Hybrid Option Valuation: PINN Models Informed by Black-Scholes and Alternative Data Sources

Kiran Bodipati
bodipati@umich.edu
University of Michigan
USA

Dhanush Kumar Mallu
dhanushm@umich.edu
University of Michigan
USA

Jehan Patel
pajehan@umich.edu
University of Michigan
USA

ABSTRACT

Predicting the value of stock options remains a critical challenge in financial markets, particularly as data availability and computational capabilities quickly evolve. Traditional valuation models, such as those based on the Black-Scholes equation, use classical financial data that provides limited advantages to investors with access only to public information. However, advances in mobile technology, cloud computing, and machine learning have made alternative data – including social media sentiment, web traffic, and high-frequency transaction data – much more accessible. This project proposes a novel framework that integrates both traditional financial metrics and alternative data sources to predict option values. We implement a knowledge-transfer mechanism where a Black-Scholes-informed source model guides data-driven target models. By embedding domain knowledge within a data-driven architecture, our approach aims to enhance predictive accuracy and to enable investors to capture incremental informational advantages at a lower cost. We have released our implementation and datasets open-source.[1 2]

KEYWORDS

Stock/options Pricing, Alternative data, Transfer Learning

1 INTRODUCTION

Valuation of stock options is central to finance, with implications for portfolio management, risk assessment, and market efficiency. Traditionally, models such as the Black-Scholes equation [4] provide a theoretical framework for option pricing but rely only on historical price data and standard financial disclosures. These approaches often fail to reflect real-time enterprise conditions or integrate diverse information flows.

Recent advances in mobile technology, data storage, and machine learning have made "alternative data" sources, such as social media sentiment, web traffic, sensor data, and real-time transaction logs, more accessible for financial modeling [12]. These unconventional datasets contain unique attributes and are evolving fast, creating opportunities to discover insights unavailable from traditional financial data. Studies show that integrating alternative data can improve predictions of market risk factors and stock trends [6].

The broader challenge is that traditional models are limited by their inability to absorb and learn from high-frequency and predictive information present in alternative datasets. This motivates our project's central objective: to leverage hybrid valuation models that utilize both classical financial data and alternative data sources for more accurate, market-responsive option pricing.

[1]Project Website: https://cse-598-ai-for-science.github.io/project-website/
[2]GitHub: https://github.com/orgs/cse-598-ai-for-science/repositories

Predicting stock option prices — contracts whose worth depends on future value of a stock — is a fundamental challenge. Traditional methods rely on financial data and mathematical models but often miss out on timely or hidden market signals, especially as new forms of data (such as social media and real-time transactions) become widely available. The key problem is how we can combine traditional financial data and newer "alternative data" for improved pricing accuracy, while retaining conventional mathematical principles.

Let $V(t, S_t)$ be the theoretical value of a European stock option as defined by the Black-Scholes partial differential equation (PDE):

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S_t^2 \frac{\partial V}{\partial S_t} - rV = 0, \tag{1}$$

where r is the risk-free rate and σ is volatility. For training, we use both traditional financial metrics and features from alternative data, such as search trends or sentiment scores.

Given historical time series data, the objective is to learn a predictive mapping $f : (S_t, X_t) \to \hat{V}(t, S_t)$ that:

- Approximates observed market prices
- Adheres to financial priors by enforcing the Black-Scholes PDE
- Integrates signals from both financial and high-frequency alternative data

This will be achieved by designing and training a Physics-Informed Neural Network (PINN) [10] that incorporates the Black-Scholes PDE into its loss function, and jointly leverages both traditional and alternative data X_t as input features.

We assume access to sufficient labeled data, including option prices, stock prices, and synchronized alternative data streams (e.g., sentiment, search trends, transactions) during training. Markets are liquid enough for reliable pricing. The Black-Scholes model serves as a useful regularization, and data pre-processing manages missing values and heterogeneous inputs.

2 RELATED WORK

Our work sits at the confluence of three areas: classical financial option-pricing models, alternative data in finance, and knowledge-guided machine learning. This section reviews the most relevant prior work and positions our approach in relation to them.

2.1 Financial Option Pricing Models

The Black–Scholes model [4] is the foundational framework for pricing European-style options. It assumes the underlying asset follows a geometric Brownian motion with constant volatility and a risk-free interest rate. This leads to a partial differential equation (PDE) with a closed-form solution for call and put options. However, its simplifying assumptions such as constant volatility and frictionless markets limit its accuracy in turbulent or information-rich market conditions.

2.2 Alternative Data in Finance

Sun et al. (2024). Recent years have seen a surge in the use of alternative data sources — such as social media sentiment, web traffic, product consumption data, geolocation, and satellite imagery — to complement traditional financial statements, SEC filings, or analyst reports. These unconventional datasets provide timeliness, granularity, and a potential informational advantages. Sun et al. [12] provide a comprehensive review of emerging applications, advantages, and limitations of alternative data in finance.

Kraaijeveld & De Smedt (2020). Analyzing cryptocurrencies, they demonstrated that public Twitter sentiment exhibits predictive power over price movements, especially when sentiment is extreme or trending [6]. This work establishes social media as a meaningful early signal of market shifts, albeit in a narrower asset-class context.

Liu et al. (2020) — FinBERT. FinBERT [7] is a pre-trained BERT variant fine-tuned on financial text corpora, enabling robust sentiment extraction from news, tweets, earnings calls, and more. It provides a bridge between unstructured natural language and quantitative financial modeling.

2.3 Knowledge-Guided and Physics-Informed Machine Learning

Physics-Informed Neural Networks (PINNs). PINNs incorporate known differential equations (such as PDEs) directly into the loss function, forcing the learned function to satisfy theoretical constraints while fitting data. This improves data efficiency, interpretability, [10] and generalization.

PINNs for Options and Stock Pricing. Recent work has extended PINNs to enforce option-pricing PDEs like Black–Scholes and Heston as structural priors. [2]. Another paper proposes a hybrid PINN/FNO approach with an estimated 50% improvement in MSE over traditional approaches. [5] A PINN-based pricer for American options on the S&P500 accommodates early exercise by embedding free-boundary conditions into the loss function [3]. While effective for structured financial data, such models do not integrate heterogeneous alternative data inputs.

EINNs (Rodríguez et al., 2023). EINNs were developed for epidemiological modeling and combine mechanistic PDE models (such as SIR) with deep networks via gradient matching and embedding alignment between source (theory) and target (data) models [11]. This enables structured knowledge transfer from mechanistic models to flexible neural architectures.

Relation to Our Approach. We adopt the PINN paradigm but extend it to accept heterogeneous input modalities (traditional plus alternative data). We adapt the EINNs-style gradient and representation alignment method to finance: our source is a Black–Scholes-informed PINN, and our targets are neural modules trained on traditional and alternative data inputs. This gives us a mechanism to inject theoretical structure while allowing flexibility and empirical learning from diverse data.

2.4 Gaps in Existing Literature and Our Contributions

Despite these advances, several gaps remain:

(1) Most financial PINNs or theory-aware networks use only structured historical financial variables and rarely incorporate unstructured or high-frequency alternative data.
(2) Gradient-matching and representation-alignment frameworks (like EINNs) have not yet been applied to financial PDE/data coupling.
(3) Existing work seldom quantifies the marginal utility of alternative data when blended with theory through systematic ablation studies.
(4) Alternative data streams can be noisy or contradictory; how to integrate them without undermining PDE consistency is underexplored.

Our Key Contributions. We present a variant of financial PINNs that fuses alternative data streams in addition to traditional variables, and we believe we are among the first to do so. We introduce a EINN-inspired gradient alignment in a financial setting to anchor empirical learning to theoretical structure. As such, we aim to study the effect of combining knowledge guided approaches and alternate data with the traditional data driven approaches.

3 METHOD

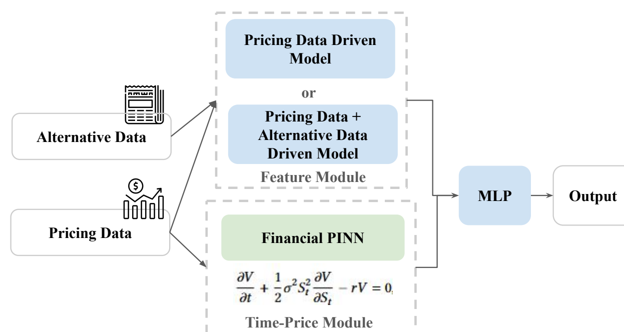


Figure 1: Proposed framework for our Option Valuation model. We aim to replicate the EINN framework to have a source and target module, where knowledge transfer takes place.

The proposed methodology adapts the methodology from the Epidemiologically Informed Neural Networks (EINNs) framework, a state-of-the-art approach designed to bridge mechanistic models with data-driven neural networks [11]. We will implement a heterogeneous domain transfer learning setup where knowledge is transferred from a single theory-driven Source Model to two independent, data-driven Target Models. This structure allows for a rigorous, direct evaluation of the incremental value of alternative data in option pricing.

3.1 The Time-Price Module (PINN)

The function of the source model is to learn the latent dynamics of the Black-Scholes world, creating a continuous and differentiable neural representation of the idealized option pricing model. We use inspiration from EINN and Financial PINNs [3, 11]

- **Architecture:** The architecture is a Physics-Informed Neural Network, with the implementation adapted from the Financial PINNs [3].
  The core network is a 6-layer MLP with 128 hidden units per layer, using ReLU activations and Dropout(0.1) for regularization. We avoid BatchNorm to ensure stability. The multi-component forward pass constructs the option value $u(S,t)$ by explicitly building in the intrinsic value. The network's normalized inputs (S_{norm}, t_{norm}) produce a `raw_output` which is then combined with the intrinsic value:

$$u(S,t) = \max(\underbrace{\max(S-K,0)}_{\text{Intrinsic Value}} + \underbrace{f_\theta(S,t,\texttt{raw\_output})}_{\text{Time Value}}, \text{min\_value}) \tag{2}$$

  where f_θ is a complex, learnable function representing the time value, ensuring the American constraint $u(S,t) \geq \max(S-K,0)$ is met *by design*.
- **Training Data & Process:** The module is trained on a large set of collocation points generated via **strategic sampling** across the problem domain, with specific point sets for the PDE interior, the boundaries ($S \approx 0$), the terminal condition ($t = T$), and the American constraint region. The training objective is to minimize a weighted, multi-component loss function:

$$\mathcal{L}_{\text{Total}} = \lambda_{\text{pde}}\mathcal{L}_{\text{PDE}} + \lambda_{\text{boundary}}\mathcal{L}_{\text{Boundary}} + \lambda_{\text{initial}}\mathcal{L}_{\text{Initial}} \tag{3}$$

  where:
  - $\mathcal{L}_{\text{PDE}}$ is the unsupervised physics loss from the residual of the Black-Scholes PDE.
  - $\mathcal{L}_{\text{Boundary}}$ enforces the boundary condition for $S \to 0$.
  - $\mathcal{L}_{\text{Initial}}$ is the supervised loss at the terminal time T, enforcing the payoff $u(S,T) = \max(S-K,0)$.

  We also implement a penalty for violating the early exercise constraint of an American Option
- **Outcome:** A trained source model, $N(S,t)$, that acts as a fast, continuous, and differentiable solver for the American option price. It has learned the latent dynamics of the Black-Scholes PDE while simultaneously respecting the no-arbitrage, early exercise boundary.

3.2 Feature Modules

We develop two independent target models, both based on a deep sequential architecture, for both the pricing data driven model as well as the alternative data model. We make use of a basic Multi Layer Perceptron (MLP) model as well as the Gated LSTM architecture [8].

3.2.1 Baseline Feature Module (Pricing Data Driven). One of our baseline module utilizes a Gated LSTM architecture as a robust benchmark for option pricing [8]. The other uses a small multi layer perceptron model.

- **Inputs:** Time-sequenced, normalized financial signals including option price, underlying price, strike, time to maturity, trading volume, risk-free rate, and 90-day moving average volatility. Scaling statistics are derived solely from training data.
- **Model:** Consists of five stacked LSTM layers each followed by batch normalization and ReLU activations, as discussed in the original paper. Adam optimizer and early stopping are used to maintain training stability.
- **Training:** We use the size and techniques used in literature, and train it using an ADAM optimizer, for 100 epochs, with a batch size of 32.

3.2.2 Enhanced Feature Module (Pricing + Alternative Data). We will extend the feature module to use the same architectures, but will make the necessary modifications to include the two columns of alternative data sources.

- **Inputs:** An enriched, normalized time series ($\mathcal{X}_{\text{traditional}}$) containing all signals from the baseline module plus engineered features from alternative data sources - the sentiment scores on technological and market decisions, as well as the geopolitical data encoded in the form of the EPU index.($\mathcal{X}_{\text{alternative}}$)
- **Architecture:** We extend the architecture of the baseline feature module. The LSTM will process the financial time series and will include additional MLP layers which will process contemporaneous alternative data. Their latent representations will be concatenated before the final prediction step.

3.3 The Knowledge Transfer Mechanism

The core of the methodology is transferring knowledge from the source to the target modules. Applying the PDE loss directly to the RNN-based target modules is intractable because their inputs are feature vectors, not the PDE variables. We will therefore adopt a variation of the technique from Rodriguez et al. [11] by using the PDE parameter calibration and regularization in the feature module.

- The key parameters of the Black-Scholes PDE (r, σ) are first calibrated using either traditional numerical methods or physics-informed neural networks (PINNs), leveraging established domain knowledge alongside observed market data.
- Once identified, these parameters are held fixed, and the PDE itself is incorporated as a regularization term for the feature module. This ensures that feature learning remains consistent with financial theory while drawing on both traditional and alternative data sources.
- The loss function guiding feature learning is defined as

$$L = L_{\text{data}} + \lambda L_{\text{PDE}} \tag{4}$$

  where L_{PDE} penalizes deviations from the calibrated PDE structure during training and λ is a balancing hyperparameter.

The training will proceed in a two-step process: first, a phase focused on aligning the embeddings, followed by a fine-tuning phase that incorporates all loss terms to achieve full knowledge transfer.

4 EXPERIMENTS

4.1 Data Sources

4.1.1 Traditional Financial Data: We acquire historical daily data for stocks and options related to the S&P 500 index using the Yahoo Finance API.

For the purpose of this research, we use **SPY**, the most popular ETF based on the S&P 500 index. We use 30 years of daily close prices, trading volumes, and for the options, implied volatility, strike prices, and expiration dates. The dataset will consist of approximately 7,500 trading days of data. We extract the current Risk free Rate of Return using an API call to Alpha Vantage.

4.1.2 Alternative Data: We initially aimed to scrape the web to collect a high volume of relevant posts. However, given the complexities in collecting the data, we shifted to finding pre curated datasets that reflected our initial dataset choices.

(1) **Geopolitical News:** We use the Economic Policy Uncertainty Index (EPU) [1] to track geopolitical impacts. The EPU tracks the mention of geopolitical events, uncertainties, economic policies and their sentiments in thousands of global, national and local news articles, and normalizes them to provide a score known as the EPU index. A detailed description of their data curation policies can be found on their official website.
(2) **Tech Advancements:** Key advancements in technologies business decisions of companies related to S&P500, as well as general public sentiment is captured through a pre-collected dataset [9] with over 290000 news articles as well as commentaries on S&P or its constituent stocks. We process this data using a pretrained financial sentiment analysis model (e.g., FinBERT) [7] to generate a sentiment score for each topic.

We test the validity of the datasets we have collected, and provide discussions in Section 6.

4.2 Testing Methodology and Benchmarking

The primary method for testing is a rigorous backtesting over the last three months of the collected dataset. This approach simulates methods that are traditionally accepted in the industry. While a rolling window backtest can be more rigorous, we do not consider the approach given that it will be computationally expensive. As such, we benchmark and discuss the performance across the following models and criteria:

- The standalone **Time-Price Module (PINN)** to evaluate the pure theory.
- The standalone **Baseline Feature Modules** to evaluate a pure traditional data-driven approach. We discuss evaluations against both the pure MLP, as well as the Gated LSTM architecture.
- The standalone **Enhanced Feature Module** to evaluate a pure data-driven approach with alternative data. Again, we discuss evaluations across both kinds of feature module architectures.
- **T**he Baseline Feature Module + Time-Price PINN Model. which takes into account only the pure financial data.
- Enhanced Feature model + Time-Price PINN Model that incorporates alternative data

When using the alternative data, we train the model individually with either sources and with both sources as a whole. Success will be measured by comparing the performance of our five benchmark models across a set of well-defined metrics.

Accuracy: We will measure predictive accuracy using Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE).

4.3 Hypothesis and Success Criteria

Our primary hypothesis is that incorporating the theoretical knowledge guided approaches, as well as alternative sources of data, in addition to training a model on traditional data will lead to improved results. Experimentally, we want to show:

(1) The hybrid models which combine the feature and time modules demonstrate statistically significant improvements in accuracy and trend correlation over their pure data-driven counterparts (Standalone Baseline and Enhanced models, respectively).
(2) **T**he feature module with alternative data shows a clear and statistically significant performance improvement over the feature module without it, thereby quantitatively proving the incremental value of incorporating alternative data.

We hope to show that a knowledge driven architecture as well as incorporating alternative data sources can show an improvement over traditional stock and option pricing prediction models.

4.4 Resources

Training the proposed deep learning models, is computationally intensive and requires a high-performance computing workstation equipped with a NVIDIA GPU.

5 RESULTS AND DISCUSSION

The experimental evaluation focused on two architectural improvements to the baseline MLP option pricing model: (1) the gated LSTM architecture with alternative data incorporated into the models and (2) physics-informed training with Black-Scholes PDE constraints. All experiments were conducted using SPY options data fetched from Yahoo Finance, with 3 expiration dates and 3 repeated runs per configuration for statistical validity.

5.1 PINN Pricing Comparison

Figure 2 shows that that the PINN model (blue line) has successfully learned to price an American call option. Its predicted price closely tracks the theoretical Black-Scholes price (red line) and correctly respects the "ground truth" floor of the Intrinsic Value (green line). The plot shows the option's value (Y-axis) at a fixed point in time ($t = 0.12y$, or about 1.5 months to expiry) across a range of possible stock prices (X-axis).

We can split the graph into three key zones based on the Strike Price ($K = \$670$), which is the vertical dashed line.

Out-of-the-Money (Price < \$670): The option has only a small "time value" (or hope value), so all lines are near zero.

At-the-Money (Price = \$670): The option's time value is at its maximum, shown by the largest gap between the PINN price (blue) and the intrinsic value (green).

In-the-Money (Price > \$670): The option's value moves 1-for-1 with the stock, closely following the intrinsic value line.

The most important insight is that the PINN's price (blue) is nearly identical to the Black-Scholes price (red). This is the correct theoretical result, as there's no financial benefit to exercising a call option early in this scenario, proving the model learned the correct financial principle.

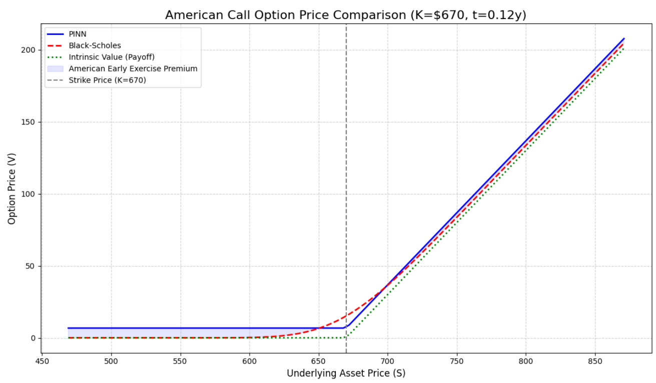


Figure 2: Comparison of performance of the Time Module (PINN) against the theoretical Black Scholes Pricing and the Intrinsic (True) Value.

5.2 Feature Module Evaluation

5.2.1 Experimental Setup. The gated LSTM architecture was evaluated against a baseline MLP across four data configurations based on SPY options data from Yahoo Finance: `yf_only` (base features r, K/S, Maturity, IV, conditional volatility), `epu` (base + Economic Policy Uncertainty index), `sentiment` (base + news sentiment scores), and `both` (base + EPU + sentiment). Each configuration was trained for 100 epochs with batch size 32 using Adam; the gated model used hidden width 64 and 4 layers.

Table 1: RMSE comparison for the baseline and enhanced feature modules by data configuration.

Config	MLP	Gated	Abs. Δ	Rel. Δ
yf_only	0.01020	0.00714	0.00306	30.0%
epu	0.01165	0.00801	0.00364	31.3%
sentiment	0.01348	0.00725	0.00623	46.2%
both	0.01220	0.01143	0.00077	6.3%

5.2.2 Aggregated Performance Results. The gated architecture improves RMSE across all configurations, with the largest gain in the sentiment setup (46.2% reduction vs. MLP) and substantial gains for EPU (31.3%) and base features (30.0%). The smallest benefit appears in the combined configuration (6.3%), suggesting diminishing returns and potential overfitting when multiple heterogeneous sources are merged.

Table 2: Variance reduction with gated architecture.

Config	MLP Std. Dev.	Gated Std. Dev.
yf_only	0.00405	0.00366
epu	0.00863	0.00530
sentiment	0.00489	0.00249
both	0.00435	0.00556

The gated architecture generally reduces run-to-run variability, especially for EPU and sentiment configurations, which show variance ratios of 0.61× and 0.51× respectively. In contrast, the combined configuration exhibits increased variance (1.28×), again highlighting instability when multiple external sources are merged.

5.2.3 Interpretation. Overall, the gated residual architecture delivers consistent RMSE gains (6–46%) over the baseline MLP across all configurations. The gains are most pronounced when integrating a single external data source, supporting the view that gating helps adaptively weight heterogeneous features; when multiple sources are combined, the benefits shrink and variance increases.

5.3 Physics-Informed Feature Module Training Evaluation

5.3.1 Experimental Setup. Physics-informed training augments the data loss with a Black–Scholes PDE residual term evaluated at physics collocation points. The total loss is

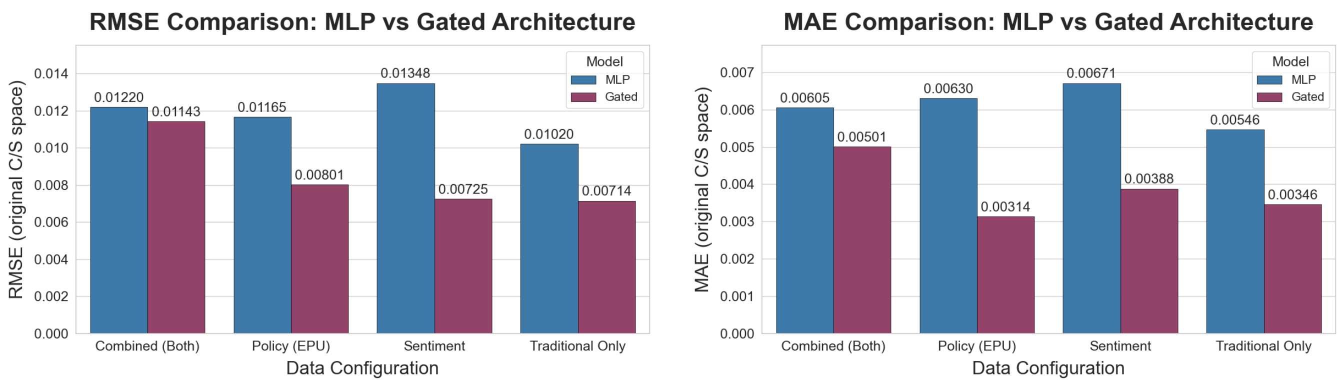
$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{data}} + \lambda \, \mathcal{L}_{\text{physics}},$$

where λ controls the strength of the physics regularization and $\mathcal{L}_{\text{physics}}$ is the mean squared PDE residual over 1000 collocation points per batch; experiments varied $\lambda \in \{0.0, 0.1, 1.0, 5.0, 10.0\}$ on the `yf_only` configuration with the gated architecture.

5.3.2 Physics Weight Sensitivity Results. The best RMSE occurs at a small physics weight ($\lambda = 0.1$), but the improvement over the baseline is marginal (1.3%). Larger physics weights ($\lambda \geq 1.0$) degrade RMSE to a certain level indicating tension between fitting empirical prices and strictly enforcing the Black–Scholes PDE. A summary of the results can be found in Table 3

5.3.3 Physics-Informed Training with External Data. Results vary sharply by configuration: the combined time and feature model with market sentiment data achieves the lowest RMSE and highest R^2, while the ones with only EPU and combined alternative data configurations perform markedly worse, suggesting configuration-dependent interactions between physics constraints and external macro and sentiment features. A summary of the results can be found in Table 4

5.3.4 Architecture Comparison Under Physics Constraints. Even with physics constraints, the feature module with the gated architecture outperforms the MLP, achieving roughly 31% lower RMSE on the same data, which shows that architectural improvements and physics regularization are complementary, with the former delivering the larger gains. The summary of data can be found in table 5



(a) RMSE by Data Source **(b) MAE by Data Source**

Figure 3: Performance comparison of the various Feature Module architectures with no alternative data sources, only Economic Policy Uncertainty Index (EPU), only social media tech sentiment, and both alternative data sources. Figure (a) compares Root Mean Square Error (RMSE) by data source, and figure (b) compares Mean Absolute Error (MAE) by data source.

Table 3: Effect of physics weight on performance of Hybrid model. (traditional data)

Physics Weight	RMSE (orig. space)	R-squared
0.0 (baseline)	0.00630	0.9939
0.1	0.00622	0.9947
1.0	0.00729	0.9942
5.0	0.00683	0.9922
10.0	0.00634	0.9920

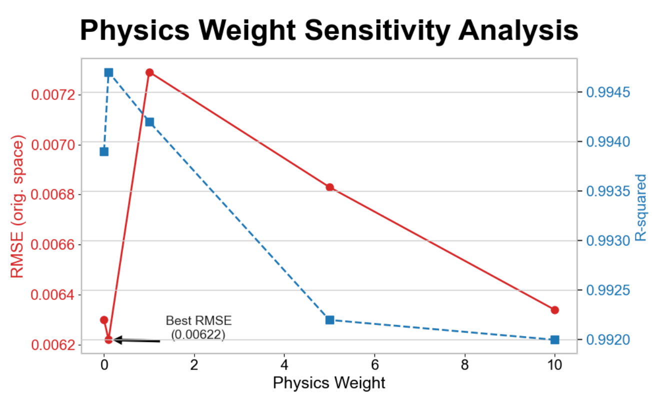


Figure 4: Effect of physics weight on performance of hybrid model. We find that a small physics weight has the best performance.

5.3.5 Interpretation. The physics-informed experiments show limited accuracy benefit in this well-sampled SPY options setting: a small physics weight can slightly improve RMSE, but strong enforcement of the PDE harms fit, consistent with real markets deviating from idealized Black–Scholes assumptions. These findings support using physics constraints cautiously and potentially reserving higher weights for low-data or extrapolative regimes rather than dense, in-sample pricing tasks.

Table 4: Physics-informed training across data sources, including Economic Policy Uncertainty Index (EPU) and social media tech sentiment.

Config	RMSE (orig. space)	Feature Count	R-squared
yf_only	0.00622	5	0.9947
epu	0.01718	6	0.9930
sentiment	0.00566	8	0.9967
both	0.01846	9	0.9918

Table 5: Architecture comparison with physics loss (yf_only).

Model	RMSE (orig. space)	R-squared
MLP	0.00987	0.9862
Gated	0.00622	0.9947

5.4 Overall Results

We compare the best data driven model (feature model), the Gated LSTM model with the Hybrid model which combines the PINN along with the gated LSTM model. As such we provide a comparison of the RMSE scores across various data sources, and summarize these results, along with the details of the architectures and data sources in Table 6.

Our analysis shows that the data-driven and hybrid approaches consistently outperform the analytical Black Scholes baseline. The Hybrid Gated LSTM + PINN model configured with Traditional and Sentiment data achieves the global minimum RMSE of **0.00566**,

Table 6: Comparison of Model Performance and Data Configurations. We compare the hybrid model with the best feature model (data driven), as well as the analytical solution.

Model	Model Description	Data Description	Data Config	Feature Count	RMSE (orig. space)
Black Scholes	Analytical	Traditional	yf_only	5	0.01667
Gated LSTM Only	Feature (data driven)	Traditional	yf_only	5	0.00714
Gated LSTM Only	Feature (data driven)	Traditional + Policy	epu	6	0.00801
Gated LSTM Only	Feature (data driven)	Traditional + Sentiment	sentiment	8	0.00720
Gated LSTM Only	Feature (data driven)	all	all	9	0.01143
Gated LSTM + PINN	Hybrid	Traditional	yf_only	5	0.00622
Gated LSTM + PINN	Hybrid	Traditional+Policy	epu	6	0.01718
Gated LSTM + PINN	**Hybrid**	**Traditional+Sentiment**	**sentiment**	**8**	**0.00566**
Gated LSTM + PINN	Hybrid	all	all	9	0.01846

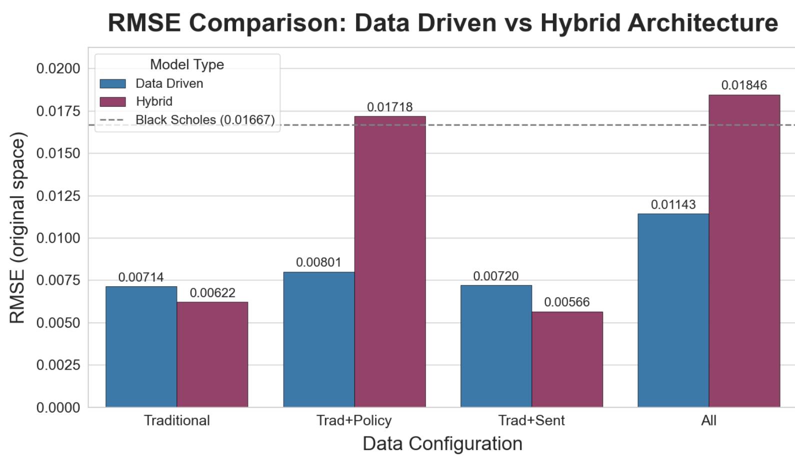


Figure 5: Performance comparison of the Data driven models with the Hybrid Models. We compare different data configurations, including with no alternative data sources, only Economic Policy Uncertainty Index (EPU), only social media tech sentiment, and both alternative data sources. We overlay the analytical Black Scholes RMSE for reference. We observe that the Hybrid model with just sentiment as a alternative signal outperforms other data configurations.

demonstrating the efficacy of physically informed regularization when combined with high-signal alternative data. However, we observe that increasing feature dimensionality does not strictly correlate with performance gains; the inclusion of economic policy uncertainty (EPU) or the aggregation of all features resulted in degraded performance for the hybrid model compared to the sentiment-augmented configuration.

5.5 Summary of Findings

The gated residual architecture consistently reduces RMSE, especially when adding a single external data source such as sentiment or EPU. Physics-informed training offers marginal benefits in this data-rich setting, however, when we combine the Physics model with the feature model, we see a slightly improved performance over the original time series model. Architectural choices remain the dominant driver of performance in the experiments. We summarize the key research questions and their findings below:

Do the hybrid models which combine the feature and time modules demonstrate statistically significant improvements over their data driven counterparts. **Answer: Yes.** When we compare the different models in Table 6 , we observe an improvement in the metrics for both the models. This validates the approach of combining physics and data driven approaches.

Does alternative data play an important role in improving the results. **Answer: Somewhat.** The best model we observe is the combined Time and Feature module (gated LSTM architecture) with the use of market sentiment data as the alternative data source. However EPU did not give as good of results, and the model with both sources of alternative data is not great either.

Is multiple alternative data sources better, as compared to a single alternate source. **Answer: No.** Consistent experiments across just the feature modules as well as combined modules showed that using a single alternative data source helped in better results. Furthermore, we notice that a single alternative data source leads to lower variance, as compared to multiple alternate data sources. We postulate that this is due to the competing interactions when having multiple data sources.

5.6 Limitations and Future Work

The current evaluation focuses on SPY options over a single time period, so robustness across assets and regimes remains untested. The physics loss is computed in a normalized feature space rather than raw (t, S) coordinates, and training variance — particularly for combined external features — suggests further work on regularization, ensembling, and out-of-sample stress testing.

Furthermore, our study is limited by the short backtesting window of three months. While this simulates recent market conditions, it does not capture the model's performance during major financial crashes or prolonged periods of low volatility. Additionally, the computational cost of training Physics-Informed Neural Networks is significantly higher than analytical models, requiring GPU acceleration for feasible training times.

A critical limitation observed in our results is the degradation of performance when combining multiple alternative data sources (EPU and Sentiment together). This suggests that our current concatenation strategy may introduce noise or conflicting signals. Future work should explore more sophisticated fusion mechanisms, such as attention-based networks, to dynamically weigh the importance of conflicting alternative data signals.

We also propose developing an adaptive lambda scheme to dynamically balance the influence of Black-Scholes PDE loss. An adaptive approach could adjust lambda throughout training based on model performance, uncertainty, or market shifts. For example, lambda could be automatically increased when market data becomes highly volatile, or decreased when data is reliable. This could be implemented by periodically evaluating predictive accuracy and PDE residuals, and updating lambda to minimize out-of-sample error. Reinforcement learning or meta-learning techniques could also be used to learn the optimal lambda schedule.

Finally, while we focused on the liquid S&P 500 market, future research should test this hybrid framework on illiquid assets (e.g., real estate, private equity, fine art). Data is scarce for these types of assets, so the regularization benefits of the PINN architecture may be more pronounced.

6 CONCLUSION

This project proposed and evaluated a novel hybrid framework for option valuation that integrates Physics-Informed Neural Networks (PINNs) with data driven approaches and alternative data sources, inspired by the EINN knowledge transfer methodology. By combining the theoretical rigor of the Black-Scholes equation with the flexibility of data-driven deep learning, we aimed to capture market nuances that traditional models miss.

Our experiments yield three key conclusions:

- **Hybrid Models outperform Data Driven Models:** The models which transfer the knowledge of physics constraints to data driven approaches outperform the pure data driven models.
- **Architecture of Feature Module Matters:** The Gated LSTM architecture proved superior to standard MLPs, acting as a robust feature module. It consistently reduced RMSE across all data configurations, with the most significant improvement observed when integrating market sentiment data.
- **The Value of Alternative Data:** We demonstrated that alternative data provides a tangible information edge. Specifically, incorporating news sentiment scores significantly outperformed models relying solely on financial metrics. However, "more is not always better"; combining multiple alternative sources (Sentiment + EPU) led to overfitting and increased variance, highlighting the need for careful feature selection and integration strategies.
- **Role of Physics Constraints:** The PINN successfully learned the latent dynamics of the Black-Scholes PDE and enforced American option constraints. While strict physics enforcement ($\lambda \geq 1.0$) slightly degraded empirical fit in this data-rich environment, a light physics regularization ($\lambda = 0.1$) provided stability. This suggests that theoretical priors act best as a "soft" guide rather than a hard constraint in liquid markets.

In summary, this work validates that a "grey box" approach—fusing financial theory with alternative data streams can enhance pricing accuracy. While challenges remain in multi-source data fusion, the framework offers a promising path for investors seeking to leverage unstructured data without abandoning established financial principles.

CODE AVAILABILITY

The project website, including documentation and visualizations, can be found here: Project Website. [3]

The source code is available on GitHub: CSE-598 AI for Science Repositories. [4]

REFERENCES

[1] [n. d.]. https://www.policyuncertainty.com/us_monthly.html

[2] Amine M. Aboussalah, Xuanze Li, Cheng Chi, and Raj Patel. 2024. The AI Black-Scholes: Finance-Informed Neural Network. arXiv:2412.12213 [cs.LG] https://arxiv.org/abs/2412.12213

[3] Navnoor Bawa. 2024. American Options PINN Pricer: Physics-Informed Neural Networks for S&P 500 Derivatives. Available at: https://medium.com/@navnoorbawa/american-options-pinn-pricer-physics-informed-neural-networks-for-s-p-500-derivatives-real-30038bac3672.

[4] Fischer Black and Myron Scholes. 1973. The Pricing of Options and Corporate Liabilities. *Journal of Political Economy* 81, 3 (1973), 637–654. https://doi.org/10.1086/260062

[5] Zakaria Elbayed and Abdelmjid Qadi El Idrissi. 2025. A Hybrid PINNs-FNO Approach for Advanced Option Pricing Models. *Preprints* (January 2025). https://doi.org/10.20944/preprints202501.0629.v1

[6] Owen Kraaijeveld and Johan De Smedt. 2020. The Predictive Power of Public Twitter Sentiment for Forecasting Cryptocurrency Prices. *Journal of International Financial Markets, Institutions and Money* 65 (2020), 101188. https://doi.org/10.1016/j.intfin.2020.101188

[7] Zhuang Liu, Degen Huang, Kaiyu Huang, Zhuang Li, and Jun Zhao. 2020. FinBERT: A Pre-trained Financial Language Representation Model for Financial Text Mining. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20*, Christian Bessiere (Ed.). International Joint Conferences on Artificial Intelligence Organization, 4513–4519. https://doi.org/10.24963/ijcai.2020/622 Special Track on AI in FinTech.

[8] Rita Pimentel, Morten Risstad, Sondre Rogde, Erlend S. Rygg, Jacob Vinje, Sjur Westgaard, and Cassandra Wu. 2025. Option pricing with deep learning: a long short-term memory approach. *Decisions in Economics and Finance* (2025). https://doi.org/10.1007/s10203-025-00518-9

[9] pmoe7. [n. d.]. PMOE7/stock-market-data: Contains various data for the stock market and stocks including news, Price and Fundamentals (ratios). https://github.com/pmoe7/Stock-Market-Data

[10] M. Raissi, P. Perdikaris, and G.E. Karniadakis. 2019. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems

[3]Project Website: https://cse-598-ai-for-science.github.io/project-website/

[4]GitHub: https://github.com/orgs/cse-598-ai-for-science/repositories

involving nonlinear partial differential equations. *J. Comput. Phys.* 378 (2019), 686–707. https://doi.org/10.1016/j.jcp.2018.10.045

[11] Alexander Rodríguez, Jiaming Cui, Naren Ramakrishnan, Bijaya Adhikari, and B. Aditya Prakash. 2023. EINNs: epidemiologically-informed neural networks. In *Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence and Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence and Thirteenth Symposium on Educational Advances in Artificial Intelligence (AAAI'23/IAAI'23/EAAI'23)*. AAAI Press, Article 1621, 8 pages. https://doi.org/10.1609/aaai.v37i12.26690

[12] Yunchuan Sun, Lu Liu, Ying Xu, Xiaoping Zeng, Yufeng Shi, Haifeng Hu, Jie Jiang, and Ajith Abraham. 2024. Alternative data in finance and business: emerging applications and theory analysis (review). *Financial Innovation* 10, 1 (2024), 127. https://doi.org/10.1186/s40854-024-00652-0

A VALIDITY OF ALTERNATIVE DATA

A.1 Sentiment Analysis of Market Developments

The sentiment analysis scores of tech developments directly reflect the people's willingness to buy or sell stocks, which automatically reflect into the pricing of a stock. Hence the market sentiment is a valid source of "alternative data" to consider.

A.2 Economic Policy Uncertainty of Geopolitical Events

One of the hypothesis we hold is that geopolitical events and government policies may lead to economic uncertainty which could affect the volatility of the stock prices. We examine this by computing and comparing the rolling 30 day averages of market volatility as well as that of the EPU. Figure 6 shows a time series comparison of the EPU and stock volatility. We observed an average of 52% correlation between the EPU and volatility in the last 10 years. This figure increases to 67% since the start of 2025. A high EPU is generally an indicator of high volatility of the stock, and can be indicative in pricing the stock.

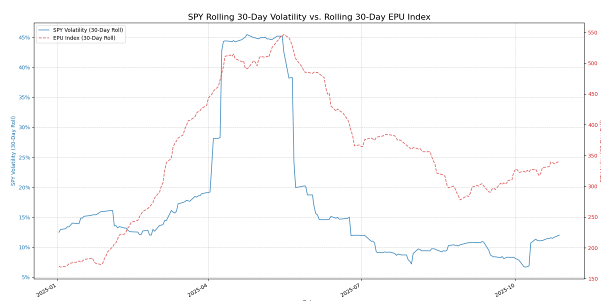


Figure 6: A comparison plot of the EPU (red) vs volatility of S&P 500 (Blue) since the start of 2025. A moderately strong correlation is observed between EPU and Volatility.