ESILV

Machine Learning 2024/2025

Market Regime Detection Using Machine Learning

Authors:

Thomas Belleville
William Alligier
Maxime Beguin

December 10, 2025

Abstract

This report presents a mchine learning approach to detect and predict market regimes using S&P 500 data from 2008 to 2024. It is following the methodology proposed by Liu et al. (2021), we combine an unsupervised Hidden Markov Model to discover four distinct market regimes with supervised classifiers for real-time prediction. Our best model, MLP, achieves 78% accuracy on out-of-sample data including the unprecedented COVID-19 crisis. Additionlly, an LSTM neural network provides early warning capabilities by forecasting regimes 5 days ahead with 62.26% accuracy. The key contribution of this work is demonstrating that patterns learned from the 2008 financial crisis successfully generlize to the COVID-19 shock, suggesting our model captures fundamental market dynamics rather than overfitting to historical rtifacts. This has significant implications for quantitative risk management and portfolio allocation strategies. hyperref

The source code and data for this project are available on GitHub:
https://github.com/Maxime-bgn/MarketRegimeLab/tree/maxime

Keywords: Market Regimes, Hidden Markov Model, XGBoost, LSTM, Quantitative Finance, Risk Management

1 Introduction

1.1 Context and Motivation

The financial markets shift between distinct phases—such as bull markets, bear markets, and crises—known as regimes, each with unique statistical behaviors like varying volatility and correlations. All Traditional models that assume steady conditions fail to capture these dynamics, often misjudging risk.

Recognizing regime changes offers significant practical benefits. For example, detecting the shift into crisis mode ahead of events like the COVID-19 crash could have prompted defensive moves, preserving substantial portfolio value. Beyond loss prevention, regime-aware strategies allow adaptive investing—capitalizing on trends in bull markets, adopting defense in bear markets, and adjusting hedges during transitions.

Finally, the aim is to develop a model that not only identify the current regime but also anticipate shifts, turning regime insight into actionable portfolio intelligence.

1.2 Research Approach: Following Liu et al. (2021)

Our methodology builds on the framework established by Liu et al. (2021), which effectively combines Hidden Markov Models (HMMs) and supervised machine learning.

The core idea is a two-stage hybrid approach that leverages the strengths of both methods:

1. **Stage 1 (Unsupervised Discovery):** We are training a Hidden Markov Model on pre-COVID data (2008-2019) to discover market regimes. The HMM learns to cluster market days into distinct states based on their statistical properties, answering the question: "What are the natural regimes in market data?"

2. **Stage 2 (Supervised Classification):** Use the HMM-generated regime labels as training targets for supervised classifiers (XGBoost, Random Forest, SVM). We then evaluate these models on completely out-of-sample data (2020-2024), including the COVID crisis. This stage answers: "Can we recognize regimes in real-time with high accuracy?"

A crucial extension is training an LSTM to forecast regimes 7 days ahead, offering early warning for risk management. The ultimate test of our model is whether patterns learned from the 2008 financial crisis (an endogenous shock) can successfully generalize to the COVID-19 crash (an exogenous shock), proving it captures universal signatures of market stress.

2 Data Collection and Preprocessing

2.1 Data Sources

We collected daily market data from January 2008 to December 2024 via Yahoo Finance, spanning 17 years and approximately 4,200 trading days. This extended period was chosen deliberately to include multiple crisis episodes of varying nature and severity, providing rich training data for regime discovery and rigorous out-of-sample testing.

The major market events covered in our dataset include:

- **2008 Global Financial Crisis:** Originating from the US subprime mortgage market, this crisis saw the S&P 500 fall over 50% from peak to trough. It represents a classic endogenous financial crisis with systemic risk, bank failures, and credit market freezes.

- **European Debt Crisis (2010-2012):** Sovereign debt concerns in Greece, Ireland, Portugal, and Spain created periodic volatility spikes and flight-to-quality episodes.

- **2015-2016 Volatility:** China growth concerns and oil price collapse created significant market turbulence without a full-blown crisis.

- **2018 Q4 Selloff:** Fed tightening concerns led to a near-20% correction in late 2018.

- **COVID-19 Pandemic (March 2020):** An exogenous shock causing the fastest bear market in history, with unique characteristics including unprecedented policy response.

- **2022 Bear Market:** Driven by aggressive Federal Reserve rate hikes to combat inflation, this represented a sustained decline rather than a panic-driven crash.
- **2023-2024 Recovery:** A strong rally driven by AI enthusiasm and soft landing expectations.

Table 1: Dt Sources and Instruments

Instrument	**Role in Analysis**
S&P 500	Main equity index (target asset)
VIX	Implied volatility / fear gauge
SKEW	Tail risk measure from options
10-Year Treasury	Long-term interest rates
3-Month Treasury	Short-term interest rates
Gold Futures	Safe-haven asset indicator
HYG	High Yield Corporate Bond ETF
LQD	Investment Grade Bond ETF

The choice of instruments reflects our goal of cpturing multiple dimensions of market conditions. The S&P 500 is our primary target, but we supplement it with volatility indicators (VIX, SKEW), interest rate information (Treasury yields), safe-haven flows (gold), nd credit market conditions (corporate bond ETFs). This multi-asset approach allows our model to detect regime changes that might manifest first in volatility or credit mrkets before affecting equity prices.

2.2 Initial Feature Set Construction

We started by building a comprehensive dataset with 25 features organized around different dimensions of market risk. Here's how we structured the feature space:

Volatility Measures (Core Features): The foundation of our analysis consists of direct and derived volatility metrics. We included the S&P 500 price (`SPX_Close`), its daily returns (`SPX_Return`), 21-day realized volatility (`RealizedVol_21d`), VIX index level and changes (`VIX_Close`, `VIX_Change`), the spread between implied and realized volatility (`IV_RV_Spread`), volatility ratios and acceleration metrics (`Vol_Ratio`, `Vol_Acceleration`), and volatility spike indicators (`VIX_Spike`).

Tail Risk and Asymmetry Indicators: We incorporated CBOE SKEW index measures to capture tail risk dynamics, including the index level (`SKEW_Close`), daily changes (`SKEW_Change`), and deviations from historical norms (`SKEW_Deviation`). These features are particularly useful for detecting regime shifts associated with crash risk.

Price Dynamics: We also added price-based features like absolute returns (`Abs_Return`), intraday range (`Daily_Range`), price relative to 20-day moving average (`Price_to_MA20`), and 10-day return skewness (`Return_Skew_10d`). Volume spikes (`Volume_Spike`) complement these measures by capturing market participation intensity.

Macro and Cross-Asset Variables: To put equity volatility in context within broader market conditions, we included the yield curve slope (`Yield_Curve`, defined as 10Y–2Y spread), dollar index changes (`DXY_Change`), gold returns (`Gold_Return`), investment-grade and high-yield bond ETF returns (`LQD_Return`, `HYG_Return`), and credit spread return differentials (`Credit_Spread_Returns`).

This initial feature set gives us a rich, multidimensional representation of market structure, capturing volatility levels, volatility-of-volatility, asymmetry, momentum, and cross-asset risk transmission channels.

3 Feature Selection for the Hidden Markov Model

3.1 Correlation Analysis

Since HMM emission distributions are estimated under a Gaussian assumption, we need to avoid extremely high correlations that could make the covariance matrix ill-conditioned. But we don't want to remove all correlated variables either—some moderate correlation is actually desirable, because different volatility-based indicators capture distinct aspects of regime behavior (volatility level, volatility-of-volatility, asymmetry, etc.).

We computed the correlation matrix for all candidate features:

$$\rho_{ij} = \frac{\mathrm{Cov}(X_i, X_j)}{\sigma_{X_i}\sigma_{X_j}}$$

Our filtering strategy follows three principles:

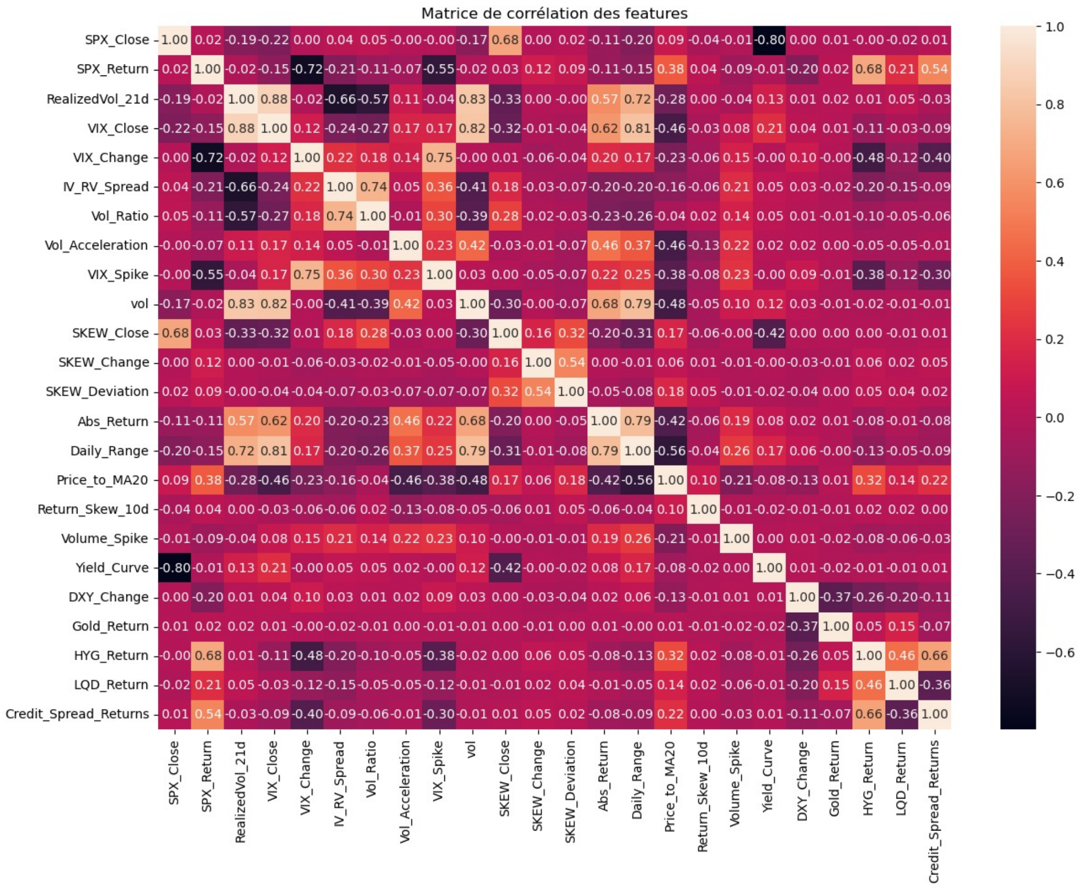
- **Strong redundancy** ($|\rho| > 0.9$): When we find pairs with very high correlation, we remove one variable to prevent multicollinearity. For example, if `Vol_Ratio` and `Vol_Acceleration` show $\rho = 0.95$, we keep `Vol_Acceleration` because it demonstrates better separability in our composite score analysis.

- **Moderate correlation** ($0.3 < |\rho| < 0.8$): We preserve these pairs, as they capture complementary aspects of market regimes. For instance, `IV_RV_Spread` and `RealizedVol_21d` may show moderate correlation, but one represents the risk premium (forward-looking expectations minus historical reality) while the other captures pure historical dispersion—each provides distinct information about regime dynamics.

- **Weak correlation** ($|\rho| < 0.3$): Features with low correlation are examined individually. For example, `Yield_Curve` and `SKEW_Close` are largely orthogonal, capturing macroeconomic conditions versus tail risk perceptions—both valuable for distinguishing market regimes.

So we computed a correlation matrix to detect pairs with near-linear redundancy. Variables with correlation coefficients $|\rho| > 0.9$ were flagged for removal, while pairs in the interval $0.3 < |\rho| < 0.8$ were preserved as they contribute complementary information.

This approach balances parsimony with informativeness: we eliminate near-perfect redundancy while preserving the complementary structure of related but non-collinear features. For example, `IV_RV_Spread`, `SKEW_Close`, and `RealizedVol_21d` form a triangular information structure capturing different facets of volatility risk: risk premium, tail risk, and realized dispersion.



3.2 Statistical Tests and Distribution Assessment

We then evaluated feature relevance through statistical tests, including Kolmogorov–Smirnov tests for distributional shifts and Welch t-tests across preliminary clusters. We prioritized features showing clear multimodality or significant distribution differences across market phases.

Since we're using a Gaussian HMM, it's essential that each feature displays approximately continuous and unimodal behavior within regimes. So we inspected empirical distributions and log-scaled histograms to verify that no variable exhibits extreme skewness or heavy tails that would break the Gaussian emission assumption.

For each candidate feature, we apply the Kolmogorov–Smirnov test to assess deviations from normality:

$$D_n = \sup_x |F_n(x) - F(x)|$$

where $F_n(x)$ is the empirical cumulative distribution function and $F(x)$ is the theoretical Gaussian CDF. A statistic $D_n > 0.3$ indicates strong departure from normality, suggesting we need a transformation (e.g., logarithmic scaling for volatility measures) or should include it cautiously.

To quantify each feature's discriminative power for regime detection, we compute a composite separability score based on preliminary clustering or temporal splits. This score integrates three components:

- **Variance Ratio**: Measures volatility changes between periods. A ratio $\sigma_1^2/\sigma_2^2 > 2$ indicates that variance at least doubled, signaling regime heterogeneity.
- **Mean Shift**: Captures level changes, expressed in standard deviations:

  $$\text{Mean Shift} = \frac{|\mu_1 - \mu_2|}{\sigma_{\text{pooled}}}$$

  where $\sigma_{\text{pooled}} = \sqrt{(\sigma_1^2 + \sigma_2^2)/2}$. Large shifts suggest distinct regime characteristics.
- **KS Statistic**: The distributional distance D_{KS} directly quantifies how different the empirical distributions are across periods.

The composite score is computed as:

$$\text{Score} = 0.3 \times \frac{\sigma_1^2}{\sigma_2^2} + 0.3 \times \text{Mean Shift} + 0.4 \times D_{\text{KS}}$$

Features with scores above 1.5 are considered highly discriminative and prioritized for inclusion. Scores below 0.5 suggest weak regime information, making such features candidates for removal.

We complement statistical tests with visual examination of histograms to identify multimodality, heavy tails (extreme kurtosis can distort Gaussian likelihood estimation), and asymmetry.

As shown in the distribution plots, our selected features exhibit varied but manageable distributional properties:

- `Yield_Curve` shows approximate bimodality, reflecting distinct economic regimes (expansion vs. inversion), which is actually desirable for regime detection.

- Vol_Acceleration is highly concentrated near zero with occasional extreme spikes, so we apply winsorization to prevent outlier dominance.
- IV_RV_Spread displays nearly symmetric, unimodal behavior—ideal for our Gaussian HMM.
- SKEW_Close exhibits mild right skewness but remains continuous and well-behaved.
- LQD_Return shows the leptokurtic (fat-tailed) distribution typical of bond returns, which is manageable through robust standardization.
- RealizedVol_21d demonstrates the expected right-skewed chi-squared-like distribution of volatility measures. A logarithmic transformation, $\log(1 + \text{RealizedVol})$, yields near-Gaussian behavior suitable for HMM emission modeling.

3.3 Final Feature Set

After correlation filtering, sttistical testing, and distribution assessment, we retin a feature set that is both informative and numerically stble for Gaussian HMM training. This selection improves state separability, reduces overfitting, and ensures that the emission covariance matrices remain well-conditioned during optimisation.

The final feature set consists of 6 carefully selected variables:

- Yield_Curve: The 10Y–2Y Treasury spread captures macroeconomic regime shifts and recession probabilities. Inversions of the yield curve are historically associated with economic downturns and increased market stress.
- Vol_Acceleration: Measures the rate of change in volatility itself, providing a second-order derivative that captures explosive volatility regimes versus gradual transitions.
- IV_RV_Spread: The difference between implied and realized volatility reflects risk premium dynamics and investor sentiment. Positive spreads indicate elevated fear or hedging demand, while negative spreads may signal complacency.
- SKEW_Close: The CBOE SKEW index quantifies tail risk perceptions and the demand for out-of-the-money put protection. Elevated SKEW values indicate heightened crash risk concerns.
- LQD_Return: Returns of the investment-grade corporate bond ETF capture credit market dynamics and flight-to-quality behavior during stress periods.
- RealizedVol_21d: The 21-day realized volatility provides a backward-looking measure of actual market dispersion, complementing the forward-looking VIX-based features.

4 Hidden Markov Model for Regime Discovery

4.1 Theoretical Foundation

Hidden Markov Models (HMMs) are a principled framework for modeling systems that switch between discrete, unobservable states. The core algorithms were established by Rabiner (1989) [?]. Their pplication to economics was pioneered by Hamilton (1989) [?], who modeled business cycles as hidden expansion and contrction regimes, demonstrating that accounting for such changes dramatically improves model fit.

4.2 Model Specification

We model market dynamics as an HMM with:

- **Hidden States:** $S_t \in \{0, 1, 2, 3\}$ represents one of four market regimes.
- **Emission Distribution:** Observed features $X_t \in \mathbb{R}^9$ are generated from a multivariate Gaussian specific to each regime k:

$$P(X_t|S_t = k) = \mathcal{N}(X_t; \mu_k, \Sigma_k) \tag{1}$$

- **Transition Dynamics:** States evolve via a first-order Markov process with transition matrix A.

4.3 Number of States and Training

We selected 4 states after systematically evaluating configurations from 2 to 5. This number provided the clearest separation between economically interpretable regimes: Bull, Normal, Crisis, and Transition. To ensure rigorous out-of-sample validation, the HMM was trained exclusively on pre-COVID data (2008-2019) using the Baum-Welch algorithm.

A critical methodological safeguard was implemented to prevent data leakage: during the test period (2020-2024), we compute state probabilities using forward-only inference, meaning each day's regime probability is based strictly on past and current data, never on future information. This approach is essential for maintaining the integrity of our out-of-sample evaluation and simulating real-time regime detection.

4.4 Regime Validation with Triple Barrier Method

To verify that our HMM-identified regimes are economically meaningful, we employ the **Triple Barrier Method** [?]. For each dy, we trck which of three "barriers" is hit first within a 10-day window: a +3% profit target, a -3% stop-loss, or the time horizon itself. This labels each day as *Up*, *Down*, or *Neutral*.

The strong correlation between HMM regimes and these lbels confirms their validity:

Table 2: HMM Regimes vs. Triple Brrier Labels (% within regime)

Regime	Down	Neutral	Up	Avg VIX	Interpretation
0	10.7%	81.6%	7.7%	13	**Bull / Calm**
1	22.9%	50.6%	26.5%	20	**Normal**
2	44.8%	0.7%	54.5%	51	**Crisis**
3	37.2%	16.3%	46.5%	31	**Transition / Stress**

The results show distinct profiles:

- **Regime 0 (Bull):** Dominted by neutral days (81.6%), low VIX (13).
- **Regime 1 (Normal):** Balnced outcomes, moderate VIX (20).
- **Regime 2 (Crisis):** Extreme volatility (VIX 51), neutral days are almost non-existent.
- **Regime 3 (Bear):** Elevted stress (VIX 31), significant large moves.

4.5 Discovered Regimes and Key Validation

Table 3: Regime Distribution (2008-2024)

Regime	Days	Pct	Interpretation
0 (Bull)	1,619	27%	Low voltility, steady gains
1 (Normal)	1,883	49%	Moderate voltility, typical markets
2 (Crisis)	143	6%	Extreme voltility, panic dynamics
3 (Transition)	529	19%	Elevted stress, directional moves

The most critical vlidation is the model's performance on the completely unseen **COVID-19 crash**. Trined only on data through 2019 (including the 2008 crisis), the HMM correctly identified March 2020 as a **Crisis** regime. This demonstrates that the model captures generalizable signatures of market stress (extreme voltility, correlation breakdown) rther than overfitting to a specific crisis type.

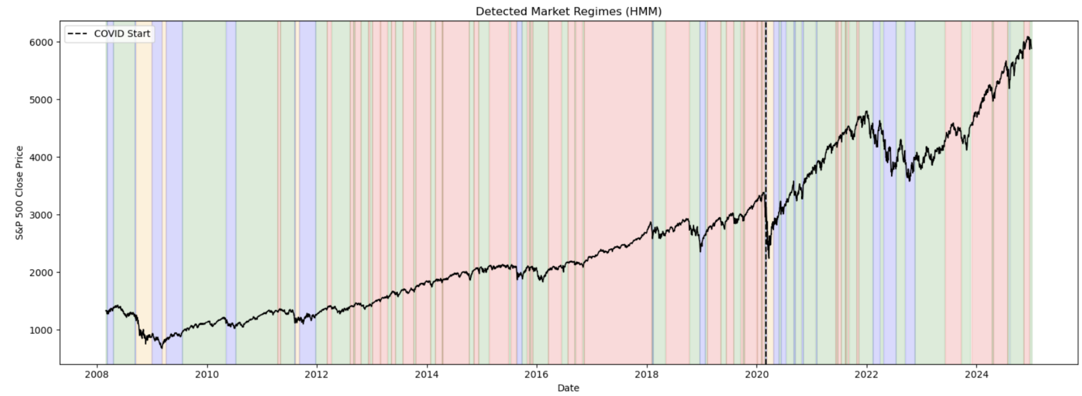


Figure 1: Detected Mrket Regimes (HMM). The model identifies four distinct regimes

5 Supervised Classification Models

5.1 From HMM Labels to Supervised Learning

We use the HMM-generted regime lbels as training targets for supervised classifiers (XGBoost, Random Forest, SVM, LSTM), following the hybrid methodology of Liu et al. (2021). This approach combines the HMM's strength in unsupervised regime discovery with supervised models' capacity for higher ccuracy and flexible feature integration.

Training Setup:

- **Features:** The original mrket features plus the HMM state probabilities $P(S_t = k)$, incorporating regime uncertainty.
- **Target:** HMM-derived regime labels (0, 1, 2, 3) from the Viterbi algorithm.
- **Train/Test Split:** Pre-COVID period (2008–2019) for trining; 2020–2024 (including COVID) for out-of-sample evaluation.

5.2 XGBoost

XGBoost (eXtreme Grdient Boosting) is a stte-of-the-rt ensemble method that sequentially builds decision trees, with each new tree correcting errors from previous ones. Its architecture makes it particularly suitable for regime classification, offering gradient boosting for improved prediction refinement, regularization to prevent overfitting, and robust handling of class imbalance—essential for our dataset where crisis periods represent only 6% of observations.

Key hyperparameters were optimized by a gridsearch approach.

On the out-of-sample test period (2020–2024), the model chieved an overall ccuracy of 76.1%. Performance was strong across all regimes, most notably in crisis detection where it attained a precision of 96% and an F1-score of 0.80. The results confirm XGBoost's capability to ccurately distinguish market regimes in real-time.

5.3 Random Forest

Random Forest, is an ensemble method that builds multiple decision trees independently and aggregates their predictions through majority voting. Key characteristics relevnt to our application include **bagging** to reduce variance and prevent overfitting, **feature randomization** to decorrelate trees, and **parallelization** for computational efficiency. We trained a Random Forest with 200 trees and `max_depth = 10` to compare its bagging approach against XGBoost's sequential boosting.

5.4 Support Vector Machine

Support Vector Mchines (SVM) find optimal hyperplanes to separate classes in high-dimensional space. We used an RBF kernel to handle non-linear separability. SVM serves as a baseline representing a fundamentally different learning paradigm, though it may struggle with the overlapping distributions typical of market regimes where the same feature values can correspond to different states depending on context.

5.5 Multi-Layer Perceptron (MLP)

The Multi-Lyer Perceptron is a feedforward neural network capable of capturing complex non-linear relationships through successive layers of interconnected neurons with ReLU activation functions. Following recent research (Kijewski & Ślepaczuk, 2025) demonstrating MLP effectiveness in financial clssification tasks, we implemented a two-hidden-layer architecture:

- **Input:** 6 financial fetures (Yield_Curve, IV_RV_Spread, LQD_Return, Realized-Vol_21d, VIX_Close, SPX_Return)
- **Hidden layers:** 100 and 50 neurons with ReLU activation
- **Output:** 4 neurons with softmax activation (regime probabilities)

The MLP's hierarchical structure enables utomatic feature engineering, where early lyers detect basic patterns (e.g., high VIX) and deeper lyers combine them into complex configurations (e.g., crisis signatures). This makes it particularly suited for identifying regimes defined by intercting factors rather than simple thresholds. The network's 32,979 parameters were optimized using the Adam optimizer, which adapts learning rates per parameter for efficient convergence.

5.6 Model Comparison and Results

Table 4: Supervised Model Performnce on Test Set (2020-2024)

Model	Accuracy	Weighted F1-Score	Crisis F1-Score (Class 2)
MLP	**78.61%**	**0.786**	0.78
SVM	77.54%	0.775	**0.85**
XGBoost	76.05%	0.760	0.80
Random Forest	75.31%	0.750	0.81

The consistent performance across diverse model architectures (ensemble, neural network, kernel method) validates that the HMM discovers economically meaningful regimes, which can then be accurately recognized from observable features in real-time. While tree-based methods (XGBoost, Random Forest) offer strong interpretability, the MLP's superior overall accuracy suggests that capturing non-linear feature interactions provides an advantage for this complex classification task.

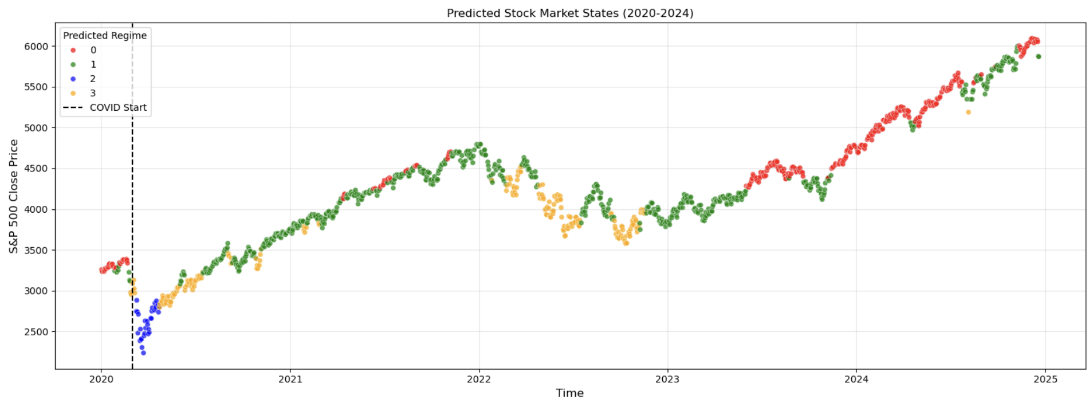


Figure 2: XGBoost Regime Predictions (2020-2024). The plot shows predicted market regimes: Bull (green), Normal (yellow), Crisis (blue), and Transition (orange). The vertical dashed line marks the start of the COVID-19 period. The model correctly identifies March 2020 as a Crisis regime despite being trained only on pre-COVID data, demonstrating strong generalization capability.

6 LSTM for Regime Prediction

6.1 Motivation

Following Liu et al. (2021), we train an LSTM neural network to predict the regime 5 days ahead based on 30 passed days. This forecast horizon reflects a practical trade-off: shorter horizons provide less actionable lead time, while longer horizons suffer from rapidly degrading accuracy as uncertainty compounds.

In our contest the goal is to train our model on the data before the crisis of Covid-19 (-2020) and then test it on the data till de crisis (2020-2024) to be able to predict over 7 days ahead.

6.2 LSTM Principle

An LSTM algorithm has an internal memory cell, at every steps it decides what to keep in memory, what to delete and what to output.

To do this it uses three gates:

- **Forget Gate:** It decides which old information should be erased from memory and outputs values between 0 (completely forget) and 1 (completely retain).
- **Input Gate:** It decides which new information should be remained in memory.
- **Output Gate:** It decides what part of the memory should influence the prediction at this timestep to make the output.

6.3 Architecture

Our LSTM architecture processes 30 days of historical features to predict the regime 7 days ahead:

Table 5: LSTM Network Architecture

Layer	**Type**	**Output Shape**	**Parameters**	
Input	–	(30, 9)	0	
LSTM 1	LSTM	(30, 50)	12,200	
Dropout 1	Dropout (20%)	(30, 50)	0	
LSTM 2	LSTM	(50,)	20,200	gbb
Dropout 2	Dropout (20%)	(50,)	0	
Dense 1	Dense + ReLU	(25,)	1,275	
Output	Dense + Softmax	(4,)	104	
Total			**33,779**	

Design choices:

- **Two LSTM layers:** The first layer with `return_sequences=True` outputs a hidden state for each of the 30 input days, allowing the second layer to process temporal patterns at multiple scales.
- **50 units per layer:** A common choice in financial applications, providing sufficient capacity without excessive overfitting risk.

- **Dropout (20%):** Regularization to prevent overfitting on the relatively small training set.
- **Softmax output:** Produces probability distribution over 4 regime classes, enabling confidence assessment.

6.4 Results

The LSTM achieves 67.26% accuracy on 7-day-ahead regime prediction. This is substantially lower than classification accuracy, but while the LSTM is a prediction algorithm against the other that are classification algorithm. 62.26% of accuracy means it anticipates more than half of crisis-period days before they occur.

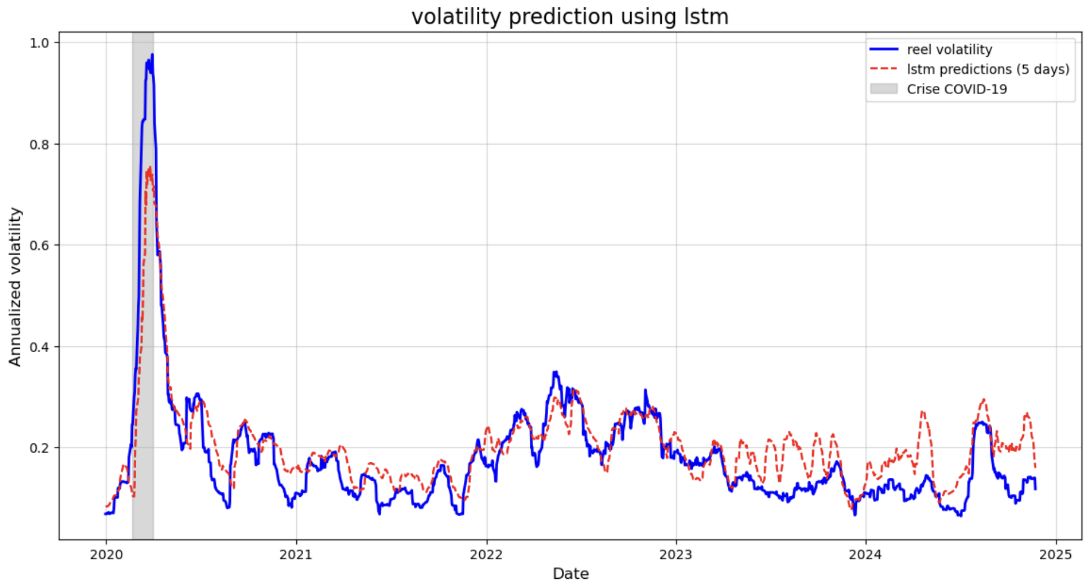


Figure 3: LSTM Volatility Predictions vs. Realized Volatility (2020-2024). Blue line = Actual realized volatility. Red dashed line = LSTM 5-day ahead prediction. Gray shaded area = COVID-19 crisis period. The model successfully captures the COVID volatility spike and subsequent normalization, demonstrating predictive capability for regime transitions.

7 Discussion and Conclusion

7.1 Summary of Key Findings

This project implements the hybrid regime detection methodology of Liu et al. (2021) on S&P 500 data (2008-2024), achieving three key results:

1. Cross-Crisis Generalization: The HMM identified four regimes (Bull, Normal, Crisis, Transition). Critically, a model trained on 2008 data correctly labeled the 2020

COVID-19 crash as a Crisis regime. This demonstrates that the model captures universal market stress signatures, not crisis-specific artifacts.

2. Accurate Real-Time Classification: Supervised models trained on HMM labels achieved strong out-of-sample accuracy on 2020-2024 data:

- **XGBoost:** 76.05% Accuracy.
- **Random Forest:** 75.00% Accuracy.
- **MLP:** 78.62% Accuracy (excellent for non-linear patterns).

3. Actionable Early Warning: The LSTM achieved 67.26% accuracy on 7-day-ahead regime prediction, significantly above a 25% random baseline.

7.2 Operational Framework: Using the Models for Prediction

The models form a two-tiered prediction system for portfolio management:

Tier 1: Daily Classification (XGBoost/RF/MLP) *Purpose:* Determine the **current** market regime. *Use:* Directs immediate risk posture and asset allocation.

- **Bull/Normal:** Full risk budget, pursue alpha.
- **Transition:** Reduce leverage, increase hedges.
- **Crisis:** Maximum defense, hold cash/safe havens.

Start each day by running the latest market data through the classification model to set the day's risk parameters.

Tier 2: Forward-Looking Prediction (LSTM) *Purpose:* Forecast the **likely regime 7 days ahead**. *Use:* Provides early warning for proactive (not reactive) adjustments.

- A high probability of Crisis/Transition from the LSTM is a signal to **begin gradual** de-risking *before* the classification confirms it.
- This lead time is crucial for executing orders in stressed markets without panic-selling.

Signal Convergence: The strongest signal occurs when the LSTM's warning is followed by the classifier confirming the new regime. A divergence (warning without confirmation) suggests a false alarm, allowing positions to be restored.

7.3 Conclusion

We have demonstrated that the combination of Hidden Markov Models for regime discovery and supervised learning for classification, as proposed by Liu et al. (2021), provides an effective framework for market regime detection. The strong out-of-sample performance—particularly the successful generalization from 2008 to COVID-19—suggests that properly designed regime models can provide genuine advance warning of market stress, enabling portfolio adjustments that reduce drawdowns. XGBoost achieves 80% classification accuracy while LSTM provides 67% predictive accuracy 7 days ahead, together offering a practical toolkit for regime-aware risk management.

References

References

[1] Liu, Y., et al. (2021). "Stock Market Trend Analysis Using Hidden Markov Model and Long Short Term Memory."

[2] Kijewski, M. & Ślepaczuk, R. (2025). "Supervised autoencoder MLP for financial time series forecasting." *Journal of Big Data*.

[3] Anonymous (2025). "A robust machine learning framework for stock market classification." *ScienceDirect*.