



Entropy-Assisted Quality Pattern Identification in Finance

Rishabh Gupta ^{1,†} , Shivam Gupta ^{2,†}, Jaskirat Singh ³  and Sabre Kais ^{1,4,*}

¹ Department of Chemistry, Purdue University, West Lafayette, IN 47907, USA; rishabhgupta2806@gmail.com

² EntropyX Labs Pvt. Ltd., Ghaziabad 201010, Uttar Pradesh, India; shivam@entropyxlabs.com

³ Softure Solutions Pvt. Ltd., New Delhi 110059, Delhi, India

⁴ Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, NC 27606, USA

* Correspondence: skais@ncsu.edu

† These authors contributed equally to this work.

Abstract: Short-term patterns in financial time series form the cornerstone of many algorithmic trading strategies, yet extracting these patterns reliably from noisy market data remains a formidable challenge. In this paper, we propose an entropy-assisted framework for identifying high-quality, non-overlapping patterns that exhibit consistent behavior over time. We ground our approach in the premise that historical patterns, when accurately clustered and pruned, can yield substantial predictive power for short-term price movements. To achieve this, we incorporate an entropy-based measure as a proxy for information gain: patterns that lead to high one-sided movements in historical data yet retain low local entropy are more “informative” in signaling future market direction. Compared to conventional clustering techniques such as K-means and Gaussian Mixture Models (GMMs), which often yield biased or unbalanced groupings, our approach emphasizes balance over a forced visual boundary, ensuring that quality patterns are not lost due to over-segmentation. By emphasizing both predictive purity (low local entropy) and historical profitability, our method achieves a balanced representation of Buy and Sell patterns, making it better suited for short-term algorithmic trading strategies. This paper offers an in-depth illustration of our entropy-assisted framework through two case studies on Gold vs. USD and GBPUSD. While these examples demonstrate the method’s potential for extracting high-quality patterns, they do not constitute an exhaustive survey of all possible asset classes.



Academic Editor: José María Amigó

Received: 10 March 2025

Revised: 9 April 2025

Accepted: 10 April 2025

Published: 16 April 2025

Citation: Gupta, R.; Gupta, S.; Singh, J.; Kais, S. Entropy-Assisted Quality Pattern Identification in Finance. *Entropy* **2025**, *27*, 430. <https://doi.org/10.3390/e27040430>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: entropy; pattern identification; algorithmic trading; finance

1. Introduction

Algorithmic trading now constitutes a major share of global market volume across equities, foreign exchange, and futures. While sub-second high-frequency trading has drawn considerable attention, many short-term and intraday models operate at minute- or hour-level intervals. As the electronic trading infrastructure continues to expand, robust, data-driven methodologies for filtering out noise and detecting recurring price patterns remain essential for practitioners [1–3]. In modern finance, the ability to recognize recurring short-term patterns has become increasingly crucial for designing and executing algorithmic trading strategies. From high-frequency trading desks to retail investors applying swing-trading techniques, the recurring assumption is that historical price behavior repeats over time, and the profitability of these techniques is directly proportional to the quality of these patterns. In fact, substantial volumes of historical data are routinely collected and mined to find these subtle yet exploitable patterns. However, the sheer

level of noise in the price series, driven by complex microstructure effects of the market and exogenous shocks, poses a constant obstacle to reliably distilling true signals from ephemeral artifacts. When attempting to exploit these signals in a systematic fashion, it becomes apparent that the success of any algorithmic model hinges on the quality of the underlying patterns.

A promising way to fortify pattern identification in such noisy contexts is to incorporate entropy as a measure of the information content. Entropy is a versatile concept that has found applications in a wide range of fields, from information theory and statistical mechanics [4,5] to biology [6] and economics [7,8]. In finance, entropy is increasingly employed to quantify uncertainty in market behavior, assess risk, and enhance the robustness of trading models [9]. Entropy, as defined in information theory, quantifies uncertainty or randomness within a probability distribution. For example, the Shannon entropy [10–12] is defined as

$$H = - \sum_{i=1}^N p_i \ln p_i, \quad (1)$$

where p_i represents the probability of the occurrence of a particular event (or outcome). In the context of financial data, each 'event' could correspond to the future price movement following a specific short-term pattern in the data. When we apply this concept locally, looking at segments or groups within the historical feature space, we can measure how "pure" or consistent the outcomes are for similar patterns. A "pure" neighborhood is one where the outcomes are highly concentrated; for example, if nearly all occurrences of a given pattern lead to a large upward move, the local probability distribution is skewed toward that outcome, and the resulting entropy is low. Thus, a low entropy value indicates that there is less uncertainty about what will happen next. Conversely, if a pattern is found in a region where the outcomes are evenly split between upward and downward moves, the entropy is high, signaling greater uncertainty and suggesting that the pattern is less informative.

This idea of leveraging entropy to derive inference in financial markets is not new [13–15]. In our previous work, the EC-GBM (Entropy-Corrected Geometric Brownian Motion) [16], we demonstrated that by incorporating an entropy constraint into the model, one can effectively narrow down the forecast trajectories to those that reduce the overall uncertainty of the system. In the EC-GBM method, the predicted trajectories by the standard GBM [17–20] are appended to the historical distribution, and the resulting change in entropy is computed. If a trajectory causes a significant drop in entropy relative to the reference state, it means that the trajectory improves the dominant features of the underlying distribution, effectively 'sharpening' the prediction by reducing uncertainty. The selected trajectories are then considered more reliable because they reflect a higher information gain or, equivalently, a more deterministic evolution of the price of the underlying asset.

By integrating this entropy-based filtering mechanism into our pattern identification process, we not only eliminate overlapping or contradictory patterns but also preserve those that provide a clear, consistent signal. In our work, lower entropy is synonymous with higher informational content: patterns in low-entropy regions imply that the historical data exhibit a strong, unambiguous trend, which can be harnessed to improve the predictive power of algorithmic trading models. This stands in contrast to purely machine learning-based clustering methods, which could group patterns based solely on distance metrics without assessing their predictive clarity. In noisy financial markets, where overfitting and misclassification are constant threats, using entropy as an additional criterion ensures that our pattern selection process remains robust and focused on genuinely informative structures in the data. By maintaining only patterns that simultaneously

exhibit low entropy and demonstrable historical profitability, we effectively increase the signal-to-noise ratio.

In summary, the use of entropy in our method enables us to quantify and enhance the quality of pattern identification. It acts as a natural filter that retains only those patterns that not only match well in a geometric sense but also carry a high degree of predictive information, much like the entropy reduction principle demonstrated in EC-GBM for filtering out less relevant forecast trajectories. In this paper, we detail an end-to-end pipeline for entropy-assisted quality pattern identification and discuss how short-term trading strategies can profit from explicitly integrating entropy measures. In this study, our ‘short-term trading patterns’ refer to short (e.g., a four-hour or eight 30 min) segments of historical OHLC (open, high, low, and close) data that meet specific threshold criteria for subsequent price movements (e.g., a ± 15 point move in gold). These patterns are not fixed or ‘pre-imposed’ by the authors; rather, they are automatically extracted from the historical record whenever a short-term price swing beyond a chosen threshold is observed. The extracted segments are then labeled ‘Buy’ or ‘Sell’ based on the predominant direction of the ensuing price movement. This approach can be viewed as discovering typical short-term momentum patterns directly from the data, as opposed to using canonical technical chart formations. In practice, this produces a large, diverse set of potential entry signals that may or may not resemble well-known trading motifs (like breakouts or retracements) but that are grounded in the actual price history. We also compare this approach to standard clustering-based pattern detection, highlighting the role entropy plays in mitigating overfitting in volatile environments. Ultimately, we show that the focus of our method on non-overlapping high-information-gain patterns can lead to more reliable forecasts and improved performance in real-time trading.

2. Methodology

The core objective of this work is to transform a large and noisy set of labeled short-term trading patterns into two coherent clusters, one for Buy and one for Sell, such that no pattern in the Buy cluster overlaps with any pattern in the Sell cluster. As illustrated in Figure 1, raw patterns often exhibit significant overlap or near-duplicates across conflicting labels, undermining their practical value in high-volatility trading. We begin by collecting high-resolution OHLC data for a target asset (for example, Gold vs. USD) and identifying time segments that precede significant price swings (e.g., ± 15 points within the subsequent two hours). Each extracted pattern is assigned a label (Buy or Sell) based on the direction of the ensuing market movement. This process results in a large pool of raw patterns, often numbering in the thousands, where many patterns overlap or nearly duplicate, even across opposing directions, which poses a challenge for robust signal extraction in algorithmic trading.

To address this challenge, we propose an entropy-assisted filtering framework that operates in two main stages. First, we evaluate each pattern using a dual scoring system that combines a measure of local entropy with a historical profitability metric (PnL). The local entropy is computed by analyzing the immediate neighborhood of a given pattern in the high-dimensional feature space, derived from indicators such as H-L, C-O, H-O, and O-L over the pattern window. If the majority of neighboring patterns share the same label, the local entropy is low, indicating a “pure” or unambiguous signal. Conversely, a high entropy value implies a mixed neighborhood, where the pattern’s outcome is less predictable. Simultaneously, we compute a PnL metric for each pattern, reflecting the average or maximum profit that could have been historically realized if a trade had been initiated when that pattern appeared. By normalizing the PnL values and combining them with the information gain (defined as the global entropy minus the local entropy) using a

weighted sum, we obtain a final score for each pattern. This scoring mechanism ensures that the most valuable patterns are those that not only have high predictive certainty (low entropy) but also have demonstrated historical profitability.

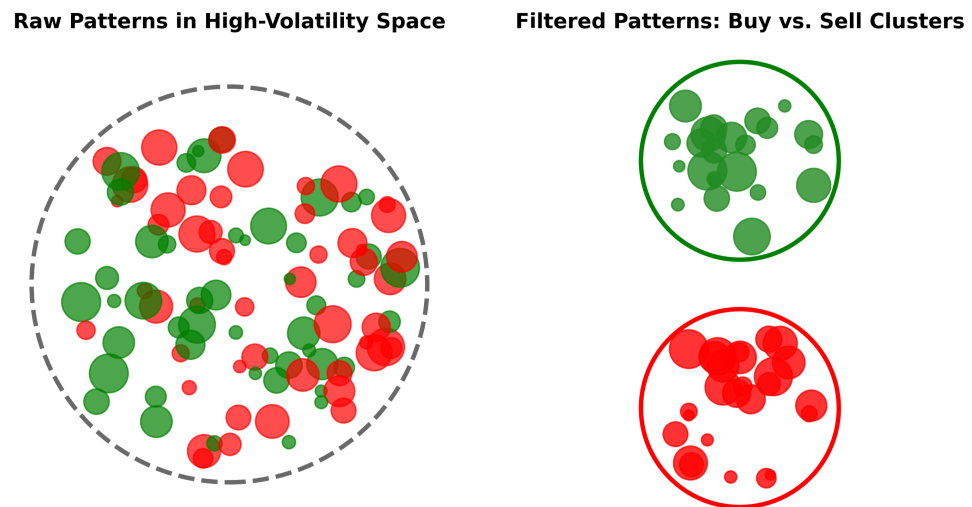


Figure 1. The left panel illustrates raw trading patterns in a high-volatility space, with each circle's color indicating Buy (green) or Sell (red) and its size reflecting local entropy (smaller circles denote lower entropy and more consistent outcomes). These raw patterns exhibit extensive overlap and near-duplicates across conflicting labels. In contrast, the right panel shows two non-overlapping clusters following entropy-based filtering, underscoring the method's ability to isolate high-quality, non-ambiguous signals by removing contradictory patterns.

In the second stage, we filter the raw patterns using a distance-based overlap criterion. Specifically, we define a distance threshold (based on the L1 (Manhattan) norm in the feature space) such that if a Buy pattern and a Sell pattern are found to be closer than this threshold, they are deemed to be overlapping and contradictory. In these cases, only the pattern with the higher combined score is retained. The end result is two refined sets—denoted B' for Buy and S' for Sell—that are non-overlapping across clusters may yet exhibit some controlled overlap within each cluster to capture the natural variation in similar trading scenarios. The core workflow of our entropy-assisted filtering method is presented in Algorithm 1.

In a real-time setting, once the 'quality' pattern library is established, each newly formed short-term pattern in live market data is compared (e.g., via the L1 distance) against these high-quality Buy and Sell templates. Because our filtering process ensures that no Buy pattern overlaps with any Sell pattern in the feature space, conflicting signals are minimized. Specifically, every 30 min, the algorithm gathers the relevant OHLC data for the last eight time segments (covering a 4-h window) and computes the same set of features (H-L, C-O, H-O, O-L, etc.). It then checks each feature vector's L1 distance against the stored Buy and Sell patterns. If the distance to a Buy pattern is below a chosen threshold, the system opens a long trade; similarly, a match to a Sell template triggers a short sale. Each trade is governed by a fixed stop-loss (e.g., 12–18 points below/above the entry) and an associated profit target, ensuring a clear risk–reward structure. Once either the target or the stop-loss is hit, the position is closed, preventing unchecked losses. While the back-testing results suggest robust performance, we note that practical considerations such as commissions, exchange fees, and slippage are not yet incorporated; these factors would reduce net profitability and must be included in any production-grade trading system.

Algorithm 1 Entropy-assisted quality pattern identification.

Require: Raw pattern sets **B** (Buy) and **S** (Sell) extracted from OHLC data, distance threshold θ , weight parameter α , global entropy H_{global}

Ensure: Filtered Buy set **B'** and Sell set **S'** such that no pattern in **B'** overlaps with any pattern in **S'**

- 1: $\mathbf{T} \leftarrow \mathbf{B} \cup \mathbf{S}$
- 2: **for all** pattern $x \in \mathbf{T}$ **do**
- 3: Compute local entropy $H(x)$ based on the neighborhood in feature space
- 4: Set information gain $\text{IG}(x) \leftarrow H_{\text{global}} - H(x)$
- 5: Normalize historical profit/loss: $\text{PnL}_{\text{norm}}(x)$
- 6: Compute combined score: $\text{score}(x) \leftarrow \alpha \cdot \text{IG}(x) + (1 - \alpha) \cdot \text{PnL}_{\text{norm}}(x)$
- 7: **end for**
- 8: Sort **T** in descending order by $\text{score}(x)$
- 9: Initialize $\mathbf{B}' \leftarrow \emptyset, \mathbf{S}' \leftarrow \emptyset$
- 10: **for all** pattern $x \in \mathbf{T}$ in sorted order **do**
- 11: **if** x is labeled **Buy** **then**
- 12: **if** for every pattern $y \in \mathbf{S}', d(x, y) \geq \theta$ **then**
- 13: $\mathbf{B}' \leftarrow \mathbf{B}' \cup \{x\}$
- 14: **end if**
- 15: **else if** x is labeled **Sell** **then**
- 16: **if** for every pattern $y \in \mathbf{B}', d(x, y) \geq \theta$ **then**
- 17: $\mathbf{S}' \leftarrow \mathbf{S}' \cup \{x\}$
- 18: **end if**
- 19: **end if**
- 20: **end for**
- 21: **return** \mathbf{B}', \mathbf{S}'

3. Results

Our experiments were carried out on real-world data for Gold vs. USD, which span from 2017 to 2023, with 2024 reserved exclusively for testing. A “pattern” in our study is defined as an eight-30 min segment of OHLC data that is represented by 32 features—specifically, eight values each for the differences H–L, C–O, H–O, and O–L—supplemented by a profitability (PnL) measure. By focusing on such short-term patterns extracted from raw historical data, we ensure that our methodology is tested under realistic market conditions that inherently include noise and irregularities not captured in the simulated datasets. This realistic setting underscores the strength of our entropy-assisted filtering approach in distilling robust signals from noisy financial time series.

Figure 2 illustrates the effectiveness of our filtering process. The top histogram in Figure 2 shows the distribution of pairwise L1 (Manhattan) distances between over 900 Buy and 1000 Sell raw patterns obtained from historical data that lead to high volatility, revealing a substantial number of near-duplicates and overlapping instances. After applying our entropy-based filtering, which integrates local entropy (to assess pattern purity) and normalized historical profitability (PnL) into a combined score, the dataset is pruned to approximately 500 Buy and 600 Sell patterns. The bottom histogram in Figure 2 demonstrates a notable increase in both the mean and median pairwise distances. This increase confirms that the filtering process effectively removes ambiguous and overlapping patterns, thereby retaining only those patterns that are truly distinct and non-overlapping in the feature space. The results for the filtered patterns in Figure 2 are generated by keeping the alpha value at 0.8, thus giving more weight to the entropy factor compared to the PnL factor showcasing the relevance of entropy in this method. In this context, a ‘quality’ pattern is one that not only exhibits low local entropy, which implies high predictive consistency but also delivers a strong historical PnL signal.

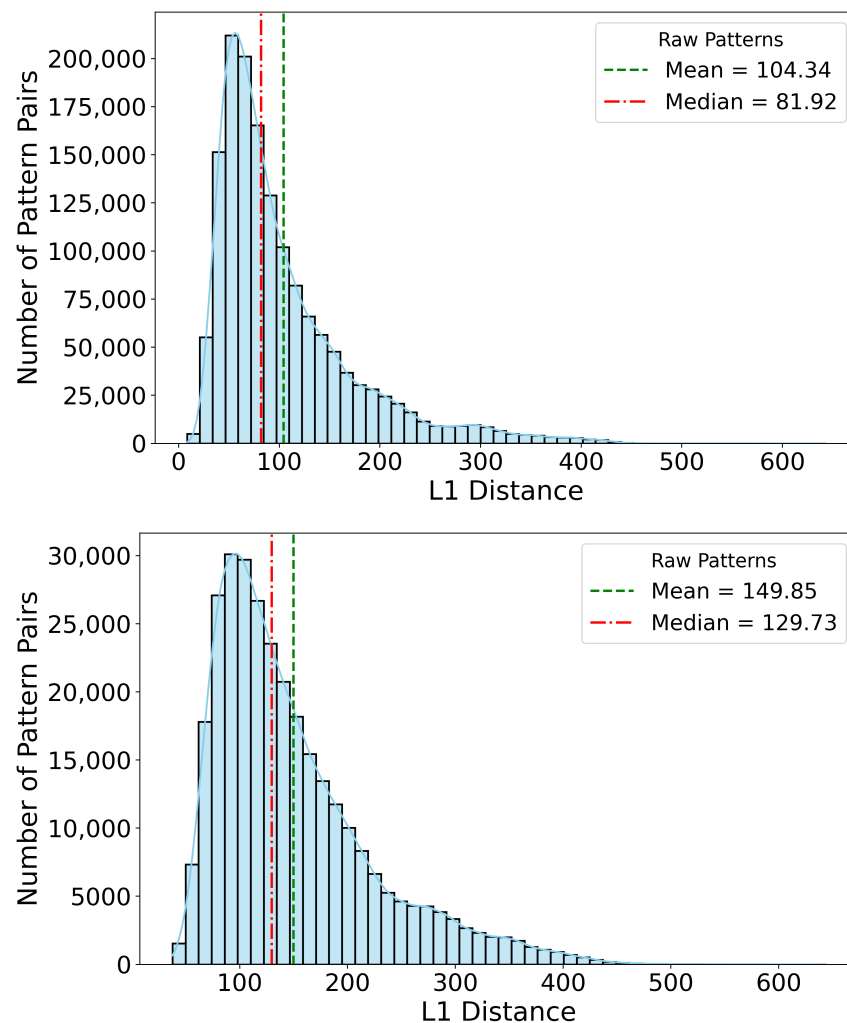


Figure 2. The top histogram shows the distribution of pairwise L1 (Manhattan) distances between 900+ Buy and 1000+ Sell raw patterns. The bottom histogram presents the distances after entropy-based filtering, reducing the dataset to approximately 500 Buy and 600 Sell patterns. The significant reduction in pattern count is accompanied by an increase in the mean and median distance, indicating that the filtering process effectively removes near-duplicates and conflicting patterns, ensuring that the remaining patterns are more distinct and non-overlapping in the feature space.

Figure 3 provides further context by depicting the monthly volatility distribution in gold prices from 2017 to 2024 using box plots of the standard deviation of open prices. Notably, 2024 exhibits the highest mean monthly volatility among all years. This escalation in volatility, particularly after the COVID period, reinforces the importance of short-term trading patterns. In an environment characterized by rapidly changing market sentiments and non-repeating long-term trends, short-term patterns serve as more reliable indicators of immediate market behavior. Their ability to capture transient market sentiments becomes even more critical as volatility increases, making the extraction of high-quality, non-overlapping patterns a vital component of robust algorithmic trading strategies. This point is further emphasized by the success of our algorithm in the real-world trading scenario, as depicted in Figure 4, which is discussed in the following.

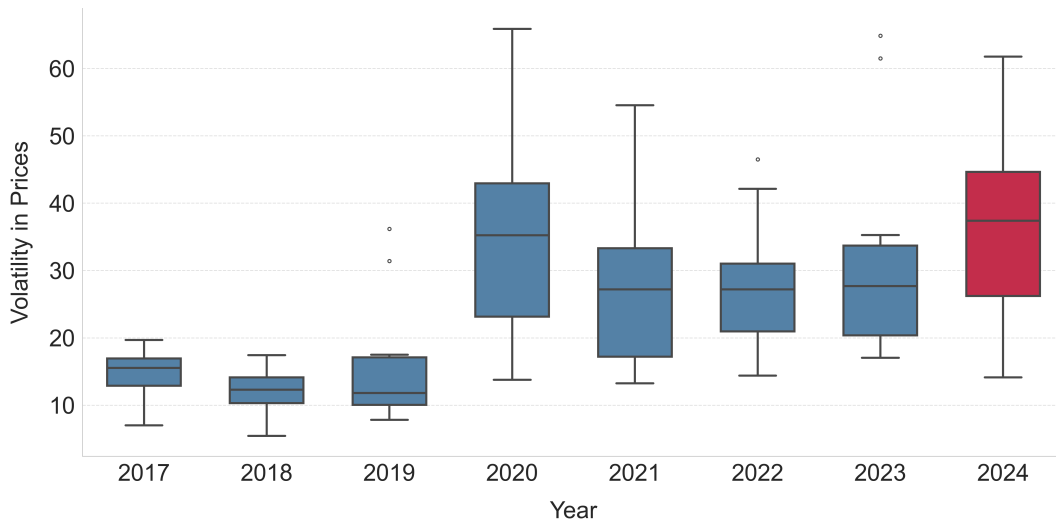


Figure 3. Distribution of monthly volatility in gold prices from 2017 to 2024. The box plots depict the spread of monthly standard deviations of open prices within each year. Notably, the mean monthly volatility for 2024 is the highest among all years, indicating increased market fluctuations in the most recent period.

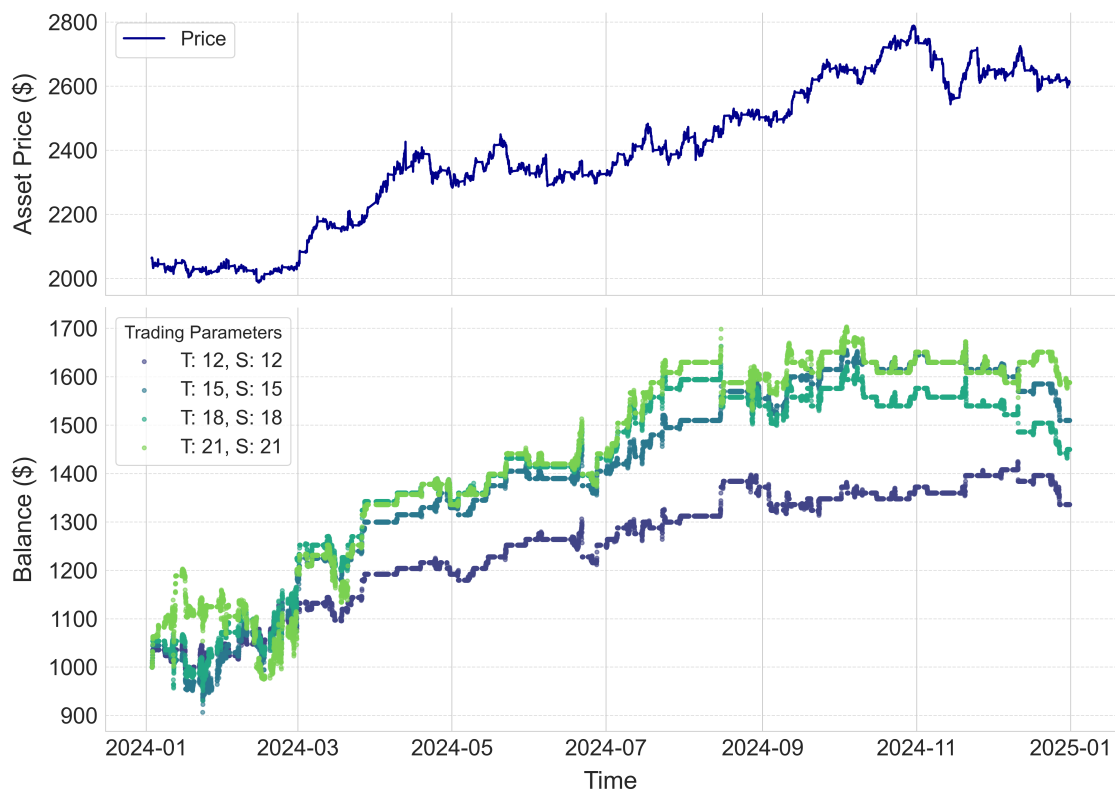


Figure 4. The top subplot shows the asset’s open price over time, while the bottom subplot tracks equity progression for different trading parameters (T: target, S: stop – loss). The model consistently yielded profits across all configurations, demonstrating adaptability to market conditions and the impact of risk–reward trade-offs.

Figure 4 shows the practical performance of our trading strategy when applied to the unseen 2024 data. The upper subplot displays the evolution of the asset price over time, while the lower subplot tracks the progression of the investment under various configurations of the target and stop-loss parameters. Our back-testing framework is designed to trigger trades whenever a pattern match occurs: each order is executed with

predefined target and stop-loss values, a usual norm in algorithmic trading strategies. The model consistently generates profits across different parameter settings, resulting in annual returns ranging from 30% to 60%. These results highlight not only the adaptability of our approach to varying market conditions but also the impact of carefully calibrated risk–reward trade-offs. In a real-time trading scenario, once a pattern match is identified, an order would be placed at the current open price, and the subsequent price action would be continuously monitored until either the target profit or the stop-loss is reached, thereby ensuring disciplined exit strategies and robust performance.

To demonstrate that our entropy-assisted method is not merely reliant on a uniformly bullish market, we apply the same workflow to GBPUSD during 2024 (Figure 5). Unlike gold, which trends predominantly upward, GBPUSD displays a variety of market behaviors—periods of uptrend, downtrend, and range-bound movement—within the same calendar year. As with gold, we use historical OHLC data from 2017 to 2023 for pattern extraction and model tuning, reserving 2024 as an out-of-sample test period. Figure 5 shows the resulting equity progression for GBPUSD, again under various target–stoploss combinations (T, S). Notably, GBPUSD exhibits more frequent short-term price declines than gold and is subject to distinct macroeconomic drivers. Despite these different market conditions, our method continues to generate consistent returns. The annualized gains for GBPUSD in 2024 range from approximately 30% to 70% (depending on the T–S configuration), demonstrating that our approach is not solely reliant on an asset in a steady upward trend. By demonstrating similar profitability patterns on GBPUSD and across multiple test years—each with distinct volatility profiles—the method’s capacity to generalize becomes clearer. While these two assets do not exhaust all market scenarios, they collectively illustrate how the entropy-based pattern-filtering approach adapts to different directional biases and varying levels of volatility, which is crucial for real-world algorithmic trading applications.

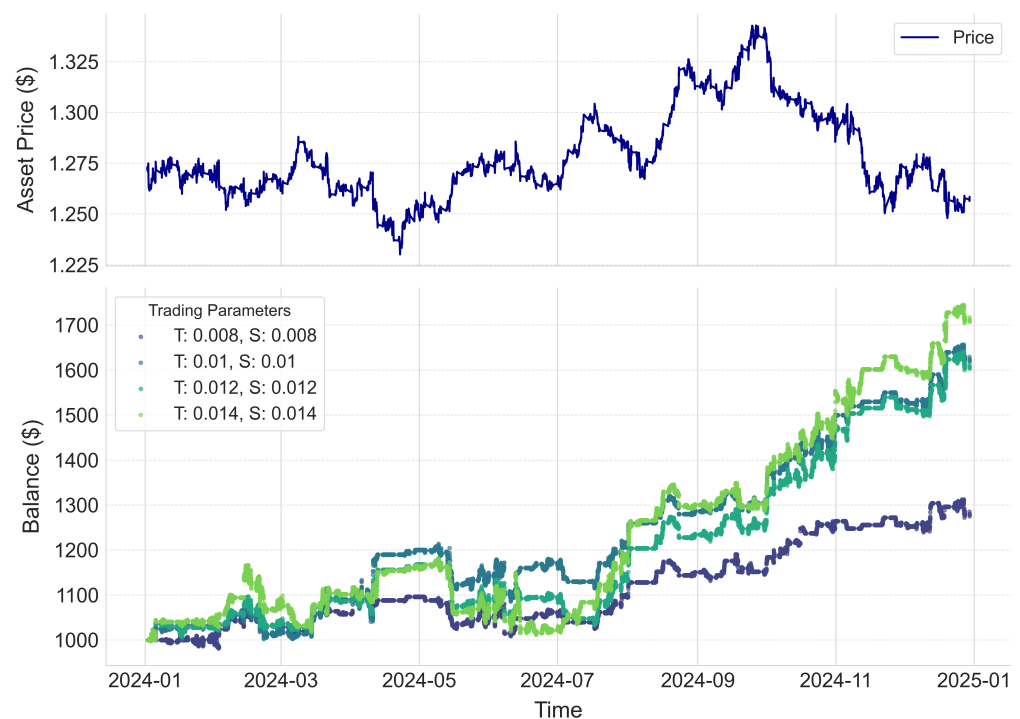


Figure 5. Equity progression of the entropy-assisted trading strategy applied to GBPUSD in 2024 under various target–stop-loss (T–S) configurations. Despite encountering both uptrends and downtrends, as well as range-bound phases, the method maintains robust profitability throughout the year, underscoring its adaptability to different market regimes.

This approach offers several advantages over traditional clustering techniques [21,22] such as K-means [23] or Gaussian Mixture Models (GMMs) [24,25] which are two widely used clustering algorithms in financial market applications such as pattern identification, risk analysis, and anomaly detection. K-Means is a centroid-based algorithm that partitions data into a predetermined number of clusters by iteratively assigning each data point to its nearest centroid and then updating these centroids. However, K-Means assumes that clusters are spherical and well separated, which limits its effectiveness when the data exhibit significant overlap. In contrast, GMM adopts a probabilistic framework by modeling data as a mixture of multiple Gaussian distributions, thereby assigning each data point a probability of membership in each cluster. This soft clustering approach is better suited to financial scenarios where data distributions are complex and overlapping, such as in stock return modeling, risk-based portfolio optimization, and fraud detection. However, these standard clustering methods typically partition the data based solely on geometric distances, often resulting in imbalanced clusters (for instance, an excessively large Buy cluster and a very small Sell cluster) and fail to account for the directional consistency and historical profitability of the patterns. In Figure 6, we visualize the Buy vs. Sell patterns both before and after entropy-assisted filtering using a two-dimensional principal component analysis (PCA) [26–29]. Although the raw patterns appear heavily intermixed, the filtered set retains fewer, more distinctive patterns. Notably, the PCA projection does not exhibit a clear boundary separating Buy and Sell points. This outcome does not imply that the method fails to distinguish between the two sides in the higher-dimensional feature space; rather, it reflects the inherent limitations of projecting complex financial data onto just two principal components. Figure 7, which compares K-means and GMM results, illustrates that conventional clustering methods can indeed generate seemingly more distinct clusters in a 2D projection; however, these algorithms often yield disproportionately large clusters on one side, indicating potential bias, which is evident from the large number of patterns moved to one cluster in both of these clustering methods. Our approach avoids such pitfalls by enforcing non-overlapping Buy and Sell sets without sacrificing a balanced representation of patterns. Consequently, while a crisp PCA boundary may not be visible, the underlying separation in the full feature space is preserved, leading to more reliable and equitable pattern sets for short-term trading strategies.

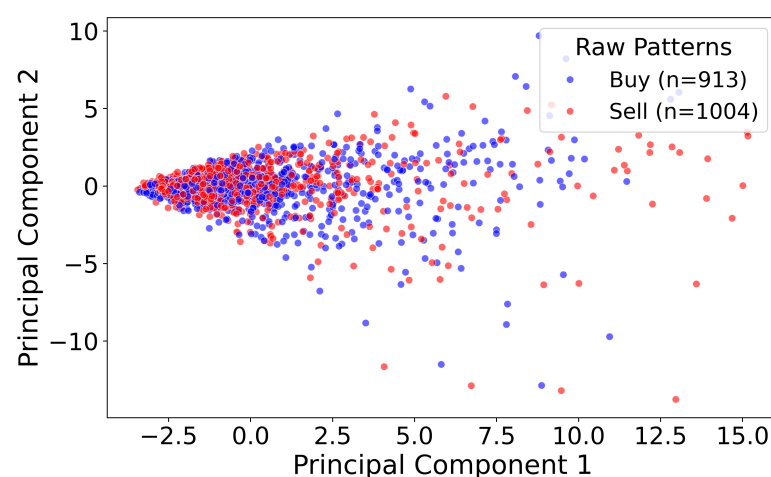


Figure 6. Cont.

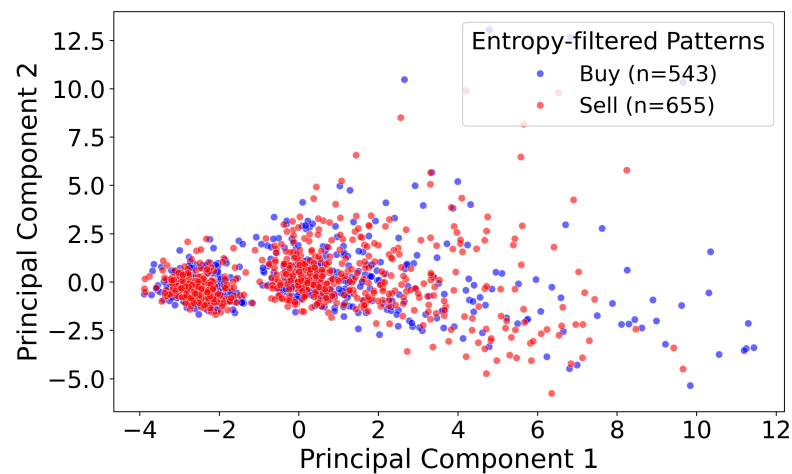


Figure 6. Visualization of Buy (blue) and Sell (red) patterns in a two-dimensional PCA projection. The top subplot depicts the raw dataset (Buy = 913, Sell = 1004), which is highly intermixed in PCA space. The bottom subplot shows the entropy-filtered dataset (Buy = 543, Sell = 655), where ambiguous, overlapping patterns have been pruned. Although there is no crisp boundary in 2D, the filtering preserves a balanced Buy–Sell ratio similar to the raw data and removes contradictory patterns in the higher-dimensional feature space.

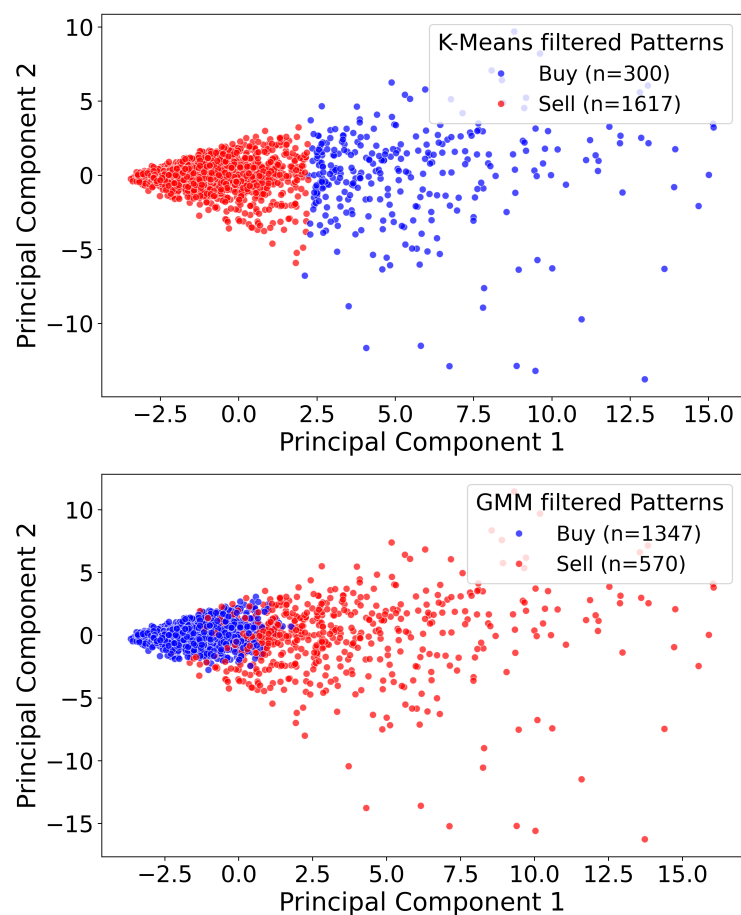


Figure 7. Comparison of patterns produced by two standard clustering methods in a two-dimensional PCA projection. The top subplot illustrates K-means-filtered patterns (Buy = 300, Sell = 1617), and the bottom subplot shows GMM-filtered patterns (Buy = 1347, Sell = 570). While both algorithms yield visually distinct clusters, they produce heavily skewed Buy–Sell partitions. In contrast, the entropy-assisted approach (Figure 6) maintains a more balanced ratio of Buy vs. Sell patterns by explicitly accounting for historical profitability and directional consistency rather than relying solely on geometric proximity.

The entropy-assisted method inherently prioritizes low-entropy patterns, i.e., those with a high level of outcome consistency, while also incorporating profitability metrics, thus producing a more balanced and informative set of signals. Furthermore, by explicitly removing overlapping patterns between Buy and Sell, our method avoids the confusion that can arise when ambiguous patterns are assigned to both clusters, leading to more robust and interpretable trading signals. Together, these results demonstrate that our entropy-assisted method not only yields a more balanced and high-quality pattern set compared to conventional clustering approaches (which tend to produce skewed clusters) but also translates into tangible trading performance improvements in a real, high-volatility market environment.

4. Conclusions

In this paper, we present an entropy-assisted framework that systematically transforms a raw, noisy set of short-term trading patterns into two high-quality, non-overlapping clusters representing Buy and Sell signals. Our approach begins by extracting short-term patterns from high-resolution OHLC data, where each pattern—defined over a four-hour window—is represented by 32 engineered features along with an associated profit/loss (PnL) measure. By quantifying the local entropy of each pattern, we assess its predictive purity: patterns with low local entropy consistently lead to one directional move, while those with high entropy are ambiguous and less reliable. When combined with normalized PnL, this dual scoring mechanism enables us to retain only those patterns that are historically profitable and directionally consistent.

The effectiveness of our methodology is evidenced by the substantial reduction in the total number of patterns, from thousands of raw overlapping signals to a balanced set of approximately 500 Buy and 600 Sell patterns, while simultaneously increasing the average pairwise distance between patterns across clusters. This indicates that our filtering process effectively removes near-duplicates and conflicting signals. Unlike conventional clustering methods such as k-means or Gaussian Mixture Models, which often yield imbalanced or biased clusters due to their reliance on geometric proximity alone, our entropy-based approach emphasizes a balanced and interpretable representation of market signals.

Overall, our methodology provides a systematic, quantitative framework for transforming a raw, noisy set of short-term trading patterns into two high-quality, non-overlapping clusters that are both predictive and profitable. This dual filtering strategy, which combines entropy-based information gain with PnL normalization, ensures that the final pattern library is well suited for algorithmic trading applications, particularly in volatile market conditions where the clarity and reliability of signals are paramount. Looking ahead, our framework offers promising avenues for future research. In particular, we propose exploring the substitution of Shannon entropy with Tsallis entropy as an alternative measure of uncertainty. Tsallis entropy [30], with its adjustable parameter that can tune the sensitivity to rare events, may offer enhanced flexibility in capturing the complex, multifractal nature of financial time series, potentially leading to further improvements in pattern quality and trading performance. Moreover, drawing inspiration from global optimization techniques, such as the pivot methods [31], future work can explore pivot moves through phase space guided by a q-distribution based on Tsallis entropy. This hybrid approach could enhance our ability to identify and relocate patterns within the feature space, potentially leading to even more efficient and adaptive algorithmic trading strategies.

Limitations and Future Research

In this paper, we demonstrate the proposed entropy-assisted method primarily on two assets—Gold vs. USD and GBPUSD—to illustrate its potential for discovering robust short-term patterns. While these two assets exhibit distinct market behaviors, we note that further exploration across additional equities, currencies, or commodities is needed to confirm broader applicability. Although we have presented results for two assets (Gold vs. USD and GBPUSD), these serve primarily as illustrative case studies of the method's efficacy. Their price movements differed notably in 2024 but do not encompass the full breadth of market behaviors found in other assets such as lower-volatility equities or highly illiquid instruments. Future research should extend these tests to a broader range of assets and time periods, and incorporate more exhaustive robustness checks to fully establish the generalizability of the approach.

While our work draws inspiration from the high-frequency trading (HFT) landscape, we acknowledge that our chosen data resolution—30 min OHLC candles—does not fully capture the microsecond scale typical of HFT. Instead, this study focuses on an intraday but more moderate frequency domain, where patterns remain tractable and computationally feasible with conventional resources. The principles behind our entropy-based approach are, however, extendable to higher-frequency data, provided a suitably fine-grained dataset is available. Nonetheless, caution is advised when extrapolating these findings to true HFT contexts, as the market microstructure and latency considerations at sub-second timescales differ significantly from those present in 30 min intervals.

Looking ahead, future research could extend our framework in several promising directions. First, applying the methodology to a broader range of assets and markets would help validate its generalizability and robustness. Second, incorporating real-time adaptive mechanisms to update the pattern library based on incoming data may further improve the performance of algorithmic trading systems. Finally, exploring hybrid models that combine entropy-based filtering with machine learning approaches could yield even more refined predictive signals, opening up new avenues for risk management and portfolio optimization in complex and dynamic market environments.

Author Contributions: Conceptualization, R.G. and S.K.; methodology, R.G.; software, R.G. and S.G.; validation, R.G., S.G. and J.S.; formal analysis, R.G.; investigation, R.G. and J.S.; resources, S.G.; data curation, S.G.; writing—original draft preparation, R.G.; writing—review and editing, R.G. and S.K.; visualization, R.G.; supervision, S.K.; project administration, S.K.; funding acquisition, S.K. All authors have read and agreed to the published version of the manuscript.

Funding: S.K. would also like to acknowledge funding from the U.S. Department of Energy (DOE) (Office of Basic Energy Sciences), under Award No. DE-SC0019215.

Institutional Review Board Statement: Not applicable

Data Availability Statement: The datasets generated during the current study are available from the corresponding author on reasonable request, while the employed financial datasets are publicly available online at <https://www.histdata.com>.

Conflicts of Interest: Author Shivam Gupta was employed by the company EntropyX Labs Pvt. Ltd. and Author Jaskirat Singh was employed by the company Softure Solutions Pvt. Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Oyeniyi, L.D.; Ugochukwu, C.E.; Mhlongo, N.Z. Analyzing the impact of algorithmic trading on stock market behavior: A comprehensive review. *World J. Adv. Eng. Technol. Sci.* **2024**, *11*, 437–453. [[CrossRef](#)]
2. Arumugam, D. Algorithmic trading: Intraday profitability and trading behavior. *Econ. Model.* **2023**, *128*, 106521. [[CrossRef](#)]

3. Joiner, D.; Vezeau, A.; Wong, A.; Hains, G.; Khmelevsky, Y. Algorithmic trading and short-term forecast for financial time series with machine learning models; state of the art and perspectives. In Proceedings of the 2022 IEEE International Conference on Recent Advances in Systems Science and Engineering (RASSE 2022), Tainan, Taiwan, 7–10 November 2022; pp. 1–9.
4. Jaynes, E.T. Information Theory and Statistical Mechanics. *Phys. Rev.* **1957**, *106*, 620–630. [[CrossRef](#)]
5. Gupta, R.; Xia, R.; Levine, D.; Kais, S. Maximal entropy approach for quantum state tomography. *PRX Quantum* **2021**, *2*, 010318. [[CrossRef](#)]
6. Teschendorff, A.E.; Enver, T. Single-cell entropy for accurate estimation of differentiation potency from a cell’s transcriptome. *Nat. Commun.* **2017**, *8*, 15599. [[CrossRef](#)]
7. Drzazga-Szcześniak, E.A.; Szczepanik, P.; Kaczmarek, A.Z.; Szcześniak, D. Entropy of financial time series due to the shock of war. *Entropy* **2023**, *25*, 823. [[CrossRef](#)]
8. Touchette, H. The large deviation approach to statistical mechanics. *Phys. Rep.* **2009**, *478*, 1–69. [[CrossRef](#)]
9. Sensoy, A. The inefficiency of Bitcoin revisited: A high-frequency analysis with alternative currencies. *Financ. Res. Lett.* **2019**, *28*, 68–73. [[CrossRef](#)]
10. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 623–656. [[CrossRef](#)]
11. Lesne, A. Shannon entropy: A rigorous notion at the crossroads between probability, information theory, dynamical systems, and statistical physics. *Math. Struct. Comput. Sci.* **2014**, *24*, e240311. [[CrossRef](#)]
12. Gray, R.M. *Entropy and Information Theory*; Springer: Berlin/Heidelberg, Germany, 2011.
13. Benedetto, B.; Giunta, G.; Mastroeni, L. A maximum entropy method to assess the predictability of financial and commodity prices. *Digit. Signal Process.* **2015**, *46*, 19–31. [[CrossRef](#)]
14. Ahn, K.; Lee, D.; Sohn, S.; Yang, B. Stock market uncertainty and economic fundamentals: An entropy-based approach. *Quant. Financ.* **2019**, *19*, 1151–1163. [[CrossRef](#)]
15. Zhou, R.; Cai, R.; Tong, G. Applications of entropy in finance: A review. *Entropy* **2013**, *15*, 4909–4931. [[CrossRef](#)]
16. Gupta, R.; Drzazga-Szcześniak, E.A.; Kais, S.; Szcześniak, D. Entropy corrected geometric Brownian motion. *Sci. Rep.* **2024**, *14*, 28384. [[CrossRef](#)] [[PubMed](#)]
17. Stojkoski, V.; Utkovski, Z.; Basnarkov, L.; Kocarev, L. Cooperation dynamics in networked geometric Brownian motion. *Phys. Rev. E* **2019**, *99*, 062312. [[CrossRef](#)] [[PubMed](#)]
18. Stojkoski, V.; Sandev, T.; Basnarkov, L.; Kocarev, L.; Metzler, R. Generalised geometric Brownian motion: Theory and applications to option pricing. *Entropy* **2020**, *22*, 1432. [[CrossRef](#)]
19. Marathe, R.R.; Ryan, S.M. On the validity of the geometric Brownian motion assumption. *Eng. Econ.* **2005**, *50*, 159–192. [[CrossRef](#)]
20. Reddy, K.; Clinton, V. Simulating stock prices using geometric Brownian motion: Evidence from Australian companies. *Australas. Account. Bus. Financ. J.* **2016**, *10*, 23–47.
21. Vilela, L.F.S.; Leme, R.C.L.; Pinheiro, C.A.M.; Carpinteiro, O.A.S. Forecasting financial series using clustering methods and support vector regression. *Artif. Intell. Rev.* **2019**, *52*, 743–773. [[CrossRef](#)]
22. Cai, F.; Le-Khac, N.-A.; Kechadi, T. Clustering approaches for financial data analysis: A survey. *arXiv* **2016**, arXiv:1609.08520.
23. He, H.; Chen, J.; Jin, H.; Chen, S.-H. Trading strategies based on K-means clustering and regression models. In *Computational Intelligence in Economics and Finance*; Springer: Berlin/Heidelberg, Germany, 2007; pp. 123–134.
24. Magazzino, C.; Mele, M.; Schneider, N. Testing the convergence and the divergence in five Asian countries: From a GMM model to a new Machine Learning algorithm. *J. Econ. Stud.* **2022**, *49*, 1002–1016. [[CrossRef](#)]
25. Ngaleu, Y.J.N.; Ngongang, E. Forex Daytrading Strategy: An Application of the Gaussian Mixture Model to Marginalized Currency pairs in Africa. *Int. J. Comput. Sci. Technol.* **2023**, *7*, 149–191.
26. Jolliffe, I.T.; Cadima, J. Principal component analysis: A review and recent developments. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2016**, *374*, 20150202. [[CrossRef](#)]
27. Nobre, J.; Neves, R.F. Combining principal component analysis, discrete wavelet transform and XGBoost to trade in the financial markets. *Expert Syst. Appl.* **2019**, *125*, 181–194. [[CrossRef](#)]
28. Deutsch, H.-P.; Beinker, M.W. Principal component analysis. In *Derivatives and Internal Models: Modern Risk Management*; Springer: Cham, Switzerland, 2019; pp. 793–804.
29. Yacine Ait-Sahalia, Y.; Xiu, D. Principal component analysis of high-frequency data. *J. Am. Stat. Assoc.* **2019**, *114*, 287–303. [[CrossRef](#)]
30. Sneddon, R. The Tsallis entropy of natural information. *Phys. A* **2007**, *386*, 101–118. [[CrossRef](#)]
31. Serra, P.; Stanton, A.F.; Kais, S. Pivot method for global optimization. *Phys. Rev. E* **1997**, *55*, 1162. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.