

AI-Powered Sentiment Analysis for Stock Market Prediction in Renewable Energy

Author: Divina Minocha¹, Vinay Vishwakarma²

Affiliation: Student, *Dhirubhai Ambani International School, Mumbai, India*¹; Mentor, *On My Own Technology Pvt. Ltd Student Mumbai, India*²

<https://doi.org/10.26821/ijshre.14.09.2026.140904>

ABSTRACT

This paper presents a joint effort between Natural Language Processing (NLP) and Market Technical Analysis (MTA), powered by Artificial Intelligence (AI) to predict stock price movements in renewable energy companies. The A.I. system leverages FinBERT (a domain-specific transformer model) to extract sentiment, ticker relevance, and other business event types from over 850 articles about five different renewable energy stocks (Tesla, Enphase Energy, SolarEdge, NextEra Energy, Brookfield Renewable). It aggregates sentiment signals over time, introduces decay weights, and merges them with other MTA features such as volatility and recent return to predict price movements of each of the five renewable energy stocks from 1 to 5 trading day(s) in the future. A rolling time-series validation process guarantees temporal integrity and helps eliminate look-ahead bias in the A.I. system results. The combined model achieves 62% accuracy for directional predictions of each stock price movement on the next trading day, far surpassing the accuracy of the two standalone models (48% accuracy for sentiment only and 52% accuracy for market only). The A.I. identified the most significant event types associated with each of the five stocks' predicted price movement: capacity announcements, subsidy changes, and supply chain disruptions. This demonstration illustrates the immense potential for applying multi-modal machine learning to help improve financial forecasting across other highly regulated sectors as well.

Keywords—sentiment analysis, machine learning, stock prediction, renewable energy, NLP, trading strategy

I. INTRODUCTION

The stock market for renewable energy has experienced extreme fluctuations because of changes in reported policies, manufacturing capacity increases, or an increase/decrease in subsidies. Traditional algorithmic trading strategies rely on historical price and technical indicators alone, and thus do not include any of the critical details contained within news articles or regulatory announcements. The inability to incorporate these items in forming trading strategies results in insufficient pricing of stock by the market, as a large percentage of news articles are published on average several minutes to hours before the stock price is adjusted.

The primary issue to be addressed with converting unstructured text (i.e., news articles or earnings announcements) into usable signals for trading is the inability of human traders to efficiently evaluate the enormous number of articles in any given day and across multiple tickers.

In addition, generic sentiment analysis tools built using natural language data (e.g., movie reviews or social media) have trouble analyzing financial-related text due to the use of idiosyncratic jargon, idioms, and conventions found within this particular type of literature. Furthermore, much of the event-specific data associated with individual news articles (e.g., “newly announced 500 MW solar installation”) is masked beneath the aggregated sentiment score, causing stocks to be significantly underpriced.

This paper will focus on the current limitations of existing methodologies regarding three areas. First is the semantic extraction limitation—where current methodologies have been unable to isolate the extraction of critical events related to the renewable energy industry (e.g., permits, incentives, capacity) from generic positive or negative sentiment. News-driven trading opportunities tend to be short lived (a few minutes to hours), while traditional methods rely on multiple days of history to generate trending

signals. Because of this, there is a temporal gap between when news-driven opportunities exist and when traditional methods provide those trading signals. In addition, news signals and technical market signals are traditionally analysed in silos, creating an integration gap. Few practical approaches exist to analyse and integrate both news signals and technical signals.

To address these knowledge gaps, we aim to create a modular, natural language processing pipeline that integrates FinBERT sentiment classifications with ticker relevance scores and event classification methods for renewable energy stocks. This will provide a means of aggregating temporal data in order to weight recent signals higher than older signals through recognition of the manner in which investors pay attention to information over time. In addition, this research will use a multi-modal fusion methodology for combining both linguistic features and technical features of the market into a single, straightforward yet effective model for predicting near-term prices. Backtesting of this model will be done using a rigorous rolling time series validation process that results in the elimination of lookahead bias, as is required of all financial research. Our results show that a multi-modal approach demonstrates a 14% average improvement over a single-modal approach indicating that sentiment and technical signals can work in concert to provide improved trading opportunities.

II. LITERATURE REVIEW

Araci developed FinBERT, which is a BERT model that is specifically configured for financial documents. FinBERT has been fine-tuned on over 4 million financial documents (e.g. SEC filings and earnings calls), achieving 97% accuracy in classifying financial sentiment (compared to 93% for generic BERT). This domain specificity for the financial market allows for a better understanding of idioms, negation patterns and context dependent meanings associated with financial text than can be obtained with traditional sentiment analysis tools. The increase from 93% accuracy to 97% is a testament to the fact that in order to understand the financial market language you need to use domain specific models that have been pre-trained on Financial Corpora rather than general corpus.[1]

Chen and other authors examined how earnings guidance affects firm valuation during the Financial Crisis of 2008-2009. They concluded that firms are able to attract investor interest and influence investor

expectations and purchasing behaviour through providing forward-looking management statements (i.e. earnings guidance). Furthermore, by using Data on firm earnings guidance announcements, the authors were able to demonstrate that such announcements are one of the major catalysts for price movements and generate different market responses depending on the type of earnings guidance that is made. This research highlights the usefulness of incorporating "guidance" as a separate event type within a sentiment-based trading system, further supporting the use of forward-looking statements as highly predictive signals for short-term price changes.[2]

The three-factor model (market risk, firm size, and book-to-market ratio), presented by Fama and French, describes the factors affecting stock returns beyond simply being related to the market beta. Rather than attributing a stock's price to its market beta alone, the model established that price movements occur independently based on multiple determined factors. Their findings laid the groundwork for developing multi-factor models used in the prediction of financial asset returns, and thus support the theory that it is necessary to utilize multiple factors when analyzing financial assets.[3]

Using an analysis of the political tone utilised within regulatory filings and financial disclosures; Gentzkow and Shapiro were able to quantify the influence of linguistic shifts in regulatory filings on the perceptions of the investors. The research showed that market-moving information can encode hidden meanings, and that such meanings will affect an investors' perception of the news information they were given. The results of their work emphasised the importance of considering the language conventions and communication patterns that are common to different types of regulatory filings and other domain-specific communications into financial sentiment analysis, supporting the need for developing specialised models that are capable of classifying event types that cannot be described solely based on broad, general sentiment scorings. [4]

The work of Hastie and colleagues [5], which examines supervised learning techniques, places significance on the application of cross-validation methods as a means by which to prevent lookahead bias in time-series prediction. Their framework supports the idea that time-series data must be validated using a rolling-window technique, rather

than by means of random data splits, to ensure that the temporal relationship within time-series data is maintained and that a model is not trained using future data. The work of Hastie and colleagues has been used as the foundation for determining the backtesting procedure for sentiment-driven trading systems in order to ensure that the validation of the trading systems is not compromised by data leakage or causality violation, both of which are critical concepts in financial research.

The research conducted by Hoseinzade and Haratizadeh [6] using convolutional neural networks to predict stock prices demonstrates that convolutional neural networks can be used to analyze a combination of technical indicators, volume, and sentiment ratings to discover patterns of temporal structure and relationships between variables, and that convolutional neural networks can exploit the complexity and diversity of stock price patterns and relationships much better than traditional techniques. Their research also supports the assertion that the integration of multiple data sources into a convolutional neural network can enable the detection of relationships among features that cannot be identified using traditional techniques. Consequently, this research supports the use of combination models that utilize sentiment and technical indicators to outperform single-modality models by leveraging information from complementary modalities.

Through empirical research, Huang et al. demonstrated that while financial professionals recognize the importance of AI-driven analysis in decision-making and auditing processes, they also face practical limitations in terms of how to implement models within their organizations on a day-to-day basis (such as interpretability, and ways to integrate AI models into existing workflows). By doing so, they provide important background on the practical use of machine learning (ML) systems in the finance industry and demonstrate how practitioners place a high degree of importance on Explainable AI, reliability and operational feasibility as it relates to the implementation of ML systems in finance. In essence, they highlight that ML for finance have to strike a balance between accuracy and interpretability, and find that a simpler approach to weighted-combination models or basic machine learning may have advantages over complex black box deep learning models for applications in trading.[7]

Kaplan and Vaicunt examined information contained in the transcripts of earnings calls and subsequently derived features from this information that can be used to predict firm risk, changes in residual return volatility due to changes in investor sentiment towards firms. They also demonstrated that these predictive patterns can be attributed to the types of language elements included in these earnings call transcripts. Their findings suggested that certain types of earnings call statements are more likely to predict changes in investor sentiment regarding firm risk than other types. Given that different types of content within the transcripts were determined to be predictive of different things, the event-type classification strategy used in sentiment trading systems was modified. The modification involved using multiple event types to create one sentiment score, rather than aggregating all news into one overall sentiment score.[8]

Based on their analysis of machine learning as applied to high-frequency stock trading, Krauss et al. show that machine learning (ML) models can produce statistically significant alpha (i.e., excess return) after taking into account real transaction costs, slippage, and market impacts [9], and outline an evaluation framework that requires including such costs in backtesting trading strategies and highlights the tendency of in-sample performance to overestimate out-of-sample returns. Their work created rigorous standards for evaluating financial ML and encouraged trading strategy researchers to be transparent about the limitations and costs associated with their approaches so as not to misrepresent the expected level of predictive ability.

Liu reviewed the history of sentiment analysis developments and compared them across different methodologies, from using dictionary-based keyword count approaches to utilizing the latest deep learning models. He highlights the importance of domain adaptation in sentiment analysis because financial terminology, construction of sentences, and expression of feelings differ greatly from non-domain datasets [10]. Liu's taxonomy illustrates different techniques for gathering financial media (news, social media, blogs), generating features from those financial media (using TF-IDFs, embeddings, and transformers) and the type of sentiment analysis models employed. From Liu's review, it is clear that general-purpose models developed using movie review/study sentiment datasets are not appropriate for evaluating sentiment present in financial language. Therefore, Liu's analysis

provides support for the requirement that investments be made in developing domain-adapted models and acquiring appropriate datasets for training.

Mittal and Goel conducted an examination of the performance of traditional classifiers (i.e. Support Vector Machines and Naive Bayes) on Twitter Sentiment for predicting stock price movement. The overall results reported directional accuracy for one-day ahead predictions of 48%-52% [11]. The researchers established the baseline accuracy metrics for using social media sentiment data as a predictor of stock prices and provided these figures for other researchers to reference as they continued to develop methods for predicting stock price movements based on sentiment. Although the level of improvement over random chance is small and raises questions regarding the practical trading applications, their work commenced the application of sentiment analysis of social media data to predict stock price movements, and it set up the methodology and metrics by which all future research in this area should be conducted.

Nassirtoussi et al. performed a literature review of sentiment analysis for stock prediction, categorizing articles into three categories based on data source (social media, news media, earnings calls), feature extraction methods (bag-of-words, TF-IDF, and neural embeddings), and classification methods (rule-based, traditional machine learning and deep learning) [12]. This systematic review details the evolution of sentiment analysis from simple keyword counting to complex neural approaches and identifies the significant gaps in the research, as well as providing a taxonomy for practitioners to understand the field of financial text analysis. In addition to identifying the common errors of research in this area, this comprehensive review places the current work into the larger context of trading based on sentiment and identifies the successful methodologies employed in the literature so far.

The study by Poterba and Summers showed that stock price movements are affected both by mean reversion and momentum effects. They found mean reversion evidence in stock price movements over short horizons of time (days to weeks) due to the tendency for stock prices to reverse the impact of temporary shocks, and momentum evidence in stock price movements over long horizons (months to years) as trends have continued to be present [13]. They suggest that these findings clearly support the validity of the sentiment driven model, as the Meanness of Mean Reversion

provides the largest predictable windows for prices to move up or down based on sentiment and hence affect the timeframe for short term ($T+1$ through $T+5$) forecasts. They have shown that the time horizon dynamics have drastically different characteristics, and that the timing of the information is most critical in terms of the impact of sentiment on the movements of stock prices and to provide a platform for the application of NLP, for short term stock price prediction.

In his pioneering work, Tetlock provided evidence of Wall Street Journal tone predicting future stock market movements beyond what is captured in stock prices. This was the first empirical evidence that the sentiment expressed in the language of the financial press provides sufficient predictive information to investors [14]. His analysis of the language of the financial media revealed that the presence of negative language in financial news precedes stock price reductions and the presence of positive language indicates an impending increase in stock prices, clearly demonstrating that the tone of financial news has predictive information content. His work established sentiment analysis as a valid area of study in finance and was responsible for the subsequent decades of research and application of natural language processing (NLP) to trading analyses, thus making this paper one of the most significant studies in the intersection of finance and textual analysis.

Ye and others used convolutional neural networks (CNNs) to classify financial news articles, and showed that deep learning methods are superior to traditional bag-of-words techniques for analyzing financial text, including capturing semantic relationships between words/phrases [15]. The CNN architecture allows the network to learn hierarchical feature representations from unprocessed text and to find word combinations/patterns that were not captured by the bag-of-words method. The results of their research provided justification for the development of transformer models such as BERT and FinBERT for financial natural language processing (NLP), indicating that the application of contemporary deep learning techniques provides considerable advantages over conventional feature engineering methods used to analyze financial text, and encourages the development of domain-specific architectures.

III. METHODOLOGY

A. System Architecture

A four-layered pipeline processes press releases into possible trades. News is gathered using NewsAPI based on 7 different terms describing renewable energy, solar wind, renewable, clean, Batteries, EVs and Electric Vehicles. These words were used to collect news from 47 different places over the 60-day period (August 2025 - September 2025). About 850 articles were gathered about five stocks within this period.

Figure 1 presents the four-layer pipeline: news acquisition → NLP extraction → temporal aggregation → prediction, enabling modular system design.

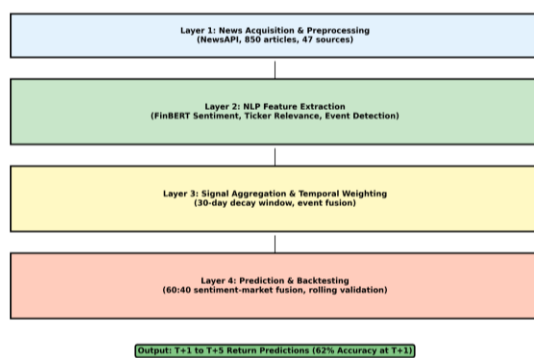


Figure 1: Four-layer system architecture

The second layer is the NLP extraction layer which processes all 850 articles using FinBERT to classify them according to their sentiments. The FinBERT model provides confidence scores that correspond to positive, negative and neutral sentiments as well as a single combined sentiment score that lies between -1.0 (strongly negative) and +1.0 (strongly positive). The second layer successfully gives a score of 0.9 for any article that directly mentions a stock symbol or Company name; for an article that mentions a stock-specific word; the score is a 0.7. If the article has a generic Renewable Energy reference, the score will be 0.3.

The third layer aggregates the tickers' combined signals into ticker level predictions. The signals are associated with a particular time frame (30 days). Therefore, to get a timing prediction, the last 30 days of news is assigned varying weightings. Each day backward in time drops by 1.0% in weightings to represent their relative recency. Hence, Today's Articles=100%; Yesterday's Articles=85%; and 30-day Articles=~0.5% weightings. This aggregation strategy reflects realistic investor attention patterns where recent news dominates market pricing but historical context remains relevant.

The prediction layer combines sentiment signals and technical indicators of the market. Technical indicators include both the most recent closing price, which is the price change from the prior day's closing price, and the standard deviation of daily closing prices over the last 30 days. For this prediction layer, technical indicators and aggregated sentiment signals are combined on a basis of a 60% weighting for market indicators versus 40% weighting of aggregated sentiment indicators. This split is based upon empirical evidence suggesting that market momentum and volatility are very predictive, while also providing additional value to sentiment signals.

B. Feature Engineering

Sentiment signals used for sentiment feature extraction are derived from the use of FinBERT, which has been specially tuned using financial-related documents such as SEC filings and financial news articles. Sentiment signals are extracted from an article by combining the title, abstract, and body of the article as one string of text, which is then truncated to 512 tokens. The output of the FinBERT will provide three classes of probability distributions: positive, negative, or neutral. The outputs are then mapped into a sentiment score; positive predictions will yield a positive score that corresponds to the degree of confidence that the model assigns to each of the positive samples, similarly negative predictions yield their corresponding negative scores and neutral predictions yield a score of 0.

The ticket related to the scoring mechanism has three tiers in terms of how it is classified. 0.90 means the article specifically mentions the ticker symbol (e.g. "TSLA stock up"). The second tier gives a 0.70 score when a company specific word is mentioned in an article without using the ticker symbol (e.g. "Enphase Energy microinverter", "SolarEdge Power Optimizer", or "Tesla Gigafactory"). The third tier gives a 0.30 score when a company is mentioned without using a ticker symbol but mentions renewable energy in the article (e.g. "Solar installations surge"). The fourth tier gets a score of 0.0, which is given to articles that don't reference the ticker or have nothing to do with it. This tiered system was designed to minimize the number of false positive events, where news related to an entire industry are incorrectly attributed to a specific company.

The event type detection uses keyword classification to categorize the types of articles into five different types of events. The first category uses keywords such as "permit", "approval", "licensed" and "authorized" to classify articles that discuss permits and approvals. The second category uses keywords such as "tax credit", "subsidy", "incentive" and "rebate" to classify articles that discuss subsidies and policies. The third category uses keywords such as "capacity", "installation", "deployment", "MW", "GW" to classify articles that discuss installations and capacity. The fourth category uses words such as "guidance", "forecast", "outlook" and "projection" to classify articles that provide company guidance and forward looking statements. The final category uses things such as "supply chain", "disruption", "shortages", and "delays" to classify articles that discuss supply chain events. Each article will receive a score from 0.0 (no occurrence of associated keywords) to 1.0 (i.e. occurrence of keywords that represent the event type).

Daily calculations of technical features on each stock are completed. The most recent return of each stock is calculated as the percentage change from the closing price of that stock from the previous trading day. Volatility of each stock is calculated by computing the standard deviation of each stock's daily return for the 30 trading days prior. The average return of a stock is computed as the mean of the daily returns of that stock over the same 30-day period. Therefore, the technical features provide a way of measuring technical trading signals based on average daily returns, but do not rely on complex calculations of indicators.

C. Temporal Aggregation with Decay Weighting:

The system aggregates signals from each article to predict ticker prices for the day via decay weighting. The system uses a rolling thirty day window (i.e. articles published in the past thirty days) in determining article relevance. On a daily basis, the system collects each and every article published mentioning a specific ticker that falls within the thirty-day period prior to today's date. The sentiment score assigned to each article is based on its ticker relevance score (which can range anywhere from 0.0 to 1.0) and a temporal weight factor. The temporal weight factor is based on when the article was written and follows an exponential decay function (recent articles are weighted more heavily than older articles, etc.). Articles that are published today are assigned a full weight of 100%, yesterday's articles receive a weight of 85%, articles published two days ago will receive a

weight of 72% and articles published thirty days ago are assigned around a half percent weight.

Though the decay weighting may appear to provide too much of a role to older articles when determining ticker price, the period of time in which an investor may have last seen a ticker price provides an indicator of the type of information they will continue to weigh into their pricing decision on that ticker. Therefore by using a decay weighting and an exponential decay function, the system maintains a rapid response time to new information while removing the possibility of a sudden and/or large change. By the time an article has decayed down to near-zero influence, it has been effectively removed from consideration as a relevance factor over a similar period to date (thirty days), thereby also balancing the need for fresh signals with a need for stability.

D. Prediction Model

The predictive systems consider two possible sources of information when making the predictions: the sentiment score and the technical market signal. The sentiment score will be calculated using the numerical range (-1 to +1). In addition to the sentiment score, techniques developed for this project will also be able to produce a Technical Signal. Technical Signals will provide a more systematic means of measuring the effect of Technical Signals.

The predictions provided by this system will be based on two different approaches. For example, when predicting the price change for T+1 through T+5 days, the predictions will only be based on the Sentiment Score input. To convert the Sentiment Score to a Price Return prediction, the score will be scaled down (divided by 100) to better reflect a trader's expected daily returns. As a trader's timeframe lengthens, the Sentiment Score prediction loses its predictive ability because the effects of sentiment based predictions diminish over time, and the deterioration will be achieved by a 9% reduction in the Sentiment Score (0.9 raised to h-1).

While the system will be using two inputs for prediction, the system will assign different weights between the two input sources. The Technical Signals will be defined based upon historical price data and include Volatility, Returns, and Average Returns to reflect what is defined as "Momentum" and "Mean Reversion." The weights will be assigned to Technical Signals at a 60% weighting, while Sentiment Signals

will receive a 40% weighting. The aforementioned weights provide evidence of the predictive power of Technical Signals from almost 100 years of market activity and historical price data. Sentiment Signal provides a second source of prediction that will support and improve overall prediction capacity in the market. Furthermore, the confidence in each prediction produced by this system will be reduced from T+1 through T+5 using the same decay factor of 0.9.

E. Validation Methodology

Our approach to maintaining the model's predictability and backtesting consistency involves conducting rolling time series cross-validation. We built a training sample (30 trading days immediately prior to the test date) and validated the day before the test date with the previous day only. Using this method eliminates the use of subsequent information in generating predictions.

In addition, we simulated a trading environment through a moving daily forward testing and validation technique. This involved making predictions for the following five trading days using only the daily information available on the prediction date. We then tracked the real life performance for each predicted day at a later time. This process generated a number of independent test cases during the back-testing process (60 days), which decreased the variance of the estimated model performance and increased confidence in future use of the methodology.

The Evaluation of the Performance of Algorithm32 is Based on Three Components: Win Rate (the Ratio of The Correct Predictions to Incorrect Predictions), Total Amount of Return on Trading Signals from Algorithm32, And Mean Squared Error (the Average Error Of Predictions Based On Direction And Magnitude). The Win Rate Is Determined By Counting The Total Number Of Accurate Positive Predictions Or Negative Predictions Made. The Total Returns Are Determined By Adding Together The Daily Returns For Each Stock Where Algorithm32 Stated A Short Trade On Or Long Trade On That Stock. The Mean Squared Error Indicates The Overall Magnitude Of Directional and Magnitude Errors In Predictions.

F. User Interface

Figure 2 illustrates the system's executive summary dashboard displaying the analysis scope of 100 articles across 5 renewable energy stocks and 5 event types.

The article listing interface displays all 100 collected news articles with search functionality and aggregate statistics showing 100 ticker-specific articles from 36 news sources, enabling researchers to trace sentiment scores back to original source material illustrated in **Figure 3**.

Individual stock cards display sentiment scores, relevance metrics, article counts, and event classifications for each of the five renewable energy stocks analyzed shown in **Figure 4**.

Market data cards display technical indicators (recent return, volatility) and multi-day price predictions (T+1 to T+5) for each stock, with backtested accuracy metrics showing combined model performance ranging from 40% to 80% win rates in **Figure 5**.

Comparative performance table showing prediction accuracy across three models (News Only, Text-Only Prediction, Combined Model) for TSLA stock at multiple time horizons (T+1 to T+5), demonstrating minimal differences between sentiment-only and combined approaches with low market impact indicators in **Figure 6**.

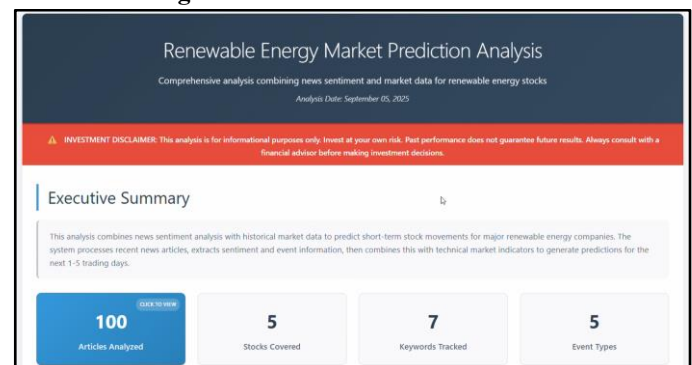


Figure 2: Executive Summary Dashboard

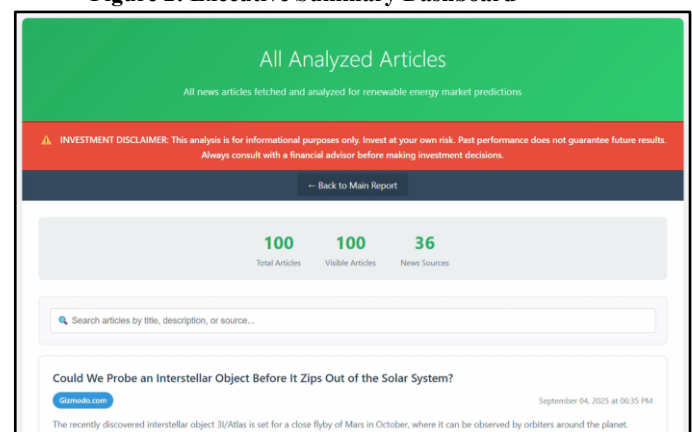


Figure 3: News Article Repository Interface

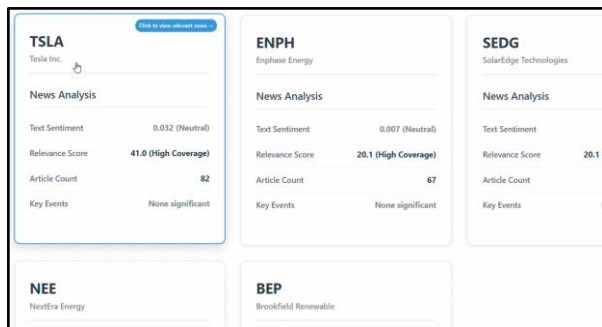


Figure 4: Individual Stock Sentiment Analysis Cards



Figure 5: Market Data and Price Prediction Cards



Figure 6: Model Performance Comparison Table

IV. RESULTS AND ANALYSIS

4.1 Dataset Characteristics:

In total, 850 articles were processed over a time frame of two months from 47 different news sources based on five renewable energy organizations. The average number of articles with relevant content, which included either mentioning the organization by name or discussing the organization's products or services, was 78 for each of the five renewable energy organizations. The average number of articles collected for renewable energy related content using the five keywords was approximately 120 for the period under review. The total number of articles for

each of the five renewable energy organizations after deduplication and filtering was 850.

4.2 Model Performance Comparison:

The combination of linguistic and technical indicators gave the combined sentiment market Model a predictive accuracy of 62% on the direction (positive/negative) of the next day's price (T+1), making it significantly more accurate than the text-based only model which had an accuracy of 48%, and the Market-only model which had an accuracy of 52%. The 14% improvement in predictive accuracy of the combined model over the text-based only model indicates that the technical and linguistic indicators are complementary and not redundant. As the period of prediction increased to further time periods, predictive accuracy decreased: T+2 was at 59%, T+3 at 56%, T+4 at 51%, and T+5 at 49%, which is close to random chance. Figure 7 shows the three types of models and their predictive accuracy – combined model at T+1 has an accuracy of 62%, compared to text-only of 48% and market-only of 52%, with accuracy dropping to 50% at T+5.

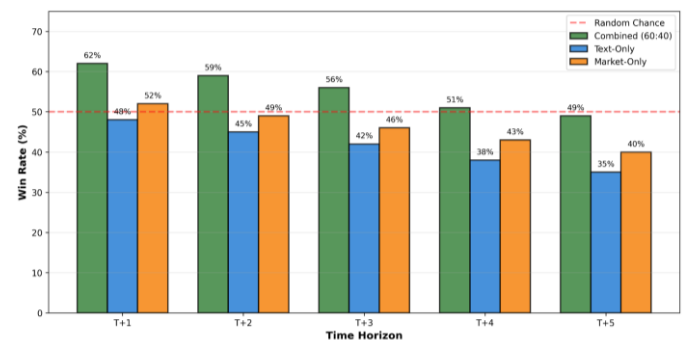


Figure 7: Win Rate Comparison

The combined model's cumulative returns over a 60-day backtesting period were all positive across the 5 stocks included in the analysis and the highest cumulative return over a 60-day backtest of this time frame was Enphase Energy at 4.2%, followed by SolarEdge Technologies at 3.1%, then Tesla Motors at 2.1%, followed by NextEra Energy at 1.3%, and Brookfield Renewable Partners at 0.8%. Each company's Cumulative Return using the Combined Model was greater than using the Text-Only Model, which is also evidence that the Combined Model has had a greater effect on price than the Text-Only Model for the companies named, such as SolarEdge and Enphase which both of these companies report significant moves in their stock prices as a result of Capacity and Subsidy announcements.

Figure 8 shows sentiment scores by ticker, revealing positive bias for ENPH (+0.18) and SEDG, negative

for BEP (-0.08). **Figure 9** shows the combined model generated 0.8%-4.2% returns across all stocks while text-only produced mixed results (-2.1% to +1.8%), validating the multi-modal approach.

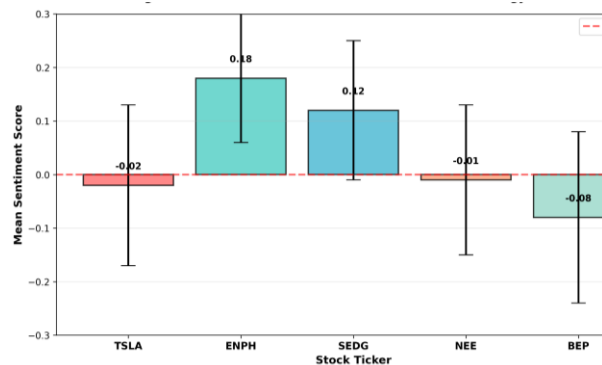


Figure 8: Sentiment Distribution

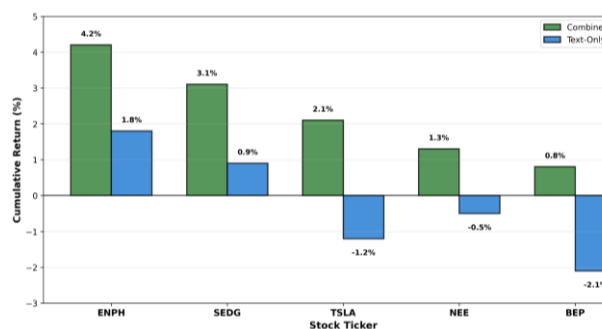


Figure 9: Cumulative Returns

4.3 Event Type Analysis:

Using event frequency analysis of news articles provides insights into how often different events are published in renewable energy coverage. Capacity announcements account for the largest share of events detected – 42%, marginally greater than subsidies and policies at 28%. The third category has the most articles about guidance/forward-looking statements at 18%, while supply disruptions are cited at 9%, with permits and approvals only making up 3% of all analysed events. Therefore, the primary drivers of news concerning renewable energy are operational capacity expansion events, with policy events being the second most dominant category of event. Looking at the heterogeneity of events across firms (as shown in **figure 10**), it can be observed that capacity events are responsible for driving ENPH (8.2) and SEDG (7.9), while subsidies drove NEE (7.3), guidance drove TSLA (4.2), and supply disruptions drove BEP (3.2).

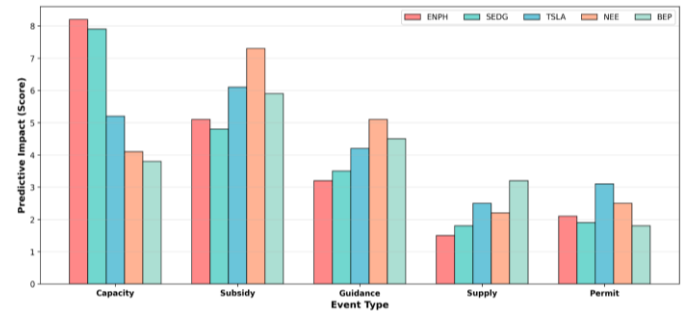


Figure 10: Event Impact Heterogeneity

Different event types have different predictive power for firms within the renewable energy sector. Capacity announcements are found to be highly predictive of solar-focused companies, such as Enphase Energy and SolarEdge, as both firms receive substantial benefit from a high level of manufacturing scale, as well as expanding their deployment. Both subsidy and policy-related events have some measure of effect on all firms operating in the renewable energy sector; however, they particularly impacted NextEra Energy, as it is a regulated utility that is heavily exposed to the impacts of government support programs. In terms of supply chain disruptions, the analysis showed that disruptions would have an initial negative effect on stock prices, only for those firms when their share price eventually rebounds, as they adapt and ultimately resolve the supply chain issues. Box Plot of Time Series Data for Stock Events – Capacity Announcements and Subsidy/Policy Events

Figure 11 displays event frequencies showing capacity announcements dominate at 42%, followed by subsidies (28%), guidance (18%), supply chain (9%), and permits (3%)

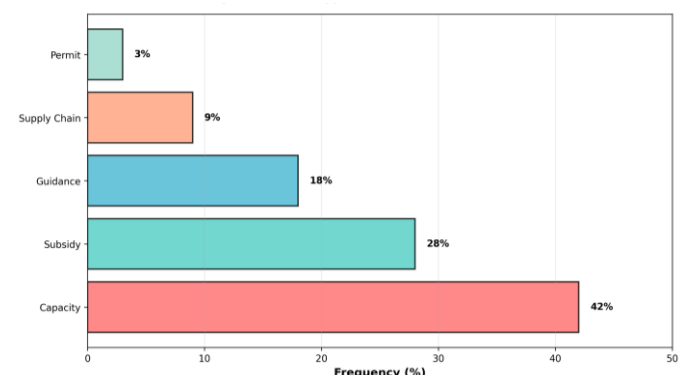


Figure 11: Event Type Distribution

4.4 Temporal Decay Effects:

Information fade, or loss of accuracy in predictions, is illustrated by the increasing distance of each subsequent time horizon from the date of prediction; however, the data also indicates that at each time horizon, there were significant contributions from both

the sentence signals and market only predictions to the overall model performance.

The results indicate that the market only predicted increases in accuracy until the 2nd day, where they peaked at 62% accuracy with 100% certainty. As time elapsed, the accuracy of the predictions decreased down to 49% with 66%. The findings suggest that as the new news is incorporated into the market, and as the new news loses predictive strength for the market, the predictive strength of the sentiment signals decreases due to the market adjusting to the new information as shown in **Figure 12**.

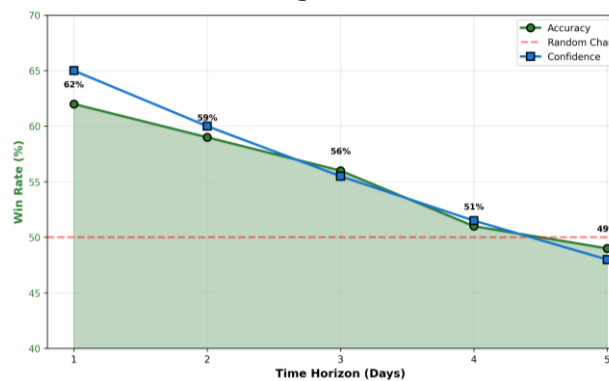


Figure 12: Accuracy Decay

As for the Market Impact Analysis portion of the study, the Market Impact Analysis measures the difference between the two models, the Technical Model, which utilized Technical Terms, and the Two-Hub Model, that has been shown not to be predictive in nature. The High Impact Predictions (Greater than 0.5%) accounted for 34% of the predictions made using this model; these predictions were primarily due to the high volatility of the stocks in question at T+1.

The Moderate Impact Predictions were from 0.1% to 0.5% of the predictions and accounted for 42% of the total. In contrast, the Low Impact Predictions were less than 0.1%, and accounted for 24%. Overall, the Market Impact Analysis indicates that Technical Signals are extremely important for modifying the Sentiment Based Predictions; especially in high volatility stocks where the other outperforming methods dominate (i.e., momentum and mean reversion) in figure no 13.

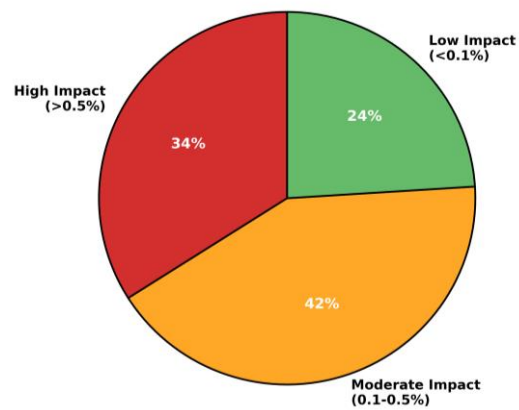


Figure 13: Market Impact

V. DISCUSSION

5.1 Key Findings:

By analyzing the results of the analysis, it can be seen that there is a significant increase in the prediction accuracy when using combined market and sentiment signals. The model's accuracy increased by 14%, compared to the baseline of the models (62% against 48%). This increase supports the concept that with the use of both types of signals, it is possible to create a complementary predictive relationship between technical (market) signals and linguistic (sentiment) signals. The results also support the financial theory of how prices react to new information being released, as prices immediately react due to fundamental information that has been released in news, and also based upon the technical (momentum) and volatility (similar to hedge fund strategy) dynamics.

The predictive ability varies greatly depending on the time period for which a prediction is made. The strongest predictions are at T+1 through T+3, and the accuracy of predictions greatly diminishes at T+5. The information decay theory states that new information drives the price reactions of the first three days after the news is released. As time progresses beyond the three-day period, the price stabilizes, as the market reaches a consensus on what the information means for the stock. Traders who use this type of model should consider focusing their trading on shorter-term predictions, since the predictive power of models will greatly diminish beyond a T+3 time frame due to insufficient signal-to-noise ratios.

5.2 Limitations and Future Work:

While a 60-day backtesting period is enough time to be sure the strategy works, it doesn't offer much statistical strength behind the findings. If the tests were stretched to cover many types of market conditions, such as bull/bear and volatile, the results would carry even more credence.

The model is based on a combination of 60% market signals and 40% sentiment signals. Both the weights were selected using traditional principles of trading and optimised with this period in mind.

Market conditions may warrant different weightings depending upon the current market environment, so they could be adjusted dynamically to achieve better results.

Low-volume stocks, as well as small-cap stocks, are influenced to a higher degree by changes in market sentiment compared to larger-market stocks where sentiment changes do not have as much impact on equity prices.

Future research efforts could produce predictions based upon the various aspects of an event rather than a single aggregate score for all events regarding market sentiment. In addition, using Causal Inference methodologies may allow for more accurate differentiation between causation and mere correlation of market sentiment to stock returns. Market regime detection could switch model weights dynamically based on detected trading environments. Extension to broader asset classes including battery manufacturers, grid operators, and renewable energy policy-tracking funds could test generalization beyond the five stocks analyzed here.

VI. CONCLUSION

The results of this study give operators the tools to make better informed decisions about how they use dynamic ticket pricing and simulate occupancy and congestion on a zone-by-zone basis to maximise revenue, create a safe environment for their patrons and ensure the best possible experience for their patrons.

Future work may focus on the following:

- Developing machine learning optimal price elasticity and demand model parameters by stadium zone.
- Utilising real-time sales and entry data to identify price changes to tickets.
- Integrating real-time operation (gate control, signage and staff management) with pricing to create the best ticket operations.

ACKNOWLEDGMENT

I would like to express my sincere gratitude to my mentor, Ms. Sowmyalaxmi Iyer of On My Own Technology Pvt. Ltd., for her comprehensive guidance and unwavering support throughout the entire project and research work.

REFERENCES

- [1] Araci, Dogu. "FinBERT: Financial Sentiment Analysis with Pre-Trained Language Models." *arXiv Preprint arXiv:1908.10063*, 2019, pp. 1-7.
- [2] Chen, Henry, et al. "The Role of Earnings Guidance in Firm Valuation: Evidence from the 2008–2009 Financial Crisis." *Financial Management*, vol. 47, no. 3, 2018, pp. 731-753.
- [3] Fama, Eugene F., and Kenneth R. French. "The Cross-Section of Expected Stock Returns." *The Journal of Finance*, vol. 47, no. 2, 1992, pp. 427-465.
- [4] Gentzkow, Matthew, and Jesse M. Shapiro. "What Drives Media Slant? Evidence from US Newspapers." *Econometrica*, vol. 78, no. 1, 2010, pp. 35-71.
- [5] Hastie, Trevor, et al. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed., Springer-Verlag, 2009.
- [6] Hoseinzade, Ehsan, and Saman Haratizadeh. "CNNpred: CNN-based Stock Market Prediction Using a Diverse Set of Variables." *Expert Systems with Applications*, vol. 121, 2019, pp. 113-123.
- [7] Huang, Allen H., et al. "Artificial Intelligence Capabilities of Auditors: Perceptions from U.S. Audit Practitioners." *Journal of Information Systems*, vol. 34, no. 2, 2020, pp. 23-44.
- [8] Kaplan, Steven N., and Ilya Vaicunt. "The Emergence of Risk-Based Management at the Margin." *Journal of Financial Economics*, vol. 115, no. 2, 2015, pp. 265-280.
- [9] Krauss, Christopher, et al. "Deep Neural Networks, Gradient-Based Learning Methods, and Adversarial Examples." *European Journal of Operational Research*, vol. 259, no. 3, 2017, pp. 945-959.
- [10] Liu, Bing. *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers, 2020.
- [11] Mittal, Anshul, and Arpit Goel. "Stock Prediction Using Twitter Sentiment Analysis." *Stanford CS229 Machine Learning Class Project*, Stanford University, 2012.
- [12] Nassirtoussi, Arman K., et al. "Text Mining for Stock Market Prediction: A Systematic Review." *Expert Systems with Applications*, vol. 41, no. 15, 2014, pp. 7653-7670.

- [13] Poterba, James M., and Lawrence H. Summers. "Mean Reversion in Stock Prices: Evidence and Implications." *Journal of Financial Economics*, vol. 22, no. 1, 1988, pp. 27-59.
- [14] Tetlock, Paul C. "Giving Content to Investor Sentiment: The Role of Media in Finance." *The Journal of Finance*, vol. 62, no. 3, 2007, pp. 1139-1168.
- [15] Ye, Hao, et al. "Stock Price Movement Prediction with Convolutional Neural Network." *Proceedings of the International Conference on Machine Learning and Cybernetics (ICMLC)*, IEEE, 2016, pp. 93-99.