AIMS Innovation Economics

InnoEcon, 1(1): 52–67.
DOI: 10.3934/InnoEcon.2026003
Received: 30 March 2026
Revised: 16 June 2026
Accepted: 02 July 2026
Published: 07 July 2026

https://www.aimspress.com/journal/InnoEcon

Research article

Harnessing big data analytics across financial markets, banking, economic forecasting, and islamic finance: A unified machine learning framework

Umar Iqbal Butt*

Swiss School of Business and Management (SSBM Geneva), Switzerland

* **Correspondence:** Email: omar.iqbal@live.com.

Abstract: This study presents a unified machine learning (ML) framework designed to bridge four interconnected domains of modern finance: financial markets, banking stability, economic forecasting, and Islamic finance. Drawing on a multi-source empirical dataset—including U.S. equity and volatility indices (2000–2024), Shariah-compliant exchange-traded funds (ETFs) (2019–2024), major banking institution stocks, macroeconomic ETF proxies, and the World Bank Global Financial Development Database (GFDD) panel (2000–2023)—I develop a layered analytical architecture combining random forest and extreme gradient boosting (XGBoost) for feature selection and predictive modelling. A central contribution of this study is the treatment of Shariah-compliance screening as a quantifiable data feature, enabling systematic ethical filtering within a big data environment. The empirical results show that the random forest model reduces out-of-sample root mean square error (RMSE) by 78.3% relative to the ordinary least squares (OLS) benchmark in forecasting the Chicago Board Options Exchange Volatility Index (CBOE VIX), and achieves an R^2 of 0.224 in cross-country banking distress prediction. Shariah-compliant ETFs record a higher Sharpe ratio (0.819 vs. 0.702 for the S&P 500) and a marginally lower maximum drawdown (−33.57% vs. −33.92%) over the 2019–2024 period, though this period-specific differential should be interpreted with caution given the short five-year sample. Bank concentration (31.0%) and interest rate spreads (25.3%) emerge as the dominant structural predictors of non-performing loan ratios across global banking systems. These findings demonstrate that big data analytics can bring together traditionally separate financial domains—including Islamic finance—within a single, coherent predictive framework, with meaningful implications for financial regulation, portfolio construction, and Shariah-compliant asset management.

Keywords: big data analytics; machine learning; islamic finance; economic forecasting; banking stability; shariah ETFs; random forest; XGBoost; unified framework

JEL Codes: C45, G10, G21, Z12

1. Introduction

The rapid digitisation of global financial systems has generated an unprecedented volume of structured and unstructured data. Financial institutions, regulators, and researchers increasingly encounter datasets that are high-dimensional, mixed-frequency, and drawn from diverse sources—conditions that stretch the limits of classical econometric methods. Machine learning (ML) and artificial intelligence (AI) offer computationally powerful alternatives that can identify complex, nonlinear patterns within these demanding financial environments (Goodell et al., 2021; Huang et al., 2020).

Despite these advances, existing research remains largely siloed. Separate modelling traditions—statistical finance for equity markets, credit-risk frameworks for banking, and vector autoregression (VAR) for macroeconomic forecasting—rarely communicate within a unified analytical architecture. This compartmentalisation carries a real cost: signals that are valuable in one domain, such as bond-market movements as indicators of banking stress, are routinely excluded from models built for another.

This gap is especially significant for Islamic finance. With an estimated USD 4.5 trillion in global assets as of 2024 and an annual growth of approximately 10%–12% (Islamic Finance Development Report, 2023), the Islamic financial sector operates under Shariah principles—including the prohibition of Riba (interest), Gharar (excessive uncertainty), and Maysir (speculation)—that give it a structurally distinct risk profile. Yet, Islamic financial assets remain poorly integrated into big data and ML research, often treated as a footnote to conventional finance rather than a domain worthy of dedicated analytical attention.

This study addresses these gaps by developing and empirically validating a unified ML framework that simultaneously covers four interconnected financial domains: (1) equity market dynamics and volatility forecasting; (2) banking system stability and non-performing loan (NPL) prediction; (3) macroeconomic regime identification through cross-asset correlation analysis; and (4) Shariah-compliant asset screening and performance analytics. Using a multi-source dataset spanning 2000–2024, I deploy random forest and extreme gradient boosting (XGBoost) models in a layered architecture and compare their performance against ordinary least squares (OLS) regression as the classical benchmark. The central research question motivating this study is: Can a unified machine learning framework, drawing on big data from multiple financial domains, generate more accurate and economically interpretable predictions than conventional domain-specific models? And can Islamic finance be systematically incorporated within this framework as a quantifiable feature rather than a parallel analytical track? Three observable gaps motivate this question: first, the siloed nature of existing ML applications in finance, which prevents cross-domain signal exploitation; second, the absence of Shariah-compliance as a formal predictive variable in quantitative big data frameworks; and third, the persistent reliance on weak benchmarks (typically OLS or naïve baselines) in existing ML finance studies, which may overstate the apparent advantage of ML methods.

This paper makes four contributions. First, I offer an integrated methodological architecture that brings together mixed-frequency data from equity markets, banking systems, macroeconomic proxies, and Islamic finance ETFs within a single analytical environment. The framework is unified at three levels: (a) a shared feature engineering pipeline applied consistently across all four domains using common macro-financial predictors processed identically; (b) a common algorithmic architecture—Random forest and XGBoost—enabling cross-domain comparability of model behaviour and feature importance; and (c) a shared interpretive lens in which Shariah-compliance functions as a cross-cutting variable that manifests differently in each domain but connects them analytically. This integration represents a form of financial technology (FinTech) innovation: by treating big data analytics as a transformative tool for financial intermediation, regulatory oversight, and inclusive Islamic finance ecosystems, the framework carries implications beyond technical modelling. Second, I provide an empirical comparison that benchmarks ensemble ML methods against not only OLS but also standard time-series and econometric models, producing a more robust and contextually honest evaluation of ML performance. Third, I establish the role of Shariah-compliance screening as a measurable risk feature—showing that Shariah-screened equity ETFs deliver superior risk-adjusted performance over the 2019–2024 period, positioning Islamic finance not merely as an ethical requirement but as a financially sound and analytically tractable investment approach. Fourth, I demonstrate that cross-domain signals—such as safe-haven asset flows as leading indicators of banking stress—are recoverable only within a unified feature architecture, not within siloed domain-specific models.

2. Literature review

2.1. Big data and machine learning in financial markets

The application of big data and ML to financial markets has accelerated substantially since 2015, driven by the capacity of complex algorithms to solve multi-source policy and prediction economic problems (Athey, 2017). Random forest and gradient-boosted tree models have demonstrated superior predictive accuracy over linear and ARIMA-class models in equity return and volatility forecasting (Gu et al., 2020; Fischer & Krauss, 2018). Specifically, Gu, Kelly, and Xiu (2020) evaluated an expansive array of machine learning methods on U.S. equity return prediction using over 900 predictor variables, finding that deep neural networks and tree-ensemble methods consistently outperform linear factor models. Gu et al. (2020) evaluated ML methods on U.S. equity return prediction using over 900 predictor variables, finding that deep neural networks and tree-ensemble methods consistently outperform linear factor models. Fischer and Krauss (2018) demonstrated that long short-term memory (LSTM) networks outperform classical econometric methods in S&P 500 directional forecasting. High-frequency data from limit-order books and options markets have further expanded the predictive horizon available to ML models (Cont, 2011; López de Prado, 2018). The emergence of alternative data sources—satellite imagery, social media sentiment, credit card transactions—has broadened the feature space beyond traditional price and volume data (Houlsby et al., 2019).

2.2. ML applications in banking stability and non-performing loan prediction

Banking system stability, traditionally modelled through early-warning systems based on financial ratios and CAMELS frameworks, has been substantially enhanced by ML approaches—a strand of the literature directly relevant to this study's unified framework. Bussière and Fratzscher (2006) developed early-warning systems for currency crises, while Berge and Jordà (2011) provided foundational frameworks for evaluating the classification of economic activity and systemic turning points into recessions and expansions. Building on these classification frameworks, Holopainen and Sarlin (2017) demonstrated that ML classifiers outperform traditional logit models in banking crisis prediction. Cross-country panel studies using the World Bank GFDD have established that domestic credit-to-GDP ratios, bank concentration, and interest rate spreads are robust structural predictors of non-performing loan accumulation (Ozili, 2019; Kjosevski & Petkovski, 2017). Importantly, research on bank-level intangible and structural drivers of performance—including intellectual capital dimensions such as human capital, structural capital, and relational capital—has shown that these factors can predict innovation-driven performance outcomes in banking institutions (Mention & Bontis, 2013), complementing traditional financial ratios as predictors of stability. The post-2008 regulatory environment—with Basel III capital adequacy requirements and macro-prudential stress testing—has created institutional demand for more granular, real-time risk modelling that integrates both structural and flow-based indicators.

2.3. Economic forecasting with ML and macroeconomic feature engineering

Macroeconomic forecasting has historically relied on structural VAR models, dynamic factor models (DFMs), and Bayesian methods. Athey and Imbens (2019) provided a comprehensive review of ML applications in economics, noting that ensemble methods are particularly valuable when the number of predictors approaches or exceeds the number of observations. Stock and Watson (2012) demonstrated gains from factor-augmented forecasting models in nowcasting GDP, extended to ML architectures by Coulombe et al. (2022). The use of Treasury bond yields, inflation-linked securities, commodity prices, and real estate proxies as macroeconomic leading indicators has a long empirical tradition (Estrella & Mishkin, 1998). Cross-asset ETF data provide daily, liquid proxies for macroeconomic signals at a frequency unavailable from quarterly national accounts.

2.4. Islamic finance, Shariah screening, and risk-adjusted performance

Islamic finance represents a rapidly growing yet analytically underserved domain within quantitative finance. The existing literature has addressed Sukuk pricing (Fathurahman & Fitriati, 2013), Islamic bank efficiency (Johnes et al., 2014), and the comparative performance of Islamic versus conventional equity indices (Ho et al., 2014; Hassan & Girard, 2011). However, the integration of Shariah-compliant instruments within ML-driven big data frameworks remains at an early stage. From an analytical perspective, Shariah-compliance screening applies a defined set of sector-based and financial-ratio filters—excluding activities that are prohibited under Islamic law and ensuring that participating companies meet specified financial thresholds. This structured screening process creates a differentiated return profile that may respond differently to financial market stress compared to conventional unconstrained indices. Hassan (2002) and Walkshäusl and

Lobe (2012) provided evidence that Islamic equity indices are broadly comparable to conventional benchmarks in performance terms, a finding that this study extends to Shariah-compliant U.S.-listed ETFs over the 2019–2024 period.

3. Data and methodology

3.1. Dataset construction

The empirical analysis draws on five principal data sources spanning daily, monthly, and annual frequencies across 2000–2024. Table 1 summarises the dataset characteristics.

Table 1. Dataset summary.

Dataset	Source	Period	Frequency	Key variables	Obs.
U.S. equity and volatility	Yahoo Finance	2000–2024	Daily	S&P 500, NASDAQ, DJIA, VIX	6288
Shariah-compliant ETFs	Yahoo Finance	2019–2024	Daily	HLAL, SPUS, UMMA	1375–748
Banking stocks	Yahoo Finance	2000–2024	Daily	JPM, BAC, C, WFC, GS, MS	6288
Macro ETFs	Yahoo Finance	2002–2024	Daily	TLT, IEF, TIP, GLD, USO, DJP, VNQ	5644
World Bank GFDD	World Bank	2000–2023	Annual	NPL, Concentration, Spread, Credit	6384

The Wahed FTSE USA Shariah ETF (HLAL), launched in July 2019, serves as the primary Shariah-compliant equity instrument in this study. HLAL tracks an FTSE Russell index that applies a two-stage screening process: (1) business activity exclusions based on Shariah principles, and (2) quantitative financial screens requiring that total debt does not exceed 33% of total assets, accounts receivable remain below 49% of total assets, and interest income stays under 5% of total revenue. The SP Funds S&P 500 Shariah ETF (SPUS, December 2019) and the Wahed Dow Jones Islamic World ETF (UMMA, 2021) provide additional Islamic benchmark coverage, though with shorter track records. The World Bank GFDD panel covers 267 countries and territories over 24 years and provides 13 structural financial indicators drawn from the International Monetary Fund (IMF) International Financial Statistics and national supervisory databases.

3.2. Methodological framework

I implement a three-layer hybrid analytical architecture. Layer 1 applies random forest (RF) and extreme gradient boosting (XGBoost) for feature importance quantification and predictive modelling. Layer 2 applies OLS regression as the classical econometric baseline, alongside ARIMA(1,1,1) and GARCH(1,1) for VIX forecasting and panel fixed-effects regression for banking stability, providing a richer set of conventional benchmarks. The OLS specification for VIX forecasting regresses the VIX level on 20-day lagged values of the full feature vector using White heteroskedasticity-consistent standard errors; full OLS coefficient estimates, standard errors, and diagnostic statistics are reported in Appendix B. The panel fixed-effects specification for NPL prediction includes country fixed effects and year dummies; full estimation output is provided in Appendix B. All estimation code—including data extraction, preprocessing, OLS, ARIMA, GARCH, Random Forest, and XGBoost—is provided in the supplementary Python notebook (Supplementary_Code.ipynb). Layer 3 computes risk-adjusted performance analytics—Sharpe ratio, maximum drawdown, annualised volatility, and cross-asset correlations—for the Islamic finance comparison domain.

I select random forest and XGBoost over other contemporary machine-learning methods for three reasons specific to this setting. First, the data are heterogeneous, of moderate size, and consist predominantly of tabular panels in which nonlinearities and feature interactions dominate; tree-ensemble methods are well-suited to such structures and require less tuning than deep architectures. Second, both methods yield native feature-importance measures, which are central to the interpretive contribution of this study; deep-learning architectures such as LSTM and transformer models typically require substantially larger, higher-frequency samples than are available here and sacrifice this interpretability, while kernel- and distance-based methods (SVM, k-NN) offer weaker feature attribution. Third, the two methods share a common architecture that can be applied identically across all four domains, which is a prerequisite for the cross-domain comparability that the unified framework requires. Deep-learning approaches for high-frequency data are identified as a direction for future work.

Random forests aggregate predictions from B decorrelated decision trees, each trained on a bootstrapped subsample with random feature subsampling at each split (Breiman, 2001). Feature importance is computed as mean decrease in impurity (Gini importance) across all trees; it is important to note that feature importance measures reflect predictive association, not causal direction. To guard against bias towards high-cardinality features, permutation importance is computed as a robustness check: each feature is randomly shuffled, and the resulting drop in out-of-sample performance is recorded. XGBoost sequentially fits additive trees to residual errors of prior trees, regularised through shrinkage and subsampling (Chen & Guestrin, 2016). Both models use 200 estimators and a strictly chronological 80/20 train/test split to prevent lookahead bias. Hyperparameters for both models are selected via 5-fold time-series cross-validation (using a rolling-window scheme that respects temporal ordering) applied to the training set only. The hyperparameter grid for random forest covers the number of trees (100, 200, 500), maximum depth (none, 10, 20), and minimum samples per leaf (1, 2, 5). For XGBoost, the grid covers the learning rate (0.01, 0.05, 0.1), maximum depth (3, 5, 7), and subsampling rate (0.7, 0.8, 1.0). Optimal values are selected by minimising mean cross-validated RMSE. Out-of-sample performance is evaluated on the held-out 20% test set; no information from the test set is used in model selection. The replication code for all models—including OLS, random forest, and XGBoost—is provided in the supplementary Python notebook (Supplementary_Code.ipynb) submitted alongside this manuscript.

For the equity volatility domain, the target is the VIX index level. The feature vector comprises a five-day rolling average of S&P 500, NASDAQ, DJIA, 20-year Treasury bonds, gold, oil, Goldman Sachs, JPMorgan, and Bank of America prices, lagged 20 trading days to avoid contemporaneous information leakage and to simulate monthly-horizon forecasting. For the banking stability domain, the target is the country-level NPL ratio from the GFDD panel. Predictors include domestic credit, bank liquid reserves, interest rate spreads, broad money-to-reserves ratios, and bank concentration. Models are estimated on 80% of country-year observations. The Sharpe ratio assumes a risk-free rate of zero, consistent with Islamic finance scholarship, where Riba-bearing instruments are excluded from the denominator (Hassan & Girard, 2011). Maximum drawdown is the largest peak-to-trough percentage decline in cumulative wealth.

3.3. The integration logic of the unified framework

The integrative logic of the framework operates at a level deeper than the shared use of common estimators, and I make it explicit here. First, all four domains are mapped into a single standardised feature representation, in which heterogeneous inputs—equity and volatility series, banking balance-sheet ratios, macroeconomic ETF proxies, and Shariah-screening attributes—are expressed as comparable features within one matrix, so that a signal generated in one domain is available as a predictor in another. Second, this shared representation is what allows genuinely cross-domain signals to be exploited: safe-haven asset flows that originate in the macro-asset domain enter the equity-volatility model as leading features, and banking-sector equity variables inform the same model—linkages that are recoverable only when the domains occupy one feature space and are invisible to siloed, single-domain models. Third, Shariah-compliance screening is operationalised as a quantifiable feature embedded directly in this representation, which is what places Islamic finance inside the same predictive architecture as conventional finance rather than treating it as a parallel, qualitative track. The unification, therefore, resides in the shared feature space and the cross-domain signal transfer it enables, not merely in the reuse of random forest and XGBoost across separate datasets.

4. Results

4.1. Descriptive statistics: Equity markets

Table 2 presents descriptive statistics for the U.S. equity dataset (2000–2024, N = 6288 daily observations). The S&P 500 exhibits a mean of 2111 with a standard deviation of 1243, reflecting the secular uptrend alongside episodic volatility clusters (2008–2009 GFC, 2020 COVID shock). The VIX records a mean of 19.87 and a maximum of 82.69—corresponding to March 2020—confirming its strongly right-skewed distributional properties.

Table 2. Descriptive statistics: U.S. equity market and volatility (2000–2024, N = 6288).

Variable	Mean	Std Dev	Min	25th Pctile	Median	75th Pctile	Max
S&P 500	2111.41	1242.50	676.53	1200.80	1491.52	2748.83	6090.27
NASDAQ	5454.34	4480.27	1114.11	2186.48	3292.61	7446.07	20,173.89
Dow Jones	17,959.69	9351.45	6547.05	10,603.10	13,364.61	24,923.23	45,014.04
VIX	19.87	8.46	9.14	13.86	17.84	23.36	82.69

4.2. Machine learning volatility forecasting

Table 3 compares out-of-sample VIX forecasting accuracy across model specifications. To ensure a robust evaluation, the ML models are benchmarked not only against OLS but also against an ARIMA(1,1,1) specification and a GARCH(1,1) model—both standard baselines in the volatility forecasting literature (Andersen et al., 2003). The OLS baseline exhibits an RMSE of 28.26; the ARIMA(1,1,1) and GARCH(1,1) produce RMSEs of 24.18 and 21.47, respectively, confirming that these time-series models outperform OLS but remain substantially weaker than ensemble ML methods. The random forest achieves an RMSE of 6.13, representing a 78.3% reduction in forecast error relative to OLS and a 71.4% reduction relative to the GARCH(1,1) benchmark. XGBoost produces an RMSE of 7.50. The negative R^2 values reflect the inherent unpredictability of the VIX level at a 20-day forward horizon—a well-documented outcome in out-of-sample volatility forecasting—but the relative RMSE improvements across all benchmarks confirm that ensemble ML methods substantially dominate linear and time-series forecasting in high-dimensional financial settings. These performance advantages should be understood as evidence of the value of cross-asset feature richness rather than solely of algorithmic superiority.

Table 3. VIX forecasting performance: OLS vs. machine learning.

Model	RMSE	R^2	RMSE improvement vs. OLS
OLS regression (baseline)	28.26	−26.78	—
Random forest (200 trees)	6.13	−0.31	78.3%
XGBoost (200 rounds)	7.50	−0.96	73.4%

Feature importance analysis (Table 4) reveals that NASDAQ composite price (28.0%) contributes the largest marginal information content to VIX prediction, followed by gold (15.9%) and 20-year Treasury bonds (15.2%). The prominence of safe-haven assets as VIX predictors has a clear economic logic: during periods of equity market stress, investors rotate capital from risk assets into gold and long-dated government bonds, generating contemporaneous price appreciation in these instruments that serves as a measurable signal of rising systemic uncertainty. This "flight-to-quality" channel—well-documented in the asset pricing literature (Baur & Lucey, 2010; Longstaff, 2004)—means that safe-haven prices are not merely correlated with VIX but are mechanically linked to the same underlying risk aversion dynamics that drive volatility. Banking stocks (Goldman Sachs, Bank of America, JPMorgan) collectively contribute 11.3% of feature importance, confirming the bidirectional financial system–market volatility channel: banking sector distress elevates system-wide uncertainty, while elevated VIX depresses bank equity through mark-to-market losses and increased funding costs.

Table 4. Random forest feature importance: VIX forecasting.

Feature	Importance	Economic interpretation
NASDAQ Composite	0.280	Technology sector stress leads VIX
Gold ETF	0.159	Safe-haven demand anticipates volatility
20-yr Treasury Bond	0.152	Flight-to-quality flow indicator
S&P 500 Index	0.102	Broad market sentiment
Dow Jones Index	0.099	Blue-chip anchor
Oil ETF	0.075	Commodity-financial stress transmission
Goldman Sachs	0.058	Systemically important institution signal
Bank of America	0.046	Banking sector stress
JPMorgan Chase	0.029	Systemic bank anchor

4.3. Banking stability prediction

The banking stability domain uses the World Bank GFDD panel (2000–2023, N = 6384 country-year observations) to predict NPL ratios (table 5). To ensure a rigorous evaluation, the RF model is benchmarked against a logistic regression classifier (using a binarised NPL threshold of 10%) and a panel fixed-effects OLS model—both standard baselines in the banking stability literature (Holopainen & Sarlin, 2017). The panel fixed-effects model achieves an R^2 of 0.091 and

RMSE of 7.43 percentage points; the logistic regression achieves an AUC of 0.68. The RF model achieves an out-of-sample R^2 of 0.224 and RMSE of 4.92 percentage points—a 33.8% reduction in RMSE relative to the fixed-effects benchmark—alongside an AUC of 0.81 in the binarised classification task. The moderate absolute R^2 of 0.224 reflects the inherent heterogeneity of cross-country NPL dynamics and does not undermine the strong relative improvement over conventional econometric baselines.

Table 5. Banking stability prediction: random forest (World Bank GFDD, 2000–2023).

Metric	Value
Model	Random forest (200 trees)
Target variable	Non-performing loans/total loans (%)
Test observations	1277 country-year obs.
Out-of-sample R^2	0.224
RMSE	4.92 percentage points
Top predictor	Bank concentration (31.0%)
2nd predictor	Interest rate spread (25.3%)
3rd predictor	Domestic credit/private sector (17.2%)

Bank concentration—the share of assets held by the five largest banks—is the dominant NPL predictor, accounting for 31.0% of RF feature importance. The economic mechanism operates through the "competition-fragility" channel: highly concentrated banking systems face reduced competitive pressure on loan pricing, which weakens monitoring incentives and encourages the extension of credit to lower-quality borrowers who benefit from implicit "too-big-to-fail" guarantees (Berger et al., 2009). Concentrated systems also suffer greater correlated loan losses during cyclical downturns because their portfolios are less geographically and sectorally diversified. This finding carries direct implications for macroprudential policy: concentration limits and structural separation requirements may be as important as countercyclical capital buffers in containing systemic credit risk. Interest rate spreads contribute 25.3% of predictive power, reflecting the credit channel mechanism: elevated lending-deposit rate differentials signal tightening credit conditions, rising funding costs for borrowers, and the progressive deterioration of debt serviceability, particularly for variable-rate loans (Bernanke & Gertler, 1995). Domestic credit to the private sector (17.2%), bank liquid reserves (13.3%), and broad money-to-reserves ratios (13.2%) complete the ranking. As with all feature importance results reported in this study, these values reflect predictive association and should not be interpreted as causal estimates.

Table 6. Global mean non-performing loan ratio by year (GFDD panel, 2014–2023).

Year	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
Mean NPL (%)	7.23	7.71	7.83	7.93	7.40	7.04	7.19	6.35	5.95	5.48

Table 6 reports aggregate NPL trends across the GFDD panel from 2014 to 2023. The declining trajectory from 7.83% (2016) to 5.48% (2023) reflects post-crisis balance-sheet repair in many banking systems, though elevated ratios in several emerging-market economies warrant ongoing supervisory attention. The panel-wide means reported in Table 6 appear elevated relative to advanced-economy norms (where NPL ratios typically range between 1% and 4%) because the GFDD panel covers 267 countries and territories, including economies in Sub-Saharan Africa, South and Southeast Asia, and the European periphery, regions where post-crisis NPL ratios have historically been considerably higher. The mean values are consistent with published World Bank aggregate statistics for this panel composition. All NPL values were extracted directly from the World Bank GFDD database and independently verified against the IMF International Financial Statistics; data sources and extraction methodology are documented in Appendix A.

4.4. Islamic finance comparative performance

Table 7 reports risk-adjusted performance metrics for HLAL against the S&P 500 over 2019–2024: an interval encompassing the COVID-19 crash (February–March 2020), post-pandemic recovery, and the 2022 Federal Reserve monetary tightening cycle.

Table 7. Risk-adjusted performance: HLAL Shariah ETF vs. S&P 500 (July 2019 to December 2024).

Metric	HLAL (Wahed Shariah ETF)	S&P 500	Difference
Annualised volatility	20.63%	20.75%	−0.12%
Sharpe ratio (rf = 0)	0.819	0.702	0.117
Maximum drawdown	−33.57%	−33.92%	0.35 pp
HLAL–SP500 correlation	0.968	—	—
Annualised return	15.89%	—	—
Observation period	Jul 2019 to Dec 2024	Jul 2019 to Dec 2024	—

HLAL generates a Sharpe ratio of 0.819, exceeding the S&P 500's 0.702 by 11.7 basis points (a 16.7% improvement in risk-adjusted terms). To assess whether this difference is statistically meaningful, a Jobson–Korkie (1981) test with bootstrap correction (following Memmel, 2003) is applied; the test yields a z-statistic of 1.84 ($p = 0.066$), indicating that the Sharpe ratio difference is marginally significant at the 10% level given the relatively short five-year observation window. A rolling-window analysis across three sub-periods—pre-pandemic (July 2019 to February 2020), pandemic and recovery (March 2020 to December 2021), and post-pandemic tightening (January 2022 to December 2024)—reveals that HLAL's Sharpe advantage is most pronounced during the recovery period (HLAL: 1.42 vs. S&P 500: 1.21) and is reversed modestly during the 2022 tightening cycle, suggesting that the performance differential is cyclically conditioned rather than structurally invariant. This superior performance is achieved with virtually identical annualised volatility (HLAL: 20.63%; S&P 500: 20.75%) and marginally lower maximum drawdown (HLAL: −33.57%; S&P 500: −33.92%). The near-zero volatility differential is consistent with prior evidence that sector-exclusion screening does not materially alter return variance for diversified indices (Walkshäusl & Lobe, 2012). Economically, the downside protection embedded in Shariah-compliant screening operates through two mechanisms: the prohibition of Gharar (excessive uncertainty) encourages exclusion of highly leveraged and financially opaque firms, reducing exposure to the most distress-prone securities; and the sector exclusions (financial institutions with interest-based income, certain commodities) reduced HLAL's exposure to the banking sector selloff of 2020 and the energy sector volatility of 2022. The HLAL–S&P 500 return correlation of 0.968 confirms broad equity market participation under Shariah-compliant screening.

The performance differential between HLAL and the S&P 500 over the study period appears to be linked, at least in part, to the portfolio composition that results from Shariah screening. The structured filtering criteria embedded in HLAL produced a portfolio with a meaningful tilt towards technology and healthcare sectors, areas that generally performed well over the 2019–2024 window. This study records and analyses this outcome empirically, without making any judgment on the suitability of these instruments from a religious or jurisprudential perspective. HLAL's annualised return of 15.89% over the observation window is consistent with the broader appreciation of screened U.S. equity indices during this period.

4.5. Macroeconomic ETF correlation structure

Table 8 presents the inter-asset correlation matrix for macroeconomic ETF proxies (2002–2024). The 20-year Treasury bond and 7-year Treasury bond exhibit a correlation of 0.912, confirming the dominant duration-risk factor in government debt pricing. The negative correlations of Treasury bonds with oil (−0.243) and commodities (−0.188) reproduce the classical flight-to-safety dynamic. Gold exhibits moderate positive correlation with commodities (0.426), consistent with its dual role as an inflation hedge and safe-haven asset. The gold–Treasury bond correlation of 0.166 suggests that gold-backed Islamic instruments provide a differentiated diversification profile relative to interest-bearing sovereign debt—an asymmetry that ML architectures can exploit through cross-domain feature interactions.

Table 8. Macroeconomic ETF pairwise correlation matrix (2002–2024).

	TBond 20 years	TBond 7 years	TIPS	Gold	Oil	Commodity	Real estate
TBond 20 y	1.000	0.912	0.702	0.166	−0.243	−0.188	−0.165
TBond 7 y	0.912	1.000	0.762	0.213	−0.213	−0.150	−0.155
TIPS	0.702	0.762	1.000	0.274	0.030	0.094	−0.051
Gold	0.166	0.213	0.274	1.000	0.197	0.426	0.057
Oil	−0.243	−0.213	0.030	0.197	1.000	0.741	0.238
Commodity	−0.188	−0.150	0.094	0.426	0.741	1.000	0.268
Real Estate	−0.165	−0.155	−0.051	0.057	0.238	0.268	1.000

5. Discussion

5.1. Cross-domain insights and comparison with the literature

The unified ML framework produces several cross-domain insights that a siloed analytical approach would systematically miss. First, the feature importance analysis for VIX forecasting reveals that safe-haven assets—gold and long-duration Treasurys—carry substantial forward-looking information about equity market volatility. This cross-asset predictive linkage is precisely the kind of structural pattern that ML architectures, trained on multi-source feature vectors, are designed to capture. Conventional financial models, calibrated on within-market data, cannot exploit this inter-market signal. My finding that gold ranks as the second most important VIX predictor is broadly consistent with Baur and Lucey (2010), who documented gold's role as a hedge against equity market stress, and with Brunnermeier and Pedersen (2009), whose liquidity spiral framework predicts that safe-haven demand rises as market-wide funding conditions tighten. The RMSE performance of my RF model (6.13) compares favourably with benchmark volatility forecasting studies: Hamid and Habib (2014) reported RMSE values in the range of 8–12 for comparable VIX forecasting horizons using GARCH-class models, and Kim and Won (2018) achieved RMSE reductions of 40%–55% over GARCH benchmarks using SVR models—a smaller gain than the 71.4% reduction over GARCH(1,1) reported here, though the difference is partly attributable to my richer cross-asset feature set.

Second, the banking stability analysis demonstrates that market structure variables (bank concentration) dominate flow variables (credit growth) in predicting NPL accumulation across countries and periods. My finding that bank concentration is the leading predictor (31.0% importance) is consistent with Ghosh (2015), who identified bank size and concentration as significant NPL determinants in a panel of U.S. commercial banks, and with Klein (2013), who found that macroeconomic deterioration interacts with structural banking system features to amplify NPL accumulation across Central, Eastern, and South-Eastern European economies. The dominance of interest rate spreads as the second predictor (25.3%) aligns with Beck et al. (2013), who found that higher net interest margins—a closely related variable—are associated with greater risk-taking in concentrated systems. My RF model's R^2 of 0.224 is comparable to the 0.19–0.28 range reported by Ozili (2019) using conventional panel regression on similar GFDD data, but my model achieves this with substantially lower RMSE, reflecting the ability of RF to capture nonlinear threshold effects in credit deterioration. These findings carry specific implications for three practitioner communities, as follows: (1) Regulators and macroprudential policymakers should consider bank concentration metrics as early warning indicators within Basel IV–aligned stress testing frameworks and countercyclical buffer calibration; the predictive dominance of structural over flow variables suggests that ex-ante structural policy (merger scrutiny, market-share caps) may be more effective than ex-post capital requirements alone. (2) Commercial bank risk managers can incorporate the RF feature importance ranking into IFRS 9 Expected Credit Loss model development: the five leading predictors identified here provide a parsimonious feature set for Stage 2 and Stage 3 migration models in cross-border portfolio contexts. (3) Sovereign debt and regulatory analysts in emerging markets should pay particular attention to interest rate spread trajectories as leading NPL indicators, given their second-ranked predictive importance across the full 267-country GFDD panel.

Third, the Islamic finance analysis reveals a notable alignment between Shariah-compliant equity screening and measured investment outcomes. This result is broadly consistent with Hassan and Girard (2011) and Ho et al. (2014), who found that Islamic equity indices are broadly comparable to—and, in several sub-periods, superior to—their conventional counterparts on risk-adjusted metrics, though earlier studies covering pre-2015 data tend to find smaller performance differentials. The relatively stronger outperformance in the present study's 2019–2024 window appears linked to the technology-sector tilt that results from Shariah screening (exclusion of conventional financial institutions, tobacco, and certain commodities), which benefited HLAL during the 2020–2021 technology equity rally. Compared with Walkshäusl and Lobe (2012), who found broadly neutral risk-adjusted performance for Islamic indices globally, my result suggests that the specific U.S. equity universe and the 2019–2024 sample period are particularly favourable for Shariah-screened instruments. For Islamic finance investors and fund managers, these findings have practical implications: the Shariah-compliance feature engineering approach demonstrated in this study—treating screening criteria as a structured binary variable within an ML feature vector—can be operationalised in ETF screening algorithms and robo-advisory platforms operating under Shariah mandates.

5.2. Islamic finance: beyond ethical constraint

The conventional framing of Islamic finance as primarily a set of prohibitions—against Riba, Gharar, and Maysir—understates the practical implications of its screening methodology. The empirical results of this study show that Shariah-compliant screening, as a structured and rule-based filtering process, correlates with superior risk-adjusted performance over the 2019–2024 period. It is important to note that this study does not evaluate or make any judgment on the permissibility of conventional financial instruments under Islamic law; its purpose is analytical.

For scholars and practitioners in Islamic finance, the empirical results of this study provide a data-driven foundation for engaging with wider debates on the performance characteristics of Shariah-compliant investment vehicles. The principle of Mizan (balance) and the concern for Zulm (exploitation or injustice) are recognised here in their analytical dimension as values embedded in screening criteria that shape portfolio composition and, in turn, its observable performance. All jurisprudential interpretation of these findings is respectfully left to qualified Islamic scholars.

5.3. Limitations and future research

Several limitations warrant acknowledgement. First, the Islamic finance ETF analysis is constrained by the short history of Shariah-compliant U.S.-listed ETFs: HLAL was launched July 2019, SPUS December 2019, and UMMA 2021, limiting the study period to approximately five years and precluding coverage of the 2008–2009 global financial crisis. Second, the banking stability models achieve a moderate R^2 of 0.224, reflecting the inherent heterogeneity of cross-country NPL dynamics; country-specific regulatory environments and NPL classification standards introduce unobserved heterogeneity not captured by the available GFDD indicators. Third, formal LSTM temporal sequence modelling was not implemented due to temporal alignment challenges between the annual GFDD panel and the high-frequency ETF datasets; future work should address this through mixed data sampling (MIDAS) methods. Finally, random forest feature importance measures reflect predictive association rather than causal identification; causal ML methods (double-debiased ML, targeted regularisation) should be applied in future extensions.

6. Conclusions

This paper has demonstrated that a unified ML framework—combining random forest and XGBoost models applied to a multi-source big data environment spanning equity markets, banking systems, macroeconomic proxies, and Islamic finance instruments—delivers empirically validated improvements over traditional econometric benchmarks across multiple financial domains.

The principal empirical findings are as follows: (1) Random forest achieves a 78.3% reduction in RMSE relative to OLS in 20-day VIX forecasting, with NASDAQ, gold, and Treasury bonds as dominant predictive features; (2) bank market concentration and interest rate spreads are the leading structural predictors of non-performing loan accumulation across 267 countries over 2000–2023, with RF achieving an out-of-sample R^2 of 0.224; (3) Shariah-compliant equity ETFs exhibit a Sharpe ratio of 0.819 vs. 0.702 for the S&P 500 over 2019–2024, a 16.7% improvement in risk-adjusted performance, with virtually identical volatility and marginally superior drawdown protection, though this differential is specific to the 2019–2024 window and is cyclically conditioned rather than structural; and (4) the macroeconomic ETF correlation structure confirms structurally negative Treasury-commodity correlations and gold diversification properties that a ML feature architecture can exploit for cross-domain signal extraction.

The broader contribution of this study is methodological: it provides a proof-of-concept for the integration of Islamic finance screening as a quantifiable, machine-readable feature within a big data analytics framework. Shariah-compliance, traditionally operationalised as a binary legal criterion, can be translated into a continuous risk dimension that ML models can process alongside conventional market signals. This translation opens a research frontier at the intersection of data science, financial economics, and Islamic political economy.

Future research should extend the temporal coverage of Islamic finance ETF data as it accumulates, incorporate formal LSTM sequence modelling for multi-domain temporal forecasting, and apply causal ML methods to disentangle predictive association from structural causation in both the banking stability and Islamic finance domains.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Conflict of interest

The authors declare no conflict of interest.

References

Andersen TG, Bollerslev T, Diebold FX, et al. (2003) Modeling and forecasting realized volatility. *Econometrica* 71: 579–625. https://doi.org/10.1111/1468-0262.00418

Athey S (2017) Beyond prediction: Using big data for policy problems. *Science* 355: 483–485. https://doi.org/10.1126/science.aal4321

Athey S, Imbens GW (2019) Machine learning methods that economists should know about. *Annu Rev Econ* 11: 685–725.

Berge TJ, Jordà Ò (2011) Evaluating the classification of economic activity into recessions and expansions. *Am Econ J-Macroecon* 3 : 246–277. https://doi.org/10.1257/mac.3.2.246Berger AN, Klapper LF, Turk-Ariss R (2009) Bank competition and financial stability. *J Financ Serv Res* 35: 99–118. https://doi.org/10.1007/s10693-008-0050-7

Bernanke B, Gertler M (1995) Inside the black box: The credit channel of monetary policy transmission. *J Econ Perspect* 9: 27–48. https://doi.org/10.1257/jep.9.4.27

Breiman L (2001) Random forests. *Mach Learn* 45: 5–32. https://doi.org/10.1023/A:1010933404324

Brunnermeier MK, Pedersen LH (2009) Market liquidity and funding liquidity. *Rev Financ Stud* 22: 2201–2238. https://doi.org/10.1093/rfs/hhn098

Bussière M, Fratzscher M (2006) Towards a new early warning system of financial crises. *J Int Money Financ* 25: 953–973. https://doi.org/10.1016/j.jimonfin.2006.07.007

Chen T, Guestrin C (2016) XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

Cont R (2011) Statistical modeling of high-frequency financial data. *Ieee Signal Proc Mag* 28: 16–25. https://doi.org/10.1109/MSP.2011.941548

Coulombe PG, Leroux M, Stevanovic D, et al. (2022) How is machine learning useful for macroeconomic forecasting? *J Appl Economet* 37: 920–964. https://doi.org/10.1002/jae.2910

Estrella A, Mishkin FS (1998) Predicting U.S. recessions: Financial variables as leading indicators. *Rev Econ Stat* 80: 45–61. https://doi.org/10.1162/003465398557320

Fathurahman H, Fitriati R (2013) Comparative analysis of return on sukuk and conventional bonds. *Am J Agr Econ* 3: 159–163.

Fischer T, Krauss C (2018) Deep learning with long short-term memory networks for financial market predictions. *Eur J Oper Res* 270: 654–669. https://doi.org/10.1016/j.ejor.2017.11.054

Goodell JW, Kumar S, Lim WM, et al. (2021) Artificial intelligence and machine learning in finance: Identifying foundations, themes, and future research directions. *Financ Res Lett* 42: 101786. https://doi.org/10.1016/j.jbef.2021.100577

Gu S, Kelly B, Xiu D (2020) Empirical asset pricing via machine learning. *Rev Financ Stud* 33: 2223-2273. https://doi.org/10.1093/rfs/hhaa009

Hassan MK (2002) Risk, return and volatility of faith-based investing: The case of the Dow Jones Islamic Index. *Proceedings of the 5th Harvard University Forum on Islamic Finance*, 43–67.

Hassan MK, Girard E (2011) Faith-based ethical investing: The case of Dow Jones Islamic indexes. *Islamic Econ Stud* 17: 1–31. https://doi.org/10.2139/ssrn.1808853

Ho CSF, Abd Rahman NA, Yusuf NHM, et al. (2014) Performance of global Islamic versus conventional share indices: International evidence. *Pac-Basin Financ J* 28: 110–121. https://doi.org/10.1016/j.pacfin.2013.09.002

Holopainen M, Sarlin P (2017) Toward robust early-warning models: A horse race, ensembles and model uncertainty. *Quant Financ* 17: 1933–1963. https://doi.org/10.1080/14697688.2017.1357972

Houlsby N, Giurgiu A, Jastrzebski S, et al. (2019) Parameter-efficient transfer learning for NLP. *Proceedings of the International Conference on Machine Learning (ICML)*, 2790–2799.

Huang W, Nakamori Y, Wang SY (2020) Forecasting stock market movement direction with support vector machine. *Comput Oper Res* 32: 2513–2522. https://doi.org/10.1016/j.cor.2004.03.016

Islamic Finance Development Report (2023) Resilience and continuity. Islamic Corporation for the Development of the Private Sector (ICD) / LSEG.

Johnes J, Izzeldin M, Pappas V (2014) A comparison of performance of Islamic and conventional banks 2004–2009. *J Econ Behav Organ* 103: S93–S107. https://doi.org/10.1016/j.jebo.2013.07.016

Khandani AE, Kim AJ, Lo AW (2010) Consumer credit-risk models via machine-learning algorithms. *J Bank Financ* 34: 2767–2787. https://doi.org/10.1016/j.jbankfin.2010.06.001

Kjosevski J, Petkovski M (2017) Non-performing loans in Baltic States: Determinants and macroeconomic effects. *Baltic J Econ* 17: 25–44. https://doi.org/10.1080/1406099X.2016.1246234

López de Prado M (2018) Advances in financial machine learning. John Wiley & Sons. https://doi.org/10.2139/ssrn.3365271

Mention AL, Bontis N (2013) Intellectual capital and performance within the banking sector of Luxembourg and Belgium. *J Intellect Cap* 14: 286–309. https://doi.org/10.1108/14691931311323896

Ozili PK (2019) Non-performing loans and financial development: New evidence. *J Risk Financ* 20: 59–81. https://doi.org/10.1108/JRF-07-2017-0112

Sherrill DE, Upton K (2018) Actively replacing passive benchmarks: A review of the U.S. market for Shariah-compliant ETFs. *Manag Financ* 44: 351–369. https://doi.org/10.1108/MF-03-2017-0067

Stock JH, Watson MW (2012) Disentangling the channels of the 2007–2009 recession. *Brookings Papers on Economic Activity*, Spring, 81–135.

Walkshäusl C, Lobe S (2012) Islamic investing. *Rev Financ Econ* 21: 53–62. https://doi.org/10.1353/eca.2012.0005

World Bank (2024) Global Financial Development Database (GFDD). Retrieved from https://www.worldbank.org/en/publication/gfdr/data.