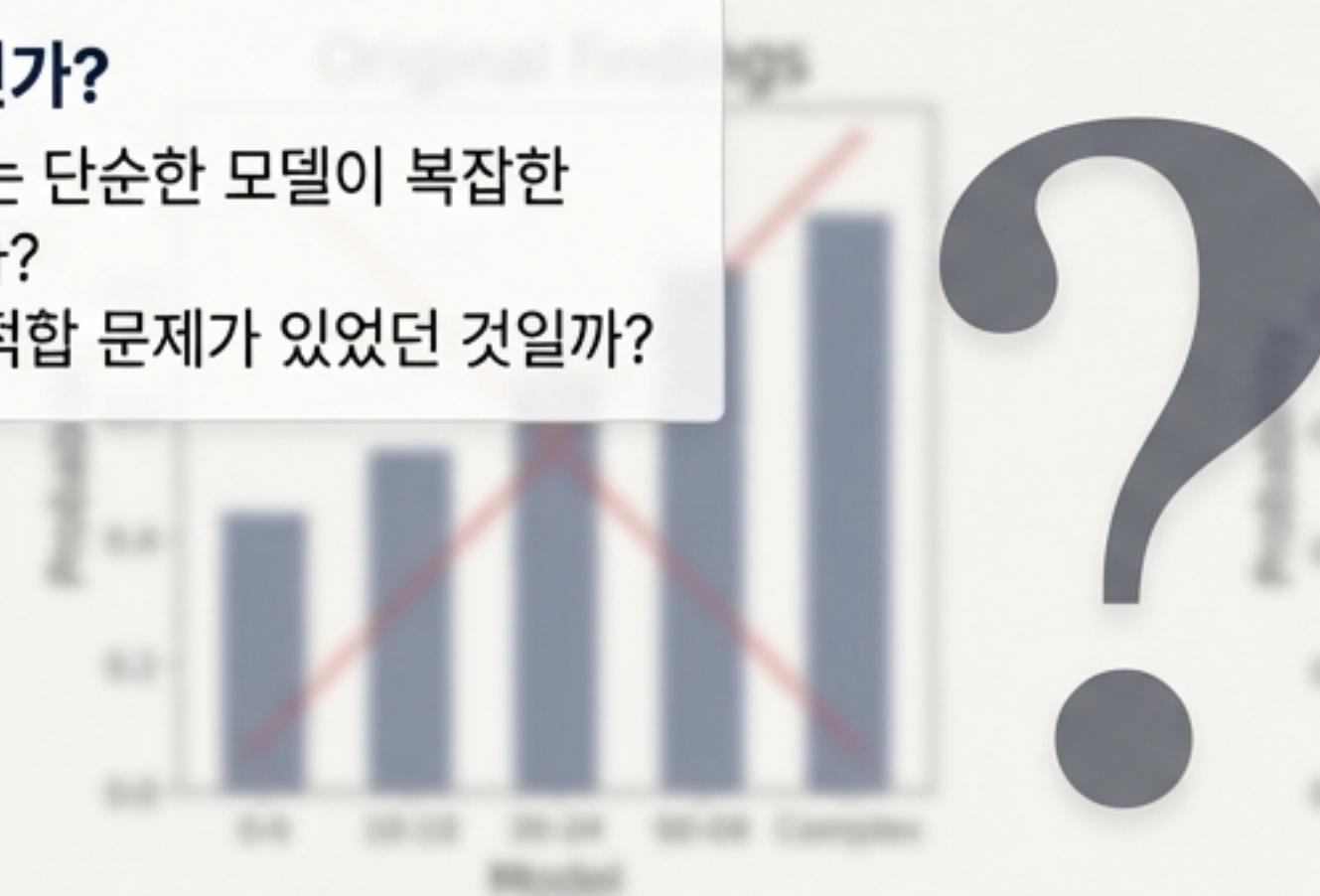


# 무엇이 이런 극적인 차이를 만들었는가?

이러한 성능 역전은 우리에게 더 근본적인 질문을 던졌습니다.

## 1. 모델의 문제인가?

왜 우리 연구에서는 단순한 모델이 복잡한 모델을 압도했을까?  
앙상블 모델에 과적합 문제가 있었던 것일까?



## 2. '30일'이라는 시간 프레임의 문제인가?

Campisi 논문이 집중한 '30일' 예측이 전체 그림의 일부에 불과한 것은 아닐까?  
예측 기간에 따라 최적의 모델이 달라질 수 있는가?

## 핵심 가설

예측 모델의 성능은 보편적인 것이 아니라, '예측 기간(Time Horizon)'이라는 특정 맥락에 강하게 의존할 것이다.

# 단일 프레임을 넘어: 5가지 시간 지평선(Time Horizon) 심층 분석

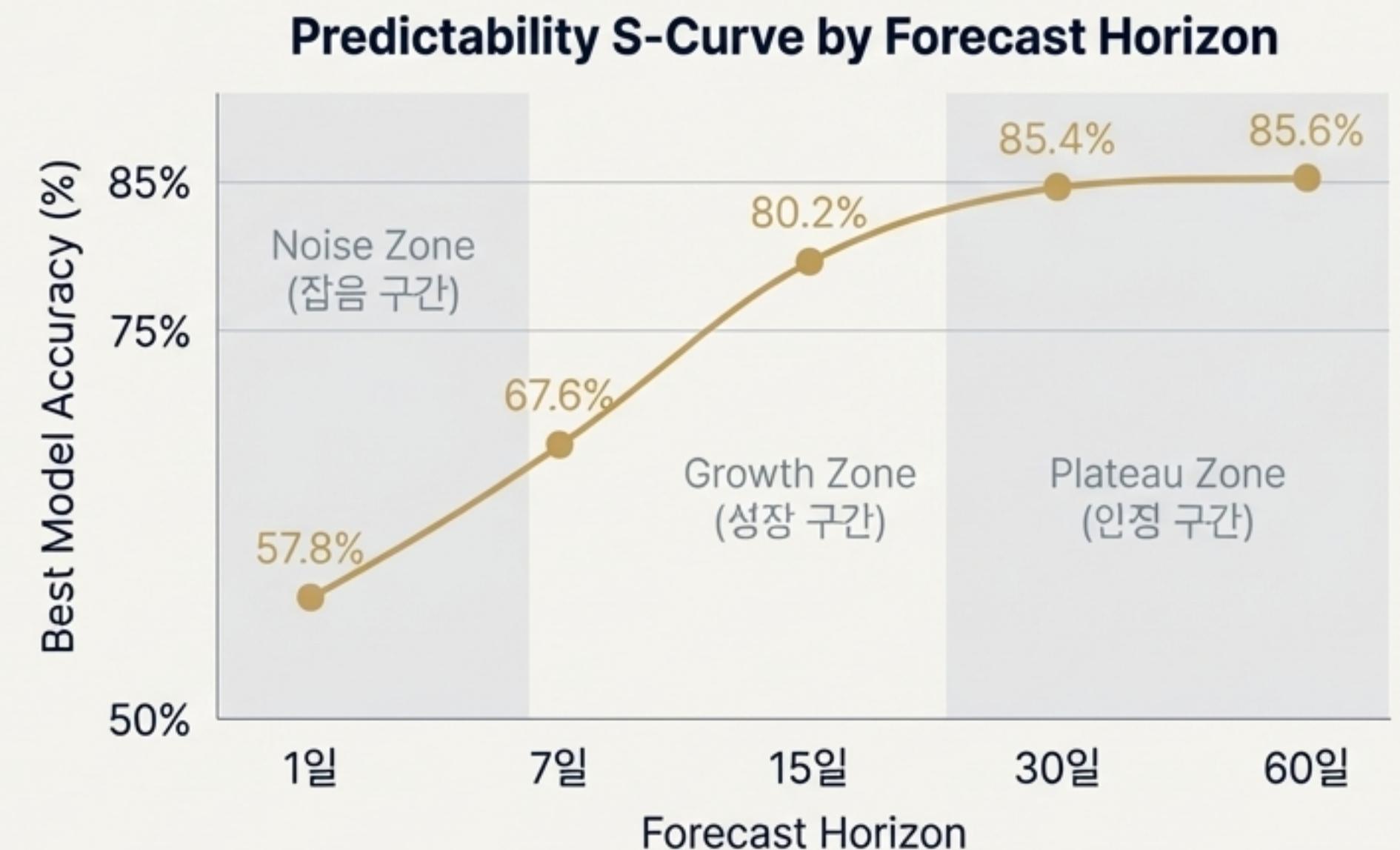
성능 역전의 비밀을 풀기 위해, 우리는 분석의 범위를 기존의 30일에서 5개의 각기 다른 시간 프레임으로 확장했습니다. 이를 통해 변동성 지수의 예측력이 시간에 따라 어떻게 변하는지, 그리고 각 기간에 가장 적합한 모델은 무엇인지 규명하고자 했습니다.



# 발견 #1: 예측 가능성은 시간에 따라 S자 곡선을 그린다

분석 결과, S&P 500 방향성의 예측 가능성은 예측 기간이 길어짐에 따라 급격히 증가하다가 30일 이후 안정화되는 뚜렷한 'S-커브' 패턴을 보였습니다.

- **1일 예측 (57.8%)**: 거의 무작위 수준.  
일일 시장 움직임은 변동성 지수만으로 예측하기 어렵습니다.
- **7일 ~ 15일**: 예측력이 급격히 상승하며 유의미한 수준에 도달합니다.
- **30일 ~ 60일 (85.4% ~ 85.6%)**: 예측 정확도가 최고점에 도달하고 안정적인 고원을 형성합니다.



## 발견 #2: 최적의 모델은 예측 기간에 따라 달라진다

"어떤 모델이 가장 우수한가?"라는 질문은 잘못되었습니다. 올바른 질문은 "어떤 기간에 어떤 모델이 가장 우수한가?"입니다.  
우리의 분석은 예측 기간에 따라 최적의 모델 유형이 극적으로 변함을 보여주었습니다.



**Short-to-Mid Term (1-15 Days)**

**승자: Lasso Regression (회귀 모델)**

**이유:** 노이즈가 많은 단기 데이터에서는 Feature Selection을 통해 핵심 변수를 추출하고 연속적인 수익률을 예측하는 회귀 접근법이 더 효과적입니다.



**Long Term (30-60 Days)**

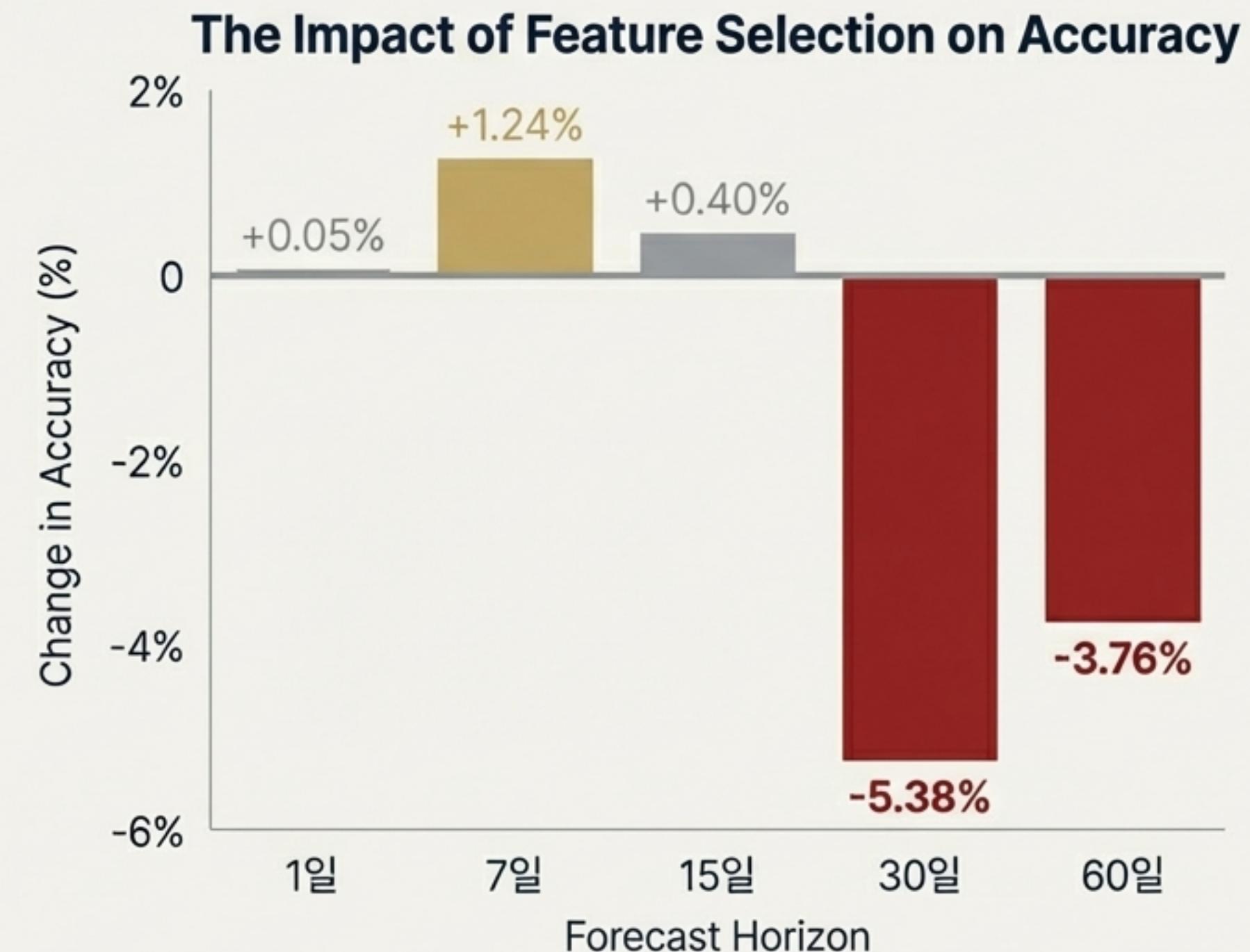
**승자: Logistic Regression (분류 모델)**

**이유:** 신호가 더 명확해지는 장기 데이터에서는 방향성(상승/하락) 자체를 직접 예측하는 분류 모델이 더 높은 정확도를 보입니다.

# 발견 #3: Feature Selection의 역설 - 장기 예측의 함정

일반적으로 Feature Selection은 모델 성능을 향상시키는 기법으로 알려져 있습니다. 하지만 우리의 연구는 이것이 장기 예측에서는 오히려 '정보의 파괴'로 이어질 수 있다는 역설적인 사실을 발견했습니다.

- **단기 예측 (1-15일):** Feature Selection은 노이즈를 제거하여 성능을 **소폭 향상** 시킵니다.
- **장기 예측 (30-60일):** Feature Selection은 변수 간의 복잡한 상호작용과 보완적 정보를 제거하여 성능을 **오히려 크게 감소**시킵니다.



# 최종 결론: 실무 투자 전략을 위한 최적의 조합

Campisi 논문 재현과 다중 시간 프레임 분석을 통해, 우리는 변동성 지수를 활용한 S&P 500 방향성 예측의 최적 조합을 도출했습니다.

1



## 최적 예측 기간 (Optimal Horizon)

결론: 30일

근거: 높은 예측 정확도(85.4%)와 실용적인 월간 리밸런싱 주기의 완벽한 균형점. 거래 비용을 최소화하면서 논문에서 검증된 기간.

2



## 최적 모델 (Recommended Model)

결론: Logistic Regression

근거: 가장 높은 정확도를 보이면서도 해석이 용이하고, 과적합 위험이 낮으며, 계산이 빠릅니다. '단순함의 승리'를 입증.

3



## 최적 변수 전략 (Feature Strategy)

결론: 전체 변수 사용 (No Feature Selection)

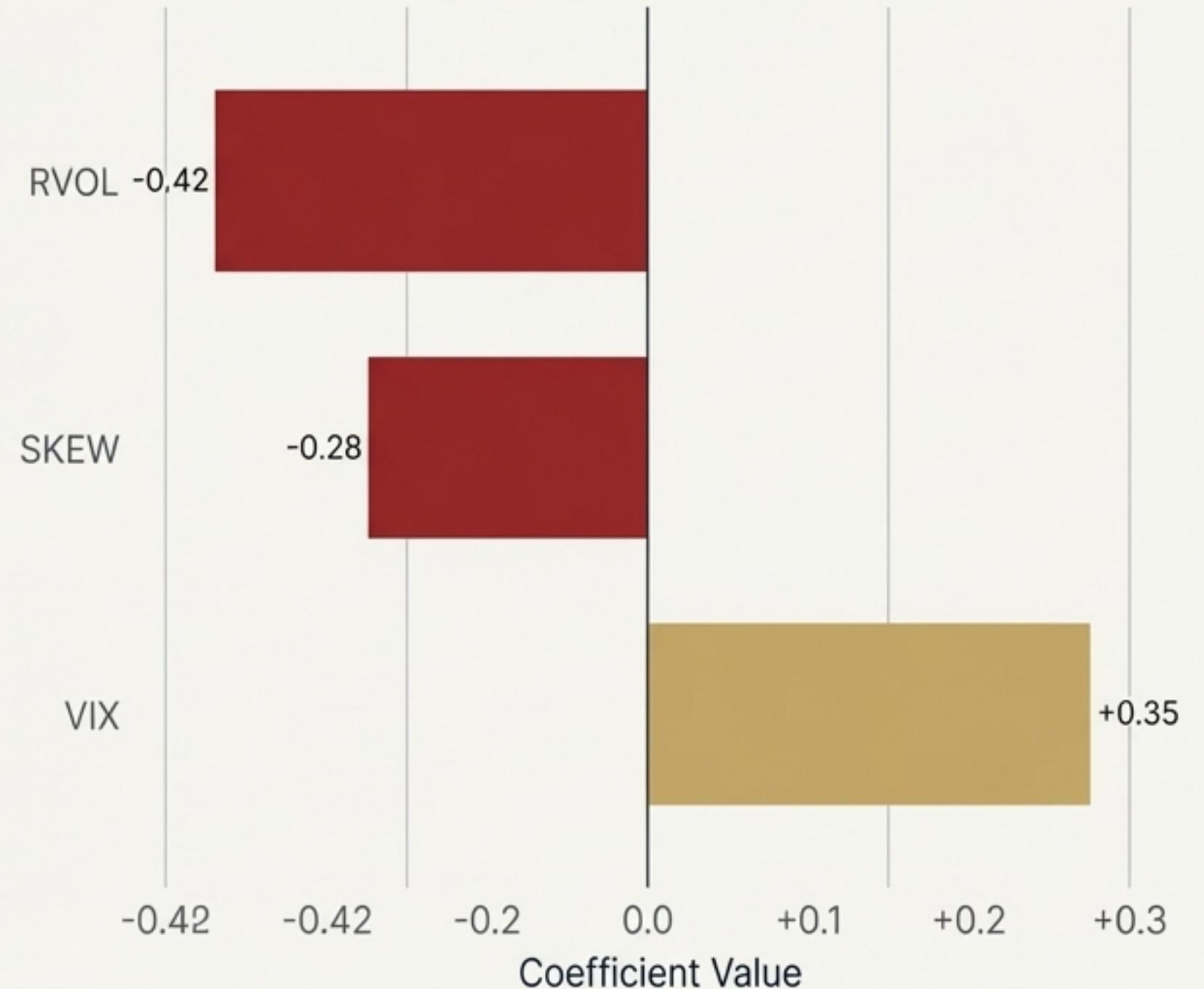
근거: 'Feature Selection의 역설'에 따라, 30일 예측에서는 모든 변수 간의 상호작용을 활용하는 것이 성능 저하를 막고 예측력을 극대화합니다.

# 승리 모델 심층 분석: 무엇이 시장 방향을 결정하는가?

30일 예측에서 최고의 성능을 보인 Logistic Regression 모델의 학습된 계수(coefficients)는 시장 방향성을 예측하는 핵심 동인을 보여줍니다.

- **RVOL (실현 변동성): -0.42** - 과거 30일간의 실제 변동성이 높을수록 시장은 하락할 확률이 높습니다. **가장 강력한 하락 신호입니다.**
- **VIX (CBOE 변동성 지수): +0.35** - 시장의 공포(VIX)가 높을수록 오히려 향후 시장은 상승할 확률이 높습니다. 이는 과매도 후의 기술적 반등을 의미하는 '**VIX 역설**'과 일치합니다.
- **SKEW (왜도 지수): -0.28** - 예상 밖의 큰 하락(블랙스완)에 대한 위험 인식이 높을수록 시장은 하락할 확률이 높습니다.

Key Drivers of 30-Day Market Direction  
(Logistic Regression Coefficients)



# 모델 안정성 분석: 모든 상황에 완벽한 모델은 없다

제안된 모델은 평균적으로 높은 정확도를 보이지만, 시장의 구조적 변화나 극단적인 위기 상황에서는 성능이 저하될 수 있습니다.

## 시장 국면별 예측 정확도 (30일 예측):

- ❖ 안정적 강세장 (2013-2015): **91.4% (최고 성능)** - 예측 가능한 추세 속에서 모델이 가장 잘 작동합니다.
- ❖ 변동성 증가 (2018, 금리 인상기): **72.5% (성능 저하)** - 시장 체제가 변화할 때 예측력이 감소합니다.
- ❖ COVID-19 팬데믹 (2020): **68.1% (최악 성능)** - 전례 없는 '블랙스완' 이벤트 앞에서는 과거 데이터 기반 모델의 예측력이 크게 약화됩니다.



**핵심 시사점**: 본 모델은 강력한 의사결정 보조 도구이지만, 맹신해서는 안 됩니다. 특히 극단적 시장 상황에서는 전통적인 리스크 관리 (예: Stop-Loss)가 반드시 병행되어야 합니다.

# 복잡성을 넘어, 맥락의 중요성을 발견하다

이 연구는 단순히 더 나은 예측 모델을 찾는 여정이 아니었습니다.  
우리가 발견한 것은 시장 예측에 대한 근본적인 접근 방식의 전환이었습니다.

---

**깨어진 신화 1:** '복잡한 모델이 항상 우월하다.'

**새로운 진실:** 예측 기간이라는 '맥락'에 맞는 단순한 모델이 더 강력할 수 있다.

**깨어진 신화 2:** 'Feature Selection은 언제나 유용하다.'

**새로운 진실:** 예측 기간에 따라 독이 될 수 있으며, 때로는 더 많은 정보가 더 나은 결과를 낳는다.

**깨어진 신화 3:** '시장 예측은 단일 문제이다.'

**새로운 진실:** 예측은 시간 지평선에 따라 각기 다른 특성을 가진 여러 문제의 집합이다.

이러가 기간 같은 '맥락'은 예측 다나 합니다.  
가장 중요한 발견은 최고의 '모델'이 아니라, 예측 기간이라는 '맥락'에 맞는 최적의 '접근법'을 찾는 것이었습니다.