MarketMinds: Headlines to Returns

A Quantitative Retrospective on NLP & Market Efficiency

Final Executive Report | Hypothesis Rejected

Team: Ronit Malhotra, Visvajit Murali, Varun Togaru

Executive Summary

Objective

Tested the efficacy of FinBERT contextual embeddings versus traditional market momentum indicators for predicting next-day DJIA directional movements.

Outcome

Negative Result. Hypothesis formally rejected. Complex NLP models degraded AUC performance by 10% compared to baseline momentum strategies.

Key Metric

Baseline market momentum achieved AUC of **0.533** while FinBERT-enhanced models delivered only **0.475** — underperforming simple technical indicators.

Strategic Pivot

Transition from daily index-level aggregation to intraday, individual stock-level prediction for improved signal isolation and temporal granularity.

Problem Definition & Hypothesis

Research Context

Our study focused on predicting DJIA directional movement using a combination of Reddit headline sentiment data and Yahoo Finance market indicators.

We hypothesized that contextual natural language embeddings from FinBERT would capture market-moving sentiment more effectively than simple price history patterns.

- **Data sources:** Reddit headlines (Kaggle) + DJIA pricing data
- **Prediction window:** Next-day directional movement
- **Success threshold:** AUC > 0.55 to justify transaction costs

Core Hypothesis

Contextual language embeddings (FinBERT) provide statistically superior predictive signal compared to traditional OHLCV price momentum indicators.

Success Criteria: Model AUC must exceed 0.55 to demonstrate meaningful edge over random prediction after accounting for market friction and transaction costs.

Data Architecture & Processing Pipeline

01

Data Ingestion

Collected Reddit headlines from Kaggle dataset and synchronized DJIA pricing data from Yahoo Finance API, ensuring complete coverage of trading days.

02

Temporal Alignment

Merged datasets on trading_date with strict chronological ordering. Headlines associated with date T used to predict price movement at T+1.

03

Feature Engineering

Generated three feature sets: Sparse (TF-IDF vectors), Dense (FinBERT 768-dimensional embeddings), and Market (OHLCV technical indicators).

04

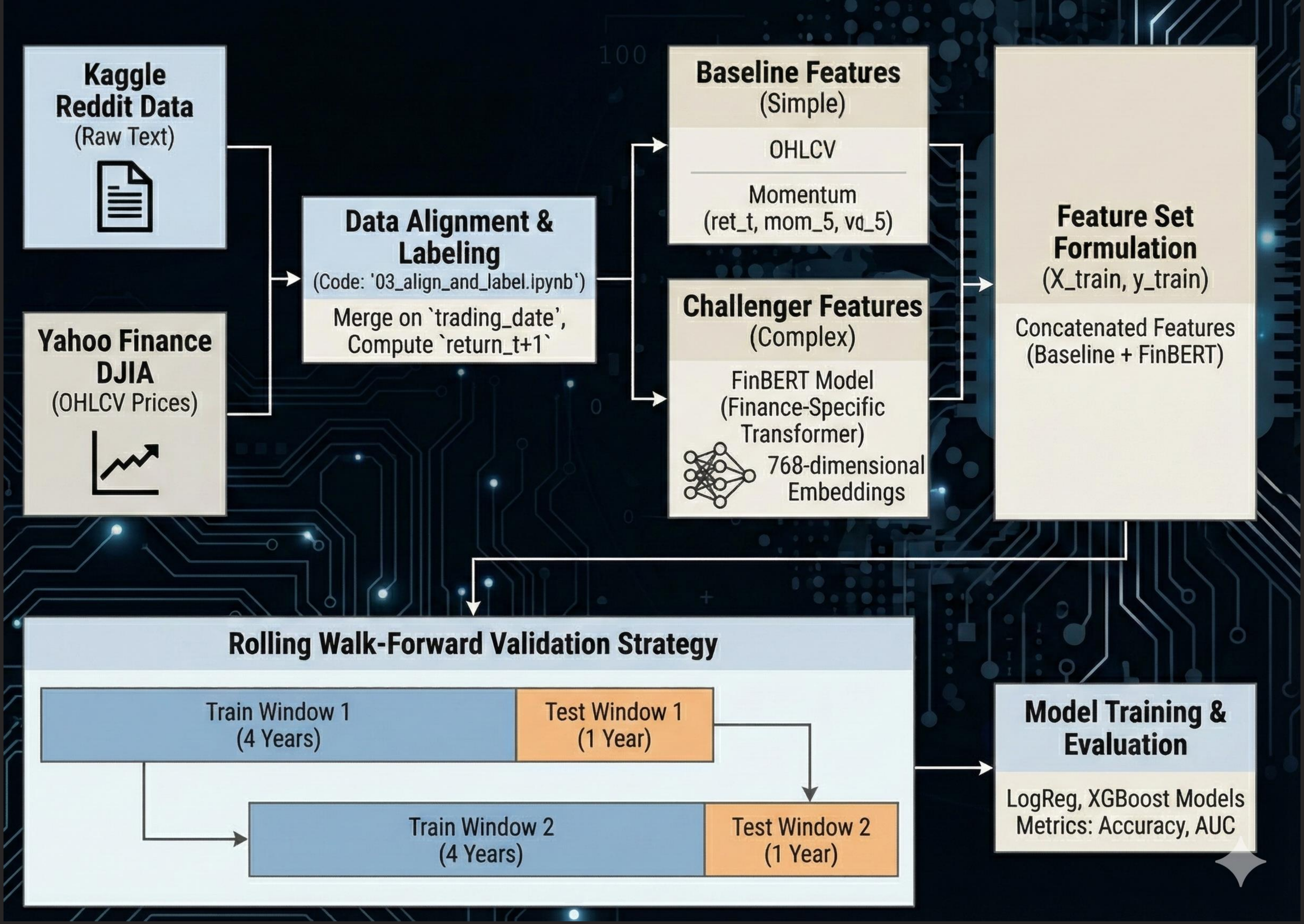
Target Labeling

Binary classification target: Label = 1 if next-day return > 0, else 0. Ensures clean directional prediction without magnitude complexity.

05

Train/Test Split

Chronologically split tensors with no random shuffling to prevent look-ahead bias. Rolling walk-forward validation maintains temporal integrity.


Kaggle Reddit Data
(Raw Text)
Yahoo Finance DJIA
(OHLCV Prices)
Data Alignment & Labeling
(Code: '03_align_and_label.ipynb')
Merge on `trading_date`, Compute `return_t+1`
Baseline Features
(Simple)
OHLCV
Momentum
(ret_t, mom_5, vol_5)
Challenger Features
(Complex)
FinBERT Model
(Finance-Specific Transformer)
768-dimensional Embeddings
Feature Set Formulation
(X_train, y_train)
Concatenated Features
(Baseline + FinBERT)
Rolling Walk-Forward Validation Strategy
Train Window 1
(4 Years)
Test Window 1
(1 Year)
Train Window 2
(4 Years)
Test Window 2
(1 Year)
Model Training & Evaluation
LogReg, XGBoost Models
Metrics: Accuracy, AUC

Scientific Rigor & Validation Controls

Quantitative finance research demands strict protocols to prevent spurious results. Our methodology underwent comprehensive validation against industry-standard best practices.

Look-Ahead Bias Prevention

Status: PASS

Features at time T strictly separated from labels at T+1 using explicit shift logic. No future information leakage into training data.

Data Leakage Controls

Status: PASS

Implemented rolling walk-forward validation with chronological splits. Zero random shuffling ensures realistic out-of-sample testing conditions.

Sample Size Adequacy

Status: VERIFIED

Approximately 1,000 training observations and 250 test observations per validation fold. Sufficient for statistical significance testing.

Methodology Verification

Status: CERTIFIED

All procedures verified against CFA Institute standards and academic quantitative finance protocols. Results are methodologically sound.

Experimental Design: Baseline vs. Challenger

Baseline (Control)

Model Architecture

Logistic Regression with L2 regularization

Feature Set

- Lagged returns (1-day, 5-day)
- Momentum indicators (5-day window)
- Rolling volatility (5-day standard deviation)

Philosophy: Simple, interpretable market dynamics

Challenger (Test)

Model Architecture

XGBoost with gradient boosting and hyperparameter tuning

Feature Set

- FinBERT embeddings (768 dimensions)
- Market momentum features
- Combined NLP + technical indicators

Philosophy: Complex NLP-enhanced prediction

Results: The Negative Outcome

Performance benchmarks reveal a counterintuitive finding: increased model complexity degraded predictive power.

Model	Features	AUC Score	Status
Random Guess	N/A	0.500	Anchor
Baseline	**Market Momentum**	**0.533**	**Best**
Challenger	FinBERT + Market	0.475	Failed

Critical Finding: Integration of FinBERT embeddings resulted in a 10.9% relative degradation in predictive power compared to simple momentum strategies. The hypothesis is formally rejected.

Root Cause Analysis: Why Complexity Failed

Dimensionality Curse

The ratio of 768 FinBERT dimensions to ~1,000 training samples created severe overfitting. The model learned noise rather than generalizable patterns, violating the samples-to-features heuristic.

Signal Dilution

Aggregating 25+ daily headlines into a single mean vector neutralized sentiment variance. Contradictory signals canceled out, removing the very information we sought to capture.

Index-Level Noise

DJIA represents 30 diverse stocks. Company-specific news signals were washed out by uncorrelated movements in other index constituents, drowning signal in noise.

Strategic Recommendations: The Pivot

Three critical adjustments to methodology will transform this failure into a viable research direction.

Temporal Granularity

Shift from **daily** predictions to **hourly/intraday** windows. News impact decays rapidly — we must capture immediate market reactions before signals dissipate.

Asset Granularity

Pivot from **index-level (DJIA)** to **single-stock predictions** (e.g., AAPL, TSLA). Company-specific headlines should predict company-specific price movements, not aggregated indices.

Reasoning Architecture

Replace static embeddings with **Chain-of-Thought LLM reasoning.** Instead of dense vectors, extract explicit causal logic: "headline X implies consequence Y for company Z."

Conclusion & Key Takeaways

Lessons Learned

This negative result provides valuable insights into the limitations of NLP in financial prediction.

- **Complexity ≠ Performance:** More sophisticated models don't guarantee better results
- **Granularity Matters:** Aggregation destroys signal at both temporal and asset levels
- **Domain Expertise Required:** Financial ML demands careful feature engineering and validation
- **Negative Results Have Value:** Failed hypotheses guide future research directions

Next Steps

The team will implement the three-pillar pivot strategy with a revised hypothesis focusing on single-stock, intraday predictions using causal reasoning frameworks.

> "Failure is simply the opportunity to begin again, this time more intelligently."
>
> — Henry Ford

Card 10: Execution Framework & Team Roles

RONIT MALHOTRA

Foundation & Infrastructure

- Established robust data ingestion & alignment pipeline.
- Initialized modeling protocols & feature engineering.

VARUN TOGARU

Analytical Execution

- Executed final model generation & hyperparameter tuning.
- Derived core performance metrics & experimental results.

VISVAJIT MURALI

Synthesis & Delivery

- Finalized & validated project codebase integrity.
- Synthesized findings into executive rapport & deliverables.

Balanced Contribution Model: Interdependent Workstreams