Predictive Performance and Interpretability of Machine Learning Models in Renewable Energy Venture Capital

Matteo Micheloni

Dissertation written under the supervision of Professor Fatima Shuwaikh

Dissertation submitted in partial fulfilment of requirements for the MSc in Finance, at the Universidade Católica Portuguesa, 10th September 2025.

Predictive Performance and Interpretability of Machine Learning Models in Renewable Energy Venture Capital

Matteo Micheloni

Abstract

This paper explores whether machine learning (ML) models can provide superior predictive accuracy compared to traditional linear approaches in forecasting the success of venture capital (VC) investments in renewable energy startups. Relying on a unique dataset covering several decades of investments, the study benchmarks a broad range of models, from classical statistical methods to advanced ensemble and gradient boosting algorithms.

The analysis shows that modern ML techniques consistently outperform linear baselines, offering more reliable and robust predictions. Ensemble and boosting methods, in particular, demonstrate strong generalization capabilities, while simpler classifiers struggle to capture the complexity and heterogeneity of venture capital data.

Interpretability is ensured through SHAP (SHapley Additive exPlanations), which highlights the structural drivers most consistently associated with startup success. Deal characteristics, especially financing structures and stages of investment, emerge as the most influential predictors, together with investee-related attributes. Notably, early-stage equity deals and unstructured venture capital investments are systematically linked to higher risk.

The findings contribute to the literature by combining predictive performance with interpretability, demonstrating that ML can uncover consistent structural patterns in renewable energy venture financing. For practitioners, the study provides evidence-based insights to improve capital allocation decisions, reduce uncertainty, and support the scaling of green technologies. By bridging methodological innovation and sustainability challenges, this research emphasizes the growing role of data-driven approaches in advancing the global energy transition.

Keywords: Venture Capital, Machine Learning, Renewable Energy.

Predictive Performance and Interpretability of Machine Learning Models in Renewable Energy Venture Capital

Matteo Micheloni

Resumo

Este trabalho investiga se modelos de aprendizado de máquina (ML) podem oferecer maior precisão preditiva em comparação com abordagens lineares tradicionais na previsão do sucesso de investimentos de venture capital (VC) em startups de energia renovável. Com base em um conjunto de dados cobrindo várias décadas, o estudo compara desde métodos estatísticos clássicos até algoritmos avançados de ensemble e gradient boosting.

A análise mostra que técnicas modernas de ML superam as lineares, oferecendo previsões mais confiáveis e robustas. Métodos de ensemble e boosting apresentam forte capacidade de generalização, enquanto classificadores simples não captam a complexidade e heterogeneidade dos dados de VC.

A interpretabilidade é assegurada por meio do SHAP (SHapley Additive exPlanations), que evidencia fatores mais associados ao sucesso das startups. Características dos negócios, sobretudo estruturas de financiamento e estágios de investimento, surgem como preditores centrais, junto com atributos das empresas investidas. Notavelmente, operações de equity em estágio inicial e investimentos de VC não estruturados estão ligados a maior risco.

Os resultados contribuem à literatura ao unir desempenho preditivo e interpretabilidade, mostrando que o ML pode revelar padrões estruturais consistentes no financiamento de energia renovável. Para profissionais, o estudo oferece insights para melhorar decisões de capital, reduzir incertezas e apoiar a expansão de tecnologias verdes. Ao integrar inovação metodológica e sustentabilidade, esta pesquisa destaca o papel crescente das abordagens orientadas por dados na transição energética global.

Palavras-chave: Venture Capital, Aprendizado de Máquina, Energia Renovável.

Table of Contents

1. Introduction and Explanation of Research Objectives

As the global economy faces escalating environmental challenges, including intensifying climate change, depletion of non-renewable resources, and more frequent extreme weather events, financial institutions are under pressure to support a sustainable development trajectory (IPCC, 2022; IEA, 2023). Venture capital (VC) has emerged as a high-risk, high-reward mechanism that channels early-stage equity into innovative companies with strong growth potential (Gompers & Lerner, 2001; Kaplan & Strömberg, 2004). Unlike bank loans or public market investments, VC targets startups lacking collateral, cash flow, or track record, relying on innovation-driven growth and long-term profitability. Such investments are illiquid, long-term, , and the active involvement of investors in governance and strategy (Da Rin et al., 2013).

Within the renewable energy sector, startups are expected to pioneer technologies from photovoltaics and grid solutions to bioenergy and clean transportation (IEA, 2023; IRENA, 2022). Despite their potential, these ventures face persistent funding gaps, as capital remains concentrated in later-stage firms and mature markets (IEA, 2023). Long development cycles, high uncertainty, and information asymmetries make early-stage financing particularly challenging (Cumming & Johan, 2009; Polzin, 2017), leaving many ventures underfunded and slowing progress toward net-zero (IEA, 2023).

Meanwhile, VC is also gradually shifting toward data-driven practices (Brynjolfsson & McElheran, 2016; Fuster et al., 2018). Traditional methods based on intuition, networks, and subjective judgment are criticized for opacity and bias (Kaplan & Strömberg, 2001; Bengtsson & Hsu, 2015). In contrast, advances in machine learning (ML), a subset of artificial intelligence that enables systems to detect patterns in data and generate predictions without explicit programming, provide a promising avenue for enhancing the efficiency and objectivity of investment evaluation (Mullainathan & Spiess, 2017; Gu et al., 2020). Already widely applied in finance for credit scoring, fraud detection, and portfolio optimization (Fuster et al., 2018; Gu et al., 2020), ML models are particularly well-suited for VC applications, as they can process large volumes of structured data and capture complex, nonlinear interactions that traditional linear statistical methods may fail to detect (Bishop, 2006; Mullainathan & Spiess, 2017).

Despite this potential, applications of ML in VC, especially in renewable energy, remain underexplored. Existing research has focused on general startup evaluation (Kleinert et al., 2020), leaving a gap at the intersection of sustainable finance and AI. This thesis aimes to contribute by benchmarking a broad set of ML and linear models on renewable energy startups. Particularly, eleven ML models are tested, including advanced gradient boosting algorithms (XGBoost, LightGBM, CatBoost), proven effective on structured, high-dimensional data (Chen & Guestrin, 2016; Ke et al., 2017; Prokhorenkova et al., 2018). These algorithms capture nonlinearities, manage imbalances, and integrate diverse variables, crucial in VC where exits are rare (Hellmann & Puri, 2002; Da Rin et al., 2013). Moreover, their compatibility with interpretability frameworks as SHAP (Lundberg & Lee, 2017) makes them not only powerful but also transparent, increasing their practical.

This study relies exclusively on structured, quantifiable variables. Unlike qualitative signals (e.g., founder traits), prone to bias and difficult to standardize, structured indicators such as investment stage, deal type, funding amounts, and investor characteristics allow systematic comparison across samples (Cumming & Johan, 2009; Kaplan & Strömberg, 2004). This enables rigorous benchmarking of ML against linear models like Logistic Regression, Probit, and Linear Discriminant Analysis, widely used for interpretability and theoretical grounding (Hosmer et al., 2013).

The research is guided by a central question: *can machine learning approaches offer superior predictive accuracy compared to traditional linear statistical models in forecasting the success or failure of renewable energy ventures financed by venture capital?* From this guiding question emerge three subsidiary lines of inquiry. First, the study compares the relative performance of different modeling paradigms, testing whether algorithms capable of capturing nonlinearities and higher-order interactions can significantly outperform classical linear methods. Second, it examines interpretability by applying SHAP to identify which features consistently drive model predictions, thereby contributing to theoretical debates on the determinants of startup success (Hellmann & Puri, 2002; Kaplan & Strömberg, 2009). Third, it evaluates practical relevance, analyzing which models are most appropriate for different stakeholders, venture capitalists, accelerators, or policymakers, depending on whether their priorities lie in minimizing false positives, avoiding false negatives, or balancing both objectives in investment decision-making.

In doing so, the thesis addresses the convergence of two urgent imperatives: the global transition to low-carbon energy systems, which the International Energy Agency (2023) estimates will require annual clean energy investment to triple by 2030, and the methodological shift in finance toward advanced, interpretable machine learning (Brynjolfsson & McElheran, 2016; Mullainathan & Spiess, 2017). By systematically testing and interpreting predictive models, this study seeks to provide both empirical evidence and actionable insights to improve capital allocation, accelerate the scaling of renewable energy technologies, and contribute to the broader field of sustainable finance.

2. Review of Relevant Literature

Understanding the success of early-stage ventures in renewable energy requires engaging three strands of literature: the role of venture capital in startup performance, the use of machine learning in evaluation and predictive analytics, and the intersection of investment with environmental innovation. Together, they provide the conceptual and methodological basis for this research.

The impact of venture capital on startup success has long been studied. Hellmann and Puri (2002), analyzing Silicon Valley startups, show that VC-backed firms are more likely to undergo professionalization through formal business processes, equity-based pay, and executive recruitment, factors linked to faster growth and higher exits. Sørensen (2007) distinguishes sorting effects, where high-potential startups attract strong investors, from influence effects, whereby investors provide expertise, governance, and networks. Kaplan and Strömberg (2004) further emphasize that outcomes depend not only on investor involvement but also on startup traits such as market size, defensibility, and managerial competence. Many determinants, however, are qualitative and judgment-based, complicating systematic evaluation. This has motivated the search for structured, quantifiable indicators, a shift directly connected to the rise of machine learning methods in venture evaluation.

A growing literature has increasingly applied ML to predicting startup success, aiming to improve on limitations of traditional approaches. Żbikowski and Antosiuk (2021) stress avoiding look-ahead bias, showing many studies used features observable only after success. They compared logistic regression, SVM, and gradient boosting, with the latter performing best (F1 around 43%) and geography/industry most predictive. Arroyo et al. (2019) broadened scope by examining 120,000 early-stage Crunchbase companies. Unlike earlier works limited to IPOs or acquisitions, they modeled multiple outcomes, including follow-on funding and closure. Ensemble methods like Random Forest and Gradient Tree Boosting provided the best precision–recall balance, particularly for acquisitions, increasing investors' odds more than tenfold despite low absolute recall. Using Crunchbase and USPTO patent data, Ross et al. (2021) introduced CapitalVX, a deep-learning ensemble achieving out-of-sample accuracies of 80–90% for exit and follow-on funding prediction. Their model integrated firm, founder, and patent data, showing up to four times the accuracy of professional VCs in identifying IPOs. Importantly, they emphasized interpretability, using SHAP to highlight investor count, last funding round, and founder profiles. Finally, Zacharakis and Meyer (1998) compared actuarial models with expert judgments in VC decision-making, finding statistical models consistently outperformed humans. Collectively, these studies show ML's promise but also note risks of data leakage, overfitting, and lack of temporal validation. This thesis addresses these gaps by restricting inputs to structured, quantifiable variables, applying multiple models, and integrating interpretability analysis.

Interpretability forms a bridge between algorithm choice and practical application. While gradient boosting and other non-linear models deliver strong predictive performance, their complexity often raises concerns of opacity and trustworthiness (Rudin, 2019; Miller, 2019). To address this, several model-agnostic frameworks have been developed. Ribeiro, Singh and Guestrin (2016) introduced LIME (Local Interpretable Model-Agnostic Explanations), which approximates complex models locally by training interpretable surrogates for individual predictions. This method provides valuable case-specific insights, for example explaining why a startup is classified as promising, but later research noted limitations of instability, sensitivity to perturbations, and inconsistency across runs (Garreau & von Luxburg, 2020). Despite these weaknesses, LIME remains widely applied in finance, e.g., in forecasting and credit risk modelling (Nallakaruppan, et al. 2024), where localized interpretability is valued by regulators and practitioners. Building on these foundations, Lundberg and Lee (2017) developed SHAP (SHapley Additive exPlanations), now the standard for tree-based models. SHAP offers globally consistent and additive feature attributions grounded in cooperative game theory, ensuring mathematical soundness and stability across models (Lundberg et al., 2020; Molnar, 2022). Its advantages over LIME, particularly in providing coherent global explanations while retaining local fidelity, are emphasized in comparative reviews of explainable AI in finance. Structured features such as funding stage, geography, and sector classification lend themselves well to SHAP explanations, producing insights both statistically rigorous and practically meaningful. In line with this evidence, the present thesis employs SHAP not only to compare predictive performance across the fourteen models tested, but also to uncover the factors most strongly influencing predictions of startup success (Lundberg & Lee, 2017; Molnar, 2022). By doing so, it addresses the dual challenge of accuracy and interpretability, while acknowledging the complementary role LIME can play in providing localized, instance-specific insights.

Another recurring theme in the literature is the limitation of venture capital datasets. Databases as Refinitiv Eikon, Crunchbase, Dealroom, and PitchBook vary in coverage, update frequency, and completeness (Dalle et al., 2017). Survivorship bias often excludes failed startups, while missing or inconsistently reported rounds distort investment trajectories (Cumming & Johan, 2009). Coverage is skewed toward North America and Western Europe, limiting generalizability to Asia, Africa, or Latin America (Dalle et al., 2017). These biases are especially pronounced in clean energy, where Ghosh and Nanda (2010) note high capital intensity, long development horizons, and regulatory dependencies that deter investment and complicate data collection.

The literature on VC and sustainability adds depth by situating these methodological issues in the green innovation context. Ghosh and Nanda (2010) highlight why clean energy ventures face investor reluctance, citing capital intensity, long cycles, and regulatory risk. Bocken (2015) highlights how sustainability-driven ventures face misalignments with VC firms' short-term return expectations and emphasizes mismatches between environmental entrepreneurs and short-termist investors, while Jiang & Liu (2024) show green VC generates positive spillovers in environmental patenting. Schabek (2020) further show listed green firms, though volatile, can rival or outperform traditional benchmarks. Collectively, these studies underscore the strategic importance of efficient capital allocation and the potential of predictive models to reduce uncertainty in early-stage investment.

In summary, prior research establishes that VC supports startup growth, that ML improves predictive accuracy in early-stage investment, and that structured indicators are valuable predictors when systematically applied. Yet gaps remain: most studies are industry-agnostic and test a limited set of models. To our knowledge, no published work has benchmarked fourteen algorithms, spanning both traditional and modern approaches, on structured data specific to renewable startups. This thesis aims to addresses that gap by rigorously evaluating multiple models, applying SHAP to identify the most influential drivers of success, and framing results in terms of both predictive performance and practical decision support.

3. Data and Methodology

3.1 Data Sources and Scope

The empirical analysis is based on a dataset constructed from Refinitiv Eikon, using its screening functionality to identify venture capital transactions in the renewable energy sector. The sample includes companies classified under Renewable Fuels and Renewable Energy Equipment & Services, as well as electric utilities and independent power producers exclusively focused on renewable energy sources. After data cleaning and the exclusion of incomplete records, the final dataset consists of 1,778 startups. While this number may seem modest compared to broader venture capital studies, it reflects the niche nature of renewable energy ventures within the VC landscape, where investment activity is concentrated in fewer firms relative to more general technology sectors. This point also constitutes an inherent limitation, discussed later.

The temporal scope spans from 1 July 1975 to 28 February 2025, providing nearly five decades of observations. Importantly, the dataset refers to the last recorded investment for each startup, rather than all financing rounds, thereby avoiding double counting and ensuring that the analysis captures the most relevant financing and deal characteristics available in the historical record. The unit of analysis is therefore the investee company at the time of its last recorded investment.

The unit of analysis is thus the investee company at the time of its last recorded investment. Success is defined as either (i) the receipt of a new round of investment within two years of this event, or (ii) a positive exit outcome, such as an acquisition, IPO, or comparable liquidity event. This operationalization reflects two concrete signals of venture viability: sustained investor confidence and realized liquidity.

3.2 Features and Variable Construction

The dataset integrates four broad categories of structured, quantifiable variables. The first relates to investee company characteristics, including funding and investment history such as total funding, number of investments, and timing of the first investment, firm attributes such as founding year and customer type, and country attractiveness as proxied by the Renewable Energy Country Attractiveness Index (RECAI). Sector alignment with the investment target of the last-round investor is also considered. The second category refers to investor firm characteristics, covering historical activity such as first and last investments, number of deals, number of companies invested, and total equity invested, as well as investor type, stage preferences, founding year, and RECAI scores of the investor's country. The third category includes fund characteristics, such as fund type (venture capital, buyout, generalist PE, other) and vintage year. The final category concerns deal characteristics, including deal type (VC, buyout, mixed), security structure such as common stock, convertible bonds, and leveraged buyout financing, financing stage such as seed, expansion, late stage, acquisition, or public market, and round-level details including round number, deal age, and number of participating funds. All categorical variables were transformed into dummy variables through one-hot encoding, enabling consistent use across classifiers. Records with missing values were excluded to preserve integrity. No scaling was applied to numerical features, since tree-based and ensemble methods are robust to differences in magnitude (Breiman et al., 1984; Breiman, 2001; Friedman, 2001).

In constructing the dataset, this thesis deliberately departs from a strictly time-aware validation framework. While many features are based on information observable at the time of the first investment, the analysis also incorporates variables from the last recorded investment, capturing the characteristics of the investee company, the investor firm and its fund, as well as the deal itself. This choice is intentional: the aim is not only to approximate an ex-ante investor decision context, but also to uncover structural patterns in financing, syndication, and deal design that have historically been associated with successful outcomes. Importantly, the inclusion of such variables does not undermine the predictive value of the models. Investors and policymakers can still apply the framework in a forward-looking manner by conditioning predictions on alternative deal characteristics or financing structures, thereby using the model to evaluate the likelihood of success under different scenarios.

3.3 Modeling Approach

A distinctive element of this research lies in the systematic evaluation of machine learning algorithms that are particularly well suited to venture capital data, especially gradient boosting families such as XGBoost, LightGBM, and CatBoost (Chen & Guestrin, 2016; Ke et al., 2017; Prokhorenkova et al., 2018). These methods have been shown to handle structured, high-dimensional datasets effectively, to capture non-linear interactions, and to remain robust under sparse and imbalanced conditions, characteristics that mirror the challenges of early-stage investment in renewable energy (Friedman, 2001). Building on this foundation, the present study expands the scope beyond boosting methods to a comprehensive benchmark of fourteen models spanning traditional linear statistical approaches, non-linear algorithms, tree-based learners, ensemble techniques, and hybrid meta-models.

Specifically, traditional linear statistical approaches include Logistic Regression, Probit Regression, and Linear Discriminant Analysis (LDA), which serve as classical baselines and remain widely used for their interpretability and theoretical rigor (Hosmer, Lemeshow & Sturdivant, 2013; Finney, 1947; Fisher, 1936). Tree-based methods comprise the Decision Tree and Random Forest, which capture hierarchical patterns in the data and account for interactions between variables in a structured manner (Breiman et al., 1984; Breiman, 2001). Ensemble boosting methods include XGBoost, LightGBM, and CatBoost, representing the state of the art in tabular learning and particularly relevant in finance and VC applications (Chen & Guestrin, 2016; Ke et al., 2017; Prokhorenkova et al., 2018). Other non-linear approaches cover Support Vector Machines (SVM) (Cortes & Vapnik, 1995), K-Nearest Neighbors (KNN) (Cover & Hart, 1967), and the Multi-Layer Perceptron (MLP) (Rumelhart, Hinton & Williams, 1986), which are suited to more flexible, non-linear decision boundaries and can model higher-order interactions (Hastie, Tibshirani & Friedman, 2009). Probabilistic statistical models such as Naïve Bayes offer a contrasting benchmark rooted in conditional independence assumptions (Domingos & Pazzani, 1997), while meta-models, the Voting Classifier and the Stacking Classifier, combine the strengths of multiple learners to enhance robustness and generalization (Dietterich, 2000; Wolpert, 1992). This broad benchmark allows for systematic comparisons both within machine learning families and against traditional approaches, directly addressing the central research question of whether advanced methods can genuinely outperform classical baselines in the renewable energy venture capital domain.

To assess performance, the dataset was split into training (80%) and test (20%) sets using a stratified split. Furthermore, the dataset includes both successful and unsuccessful ventures, with the latter accounting for roughly half of the sample. Stratification ensures that the relative balance of successful and unsuccessful startups is preserved across both sets, which is considered best practice in classification tasks (Hastie, Tibshirani & Friedman, 2009). Since the dataset is approximately balanced, no resampling or reweighting was applied. Although it cannot be guaranteed that reporting in Refinitiv Eikon is fully complete or error-free, the inclusion of unsuccessful cases reduces survivorship bias and strengthens the empirical basis of the analysis.

To further reduce overfitting, stratified k-fold cross-validation was applied during the tuning of more complex models (Random Forest, XGBoost, LightGBM, CatBoost, SVM, KNN). This procedure repeatedly partitions the training set into k folds, using k–1 folds for training and one for validation, ensuring that every observation is used for validation exactly once. Such an approach increases robustness by averaging performance across multiple folds (Kuhn & Johnson, 2013). Unlike some prior studies, this thesis does not adopt time-aware cross-validation (rolling origin or blocked folds), since variables related to the last recorded investment are deliberately included and extend beyond a purely ex-ante setting.

Finally, hyperparameter tuning was deliberately limited. Most models were trained using default parameters provided by Python's scikit-learn and specialized libraries, with manual adjustments only where needed for stability, such as increasing maximum iterations for MLP and logistic regression, or setting the number of trees in ensemble models to 100. Systematic searches such as grid search, random search, or Bayesian optimization (Optuna) were not implemented. This decision was taken to preserve comparability across a large set of models: highly tuned individual models might achieve marginally higher scores, but at the cost of introducing bias into cross-model comparisons (Bergstra & Bengio, 2012). As noted by Bergstra & Bengio (2012), hyperparameter optimization can substantially alter relative rankings, making it less suitable when the research goal is benchmarking rather than deployment.

3.4 Evaluation Metrics and Interpretability Framework

Performance was assessed using five core metrics widely recommended for binary classification in finance and venture capital (VC) contexts: Accuracy, Precision, Recall, F1 Score, and the Receiver Operating Characteristic Area Under the Curve or ROC-AUC (Powers, 2011; Fawcett, 2006). These indicators capture complementary aspects of model performance. Accuracy reflects overall correctness, indicating the proportion of correct predictions among all cases. Precision emphasizes the reliability of predicted positives, which is particularly important for investors seeking to minimize false positives when selecting startups. Recall measures the ability of a model to capture true positives, reflecting the proportion of successful ventures correctly identified. F1 Score provides a balanced measure that accounts for the trade-off between precision and recall, and is especially valuable in situations where both false positives and false negatives carry significant implications. Finally, ROC-AUC evaluates a model's ability to rank cases consistently across decision thresholds, offering a threshold-independent measure of discriminative capacity (Fawcett, 2006).

A recurring challenge in startup prediction tasks concerns class imbalance, as successful ventures typically represent only a minority of cases. In such contexts, high accuracy can be misleading: a model predicting only failure may still achieve a high score. To overcome this limitation, the literature emphasizes complementary metrics such as ROC-AUC, Precision, Recall, and F1, which provide a more nuanced view of performance (Naidu et al., 2023). In highly imbalanced applications, further tools such as the area under the precision–recall curve (AUPRC), log loss, or calibration analysis are recommended to assess probabilistic reliability (Davis & Goadrich, 2006; Niculescu-Mizil & Caruana, 2005). In this study, however, the dataset includes both successful and unsuccessful ventures in nearly equal proportions, which reduces the severity of imbalance concerns. For this reason, the evaluation framework focuses on the five primary metrics, Accuracy, Precision, Recall, F1 Score, and ROC-AUC, since these are sufficient to capture the trade-offs most relevant for investors. Other measures were considered but deemed unnecessary. For instance, log loss was employed internally as the default training objective for gradient boosting models (XGBoost, LightGBM, CatBoost), yet it was excluded from the comparative evaluation to maintain consistency across models.

Beyond predictive accuracy, interpretability represents a central concern in applying machine learning to finance, where opaque "black-box" models often face skepticism from investors and regulators (Rudin, 2019; Miller, 2019). To address this, the study adopts SHAP (SHapley Additive exPlanations) as the primary interpretability tool (Lundberg & Lee, 2017). SHAP decomposes individual predictions into additive feature contributions, providing both global insights, through bar plots of mean absolute SHAP values that identify the most influential features on average, and local insights, through summary dot plots that illustrate directionality and heterogeneity of effects across ventures (Lundberg et al., 2020; Molnar, 2022). While permutation importance was occasionally used as a robustness check, SHAP was prioritized for its cooperative game-theoretic foundation and growing adoption in financial machine-learning applications. Applying SHAP consistently across all fourteen models ensures that predictive performance is accompanied by transparency, making results more interpretable and actionable for investors, accelerators, and policymakers.

3.5 Robustness

The robustness of this study's findings rests on the interplay between dataset design, validation strategies, and the intrinsic properties of the algorithms employed. Given complexities as heterogeneity, imbalance between successful and unsuccessful ventures, and nonlinear feature interactions, robustness is a central concern in the evaluation of predictive models.

First, robustness is supported by the composition of the dataset itself. The sample includes both successful and unsuccessful ventures, with a distribution of approximately 48% successful and 52% unsuccessful. This partially mitigates survivorship bias, a common limitation in venture datasets where failed firms are often underreported (Cumming & Johan, 2009). However, since failure rates in venture capital are typically much higher in reality, the near-balanced distribution likely reflects reporting practices in Refinitiv rather than the true underlying dynamics of the market (Gompers & Lerner, 2001). As such, while the dataset offers a more representative mix than success-only samples, survivorship bias cannot be fully excluded. Moreover, the dataset was carefully cleaned and validated to minimize reporting errors, although inconsistencies in Refinitiv Eikon cannot be entirely excluded. Cross-checks were conducted to identify outliers and implausible values in key variables such as investment year, funding amounts, and deal structures. These steps reduce the likelihood that the results are driven by data quality issues rather than substantive relationships.

Second, robustness is influenced by the validation strategy. While no explicit k-fold cross-validation was applied at the study level, the majority of the algorithms embed internal resampling or regularization mechanisms that help prevent overfitting. Ensemble-based methods such as Random Forest, XGBoost, LightGBM, and CatBoost incorporate strong safeguards against variance. Random Forest achieves stability by averaging across multiple decision trees trained on bootstrapped samples, while boosting algorithms iteratively refine errors and employ regularization penalties (Breiman, 2001; Chen & Guestrin, 2016; Ke et al., 2017; Prokhorenkova et al., 2018). Mixed models, such as the Voting and Stacking Classifiers, further enhance robustness by combining predictions from multiple base learners, thereby reducing the risk that idiosyncratic weaknesses of individual algorithms dominate overall results (Dietterich, 2000; Kuncheva, 2004). In particular, Voting classifiers achieve variance reduction through majority aggregation, while Stacking leverages a meta-model to learn from the strengths and weaknesses of base models, improving stability and generalization (Wolpert, 1992).

Linear statistical models, including Logistic Regression, Probit Regression, and Linear Discriminant Analysis, provide robustness of a different nature. By enforcing strict functional forms and assumptions, such as linearity or homoscedasticity, these models avoid high variance fits and remain stable across samples (Hosmer et al., 2013; Fisher, 1936; Finney, 1947). However, this robustness comes at the cost of flexibility, making such models vulnerable when confronted with complex, nonlinear venture data (Hastie, Tibshirani & Friedman, 2009). Neural and instance-based models occupy an intermediate position: Multi-Layer Perceptrons (MLPs) achieve robustness through weight regularization and early stopping (Hastie et al., 2009), whereas K-Nearest Neighbors (KNN) lacks built-in variance control and remains highly sensitive to the choice of k and dimensionality (Beyer et al., 1999).

Third, robustness was reinforced through the interpretability analysis. The use of SHAP values not only confirmed the predictive relevance of features across multiple algorithms but also revealed a consistent set of core drivers appearing in both machine learning and linear models. This convergence across modeling paradigms enhances confidence that results are not merely artifacts of a single algorithm's inductive biases, but rather reflect genuine structural patterns within renewable energy venture financing.

4. Quantitative Performance Evaluation

4.1 Evaluation Metrics

To systematically evaluate the predictive models in this study, five widely accepted classification metrics are used: Accuracy, Precision, Recall, F1 Score, and the Area Under the Receiver Operating Characteristic Curve (ROC-AUC). Each captures a different perspective on performance, and together they provide a comprehensive assessment of predictive quality in renewable energy VC investments (Fawcett, 2006; Sokolova & Lapalme, 2009).

Accuracy measures the overall proportion of correct predictions both successful and unsuccessful investments. Formally, it is the sum of true positives and true negatives divided by all cases. While intuitive, accuracy can be misleading in imbalanced datasets where failures outnumber successes (Chawla et al., 2002). For instance, if only 20% of startups succeed, always predicting failure still yields 80% accuracy. Thus, accuracy may overstate real predictive capability in the presence of class imbalance.

Precision is the ratio of true positives to all positive predictions, effectively measuring how often a model's prediction of success is correct. In VC, high precision means the model rarely misallocates capital to failing ventures. CatBoost and the Voting Classifier, with precision above 0.90, show strong reliability in highlighting genuinely promising opportunities. However, precision does not account for missed opportunities (false negatives), which are also costly in the VC setting.

Recall, or sensitivity, captures the proportion of actual successes correctly identified, highlighting a model's ability to avoid false negatives. In VC, high recall ensures high-potential startups are not missed (Gompers & Lerner, 2001). However, maximizing recall may do so by predicting many false positives, reducing efficiency (Sokolova & Lapalme, 2009). This trade-off was evident in models like Naïve Bayes, which achieved high precision (0.8374) but extremely low recall (0.3552), suggesting that it flagged few startups as successful, but was often right when it did.

F1 Score, the harmonic mean of precision and recall, providing a balanced measure when both false positives and false negatives are costly. This is often the case in VC screening, where funding failures and overlooking successes are equally costly (Powers, 2011). XGBoost's F1 score of 0.8772 reflects strong performance in balancing precision and recall, making it a compelling candidate for investment triage systems.

ROC-AUC evaluates a model's discriminative ability across thresholds. It plots the true positive rate against the false positive rate, with the area under the curve reflecting how well the model separates successes from failures (Fawcett, 2006). AUC close to 1.0 indicates excellent separability. ROC-AUC is particularly valuable in imbalanced settings, as it is not tied to a specific threshold and considers performance across a spectrum of decision boundaries (Bradley, 1997), ROC-AUC provides a more reliable and robust view of a model's true predictive value in high-risk, asymmetrically distributed datasets like those in venture capital (Fawcett, 2006; Bradley, 1997).

In summary, while accuracy offers a general snapshot, it is insufficient alone in this domain. ROC-AUC, due to its threshold-agnostic nature, serves as the most comprehensive performance indicator, especially in early-stage green venture contexts where class imbalance and prediction uncertainty are prevalent. Precision and recall, viewed alongside the F1 Score, allow for deeper insights into each model's behavioral trade-offs, making them powerful decision tools. Exhibit 1 reports the numerical results obtained for each model across ROC-AUC, Accuracy, Precision, Recall and F1 Score.

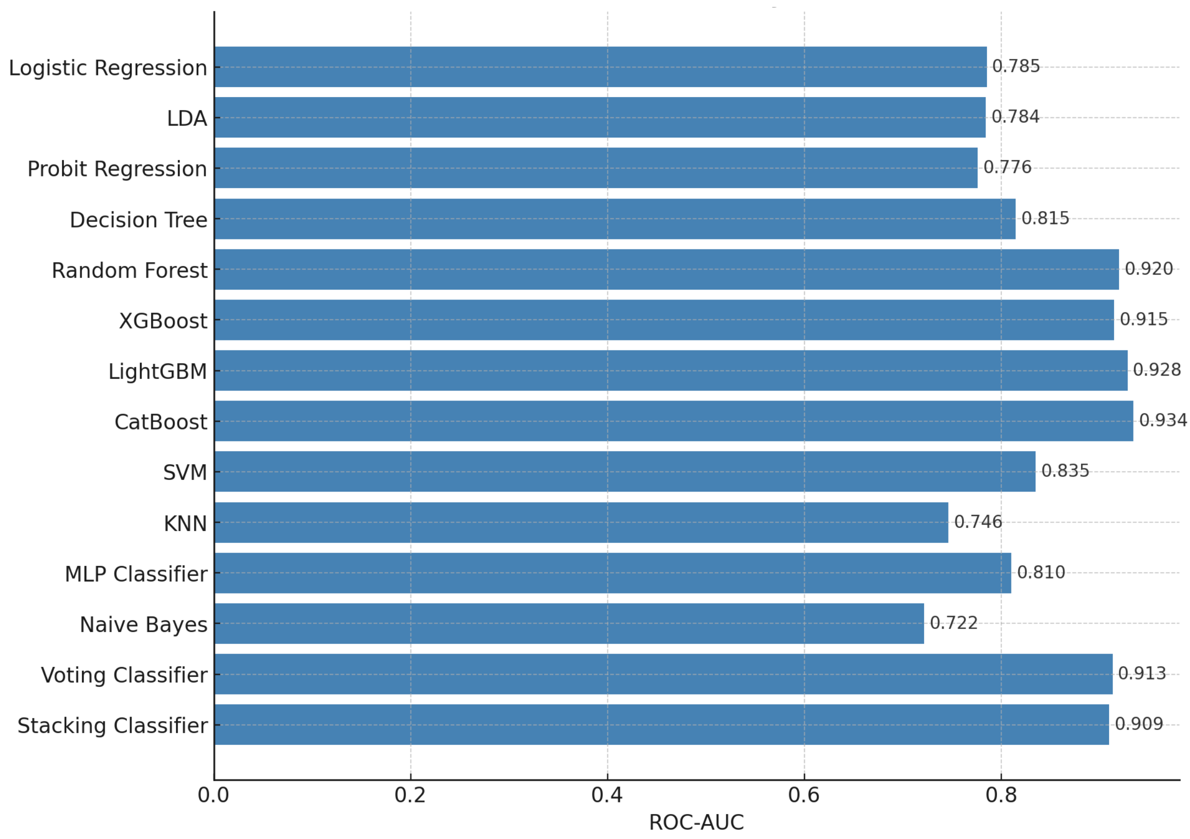
Exhibit 1: *Performance metrics across models*

Model	ROC-AUC	Accuracy	Precision	Recall	F1 Score
Logistic Regression	0.7853	0.7154	0.7363	0.7414	0.7388
LDA	0.7845	0.7247	0.7918	0.6690	0.7252
Probit Regression	0.7763	0.7116	0.7345	0.7345	0.7345
XGBoost	0.9148	0.8689	0.8929	0.8621	0.8772
CatBoost	0.9343	0.8652	0.9007	0.8448	0.8719
LightGBM	0.9283	0.8521	0.8702	0.8552	0.8626
Voting Classifier	0.9132	0.8614	0.9000	0.8379	0.8679
Stacking Classifier	0.9095	0.8521	0.8728	0.8517	0.8621
SVM	0.8350	0.7659	0.7855	0.7828	0.7841
Random Forest	0.9197	0.8390	0.8696	0.8276	0.8481
Decision Tree	0.8147	0.8146	0.8399	0.8138	0.8266
MLP Classifier	0.8102	0.7397	0.7631	0.7552	0.7591
K-Nearest Neighbors	0.7464	0.6704	0.6952	0.7000	0.6976
Naïve Bayes	0.7219	0.6124	0.8374	0.3552	0.4988

4.2 ROC-AUC Analysis

The ROC AUC (Area Under the Receiver Operating Characteristic Curve) measures the model's discriminative power across all classification thresholds (Fawcett, 2006). By evaluating how well the model ranks successful investments higher than unsuccessful ones, ROC AUC provides a robust, threshold-agnostic assessment, particularly valuable in imbalanced datasets typical of early-stage VC investments (Bradley, 1997). Exhibits 3, 4, and 5 present the ROC-AUC curves of the models. Each curve illustrates how effectively a model distinguishes successful from unsuccessful investments by plotting the true positive rate (correctly identified successes) against the false positive rate (incorrectly predicted successes) across different classification thresholds. The closer a curve lies to the top-left corner, the stronger the model's discriminative performance.

Exhibit 2: *ROC-AUC scores across mode*



4.2.1 Traditional Linear Classification Models

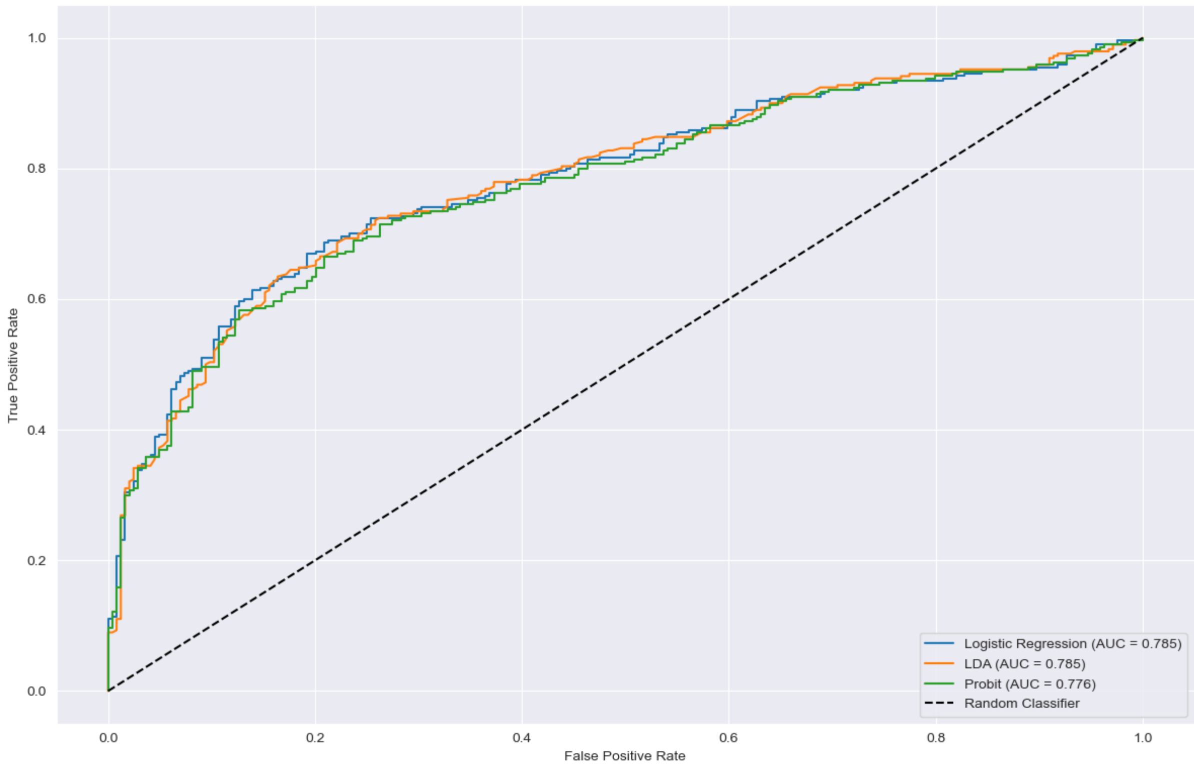
Among the traditional linear statistical classifiers, ROC-AUC scores revealed moderate yet informative discriminatory capabilities. Logistic Regression achieved an ROC-AUC of 78.53%, indicating that while the model clearly performs above chance, its ability to distinguish successful from unsuccessful investments lacks the refinement required for highly selective screening in venture capital (Hosmer et al., 2013). This is consistent with the nature of logistic regression, which models a linear decision boundary and may not capture non-linear dependencies often present in complex investment data (Menard, 2002).

Linear Discriminant Analysis (LDA) yielded a similar ROC-AUC of 78.45%. LDA maximizes the ratio of between-class variance to within-class variance, assuming multivariate normality and equal covariance across classes (Fisher, 1936). Its conservative structure tends to reduce false positives, which explains its slightly higher precision in this setting (Hastie, Tibshirani, & Friedman, 2009). However, like logistic regression, LDA's linear assumptions limit its flexibility when dealing with the high-dimensional, noisy nature of venture capital data, especially in the renewable energy sector where signals are weak and heterogeneity is high (Hastie, Tibshirani & Friedman, 2009).

Probit Regression followed closely with a ROC-AUC of 77.63%. Although based on a more complex latent variable structure (Amemiya, 1981), its practical performance in this context does not offer clear benefits over logistic models. The added complexity of Probit often does not translate into superior predictive power when applied to high-dimensional financial data. Moreover, convergence issues and the assumption of normally distributed errors limit its adaptability in datasets characterized by mixed scales and non-linearity (Wooldridge, 2010).

Linear models offer important interpretability advantages, making them attractive for baseline comparisons and explainable decision-making. Nevertheless, their inability to capture non-linear patterns limits their usefulness as standalone predictive tools in dynamic investment contexts (Breiman, 2001; Hastie, Tibshirani & Friedman, 2009).

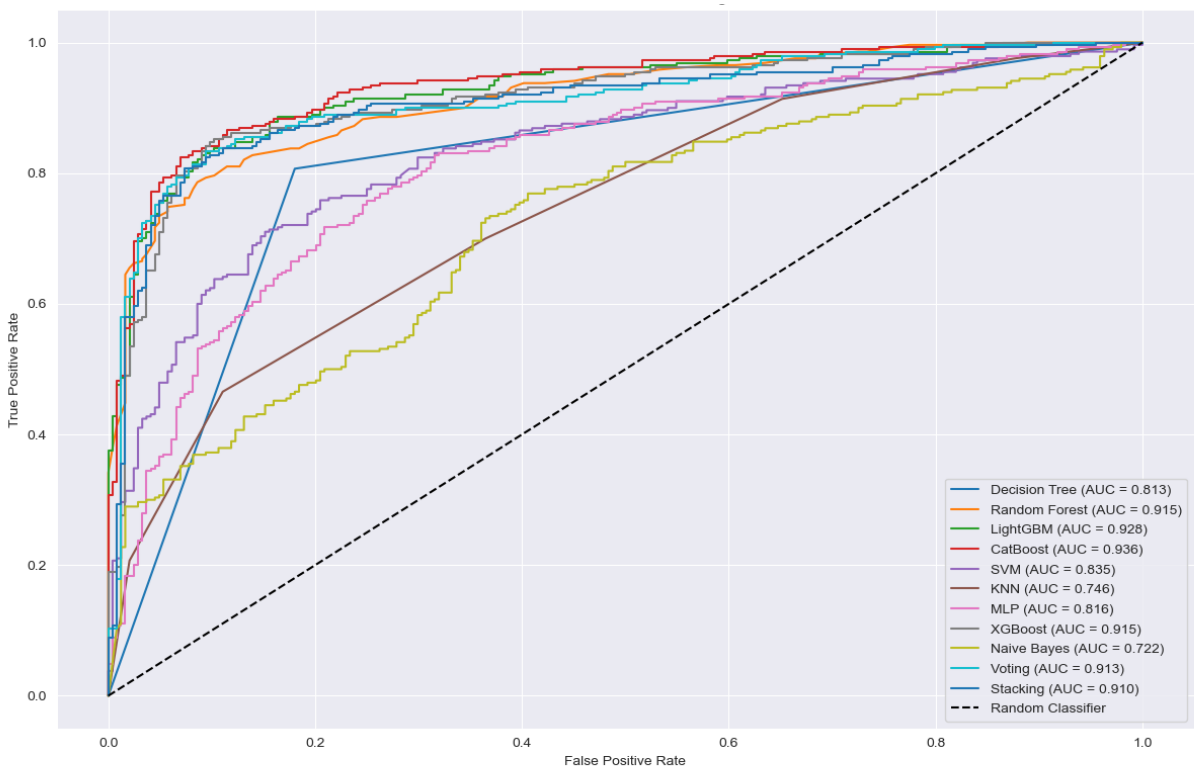
Exhibit 3: *ROC-AUC curves of classical linear statistical models*



4.2.2 Machine Learning Models

In sharp contrast, machine learning models, particularly tree-based algorithms and ensemble strategies, exhibited substantially stronger ROC-AUC performance, with several models achieving values well above 90%. This reflects their superior ability to capture latent patterns in the data that statistical models fail to exploit.

Exhibit 4: *ROC-AUC curves of ML models*



CatBoost recorded the highest ROC-AUC at 93.43%, demonstrating excellent discriminatory capacity. As highlighted by Hancock et al. (2020), CatBoost's advantage stems from its robust handling of categorical variables, efficient regularization, and native support for missing values, all of which are prevalent in real-world investment datasets. LightGBM followed closely with a ROC-AUC of 92.83%. Its histogram-based gradient boosting approach improves training speed and memory efficiency, making it highly scalable while maintaining predictive accuracy (Ke et al., 2017). Its depth-wise tree growth strategy contributes to discovering sharp splits in heterogeneous data (Ke et al., 2017), which may explain its strong performance in renewable energy ventures prediction.

XGBoost, a widely adopted gradient boosting algorithm, achieved a ROC-AUC of 91.48%. Its implementation of second-order optimization, shrinkage, and column subsampling ensures stability and regularization (Chen & Guestrin, 2016). As demonstrated in Żbikowski & Antosiuk (2021) and Arroyo et al. (2019), XGBoost consistently outperforms linear classifiers in venture capital applications due to its flexibility in modeling complex, high-dimensional interactions. Ensemble models also performed strongly: the Voting Classifier achieved a ROC-AUC of 91.32%, while the Stacking Classifier scored 90.95%. Both benefit from combining multiple learners, thereby averaging out biases and capturing broader patterns (Dietterich, 2000; Sagi & Rokach, 2018). As noted by Sollich & Krogh (1995), ensemble strategies leverage model diversity to increase generalizability and minimize overfitting, a property particularly valuable in early-stage investment screening.

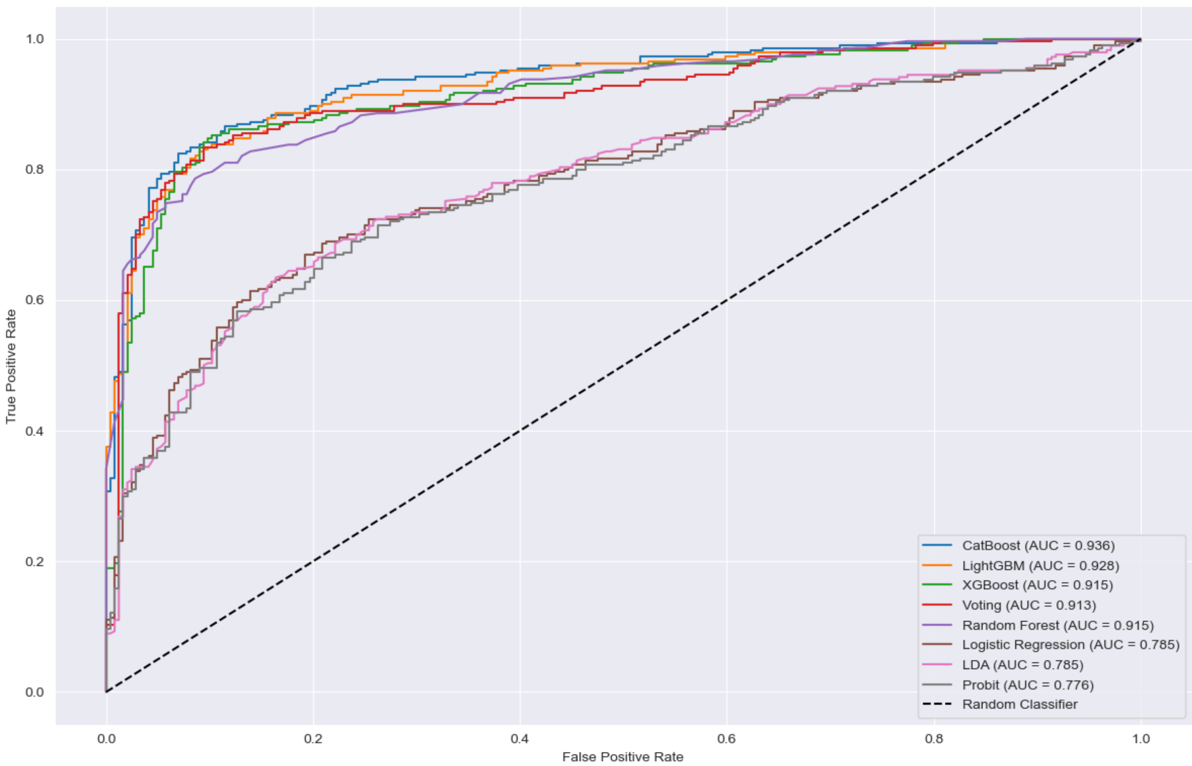
The Decision Tree, though more basic than its ensemble counterparts, achieved a respectable ROC-AUC of 81.47%. Its straightforward rule-based structure enhances interpretability but exposes it to overfitting, especially in the presence of small or correlated datasets. Random Forest, by contrast, returned a robust ROC-AUC of 91.97%. Its averaging mechanism over numerous bootstrap-sampled trees reduces variance and increases robustness to noise, aligning with prior findings that bagging techniques are particularly suited to noisy investment data (Breiman, 2001).

Performance was more modest among other non-linear classifiers. The Support Vector Machine (SVM) achieved a ROC-AUC of 83.50%. While SVMs can effectively model complex boundaries, their performance declines with overlapping class distributions and noisy features, both common in startup datasets (Hastie, Tibshirani & Friedman, 2009). The Multilayer Perceptron (MLP) reached 81.02%, confirming its generalization ability, though it remains sensitive to feature scaling and hyperparameter tuning.

The K-Nearest Neighbors (KNN) model, with a ROC-AUC of 74.64%, struggled to separate successful from unsuccessful investments, consistent with the literature noting its sensitivity to the curse of dimensionality in high-dimensional financial datasets (Pestov, 2013). Finally, Naïve Bayes scored the lowest ROC-AUC at 72.19%. Its assumption of conditional independence between features is rarely satisfied in venture data, where variables such as funding amount, investor count, and sector classification are correlated (Domingos & Pazzani, 1997).

Overall, ROC-AUC provides a threshold-independent view of model discrimination power and is particularly effective in distinguishing the strengths of advanced machine learning classifiers over statistical techniques. A visual comparison between the top five ML models and classical linear statistical models is presented in Exhibit 5, clearly illustrating the separation: while traditional models plateau around 78%, the best ML methods exceed 91%.

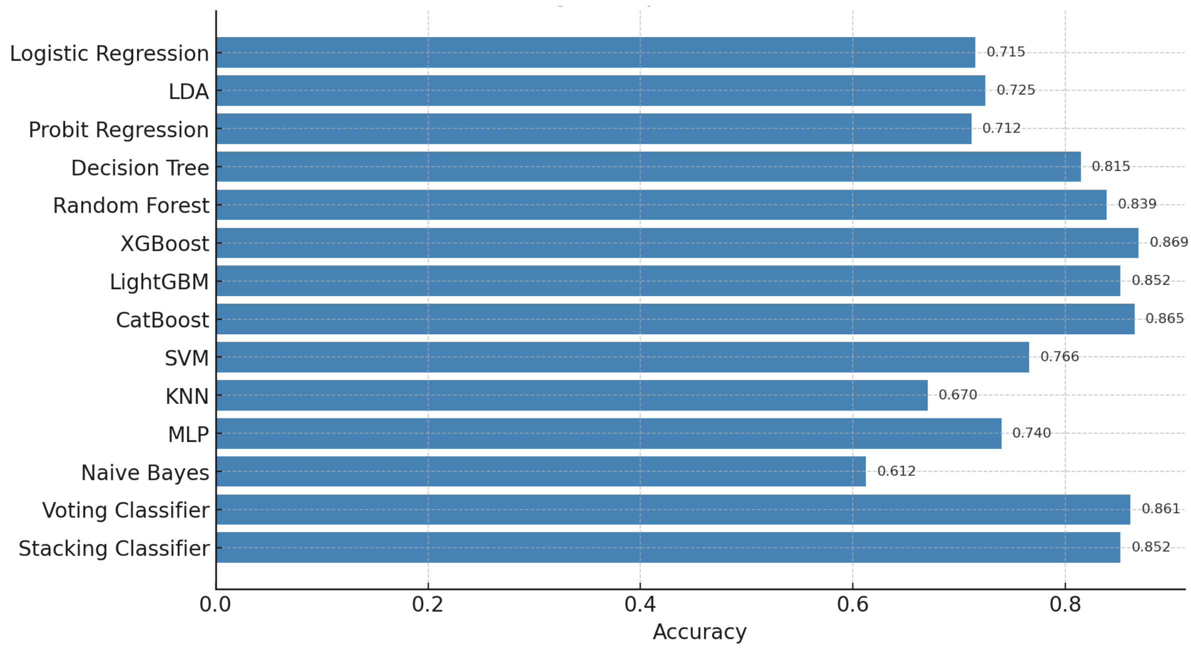
Exhibit 5: *ROC-AUC curves of classical linear statistical models and the best ML models*



4.3 Accuracy Analysis

Accuracy, the most intuitive performance metric, measures the overall proportion of correct predictions, both successful and unsuccessful investments, out of all classifications made Fawcett, 2006). While popular due to its simplicity, accuracy alone can be misleading in imbalanced datasets, such as venture capital portfolios where success rates are low (Chawla et al., 2002). Nevertheless, when interpreted alongside other metrics, it provides a useful starting point to compare classifier performance across different algorithmic paradigms. In the context of this study, where the dataset is approximately balanced between successful and unsuccessful ventures, accuracy remains a meaningful and relevant indicator. However, it is important to note that this is not the norm in most venture capital settings, where successes typically represent only a small fraction of outcomes, and in such imbalanced contexts accuracy becomes far less informative (He & Garcia, 2009).

Exhibit 6: *Accuracy scores across models*



4.3.1 Traditional Linear Classification Models

Among traditional linear classifiers, Linear Discriminant Analysis (LDA) achieved the highest accuracy at 72.47%, slightly outperforming both Logistic Regression (71.54%) and Probit Regression (71.16%). These results align with expectations from interpretable, statistically grounded methods, which offer reasonable performance when trained on structured financial data but struggle to capture complex patterns (Hosmer et al., 2013; James et al., 2013). While their accuracy levels are acceptable, they fall short of the thresholds typically desired in high-stakes investment decisions (Friedman et al., 2001; Kuhn & Johnson, 2013). This confirms the idea, consistent with prior discussions in the literature, that traditional models can function as robust baselines and offer interpretability advantages, yet their rigid assumptions and linear boundaries limit their ability to fully reflect the nuanced dynamics of venture success (Kaplan & Strömberg, 2004), particularly in renewable energy where data are noisy and highly heterogeneous. This confirms again, that traditional models can function as robust baselines and offer interpretability advantages, yet their rigid assumptions and linear boundaries limit their ability to fully reflect the nuanced dynamics of venture success (Kaplan & Strömberg, 2004; Hosmer et al., 2013).

4.3.2 Machine Learning Models

In contrast, machine learning classifiers substantially outperformed their statistical counterparts in terms of accuracy. XGBoost led the group with an accuracy of 86.89%, followed closely by CatBoost (86.52%) and the Voting Classifier (86.14%). These models demonstrate a superior capacity to generalize from complex, high-dimensional data, an essential feature in the heterogeneous space of early-stage renewable investments (Chen & Guestrin, 2016; Ke et al., 2017; Prokhorenkova et al., 2018). This finding is consistent with broader evidence in the literature, which has repeatedly shown that ensemble boosting methods tend to outperform linear approaches in venture capital prediction tasks, owing to their ability to capture non-linear interactions and adapt to diverse data structures (Żbikowski & Antosiuk, 2021; Arroyo et al., 2019).

LightGBM and the Stacking Classifier also achieved high accuracy scores (85.21%), indicating consistency across ensemble-based gradient boosting strategies. Random Forest (83.90%) and the Decision Tree (81.46%) further illustrate the strength of tree-based methods, especially in capturing hierarchical decision pattern (Breiman, 2001; Quinlan, 1993). This aligns with evidence from Arroyo et al. (2019), who found that Random Forest and Gradient Boosting Machines consistently outperform linear baselines in predicting venture outcomes.

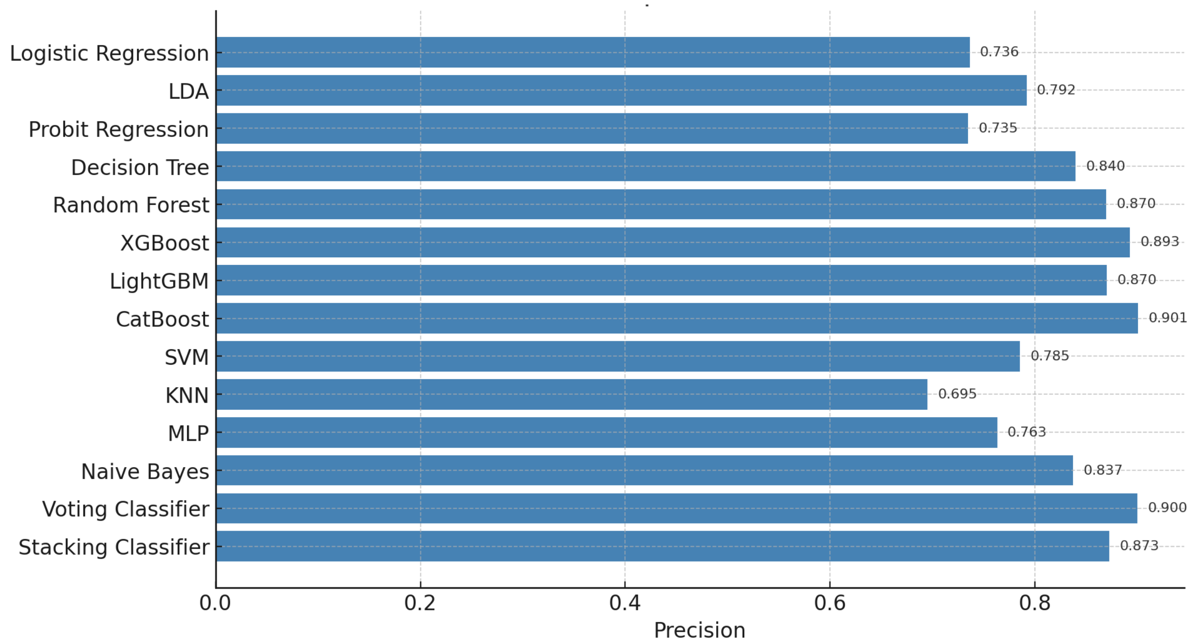
Other models such as the Support Vector Machine (76.59%) and Multilayer Perceptron (73.97%) posted moderate results, reflecting their reliance on kernel tuning and data preprocessing. In contrast, K-Nearest Neighbors (67.04%) and Naïve Bayes (61.24%) showed limited effectiveness, with the latter notably underperforming due to its strong independence assumptions (Domingos & Pazzani, 1997).

Overall, accuracy results confirm that modern machine learning classifiers, particularly ensemble techniques, are markedly better suited for the nuanced prediction challenges of VC investment.

4.4 Precision Analysis

Precision, which measures the proportion of correctly predicted positive cases among all instances classified as positive, is a key metric in the context of venture capital (Sokolova & Lapalme, 2009). In early-stage investment decisions, a high precision score ensures that capital is not misallocated to startups erroneously predicted to succeed. It is particularly relevant when false positives, i.e., funding startups that ultimately fail, carry substantial financial consequences (Gompers & Lerner, 2001; Kaplan & Strömberg, 2004).

Exhibit 7: *Precision scores across models*



4.4.1 Traditional Classification Models

Among traditional classifiers, precision scores reveal a cautiously conservative orientation toward success prediction. Linear Discriminant Analysis (LDA) leads this group with a precision of 79.18%, indicating that when the model predicts a venture to be successful, it is often correct. This cautious stance stems from LDA's reliance on maximizing class separability under assumptions of normal distribution and equal covariance, which naturally reduces the model's propensity to over-predict positives (Gompers & Lerner, 2001; Kaplan & Strömberg, 2004).

Logistic Regression and Probit Regression follow with nearly identical precision scores of 73.63% and 73.45% respectively. These results reflect their shared reliance on linear decision boundaries and assumptions about feature contributions. While these methods remain valued for their interpretability and transparency in explaining outcomes, their precision is ultimately constrained by parametric rigidity and limited ability to adapt to irregular, multidimensional VC data (Hosmer et al., 2013; Kleinert et al., 2020).

4.4.2 Machine Learning Models

The precision scores of machine learning models reveal substantial variation across algorithms, reflecting their distinctive learning paradigms and generalization strategies (Bishop, 2006; Kuhn & Johnson, 2013). Among all classifiers tested, CatBoost and the Voting Classifier achieved the highest precision, both exceeding 90%. CatBoost reached 90.07%, confirming its effectiveness in filtering successful ventures with minimal false positives. This strong performance is attributable to its robust treatment of categorical variables and missing values, which are frequent in early-stage venture data (Hancock & Khoshgoftaar, 2020). The Voting Classifier, which aggregates the predictions of Logistic Regression, Random Forest, and XGBoost through soft voting, achieved a nearly identical precision of 90.00%. Its ensemble nature mitigates the biases of individual learners, leading to more stable classifications, an advantage frequently highlighted in ensemble learning literature (Polikar, 2012). These results suggest that when such models classify a startup as successful, they are highly likely to be correct, a property of critical importance for minimizing capital misallocation in venture capital contexts.

XGBoost and the Stacking Classifier also delivered strong results, with precision scores of 89.29% and 87.28%, respectively. XGBoost's regularized objective function and shrinkage mechanisms contribute to its ability to generate precise positive classifications while limiting overfitting (Chen & Guestrin, 2016). The Stacking Classifier, which integrates predictions from Random Forest and XGBoost through a logistic regression meta-learner, benefits from the layered learning process, refining classification boundaries and improving robustness (Wolpert, 1992).

LightGBM followed closely with a precision of 87.02%. Its histogram-based gradient boosting algorithm, known for efficiency and fine-grained split discovery, has been noted in prior studies to achieve competitive precision in heterogeneous, high-dimensional settings (Ke et al., 2017). Its conservative partitioning strategy, favoring fewer but more confident classifications, likely contributed to its performance (Prokhorenkova et al., 2018). Random Forest, at 86.96%, also displayed stable reliability. By averaging across multiple decision trees trained on bootstrap samples, Random Forest reduces variance and minimizes the likelihood of spurious splits, thereby lowering the rate of false positives (Breiman, 2001). The simpler Decision Tree classifier still achieved a respectable 83.99%, suggesting that even rule-based models can yield meaningful precision. However, its well-known tendency to overfit small or noisy datasets limits its generalizability, a drawback long documented in the literature (Quinlan, 1993).

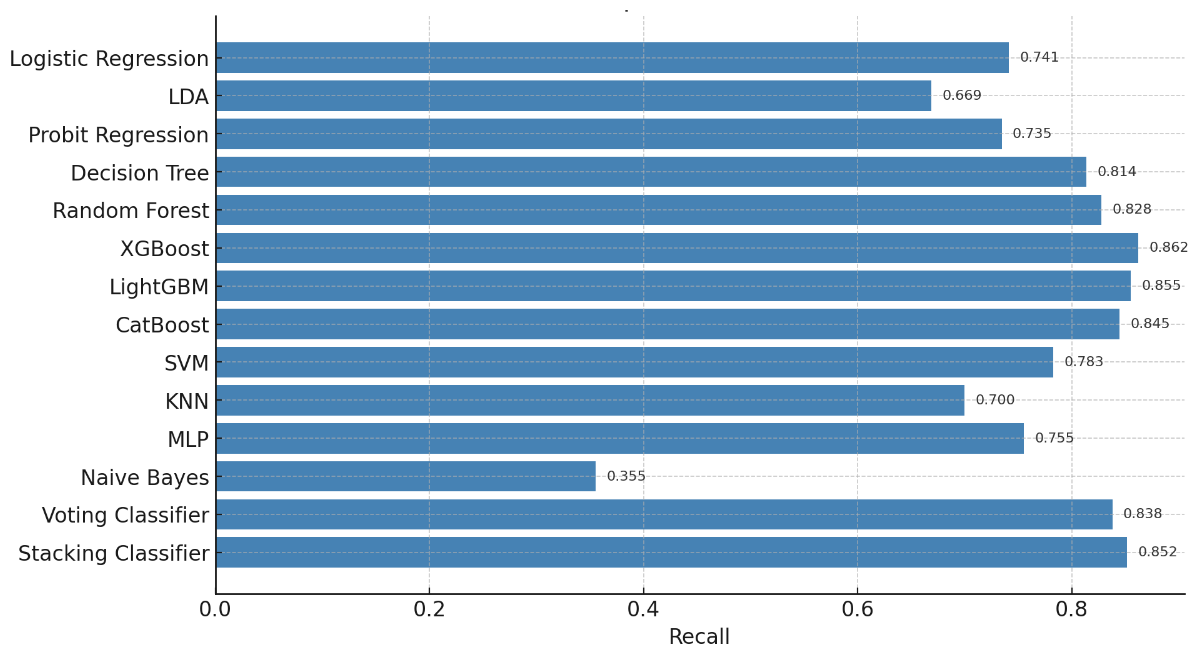
Other algorithms exhibited more modest precision. The Support Vector Machine (78.55%) and the Multilayer Perceptron (76.31%) performed reasonably, though both rely on careful kernel tuning or hyperparameter optimization to achieve higher stability in practice. Their relatively high precision but moderate recall suggests that they are selective in identifying positives, potentially at the expense of overlooking genuine successes (Hastie et al., 2009). K-Nearest Neighbors, at 69.52%, struggled to maintain robust classification boundaries, a result consistent with the well-documented challenges of analyzing and organizing data in high-dimensional spaces (Beyer et al., 1999). Finally, Naïve Bayes achieved a comparatively high precision of 83.74%, but this result is undermined by extremely poor recall, rendering the model overly conservative.

The precision metric underscores the superior ability of ensemble classifiers and boosting methods to minimize false positives. This property makes them especially attractive for venture capital investors, where avoiding misallocated capital is as critical as identifying high-potential opportunities.

4.5 Recall Analysis

Recall, also known as sensitivity, captures the proportion of actual successful investments that are correctly identified by the model (Sokolova & Lapalme, 2009). It provides an indication of the model's ability to minimize false negatives, i.e., instances where promising ventures are incorrectly classified as failures, which can lead to substantial loss of potential gains (Powers, 2011). This property is particularly important in venture capital, where a small number of highly successful startups often generate a disproportionate share of returns (Gompers & Lerner, 2001). High recall is thus especially valuable in the renewable energy domain, where capital-intensive ventures may appear risky at early stages but, if successful, deliver significant long-term payoffs and positive environmental externalities. A model with strong recall ensures that these opportunities remain within the investor's consideration set, even if this comes at the cost of also flagging some weaker firms, aligning with the venture capital strategy of casting a wide net to capture rare but impactful successes.

Exhibit 8: *Recall scores across models*



4.5.1 Traditional Classification Models

Among classical classifiers, Logistic Regression achieved a recall of 74.14%, showing a decent capacity to identify truly successful investments. Linear Discriminant Analysis (LDA), while registering the highest precision among statistical models, had a considerably lower recall of 66.90%, reflecting a more conservative classification strategy that reduces false positives at the expense of overlooking many genuine successes (Hosmer et al., 2013). Probit Regression, meanwhile, produced a recall of 73.45%, yielding relatively balanced sensitivity and specificity, albeit at the cost of additional numerical complexity (Hosmer et al., 2013).

Their reliance on linear decision boundaries means that complex interactions between features such as funding history, sector, and investor type are frequently underexplored, leading to systematic false negatives (Hastie et al., 2009). In the context of renewable energy, an industry marked by technological uncertainty, long development horizons, and heterogeneous firm profiles, such limitations can be particularly damaging, as they increase the likelihood of overlooking breakthrough ventures that could generate outsized financial and environmental returns (Gompers & Lerner, 2001).

4.5.2 Machine Learning Models

Machine learning classifiers consistently outperformed their classical counterparts in recall, showcasing their superior ability to detect subtle, non-linear patterns that distinguish successful ventures from those that ultimately fail. Among all models, CatBoost achieved the highest recall at 84.48%, closely followed by XGBoost (86.21%), Random Forest (82.76%), and LightGBM (85.52%). These tree-based algorithms rely on boosting and bagging techniques that iteratively focus on previously misclassified cases, thereby improving sensitivity to rare but important success signals (Chen & Guestrin, 2016; Ke et al., 2017; Prokhorenkova et al., 2018).

The Stacking Classifier (85.17%) and Voting Classifier (83.79%) further illustrate the advantages of ensemble learning. By combining base classifiers with different error patterns, they expand the funnel of true positives captured, ensuring broader generalization across heterogeneous investment scenarios (Zenobi et al., 2001). Such stacking strategies leverage diversity to capture a broader range of signals across the investment landscape (Mienye et al.,2022), a crucial feature in contexts where no single model perfectly captures the underlying dynamics.

Other non-linear classifiers also posted respectable recall values. The Support Vector Machine (SVM) reached 78.28%, and the Multilayer Perceptron (MLP) scored 75.52%. Both models benefited from their ability to approximate flexible decision boundaries, though they remain more dependent on feature preprocessing and careful hyperparameter tuning compared to boosting-based methods (Cortes & Vapnik, 1995; Bishop, 2006). The Decision Tree, with a recall of 81.38%, confirms that even simple, rule-based models can be effective at identifying true positives when trained on structured financial data, though its limited regularization makes it prone to overfitting (Quinlan, 1993). By contrast, K-Nearest Neighbors (KNN) achieved a recall of only 70.00%, reflecting again the challenges of high-dimensional spaces (Beyer et al., 1999).

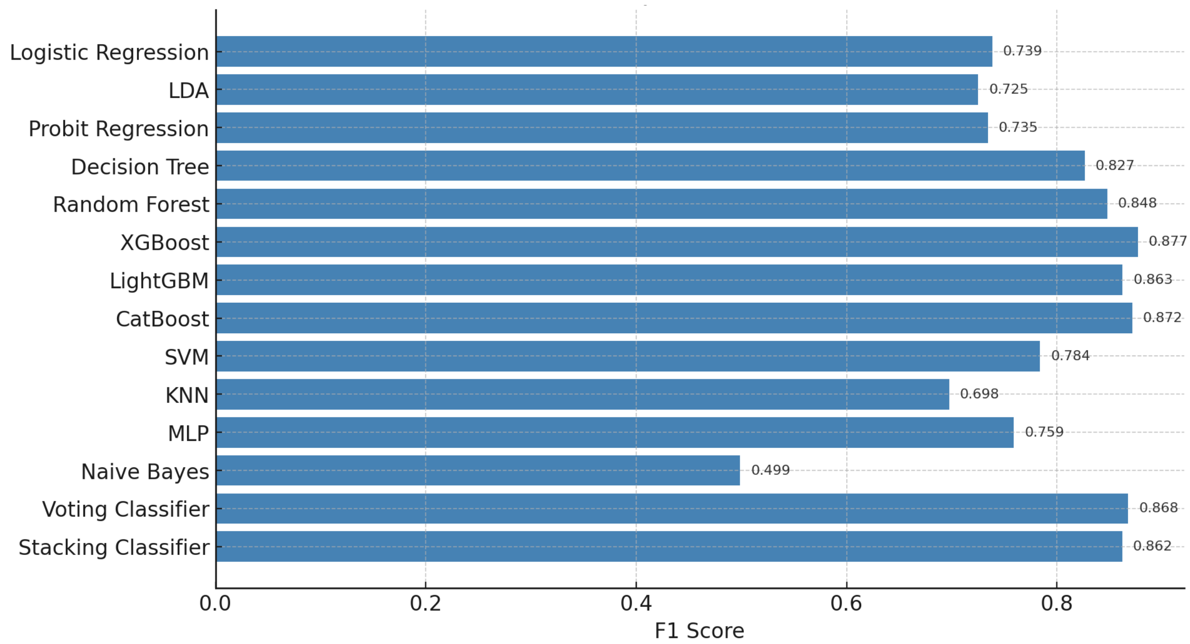
Finally, Naïve Bayes performed poorly in this dimension, with a recall of just 35.52%. Its strict independence assumption led to systematic under-detection of successful ventures, resulting in a strong bias toward predicting failure. This outcome is consistent with critiques in the literature that caution against the use of Naïve Bayes in financial datasets (Gupta, et al., 2021), where interdependencies among features, such as funding stage, geography, and investor reputation, might be present.

Results emphasize the strategic importance of recall as a metric in venture capital screening. In early-stage investing, where only a handful of “home run” deals drive the bulk of portfolio returns, missing a high-potential startup can have disproportionate consequences (Gompers & Lerner, 2001). Models that maximize recall, particularly ensemble and gradient boosting classifiers, therefore provide investors with a critical advantage: the ability to capture a wider pool of potential winners without unduly inflating false negatives.

4.6 F1 Score Analysis

F1 Score represents the harmonic mean of precision and recall and is widely regarded as one of the most informative single metrics for evaluating model performance in classification tasks where both false positives and false negatives carry high costs (Sokolova & Lapalme, 2009). In the context of VC, this trade-off is particularly relevant, balancing the risk of misallocating capital to startups that ultimately fail (false positives) against the risk of missing opportunities to invest in high-potential ventures (false negatives) (Powers, 2011).Unlike accuracy, which can mask imbalances in error types, the F1 Score provides a balanced performance indicator that explicitly penalizes both forms of misclassification. This makes it particularly suitable for the binary classification challenge posed by early-stage startup investment screening.

Exhibit 9: *F1 scores across models*



4.6.1 Traditional Classification Models

Among traditional models, Logistic Regression attained an F1 Score of 73.88%, indicating a reasonably balanced performance in recognizing both successful and unsuccessful startups. While it may not capture complex non-linear interactions, it shows good generalizability when applied to structured financial indicators (Hosmer et al., 2013). Probit Regression, with a nearly identical F1 Score of 73.45%, offered no substantive advantage in this dataset: although theoretically more sophisticated through its latent variable formulation (Amemiya, 1981), in practice its performance closely mirrors that of logistic regression, confirming that the added complexity yields only marginal gains in high-dimensional venture capital data.

Linear Discriminant Analysis (LDA), in contrast, returned a slightly lower F1 Score of 72.52%. This result reflects its trade-off between high precision and low recall: LDA is conservative in predicting success, which reduces false positives but simultaneously increases the likelihood of overlooking promising investment targets. Such conservatism can be advantageous in contexts where investors wish to minimize capital misallocation, yet detrimental in high-risk, innovation-driven sectors like renewable energy, where capturing breakthrough opportunities is often decisive for portfolio performance (Cumming & Johan, 2009).

While all three statistical models demonstrate sufficient capability to serve as transparent and explainable baselines, their performance remains constrained by assumptions of linearity, normality, and homoscedasticity, which again do not hold well in the complex, noisy landscape of venture capital (Kaplan & Strömberg, 2004). Their limited ability to adjust to non-linear relationships restricts their effectiveness in accurately capturing the nuanced dynamics of startup success (Kleinert et al., 2020).

4.6.2 Machine Learning Models

Machine learning classifiers consistently surpassed traditional models in terms of F1 Score, highlighting their ability to reconcile the trade-off between identifying successful ventures and avoiding false alarms. Among them, XGBoost achieved the highest F1 Score at 87.72%, followed closely by CatBoost (87.19%), Voting Classifier (86.79%), LightGBM (86.26%), and Stacking Classifier (86.21%). These results reflect the mentioned strengths of gradient boosting and ensemble architectures in correcting misclassifications iteratively and capturing complex inter-feature interactions (Chen & Guestrin, 2016; Ke et al., 2017; Prokhorenkova et al., 2018).

Random Forest delivered a strong F1 Score of 84.81%, confirming its reliability as a general-purpose classifier capable of maintaining high predictive accuracy across both classes given its bagging mechanism and random feature sampling (Breiman, 2001). The Decision Tree, though structurally simpler and more prone to overfitting, still achieved a respectable F1 Score of 82.66%, confirming that even basic rule-based learners can extract useful insights from structured venture data (Quinlan, 1993).

Non-tree-based classifiers exhibited more variability. The Support Vector Machine (SVM) and the Multilayer Perceptron (MLP) achieved F1 Scores of 78.41% and 75.91%, respectively. These models demonstrate moderate success in balancing sensitivity and specificity but require careful hyperparameter tuning and preprocessing (e.g., scaling) to achieve optimal performance, steps less critical for tree-based methods (Cortes & Vapnik, 1995; Bishop, 2006). K-Nearest Neighbors (KNN), with an F1 Score of 69.76%, struggled due to its reliance on distance metrics, which become less reliable in high-dimensional feature spaces (Hastie, Tibshirani & Friedman, 2009). Finally, Naïve Bayes recorded the lowest F1 Score at 49.88%, largely penalized by its extremely low recall (35.52%), despite its relatively high precision.

Overall, gradient boosting algorithms and ensemble meta-models dominate this metric, confirming their suitability for designing data-driven VC screening systems that combine accuracy with capital efficiency. In the renewable energy domain, where both false positives and false negatives are costly, their ability to sustain high F1 Scores makes them particularly attractive tools for investors and policymakers.

5. Interpretability through SHAP Analysis of Predictive Features

To better understand the mechanisms driving the success of renewable energy venture capital investments, this chapter applies SHAP values to interpret the predictions of the classification models evaluated. SHAP is a model-agnostic interpretability framework grounded in cooperative game theory (Shapley, 1953; Lundberg & Lee, 2017). It assigns each feature a contribution to the prediction by considering all possible feature combinations. Specifically, SHAP decomposes a model's output into the additive impact of each feature, offering both global and local interpretability (Lundberg & Lee, 2017). Two complementary SHAP visualization tools are employed. The Mean SHAP value bar plots summarize the average absolute contribution of each feature across predictions, ranking variables by global importance and showing which drive most of the predictive power (Molnar, 2019). By contrast, SHAP summary dot plots display the distribution of feature impacts across individual observations. Each point corresponds to a prediction, colored by the feature’s value, enabling us to assess whether high or low values increase or decrease the probability of success (Lundberg et al., 2020). This dual approach highlights not only the most influential features but also the direction and heterogeneity of their effects.

Rather than analyzing each model separately, this chapter adopts a unified, feature-centric perspective. Organizing SHAP insights around variables rather than models identifies the predictors that consistently drive success across algorithms. This cross-model view also allows us to compare how different techniques interpret the same features, contrasting traditional statistical models with machine learning methods.

5.1 Investee Company Characteristics

Features capturing the characteristics and history of investee companies emerge consistently among the strongest predictors across models. The year of the first investment received by the company stands out as one of the most important features in nearly all models, including Logistic Regression, Probit, LDA, Random Forest, CatBoost, XGBoost, LightGBM, SVM, MLP, Stacking, and Voting classifiers. SHAP dot plots show predominantly positive associations, especially in tree-based models, where more recent first investments increase the predicted probability of success. This pattern suggests that startups entering the market in later years may benefit from more favorable sectoral conditions, technological maturity, or stronger policy frameworks for renewable energy. This result is consistent with previous empirical findings, where more recent funding was associated with faster exits and more favorable market timing (Gompers et al., 2008; Chemmanur et al., 2011).

The number of investments received to date also plays a central role. In most ensemble models, higher values of this feature correspond to higher SHAP scores, indicating a positive relationship with success. Dot plots reveal relatively clean positive correlations, particularly in CatBoost, LightGBM, and Stacking, though with some noise in XGBoost and MLP. This feature likely proxies for firm resilience, repeated investor interest, and milestone achievement, consistent with theories of staged financing and cumulative advantage (Hellmann & Puri, 2002).

Other important investee-level variables include the company's founding year, which exhibits mixed effects across models. In some cases, younger companies appear associated with higher success probabilities, while in others older firms benefit, possibly reflecting nonlinearities between maturity and adaptability. Similarly, the total funding received to date is generally associated with positive SHAP contributions in tree-based models, supporting the view that financial scale enables milestone achievement and boosts investor confidence (Da Rin et al., 2013). However, dot plots indicate heterogeneity: in some models, very high funding values display diminishing or inconsistent marginal effects, a phenomenon also noted in prior empirical studies showing decreasing marginal returns to additional financing (Puri & Zarutskie, 2012).

Finally, RECAI-based country attractiveness variables appear less prominent globally but exhibit directionality in linear models. Startups located in countries ranked among the global top 10 tend to display slightly positive SHAP contributions, while those in lower-ranked countries are more often associated with negative values. These patterns suggest that macro-environmental conditions captured by RECAI contribute modestly but systematically to investment outcomes, consistent with evidence that regulatory frameworks and renewable energy policy environments shape entrepreneurial outcomes (Polzin et al., 2017).

5.2 Investor Firm Characteristics

Investor-related variables also emerge as influential, particularly in ensemble models. The total number of deals executed by the investor firm and the total number of companies invested in typically show positive SHAP contributions, indicating that more experienced investors are associated with higher startup success probabilities. This resonates with findings by Colombo & Grilli (2010), who emphasize the compound effects of investor experience. However, dot plots reveal that these relationships are sometimes noisy, especially in XGBoost and Random Forest, suggesting that investor experience interacts with other features such as deal structure (also noted in Cumming et al., 2005).

Temporal dimensions of investor activity provide additional insights. The first investment year of an investor often shows negative SHAP values in linear models, suggesting that older, more established investors may be less adaptive to emerging technologies (also noted in Bertoni et al., 2015). By contrast, the last investment year frequently exhibits a positive relationship with success, particularly in MLP and LightGBM models, implying that recently active investors bring more up-to-date knowledge, networks, and strategies aligned with current market dynamics (also noted in Ewens & Rhodes-Kropf, 2015).

Variables capturing equity committed by investors (e.g., total estimated equity invested to date) exhibit inconsistent effects. While in some ensemble models this variable contributes positively, in others the relationship appears flat or even negative. This suggests that the raw scale of equity invested may not be a straightforward predictor of success, but rather depends on interaction effects with deal structuring, syndication, and timing. These patterns are consistent with the broader literature on how investor experience and timing matter more than raw scale of financial commitment (Colombo & Grilli, 2010).

Investor type dummies generally played a minor role compared to experience and timing features. Nonetheless, some specific categories, such as incubator programs or corporate-affiliated investors, occasionally displayed localized positive contributions in dot plots, pointing to potential complementarities between strategic resources and financial capital. (Grilli & Murtinu, 2014).

5.3 Fund Characteristics

Fund-level features are comparatively less prominent in terms of mean SHAP values but still provide meaningful signals in certain models. Fund type variables, distinguishing between venture capital, buyout, and generalist private equity, occasionally emerge as relevant. Buyout-oriented funds tend to be associated with positive SHAP values, especially in tree-based models, aligning with evidence that structured funds with clear exit horizons support higher success rates (Tykvová, 2018).

The fund year feature, however, exhibits more ambiguous effects. In several linear models it shows negative SHAP contributions, while in ensemble classifiers results are inconsistent or weak. Dot plots frequently reveal noisy patterns, suggesting that fund year by itself is not a reliable predictor, but may capture unobserved heterogeneity in fund cycles or strategic orientations (Kaplan & Strömberg, 2009).

Overall, fund characteristics contribute modestly relative to investee and deal-level variables. Their influence appears conditional, amplifying or moderating the effects of other features rather than driving predictions independently.

5.4 Deal Characteristics

Deal-related features constitute some of the most influential predictors across all models, underscoring the centrality of transaction design in shaping investment outcomes. Among security types, Leveraged Buyout Financing consistently ranks among the strongest predictors, with clear positive SHAP contributions in CatBoost, LightGBM, Random Forest, and Voting Classifier. Dot plots show clear stratification, where the presence of this security type increases success probability. This suggests that structured buyout financing, typically associated with clearer governance and exit strategies, enhances venture outcomes. This finding aligns with prior work in clean energy private equity, which emphasized the importance of structured buyouts with clear governance mechanisms (Fuerst & McAllister, 2011).

By contrast, Venture Capital Equity Investment generally exhibits negative SHAP values, particularly in linear models and ensemble classifiers, reflecting the higher risks inherent in unstructured equity deals at early stages. This divergence in outcomes between buyout-driven and VC-driven structures is consistent with Tykvová (2018), who highlights the different success dynamics of venture capital versus private equity financing.

Deal type dummies reinforce this divergence. Buyout deals show robust positive associations across Decision Tree, SVM, CatBoost, and Voting models, while Pure Venture Vapital deals often contribute negatively across Probit, XGBoost, CatBoost, Random Forest, Stacking, and Voting classifiers.. This confirms that structured exit-oriented transactions are systematically more likely to succeed than early-stage, high-risk equity investments (also noted in Rosenbusch et al. 2013).

Other deal-stage variables add nuance. Public market financing rounds display positive SHAP contributions in multiple models, indicating that ventures reaching advanced stages of financing are systematically more successful. Conversely, early-stage and seed financing rounds often show negative associations, consistent with the high uncertainty characterizing firms at these stages and previous research (Gompers & Lerner, 2001). The deal age at financing in months yields heterogeneous results: in some models longer deal maturity is associated with positive contributions, while in others the relationship is flat or inconsistent, reflecting heterogeneity in deal pacing.

Round-level details, including the number of funds participating, generally display weak-to-positive SHAP values in ensemble models. A greater number of funds is often linked to higher probabilities of success, likely due to risk-sharing, reputational validation, and access to diversified resources (Lerner, 1994). However, dot plots also reveal noise in this feature, highlighting that the effect of syndication varies depending on deal structure and investor composition.

6. Discussion and Implications

The comparative analysis of all classification models evaluated in this study reveals a consistent and compelling pattern: machine learning (ML) algorithms, particularly ensemble and gradient boosting methods, outperform traditional statistical models across every key metric. These findings are not merely incremental but substantial. Top ML models achieve ROC-AUC scores above 91%, accuracy levels between 85% and 87%, precision and recall metrics in the 80–90% range, and F1 scores consistently above 85%, reflecting balanced and robust predictive performance. By contrast, traditional statistical models—Logistic Regression, Linear Discriminant Analysis (LDA), and Probit Regression—achieve more modest results, with ROC-AUC plateauing around 77–78%, accuracy between 71% and 73%, and precision and recall values generally between 70% and 75%. While these models retain interpretability advantages, their discriminatory power is markedly weaker. As highlighted by Hastie et al. (2009), they remain useful as baselines but are structurally limited by rigid linear assumptions that constrain adaptability to the complex, high-dimensional dynamics of renewable energy ventures.

When performance across all evaluation metrics is considered simultaneously, ensemble and boosting methods emerge as the most consistently superior. These models, including gradient boosting techniques and ensemble combinations, achieved 10–15 percentage point gains in Accuracy and ROC-AUC and up to 15–20 percentage point gains in F1 Score compared to both linear classifiers and other ML approaches such as SVM, MLP, or KNN. Within ML, ensemble boosting algorithms displayed the most balanced performance across all dimensions, while non-ensemble models generally scored in intermediate ranges, confirming the importance of architectures that combine multiple weak learners or leverage iterative boosting.

From a venture capital perspective, the implications of these findings can be directly linked to investor objectives. For investors primarily concerned with minimizing false positives (FPs) and avoiding capital misallocation, high-precision models are the most appropriate, with ensemble classifiers offering precision above 90%. For accelerators, policymakers, or corporate investors who prioritize inclusivity and aim to avoid overlooking potential breakthroughs, high-recall models are preferable, with several boosting techniques achieving recall above 85%, thereby maximizing coverage of genuinely successful ventures even at the cost of some false positives. For generalist venture funds or diversified portfolios, the F1 Score provides the most balanced single benchmark, with top-performing ensembles maintaining values above 85%, ensuring both capital efficiency and opportunity capture. Finally, for ESG-focused funds or public-sector allocators, transparency may outweigh raw predictive performance, leaving classical models such as Logistic Regression and Decision Trees relevant thanks to their interpretability and accountability in regulated contexts.

The SHAP-based interpretability analysis provides deeper insights into the second research objective: identifying the features that consistently drive predictive performance across models. Results show that deal characteristics are the most systematically influential group, with leveraged buyout financing, buyout deals, and public market financing exhibiting strong positive associations with success, while venture capital equity investments and early-stage financing rounds tend to show negative contributions. Investee characteristics also emerge as robust predictors, particularly the year of first investment and the number of investments received, which consistently displayed positive associations with success. Investor-related features such as activity recency and experience contributed further nuance, while fund-level variables played a more modest and often conditional role. Some features, such as the total equity invested by firms, exhibited unstable directionality across models, underlining the need for cautious interpretation. These results are broadly in line with previous research showing the importance of staged financing, investor expertise, and deal structuring in determining venture outcomes (Hellmann & Puri, 2002; Kaplan & Strömberg, 2009; Colombo & Grilli, 2010; Da Rin et al., 2013).

Importantly, the methodological choice to include last-round deal characteristics instead of adopting time-aware validation proved justified. The SHAP analysis demonstrated that variables tied to deal structuring and financing design systematically drive success, suggesting that structural characteristics, rather than purely temporal extrapolation, are central to predictive power in this context. This confirms the practical value of the models as decision-support tools capable of simulating outcomes under different financing scenarios, rather than purely ex-ante forecasts.

Taken together, these results reinforce the predictive and theoretical relevance of a small set of core drivers of venture outcomes: timing of initial investment, deal type, investor experience, and financial scale. These findings are consistent with prior literature emphasizing the importance of staged financing, structured exits, and investor expertise in venture performance (Hellmann & Puri, 2002; Kaplan & Strömberg, 2009; Da Rin et al., 2013). The contribution of this thesis lies in demonstrating that these mechanisms can be systematically detected and quantified through SHAP interpretability applied to machine learning models.

Finally, the results support a tiered modeling strategy: classical statistical models remain valuable as interpretable benchmarks and explanatory tools, particularly in contexts where transparency is paramount. However, for high-stakes decision-making where predictive accuracy is critical, advanced ensemble methods such as XGBoost, CatBoost, LightGBM, and Voting/Stacking classifiers demonstrate clear superiority. This hybrid approach ensures both accountability and performance, aligning with the dual goals of effective capital allocation and robust risk management in the evolving renewable energy venture capital ecosystem.

7. Conclusion

This research was designed to examine the following central question: *can machine learning approaches offer superior predictive accuracy compared to traditional linear statistical models in forecasting the success or failure of venture capital investments in the renewable energy sector?* By systematically benchmarking a wide array of models on structured financial and investment data, and complementing predictive evaluation with SHAP interpretability, this study has provided empirical evidence that helps address this question.

The results strongly suggest that advanced machine learning techniques, particularly gradient boosting and ensemble methods, outperform classical linear models across all key performance metrics. Gradient boosting algorithms such as XGBoost, CatBoost, and LightGBM, along with ensemble combinations, consistently achieved the highest scores in accuracy, precision, recall, F1 Score, and ROC-AUC, outperforming linear statistical models by margins of 10–20 percentage points. This demonstrates the value of algorithms capable of capturing complex, nonlinear interactions within venture capital data. At the same time, interpretability analyses reveal a consistent set of features that underpin venture outcomes across modeling paradigms. Deal characteristics such as leveraged buyout financing, buyout deals, and public market financing emerged as systematically associated with higher probabilities of success. Conversely, early-stage or unstructured venture capital equity investments were negatively associated with success, highlighting the heightened risks of such structures. At the investee level, variables such as the year of first investment and the number of investments received were consistently among the strongest predictors, while investor-related features, including experience and recency of activity, also played significant roles.

The contribution of this thesis lies in showing that predictive performance and interpretability can be jointly achieved in venture capital research, particularly in the renewable energy domain. By applying SHAP to both linear and machine learning models, the study demonstrates not only that machine learning methods offer superior predictive power, but also that their predictions can be explained in ways that align with established theoretical insights in the literature on venture financing, such as staged financing (Hellmann & Puri, 2002), structured exits (Kaplan & Strömberg, 2009), and investor expertise (Da Rin et al., 2013). This dual perspective is particularly valuable for practitioners seeking both performance and accountability in investment decision-making.

In sum, this thesis has shown that machine learning models can provide substantial improvements over traditional linear models in predicting the success of renewable energy startups. At the same time, SHAP analysis has highlighted a consistent set of features that underpin venture outcomes, reinforcing the importance of investment timing, structured deal types, and investor experience. While the findings must be interpreted with caution given the study's limitations, they point toward the growing potential of machine learning as a decision-support tool in sustainable venture capital. By enhancing predictive accuracy without sacrificing interpretability, these models offer investors, accelerators, and policymakers a promising pathway to more efficient and transparent capital allocation in the green transition.

8. Limitations and future research

Despite the promising findings of this thesis, several methodological and substantive limitations must be acknowledged, which also indicate fruitful directions for future research. The dataset includes both successful and unsuccessful ventures, with approximately 52 percent classified as unsuccessful and 48 percent as successful. This design partially mitigates survivorship bias, a common issue in venture datasets where failed firms are often underreported (Cumming & Johan, 2009). However, since failure rates in venture capital are typically much higher in reality (Gompers & Lerner, 2001), the near-balanced distribution more likely reflects reporting practices in Refinitiv Eikon rather than the true underlying dynamics of the VC market. Thus, while the dataset improves representativeness compared to success-only samples, survivorship bias cannot be considered fully eliminated. In addition, potential reporting errors or omissions in Refinitiv Eikon cannot be excluded, which may have affected the precision of the models, particularly in borderline cases of success or failure.

A further limitation relates to the geographic and sectoral composition of the sample. The ventures examined are disproportionately concentrated in North America and Western Europe, while Asia, Africa, and Latin America remain underrepresented, a pattern also noted in other venture capital studies relying on Crunchbase or PitchBook (Block et al., 2018; Huang et al., 2020). This imbalance constrains the external validity of the findings, as institutional environments, financing ecosystems, and regulatory frameworks differ significantly across regions (Cumming & Johan, 2017). The sectoral scope is also restricted to renewable energy ventures. While this focus provides thematic coherence and relevance for the green transition, it limits the generalizability of the results to other areas of sustainable innovation such as energy efficiency, sustainable mobility, or carbon capture. Expanding the dataset geographically and sectorally in future work would improve the robustness and applicability of the models.

Another constraint lies in the temporal dimension of the analysis. The models were trained and tested without explicitly incorporating time-aware validation. As a result, they may be exposed to temporal leakage and unable to capture evolving market dynamics fully. Although deal characteristics proved highly informative and offered valuable structural insights, this choice likely restricted the models' ability to identify time-dependent patterns. Incorporating longitudinal validation frameworks, such as rolling or walk-forward cross-validation, would strengthen confidence in the temporal robustness of predictions.

The inclusion of last-round investment variables presents an additional methodological limitation. While these variables enriched the analysis by revealing structural financing patterns, they depart from a strictly ex-ante decision-making context, where such information would not yet be available to investors. For this reason, the models are best interpreted as decision-support tools that help simulate outcomes under different financing scenarios, rather than pure predictors based solely on initial conditions.

Finally, the models rely exclusively on structured financial and operational indicators. Qualitative and unstructured factors such as founder characteristics, intellectual property, innovation quality, or strategic partnerships were not included, despite their centrality in venture capital decision-making (Kaplan & Strömberg, 2004). This exclusion ensures objectivity, comparability, and reproducibility, but inevitably reduces comprehensiveness. As such, the predictive outputs generated here should complement, rather than replace, the contextual judgment and due diligence of experienced investors. Future research should address these limitations by integrating unstructured and qualitative data sources, such as patent activity, media sentiment, or founder backgrounds, which could enrich the feature space and capture dimensions of innovation and market positioning not reflected in financial records. Expanding the dataset to cover a broader set of geographies and green technologies would further enhance external validity, while the adoption of longitudinal validation techniques would ensure that models remain robust under changing market conditions. Finally, combining machine learning predictions with expert-driven investment criteria may yield hybrid frameworks that balance efficiency with accountability, offering more comprehensive tools for decision-making in sustainable finance.

In sum, while this thesis has demonstrated the potential of machine learning and interpretability techniques for forecasting venture outcomes in renewable energy, its findings must be interpreted with caution. Acknowledging these limitations provides a basis for methodological refinement and empirical expansion, ensuring that subsequent research can build on this foundation to advance both academic understanding and practical applications in green venture capital.

9. Appendix

Exhibit 10: list of the dependent variables/features in the dataset, grouped into four main categories: Investee Company Characteristics, Investor Firm Characteristics, Fund Characteristics, and Deal Characteristics.

Category	Variable
Investee Company Characteristics - Funding & Investment History	Total Funding Received To Date (USD, Millions)
Investee Company Characteristics - Funding & Investment History	First Investment Received YEAR
Investee Company Characteristics - Funding & Investment History	Number Of Investments Received To Date
Investee Company Characteristics - Funding & Investment History	Number Of Investor Funds To Date
Investee Company Characteristics - Funding & Investment History	Number Of Investor Firms To Date
Investee Company Characteristics - Country Attractiveness	Country RECAI Top 10
Investee Company Characteristics - Country Attractiveness	Country RECAI 11–30
Investee Company Characteristics - Country Attractiveness	Country RECAI 31–Others
Investee Company Characteristics - Firm Characteristics	Company Founded YEAR
Investee Company Characteristics - Firm Characteristics	Primary Customer Type: Business
Investee Company Characteristics - Firm Characteristics	Primary Customer Type: Consumer
Investee Company Characteristics - Firm Characteristics	Primary Customer Type: All
Investee Company Characteristics - Firm Characteristics	Primary Customer Type: Not Specified
Investee Company Characteristics - Sector Alignment	Sector specified in investor target
Investee Company Characteristics - Sector Alignment	Sector not specified in investor target
Investee Company Characteristics - Sector Alignment	Sector diversified/undefined in investor target

Investor Firm Characteristics - Investment History	First Investment YEAR
Investor Firm Characteristics - Investment History	Last Investment YEAR
Investor Firm Characteristics - Investment History	Total Number of Companies Invested
Investor Firm Characteristics - Investment History	Total Number Of Deals
Investor Firm Characteristics - Investment History	Total Estimated Equity Invested To Date (USD, Millions)
Investor Firm Characteristics - Investment History	Firm Founded YEAR
Investor Firm Characteristics - Country Attractiveness	Country RECAI Top 10
Investor Firm Characteristics - Country Attractiveness	Country RECAI 11–30
Investor Firm Characteristics - Country Attractiveness	Country RECAI 31–Others

Investor Firm Characteristics - Investor Type	Government Affiliated Program
Investor Firm Characteristics - Investor Type	Corporate PE/Venture
Investor Firm Characteristics - Investor Type	Endowment, Foundation or Pension Fund
Investor Firm Characteristics - Investor Type	Incubator/Development Program
Investor Firm Characteristics - Investor Type	Angel Group
Investor Firm Characteristics - Investor Type	Bank Affiliated
Investor Firm Characteristics - Investor Type	Individuals
Investor Firm Characteristics - Investor Type	Insurance Firm Affiliate
Investor Firm Characteristics - Investor Type	Investment Management Firm
Investor Firm Characteristics - Investor Type	Private Equity Firm
Investor Firm Characteristics - Investor Type	Other
Investor Firm Characteristics - Stage Preferences	Seed & Start-up Stage
Investor Firm Characteristics - Stage Preferences	Growth Stage
Investor Firm Characteristics - Stage Preferences	Late Stage / Pre-Exit Stage
Investor Firm Characteristics - Stage Preferences	Fund of Funds
Investor Firm Characteristics - Stage Preferences	Generalist PE Stage
Investor Firm Characteristics - Stage Preferences	Other Stage
Investor Firm Characteristics - Stage Preferences	Buyout / M&A Focused

Fund Characteristics - Fund Type	Generalist Private Equity
Fund Characteristics - Fund Type	Venture Capital
Fund Characteristics - Fund Type	Buyout
Fund Characteristics - Fund Type	Other
Fund Characteristics	Fund Year

Deal Characteristics - General Deal Type	Venture Capital Deals (Yes/No)
Deal Characteristics - General Deal Type	Buyout Deals (Yes/No)
Deal Characteristics - General Deal Type	Pure Venture Capital Deals (Yes/No)
Deal Characteristics - General Deal Type	VC Reporter Deals (Yes/No)
Deal Characteristics - Security Type	Acquisition Financing
Deal Characteristics - Security Type	Common Stock
Deal Characteristics - Security Type	Bridge Loan
Deal Characteristics - Security Type	Debt
Deal Characteristics - Security Type	Convertible Bonds (CB)
Deal Characteristics - Security Type	Leveraged Buyout Financing
Deal Characteristics - Security Type	Series Convertible Preferred Stock/Shares
Deal Characteristics - Security Type	Other
Deal Characteristics - Security Type	Venture Capital Equity Investment
Deal Characteristics - Financing Stage	Acquisition
Deal Characteristics - Financing Stage	Early Stage
Deal Characteristics - Financing Stage	Expansion
Deal Characteristics - Financing Stage	Later Stage
Deal Characteristics - Financing Stage	Other
Deal Characteristics - Financing Stage	Public Market
Deal Characteristics - Financing Stage	Seed

Deal Characteristics - Financing Stage	VC Partnership
Deal Characteristics - Round Details	Round Number
Deal Characteristics - Round Details	Deal Age At Financing (Months)
Deal Characteristics - Round Details	Number Of Funds

Exhibit 11: reports the mathematical definitions of the evaluation metrics used in this study, including Accuracy, Precision, Recall, F1 Score, and ROC-AUC. These formulas clarify how each metric balances correct and incorrect predictions, providing the basis for comparing model performance.

Accuracy $$\frac{TP + TN}{TP + TN + FP + FN}$$

Precision $$\frac{TP}{TP + FP}$$

Recall (TPR) $$\frac{TP}{TP + FN}$$

F1 Score $$2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} = \frac{2TP}{2TP + FP + FN}$$

ROC-AUC $$\int_0^1 TPR(FPR^{-1}(x))\,dx = Pr(score(X_1) \geq score(X_0))$$

Legend:
TP = True Positives
TN = True Negatives
FP = False Positives
FN = False Negatives
TPR = True Positive Rate (Recall)
FPR = False Positive Rate

Exhibit 12: SHAP Importance of Investee Company Characteristics across Models (based on the first 20 most influential features by mean SHAP value, with model and position in brackets)

Feature: Investee Company Characteristics	***Model (position)***
First Investment Received YEAR	Logistic Regression (1), Probit (7), LDA (1), XGBoost (1), Naive Bayes (12), MLP (1), Decision Tree (1), LightGBM (1), KNN (2), Random Forest (1), SVM (1), CatBoost (1), Stacking (1), Voting (1)
Number Of Investments Received To Date	XGBoost (3), MLP (6), Decision Tree (2), LightGBM (3), Random Forest (4), CatBoost (3), Stacking (3), Voting (3)
Number Of Investor Firms To Date	Logistic Regression (3), Probit (12), LDA (3), Decision Tree (13), LightGBM (10), Naive Bayes (15), MLP (14), SVM (6), CatBoost (9), Stacking (18), Voting (7)
Number Of Investor Funds To Date	XGBoost (15), Decision Tree (14), LightGBM (11), Naive Bayes (17), Random Forest (7), CatBoost (8), Stacking (15)
Total Funding Received To Date	XGBoost (5), Naive Bayes (4), Decision Tree (6), LightGBM (6), Random Forest (8), CatBoost (6), Stacking (5), Voting (5)
Company Founded YEAR	Logistic Regression (4), Probit (16), LDA (7), XGBoost (4), Naive Bayes (18), MLP (9), Decision Tree (7), LightGBM (4), KNN (7), Random Forest (3), SVM (3), CatBoost (4), Stacking (4), Voting (4)
Primary Customer Type: Business	XGBoost (20), LightGBM (17), MLP (18), KNN (15), SVM (18)
Primary Customer Type: Consumer	KNN (18), SVM (19)
Primary Customer Type: Not Specified	Logistic Regression (20), LDA (20)
Investee Country RECAI 31-Others	Naive Bayes (16)

Exhibit 13: SHAP Importance of Investor Firm Characteristics across Models (based on the first 20 most influential features by mean SHAP value, with model and position in brackets)

Feature: Investor Firm Characteristics	***Model (position)***
Firm Investors First Investment YEAR	Logistic Regression (7), LDA (10), XGBoost (10), Decision Tree (12), LightGBM (15), Random Forest (12), SVM (17), Stacking (10), Voting (12)
Firm Investors Last Investment YEAR	XGBoost (16), MLP (3), Decision Tree (16), LightGBM (14), Random Forest (10), SVM (7), CatBoost (11), KNN (14)
Firm Investors Total Number of Companies Invested	Logistic Regression (12), Probit (4), LDA (4), XGBoost (11), LightGBM (16), CatBoost (16), Stacking (14), Voting (16)
Firm Investors Total Number Of Deals	Logistic Regression (16), Probit (6), LDA (5), XGBoost (6), Decision Tree (14), LightGBM (8), CatBoost (13), Stacking (11), Voting (10)
Firm Investors Total Estimated Equity Invested To Date	XGBoost (7), Decision Tree (11), LightGBM (12), Random Forest (15), CatBoost (14), Stacking (6), Voting (17)
Firm Investors Founded YEAR	Logistic Regression (14), XGBoost (8), Decision Tree (9), LightGBM (9), MLP (17), CatBoost (10), Stacking (16), Voting (14)
Investor Country RECAI 11–30	SVM (14)
Investor Type: Venture Capital (Firm)	LDA (12), SVM (16)
Investor Type: Private Equity Firm (Firm)	Logistic Regression (18)
Investor Type: Incubator/Development Program (Firm)	Logistic Regression (19), LDA (18), Naive Bayes (3), KNN (4), Voting (20)
Investor Type: Individuals (Firm)	Naive Bayes (5), KNN (10)
Investor Type: Investment Management Firm (Firm)	Naive Bayes (8), Stacking (20)
Investor Type: Government Affiliated Program (Firm)	Naive Bayes (20), MLP (11), KNN (17)
Investor Type: Bank Affiliated (Firm)	KNN (12)
Investor Type: Generalist Private Equity (Firm)	MLP (8)
Stage Preference: Seed & Start-up (Firm)	XGBoost (19), Decision Tree (17), LightGBM (20), MLP (20), KNN (9), Stacking (17)
Stage Preference: Late Stage / Pre-Exit (Firm)	Naive Bayes (14), KNN (20)
Stage Preference: Generalist PE (Firm)	MLP (19)
Stage Preference: Other Stage (Firm)	MLP (12)

Exhibit 14: SHAP Importance of Fund Characteristics across Models (based on the first 20 most influential features by mean SHAP value, with model and position in brackets)

Feature: Fund Characteristics	**Model (position)**
Fund Year	Probit (18), XGBoost (13), MLP (7), Decision Tree (10), LightGBM (13), Random Forest (17), SVM (12), CatBoost (20), Stacking (19)
Fund Type: Generalist Private Equity	Logistic Regression (13), LDA (13)
Fund Type: Venture Capital	Logistic Regression (10), LDA (12), Decision Tree (19), LightGBM (19), MLP (13), Stacking (8), Voting (11), CatBoost (18)

Exhibit 15: SHAP Importance of Deal Characteristics across Models (based on the first 20 most influential features by mean SHAP value, with model and position in brackets)

Feature: Deal Characteristics	**Model (position)**
Leveraged Buyout Financing	Logistic Regression (5), LDA (9), XGBoost (2), Naïve Bayes (1), MLP (2), Decision Tree (3), LightGBM (2), KNN (1), Random Forest (2), SVM (2), CatBoost (2), Stacking (2), Voting (2)
Venture Capital Equity Investment	Logistic Regression (2), Probit (2), LDA (2), XGBoost (17), MLP (4), Random Forest (19), SVM (9), CatBoost (15), Voting (15)
Acquisition Financing	Logistic Regression (9), Probit (3), LDA (8), Random Forest (18), SVM (8), Stacking (11), Voting (19)
Common Stock	Probit (8), Naïve Bayes (9)
Bridge Loan	Probit (10)
Debt	Probit (9), KNN (11)
Convertible Bonds (CB)	Probit (17)
Series Convertible Preferred Stock/Shares	Logistic Regression (8), Probit (5), LDA (6), KNN (16), Voting (13)
Other (stage)	Naïve Bayes (6)
Early Stage	Logistic Regression (11), Probit (13), LDA (11), MLP (15), KNN (19), XGBoost (18)
Expansion	Logistic Regression (14), Probit (19), LDA (17), Naïve Bayes (19), KNN (8)
Later Stage	Decision Tree (18), KNN (12)
Acquisition (stage)	Probit (15), LDA (16), Random Forest (20)
Public Market	Logistic Regression (16), Probit (1), LDA (15), XGBoost (12), MLP (10), Decision Tree (5), LightGBM (5), Random Forest (14), SVM (4), CatBoost (5), Stacking (7), Voting (6)
Seed	Naïve Bayes (13), KNN (6)

VC Partnership	Probit (11)
Buyout Deals	Logistic Regression (6), Probit (14), LDA (14), MLP (5), Decision Tree (8), Random Forest (10), SVM (10), CatBoost (12), Voting (9)
Pure Venture Capital Deals	Probit (20), XGBoost (14), Random Forest (5), SVM (13), CatBoost (7), Stacking (13), Voting (17)
Venture Capital Deals	Random Forest (13)
Round Number	Naïve Bayes (10), MLP (16), SVM (20), LightGBM (18)
Deal Age At Financing (Months)	XGBoost (9), Naïve Bayes (7), MLP (9), Decision Tree (4), LightGBM (7), Random Forest (6), CatBoost (17), Stacking (9), Voting (8)
Number Of Funds	Naïve Bayes (11), Decision Tree (20), Random Forest (16), SVM (15), CatBoost (19)

Exhibit 16: Top 20 SHAP Feature Importances – Logistic Regression (ranked by mean |SHAP| value)

Shap importance Logistic Regression	
Feature	**Mean |SHAP| value**
Investee Company First Investment Received YEAR	0.185
Primary Security Type Venture Capital Equity Investment	0.071
Investee Company Number Of Investor Firms To Date	0.059
Investee Company Founded YEAR	0.055
Primary Security Type Leveraged Buyout Financing	0.052
Buyout Deals	0.051
Firm Investors First Investment YEAR	0.049
Primary Security Type Series Convertible Preferred Stock/Shares	0.041
Primary Security Type Acquisition Financing	0.033
Fund Investors Type: Venture Capital	0.033
Deal Round Financing Stage Code 2 Early Stage	0.030
Firm Investors Total Number of Companies Invested	0.024
Fund Investors Type: Generalist Private Equity	0.022
Firm Investors Founded Year	0.020
Deal Round Financing Stage Code 2 Expansion	0.017
Firm Investors Total Number Of Deals	0.016
Deal Round Financing Stage Code 2 Public Market	0.015
Firm Investors Type: Private Equity Firm	0.015
Firm Investors Type: Incubator/Development Program	0.013
Investee Company Primary Customer Type: Not Specified	0.012

Exhibit 17: Top 20 SHAP Feature Importances – LDA (ranked by mean |SHAP| value)

Shap Importance LDA	
Feature	**Mean |SHAP| value**
Investee Company First Investment Received YEAR	0.163
Primary Security Type Venture Capital Equity Investment	0.124
Investee Company Number Of Investor Firms To Date	0.092
Firm Investors Total Number of Companies Invested	0.077
Firm Investors Total Number Of Deals	0.077
Primary Security Type Series Convertible Preferred Stock/Shares	0.056
Investee Company Founded YEAR	0.045
Primary Security Type Acquisition Financing	0.042
Primary Security Type Leveraged Buyout Financing	0.038
Firm Investors First Investment YEAR	0.029
Deal Round Financing Stage Code 2 Early Stage	0.027
Firm Investors Type: Venture Capital	0.020
Fund Investors Type: Generalist Private Equity	0.020
Buyout Deals	0.019
Deal Round Financing Stage Code 2 Public Market	0.018
Deal Round Financing Stage Code 2 Acquisition	0.017
Deal Round Financing Stage Code 2 Expansion	0.017
Firm Investor Type: Incubator/Development Program	0.016
Investee Company Total Funding Received To Date	0.016
Investee Company Primary Customer Type: Not Specified	0.015

Exhibit 18: Top 20 SHAP Feature Importances – Probit Regression (ranked by mean |SHAP| value)

Shap Importance Probit	
Feature	**Mean |SHAP| value**
Deal Round Financing Stage Code 2 Public Market	1.42
Primary Security Type Venture Capital Equity Investment	0.81
Primary Security Type Acquisition Financing	0.63
Firm Investors Total Number of Companies Invested	0.59
Primary Security Type Series Convertible Preferred Stock/Shares	0.58
Firm Investors Total Number of Deals	0.57
Investee Company First Investment Received YEAR	0.43
Primary Security Type Common Stock	0.38
Primary Security Type Debt	0.36
Primary Security Type Bridge Loan	0.31
Deal Round Financing Stage Code 2 VC Partnership	0.30
Investee Company Number Of Investor Firms To Date	0.27
Deal Round Financing Stage Code 2 Early Stage	0.25

Buyout Deals	0.25
Deal Round Financing Stage Code 2 Acquisition	0.25
Investee Company Founded YEAR	0.24
Primary Security Type Convertible Bonds (CB)	0.21
Fund Investors Fund Year	0.20
Deal Round Financing Stage Code 2 Expansion	0.19
Pure Venture Capital Deals	0.19

Exhibit 19: Top 20 SHAP Feature Importances – XGBoost (ranked by mean |SHAP| value)

Shap Importance XGBoost	
Feature	**Mean |SHAP| value**
Investee Company First Investment Received YEAR	2.85
Primary Security Type Leveraged Buyout Financing	1.05
Investee Company Number Of Investments Received To Date	0.95
Investee Company Founded YEAR	0.68
Investee Company Total Funding Received To Date	0.43
Firm Investors Total Number Of Deals	0.42
Firm Investors Total Estimated Equity Invested To Date	0.40
Firm Investors Founded YEAR	0.40
Deal Age At Financing In Months	0.40
Firm Investors First Investment YEAR	0.39
Firm Investors Total Number of Companies Invested	0.39
Deal Round Financing Stage Code 2 Public Market	0.38
Fund Investors Fund Year	0.36
Pure Venture Capital Deals	0.34
Investee Company Number Of Investor Funds To Date	0.33
Firm Investors Last Investment YEAR	0.33
Primary Security Type Venture Capital Equity Investment	0.31
Deal Round Financing Stage Code 2 Early Stage	0.30
Firm Investors Stage Investment Preference Seed & Start-up Stage	0.28
Investee Company Primary Customer Type: Business	0.27

Exhibit 20: Top 20 SHAP Feature Importances – CatBoost (ranked by mean |SHAP| value)

Shap Importance CatBoost	
Feature	**Mean |SHAP| value**
Investee Company First Investment Received YEAR	1.85
Primary Security Type Leveraged Buyout Financing	0.60
Investee Company Number Of Investments Received To Date	0.35
Investee Company Founded YEAR	0.25

Deal Round Financing Stage Code 2 Public Market	0.19
Investee Company Total Funding Received To Date	0.18
Pure Venture Capital Deals	0.12
Investee Company Number Of Investor Funds To Date	0.12
Investee Company Number Of Investor Firms To Date	0.11
Firm Investors Founded Year	0.11
Firm Investors Last Investment YEAR	0.10
Buyout Deals	0.10
Firm Investors Total Number Of Deals	0.10
Firm Investors Total Estimated Equity Invested To Date	0.09
Primary Security Type Venture Capital Equity Investment	0.08
Firm Investors Total Number of Companies Invested	0.08
Deal Age At Financing In Months	0.07
Fund Investors Type: Venture Capital	0.05
Number Of Funds	0.05
Fund Investors Fund Year	0.04

Exhibit 21: Top 20 SHAP Feature Importances – LightGBM (ranked by mean |SHAP| value)

Shap Importance LightGBM	
Feature	**Mean |SHAP| value**
Investee Company First Investment Received YEAR	0.285
Primary Security Type Leveraged Buyout Financing	0.120
Investee Company Number Of Investments Received To Date	0.090
Investee Company Founded YEAR	0.040
Deal Round Financing Stage Code 2 Public Market	0.038
Investee Company Total Funding Received To Date	0.036
Deal Age At Financing In Months	0.035
Firm Investors Total Number Of Deals	0.034
Firm Investors Founded Year	0.032
Investee Company Number Of Investor Firms To Date	0.031
Investee Company Number Of Investor Funds To Date	0.030
Firm Investors Total Estimated Equity Invested To Date	0.030
Fund Investors Fund Year	0.028
Firm Investors Last Investment YEAR	0.027
Firm Investors First Investment YEAR	0.027
Firm Investors Total Number of Companies Invested	0.025
Investee Company Primary Customer Type: Business	0.021
Round Number	0.018
Fund Investors Type: Venture Capital	0.017
Firm Investors Stage Investment Preference Seed & Start-up Stage	0.016

Exhibit 22: Top 20 SHAP Feature Importances – Voting Classifier (ranked by mean |SHAP| value)

Shap importance Voting Classifier	
Feature	**Mean |SHAP| value**
Investee Company First Investment Received YEAR	0.211
Primary Security Type Leveraged Buyout Financing	0.100
Investee Company Number Of Investments Received To Date	0.058
Investee Company Founded YEAR	0.042
Investee Company Total Funding Received To Date	0.019
Deal Round Financing Stage Code 2 Public Market	0.018
Investee Company Number Of Investor Firms To Date	0.018
Deal Age At Financing In Months	0.017
Buyout Deals	0.016
Firm Investors Total Number Of Deals	0.012
Fund Investors Type: Venture Capital	0.012
Firm Investors First Investment YEAR	0.012
Primary Security Type Series Convertible Preferred Stock/Shares	0.012
Firm Investors Founded Year	0.011
Primary Security Type Venture Capital Equity Investment	0.011
Firm Investors Total Number of Companies Invested	0.009
Firm Investors Total Estimated Equity Invested To Date	0.007
Pure Venture Capital Deals	0.006
Primary Security Type Acquisition Financing	0.006
Firm Investors Type: Incubator/Development Program	0.005

Exhibit 23: Top 20 SHAP Feature Importances – Stacking Classifier (ranked by mean |SHAP| value)

Shap Importance Stacking Classifier	
Feature	**Mean |SHAP| value**
Investee Company First Investment Received YEAR	0.208
Primary Security Type Leveraged Buyout Financing	0.115
Investee Company Number Of Investments Received To Date	0.062
Investee Company Founded YEAR	0.048
Investee Company Total Funding Received To Date	0.025
Firm Investors Total Estimated Equity Invested To Date	0.024
Deal Round Financing Stage Code 2 Public Market	0.021
Fund Investors Type: Venture Capital	0.020
Deal Age At Financing In Months	0.019
Firm Investors First Investment YEAR	0.017
Firm Investors Total Number Of Deals	0.016
Primary Security Type Acquisition Financing	0.015

Pure Venture Capital Deals	0.014
Firm Investors Total Number of Companies Invested	0.012
Investee Company Number Of Investor Funds To Date	0.012
Firm Investors Founded Year	0.010
Firm Investors Stage Investment Preference Seed & Start-up Stage	0.008
Investee Company Number Of Investor Firms To Date	0.007
Fund Investors Fund Year	0.007
Firm Investors Type: Investment Management Firm	0.005

Exhibit 24: Top 20 SHAP Feature Importances – SVM (ranked by mean |SHAP| value)

Shap Importance SVM	
Feature	**Mean |SHAP| value**
Investee Company First Investment Received YEAR	0.118
Primary Security Type Leveraged Buyout Financing	0.047
Investee Company Founded YEAR	0.018
Deal Round Financing Stage Code 2 Public Market	0.013
Investee Company Number Of Investor Funds To Date	0.011
Investee Company Number Of Investor Firms To Date	0.010
Firm Investors Last Investment YEAR	0.009
Primary Security Type Acquisition Financing	0.008
Primary Security Type Venture Capital Equity Investment	0.008
Buyout Deals	0.006
Fund Investors Type: Generalist Private Equity	0.006
Fund Investors Fund Year	0.005
Pure Venture Capital Deals	0.006
INVESTOR RECAI ranked 11-30	0.006
Number Of Funds	0.005
Firm Investors Type: Venture Capital	0.005
Firm Investors First Investment YEAR	0.004
Investee Company Total Funding Received To Date	0.004
Investee Company Primary Customer Type: Consumer	0.004
Round Number	0.004

Exhibit 25: Top 20 SHAP Feature Importances – Random Forest (ranked by mean |SHAP| value)

Shap Importance Random Forest	
Feature	**Mean |SHAP| value**
Investee Company First Investment Received YEAR	0.140
Primary Security Type Leveraged Buyout Financing	0.050
Investee Company Founded YEAR	0.048

Investee Company Number Of Investments Received To Date	0.028
Pure Venture Capital Deals	0.025
Deal Age At Financing In Months	0.025
Investee Company Number Of Investor Funds To Date	0.025
Investee Company Total Funding Received To Date	0.021
Investee Company Number Of Investor Firms To Date	0.020
Buyout Deals	0.020
Firm Investors Last Investment YEAR	0.020
Firm Investors First Investment YEAR	0.019
Venture Capital Deals	0.018
Deal Round Financing Stage Code 2 Public Market	0.017
Firm Investors Total Estimated Equity Invested To Date	0.017
Number of Funds	0.017
Fund Investors Fund Year	0.016
Primary Security Type Acquisition Financing	0.016
Primary Security Type Venture Capital Equity Investment	0.015
Deal Round Financing Stage Code 2 Acquisition	0.014

Exhibit 26: Top 20 SHAP Feature Importances – Decision Tree (ranked by mean |SHAP| value)

Shap Importance Decision Tree	
Feature	**Mean |SHAP| value**
Investee Company First Investment Received YEAR	0.235
Investee Company Number Of Investments Received To Date	0.130
Primary Security Type Leveraged Buyout Financing	0.103
Deal Age At Financing In Months	0.041
Deal Round Financing Stage Code 2 Public Market	0.040
Investee Company Total Funding Received To Date	0.027
Investee Company Founded YEAR	0.025
Buyout Deals	0.018
Firm Investors Founded Year	0.017
Fund Investors Fund Year	0.017
Firm Investors Total Estimated Equity Invested To Date	0.017
Firm Investors First Investment YEAR	0.016
Investee Company Number Of Investor Firms To Date	0.015
Investee Company Number Of Investor Funds To Date	0.015
Firm Investors Total Number Of Deals	0.014
Firm Investors Last Investment YEAR	0.014
Firm Investors Stage Investment Preference Seed & Start-up Stage	0.013
Deal Round Financing Stage Code 2 Later Stage	0.013
Fund Investors Type: Venture Capital	0.012

Number Of Funds	0.011

Exhibit 27: Top 20 SHAP Feature Importances – MLP Classifier (ranked by mean |SHAP| value)

Shap Importance MLP	
Feature	**Mean |SHAP| value**
Investee Company First Investment Received YEAR	0.122
Primary Security Type Leveraged Buyout Financing	0.066
Firm Investors Last Investment YEAR	0.024
Primary Security Type Venture Capital Equity Investment	0.021
Buyout Deals	0.020
Investee Company Number Of Investments Received To Date	0.020
Fund Investors Fund Year	0.019
Firm Investors Type: Generalist Private Equity	0.018
Investee Company Founded YEAR	0.016
Deal Round Financing Stage Code 2 Public Market	0.015
Firm Investors Type: Government Affiliated Program	0.014
Firm Investors Stage Investment Preference Other Stage	0.013
Fund Investors Type: Venture Capital	0.012
Investee Company Number Of Investor Firms To Date	0.011
Deal Round Financing Stage Code 2 Early Stage	0.010
Round Number	0.009
Firm Investors Founded YEAR	0.009
Investee Company Primary Customer Type: Business	0.008
Firm Investors Stage Investment Preference Generalist PE Stage	0.008
Firm Investors Stage Investment Preference Seed & Start-up Stage	0.007

Exhibit 28: Top 20 SHAP Feature Importances – K-Nearest Neighbors (ranked by mean |SHAP| value)

Shap Importance KNN	
Feature	**Mean |SHAP| value**
Primary Security Type Leveraged Buyout Financing	0.025
Investee Company First Investment Received YEAR	0.013
Deal Round Financing Stage Code 2 Public Market	0.012
Firm Investors Type: Incubator/Development Program	0.010
Primary Security Type Acquisition Financing	0.009
Deal Round Financing Stage Code 2 Seed	0.008
Investee Company Founded YEAR	0.007
Deal Round Financing Stage Code 2 Expansion	0.006
Firm Investors Stage Investment Preference Seed & Start-up Stage	0.006

Firm Investors Type: Individuals	0.006
Primary Security Type Debt	0.005
Deal Round Financing Stage Code 2 Later Stage	0.005
Firm Investors Type: Bank Affiliated	0.005
Firm Investors Last Investment YEAR	0.005
Investee Company Primary Customer Type: Business	0.005
Primary Security Type Series Convertible Preferred Stock/Shares	0.005
Firm Investors Type: Government Affiliated Program	0.005
Investee Company Primary Customer Type: Consumer	0.005
Deal Round Financing Stage Code 2 Early Stage	0.005
Firm Investors Stage Investment Preference Late Stage / Pre-Exit Stage	0.004

Exhibit 29: Top 20 SHAP Feature Importances – Naïve Bayes (ranked by mean |SHAP| value)

Shap Importance Naive Bayes	
Feature	**Mean |SHAP| value**
Primary Security Type Leveraged Buyout Financing	0.163
Deal Round Financing Stage Code 2 Public Market	0.040
Firm Investor Type: Incubator/Development Program	0.031
Investee Company Total Funding Received To Date	0.024
Firm Investors Type: Individuals	0.021
Deal Round Financing Stage Code 2 Other	0.016
Deal Age At Financing In Months	0.013
Firm Investors Type: Investment Management Firm	0.012
Primary Security Type Common Stock	0.008
Round Number	0.008
Number Of Funds	0.007
Investee Company First Investment Received YEAR	0.007
Deal Round Financing Stage Code 2 Seed	0.004
Firm Investors Stage Investment Preference Late Stage / Pre-Exit Stage	0.004
Investee Company Number Of Investor Firms To Date	0.003
INVESTEE RECAI 31 - OTHERS	0.002
Investee Company Number Of Investor Funds To Date	0.002
Investee Company Founded YEAR	0.002
Deal Round Financing Stage Code 2 Expansion	0.002
Firm Investors Type: Government Affiliated Program	0.001

Exhibit 30: SHAP Summary Dot Plot – Logistic Regression (20 most relevant features, showing distribution, direction, and magnitude of SHAP values)

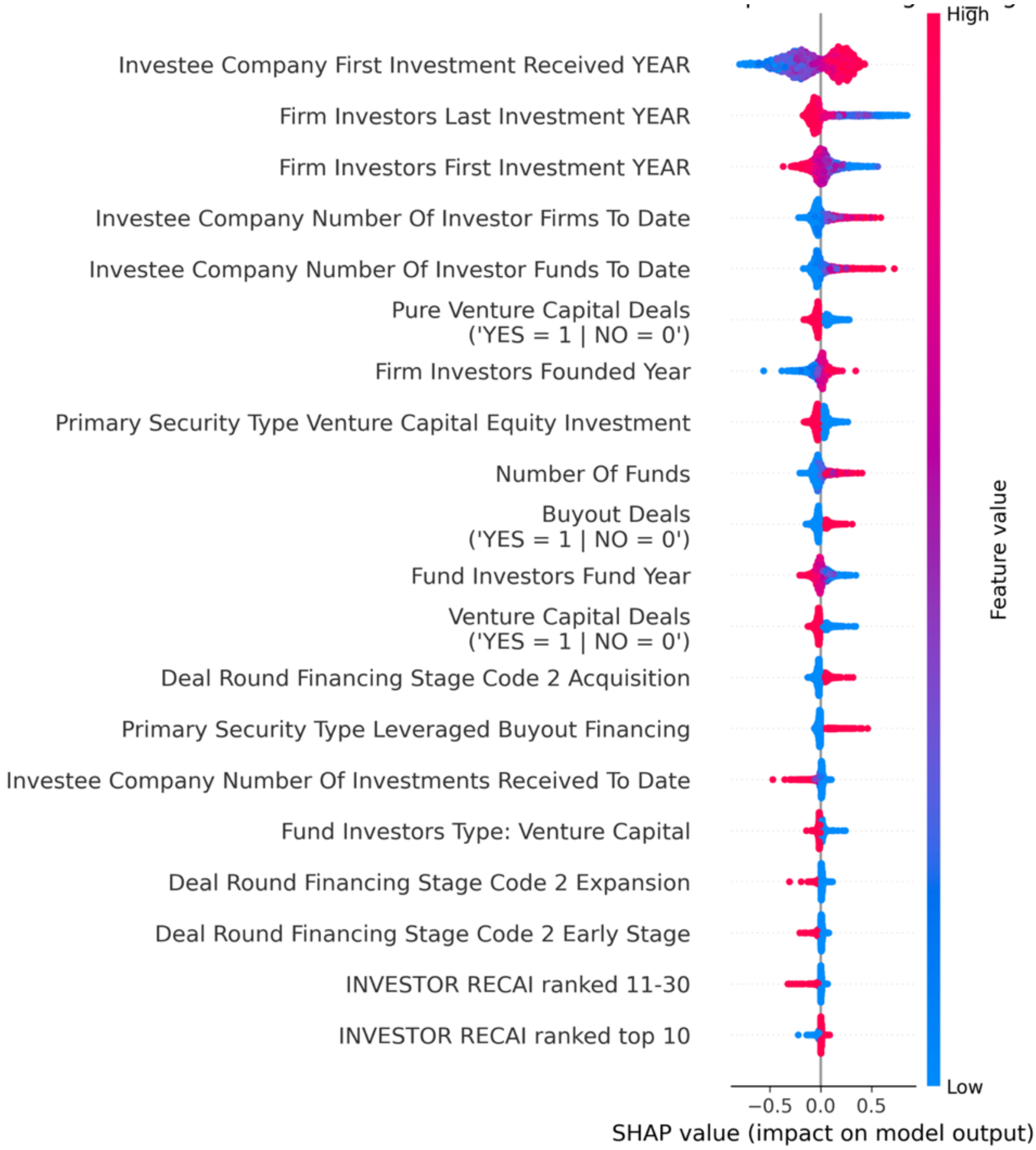


Exhibit 31: SHAP Summary Dot Plot – LDA (20 most relevant features, showing distribution, direction, and magnitude of SHAP values)

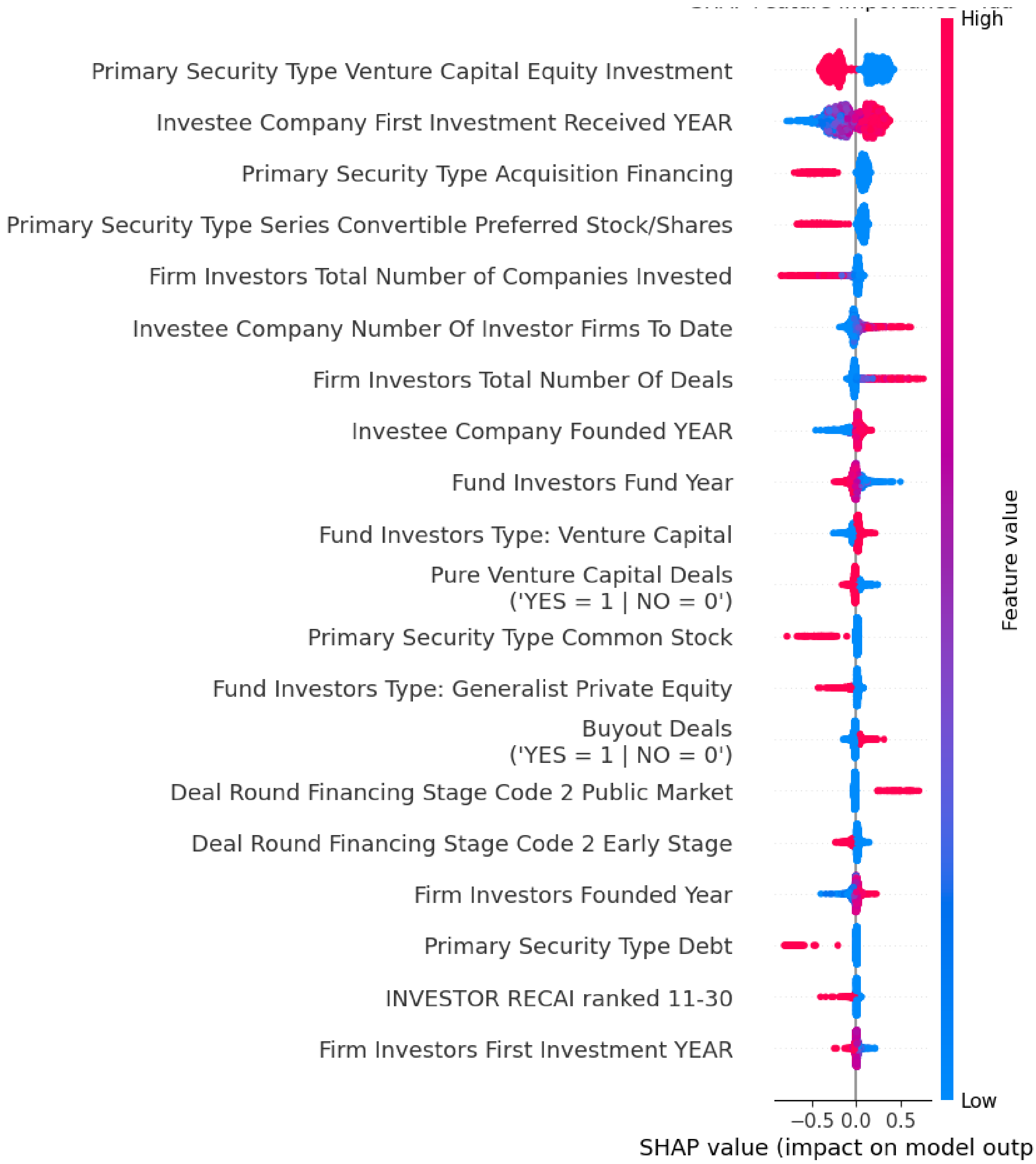


Exhibit 32: SHAP Summary Dot Plot – Probit (20 most relevant features, showing distribution, direction, and magnitude of SHAP values)

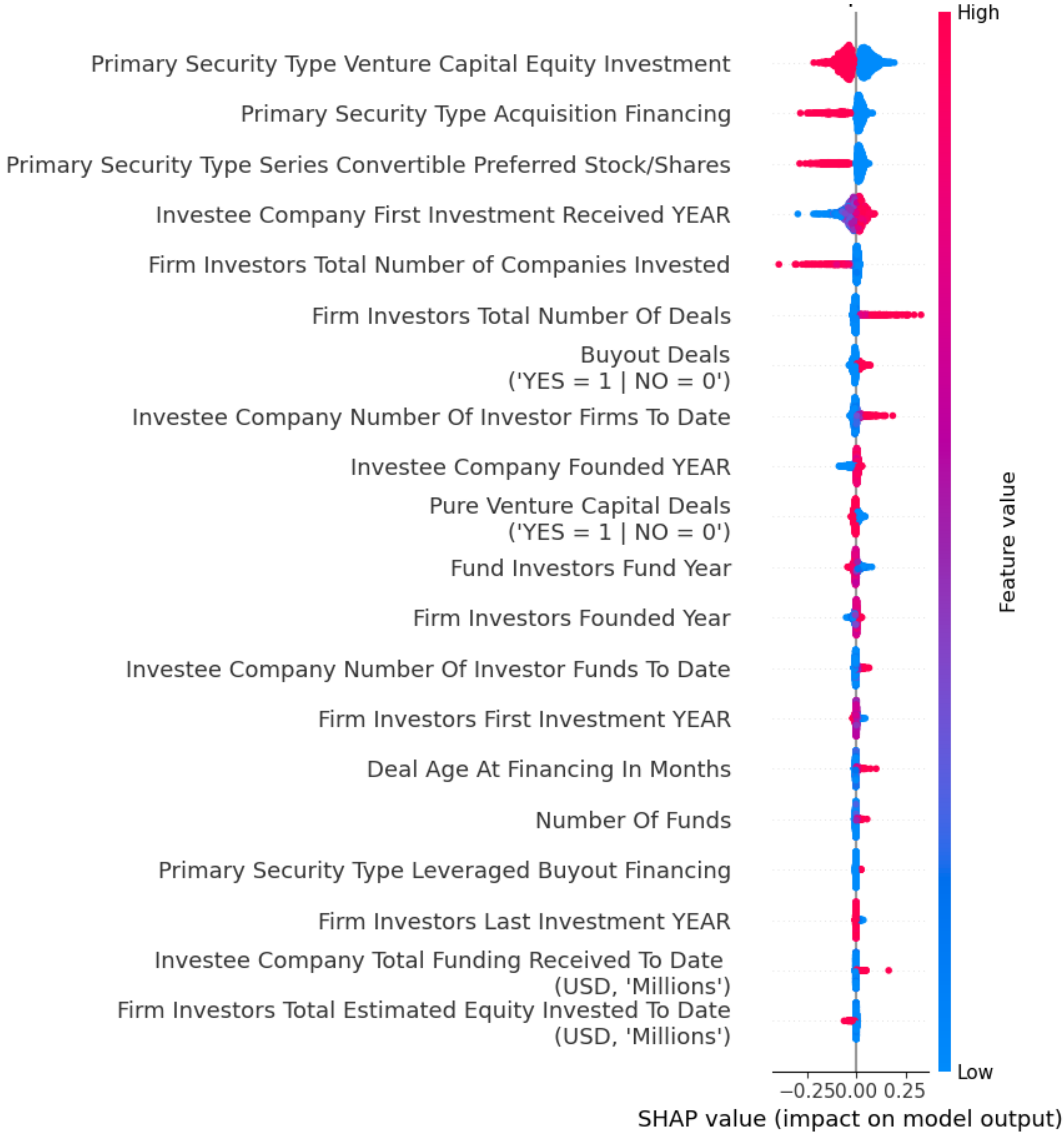


Exhibit 33: SHAP Summary Dot Plot – XGBoost (20 most relevant features, showing distribution, direction, and magnitude of SHAP values)

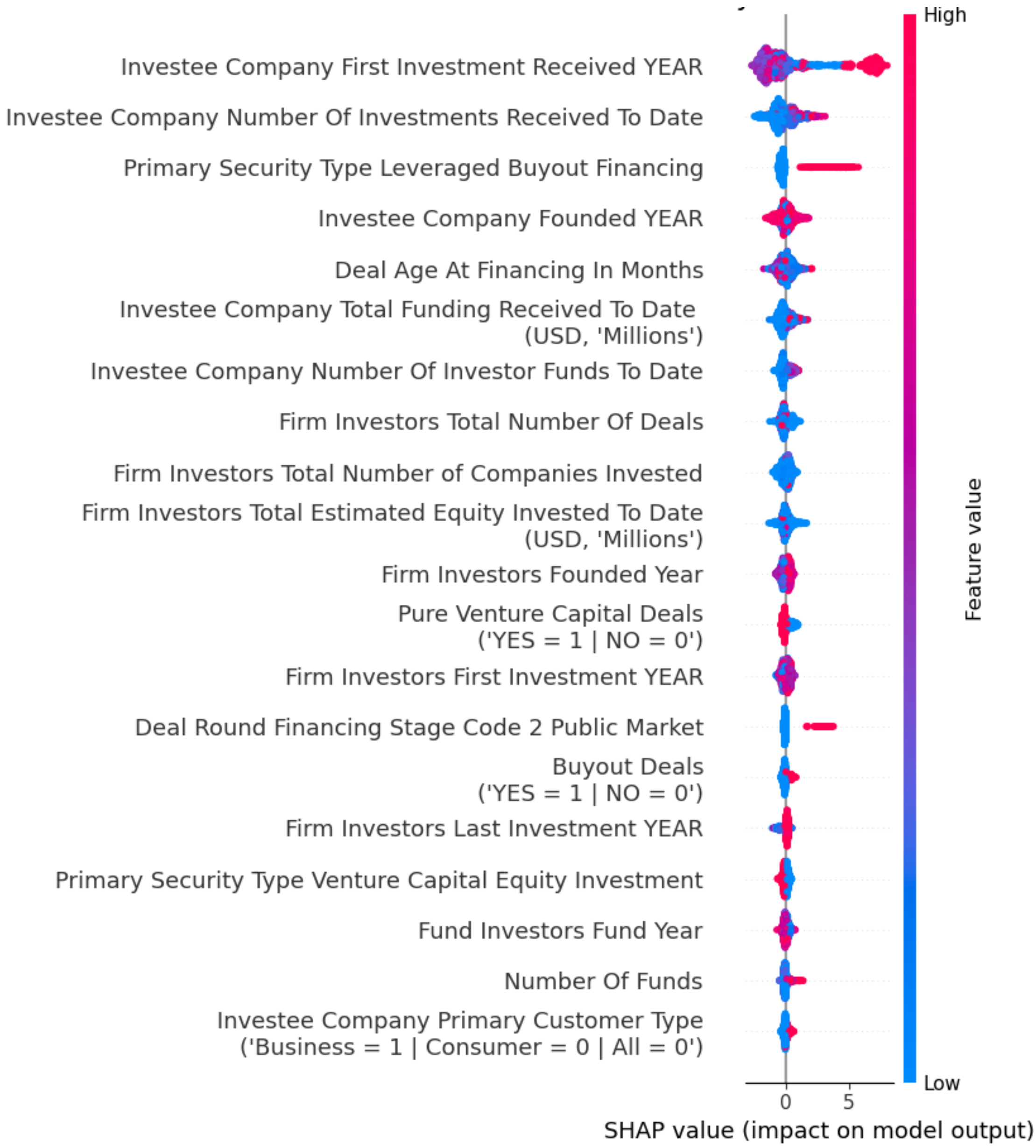


Exhibit 34: SHAP Summary Dot Plot – CatBoost (20 most relevant features, showing distribution, direction, and magnitude of SHAP values)

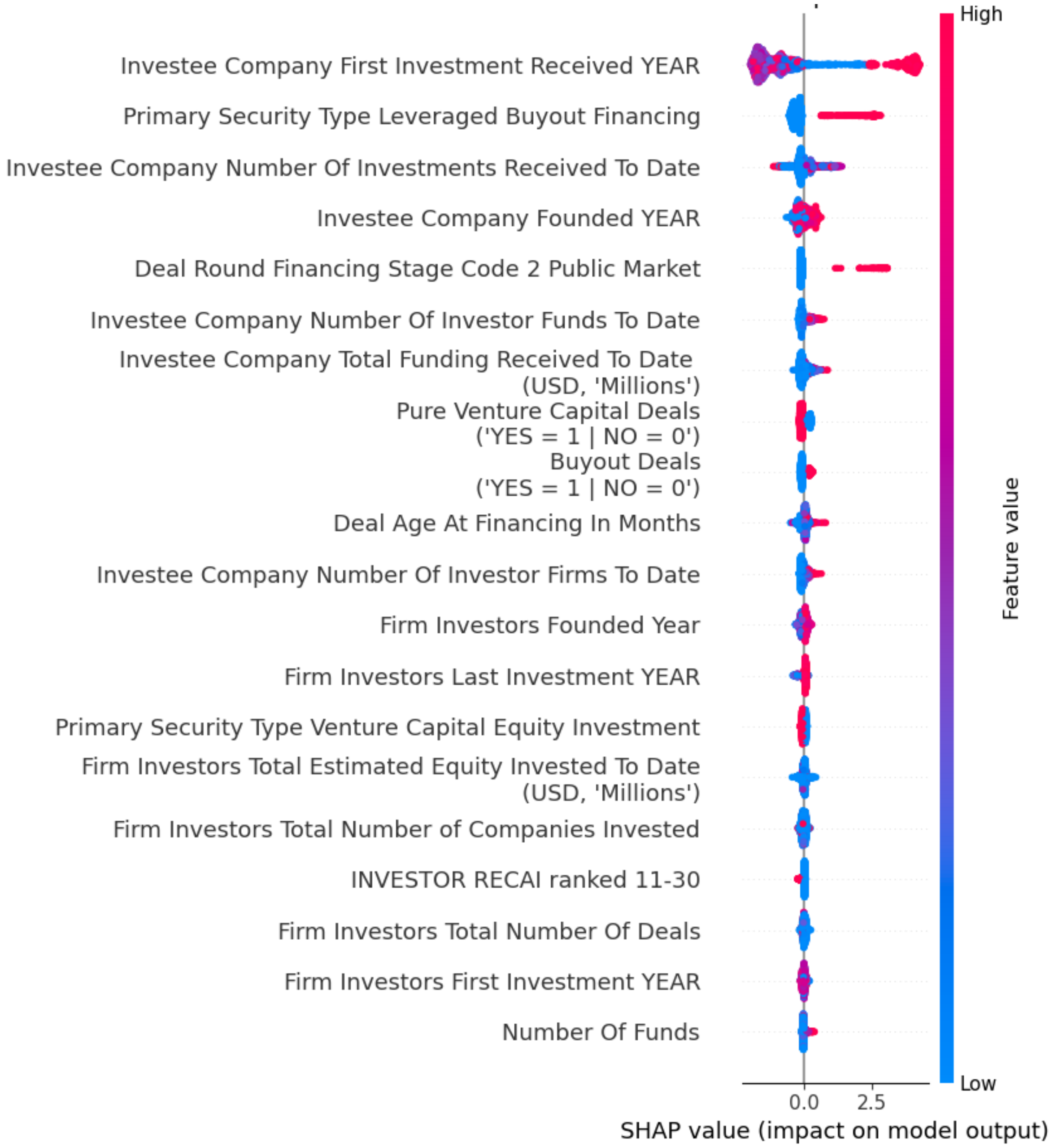


Exhibit 35: SHAP Summary Dot Plot – LightGBM (20 most relevant features, showing distribution, direction, and magnitude of SHAP values)

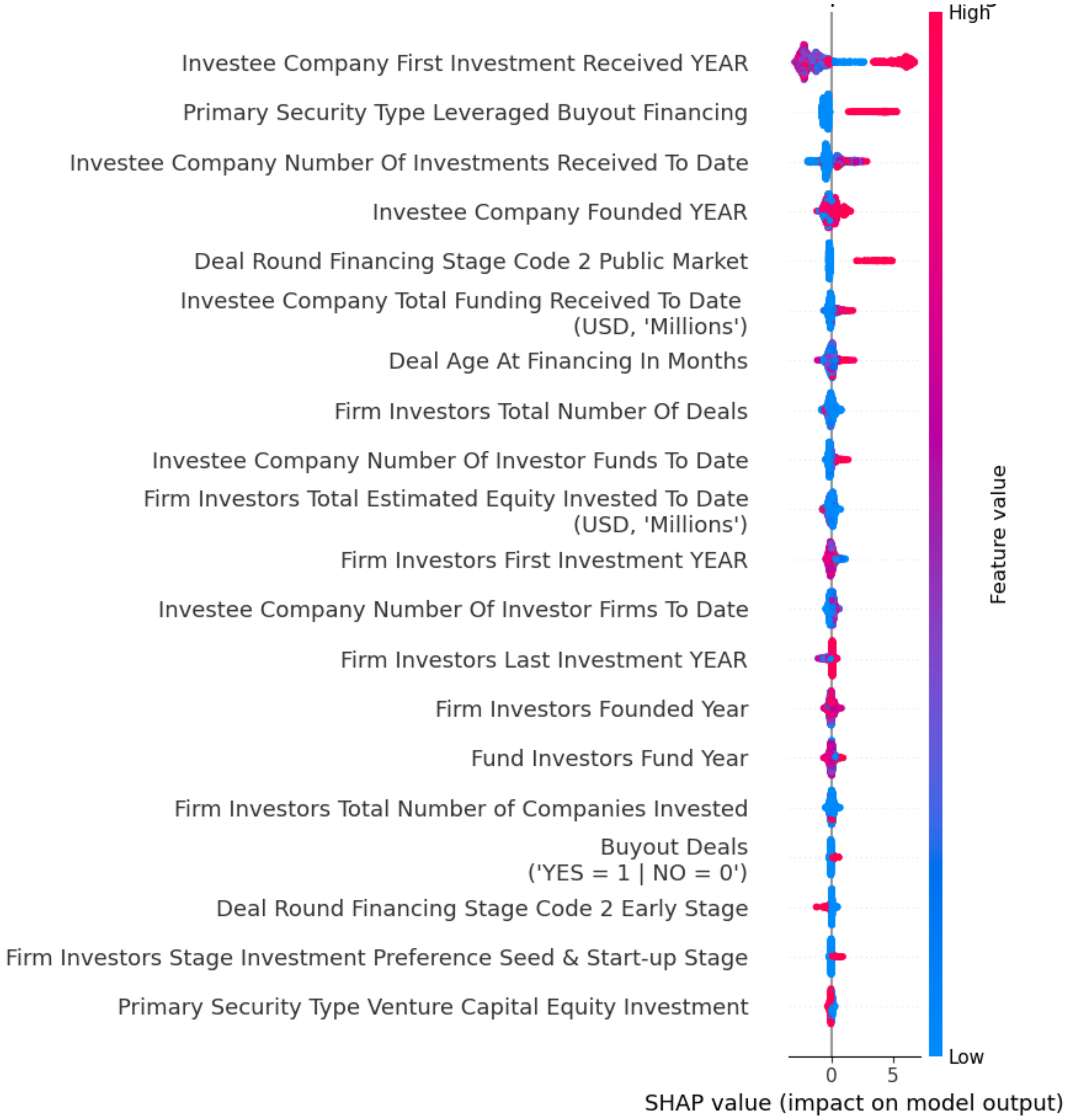


Exhibit 36: SHAP Summary Dot Plot – Voting Classifier (20 most relevant featu res, showing distribution, direction, and magnitude of SHAP values)

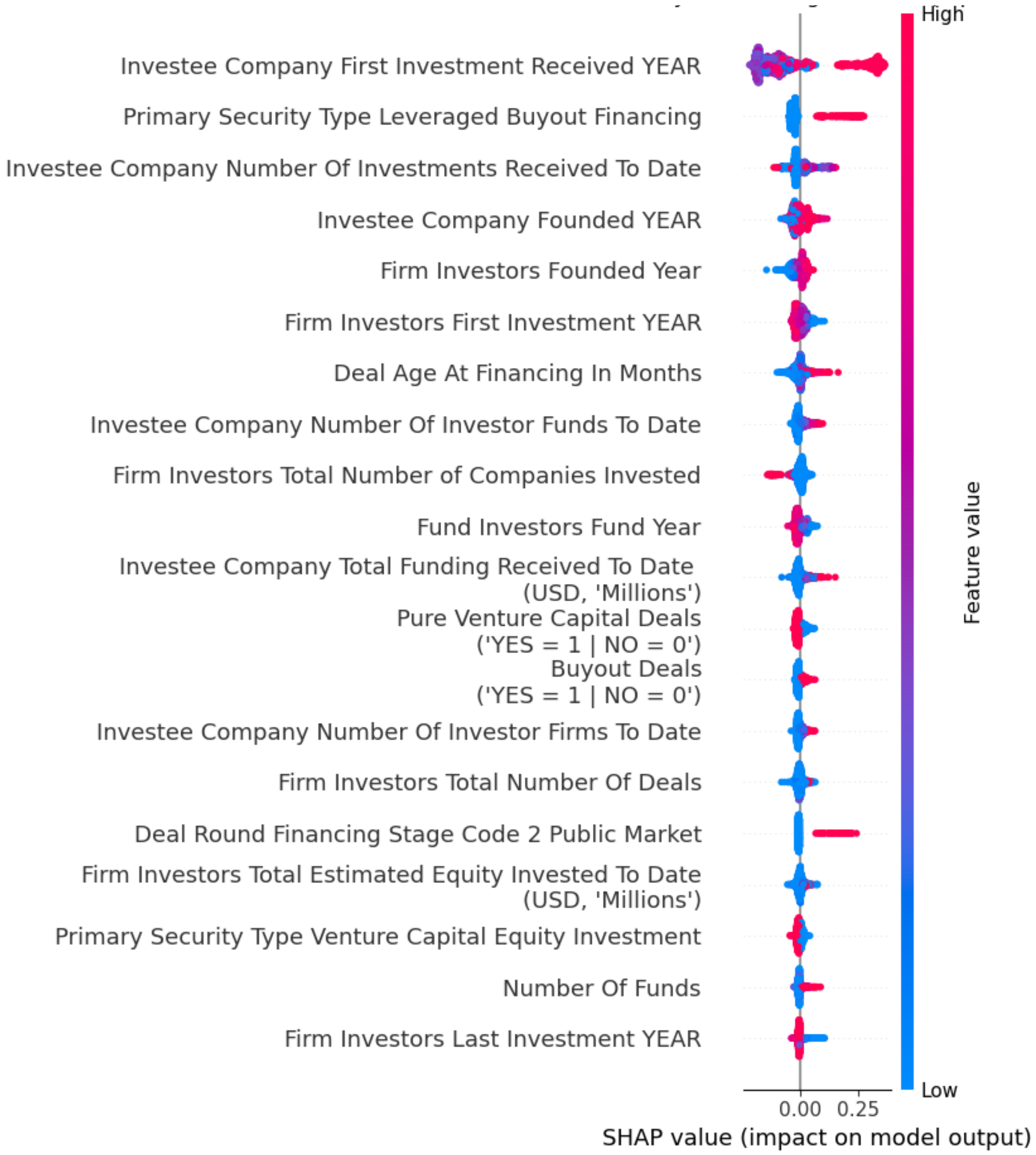


Exhibit 37: SHAP Summary Dot Plot – Stacking Classifier (20 most relevant featu res, showing distribution, direction, and magnitude of SHAP values)

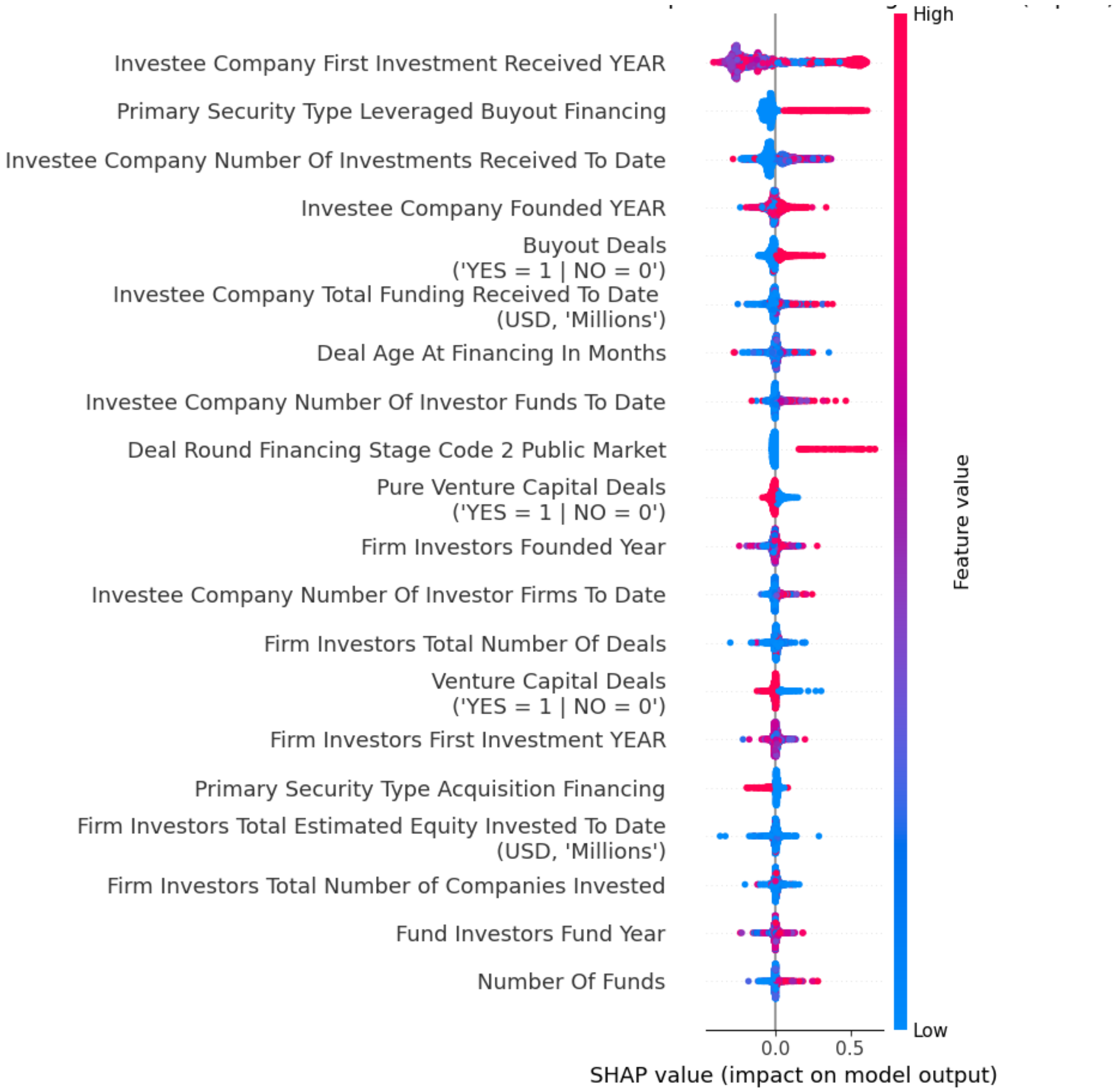


Exhibit 38: SHAP Summary Dot Plot – SVM (20 most relevant featu res, showing distribution, direction, and magnitude of SHAP values)

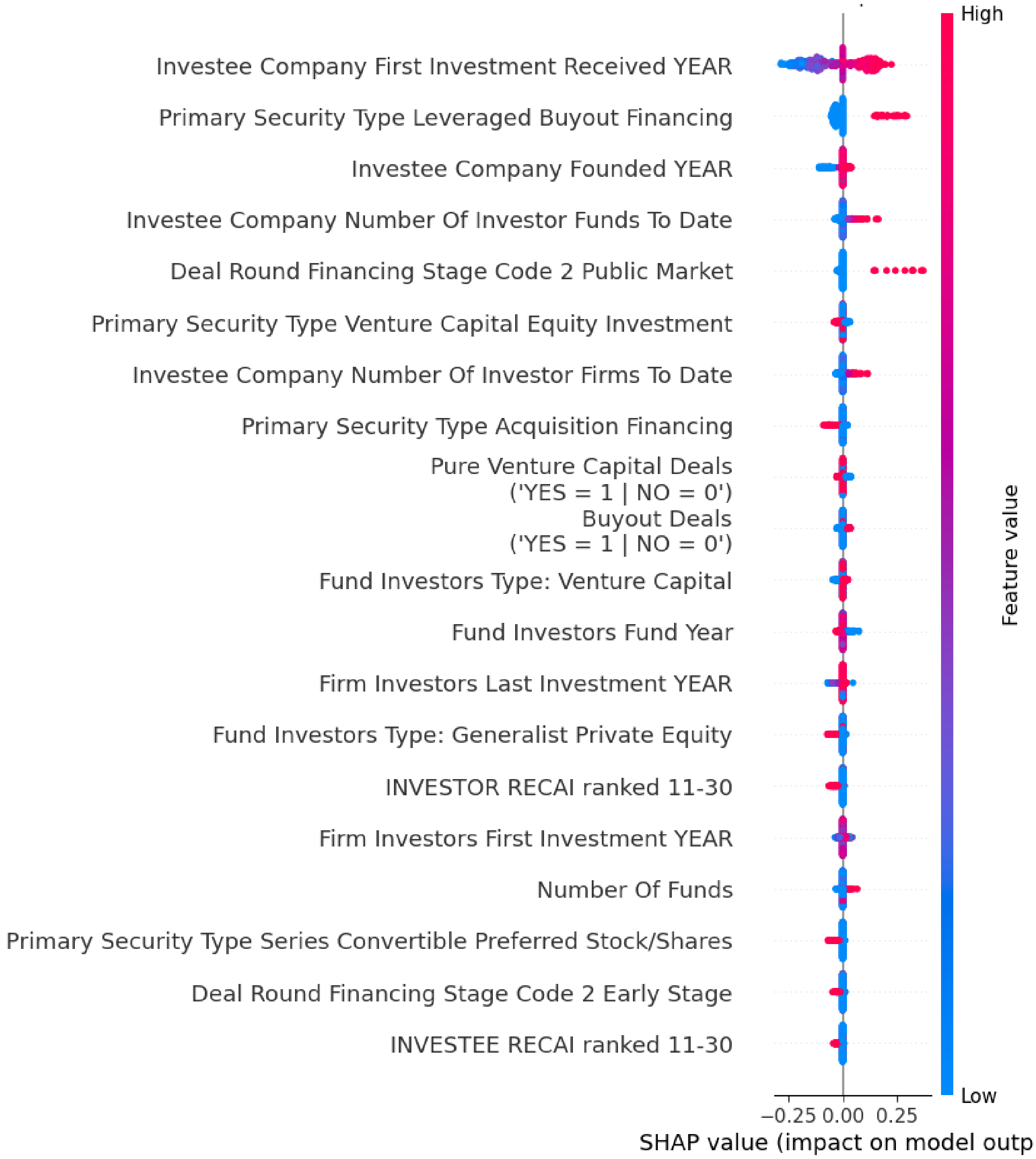


Exhibit 39: SHAP Summary Dot Plot – Random Forest (20 most relevant featu res, showing distribution, direction, and magnitude of SHAP values)

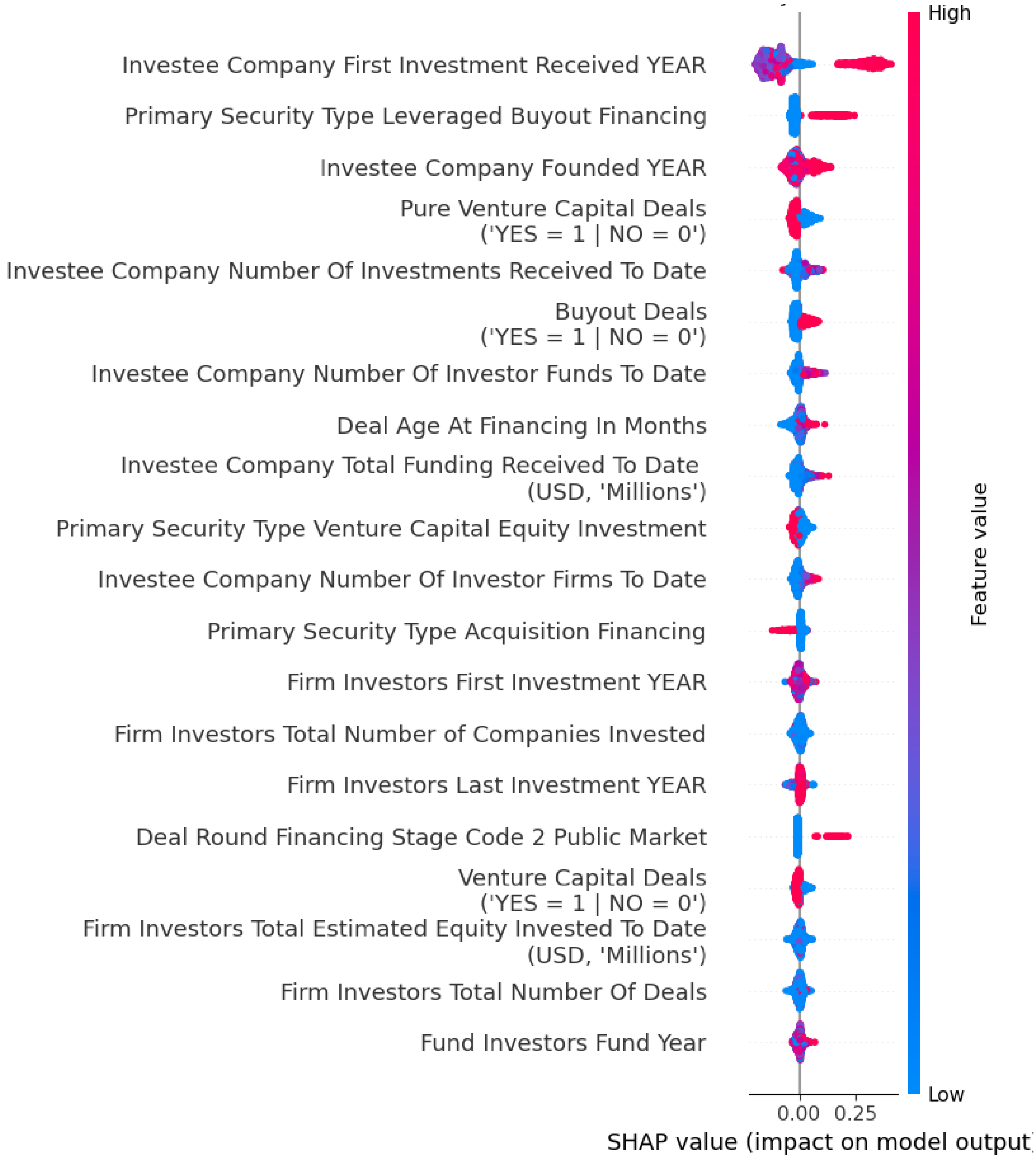


Exhibit 40: SHAP Summary Dot Plot – Decision Tree (20 most relevant featu res, showing distribution, direction, and magnitude of SHAP values)

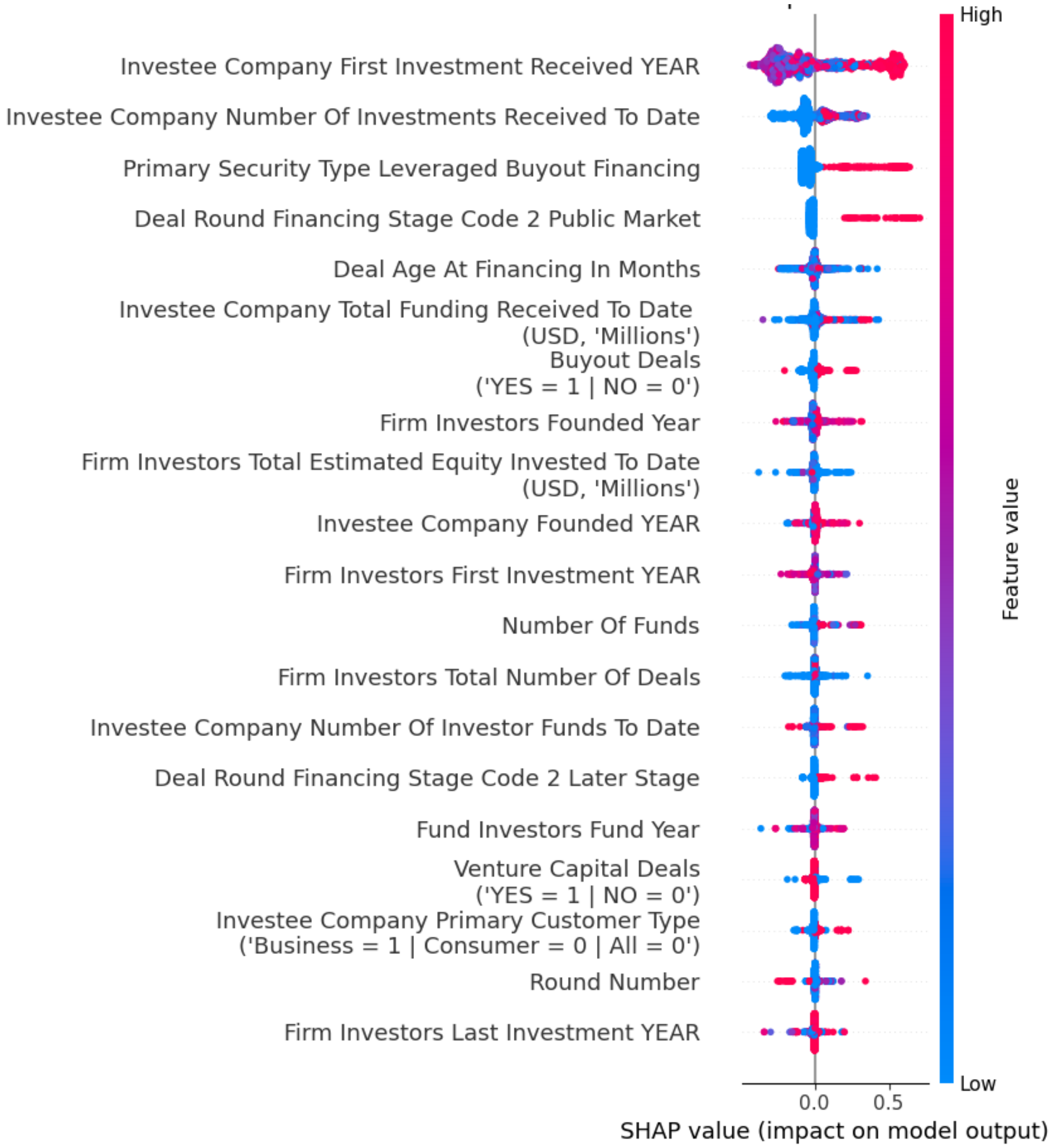


Exhibit 41: SHAP Summary Dot Plot – MLP (20 most relevant featu res, showing distribution, direction, and magnitude of SHAP values)

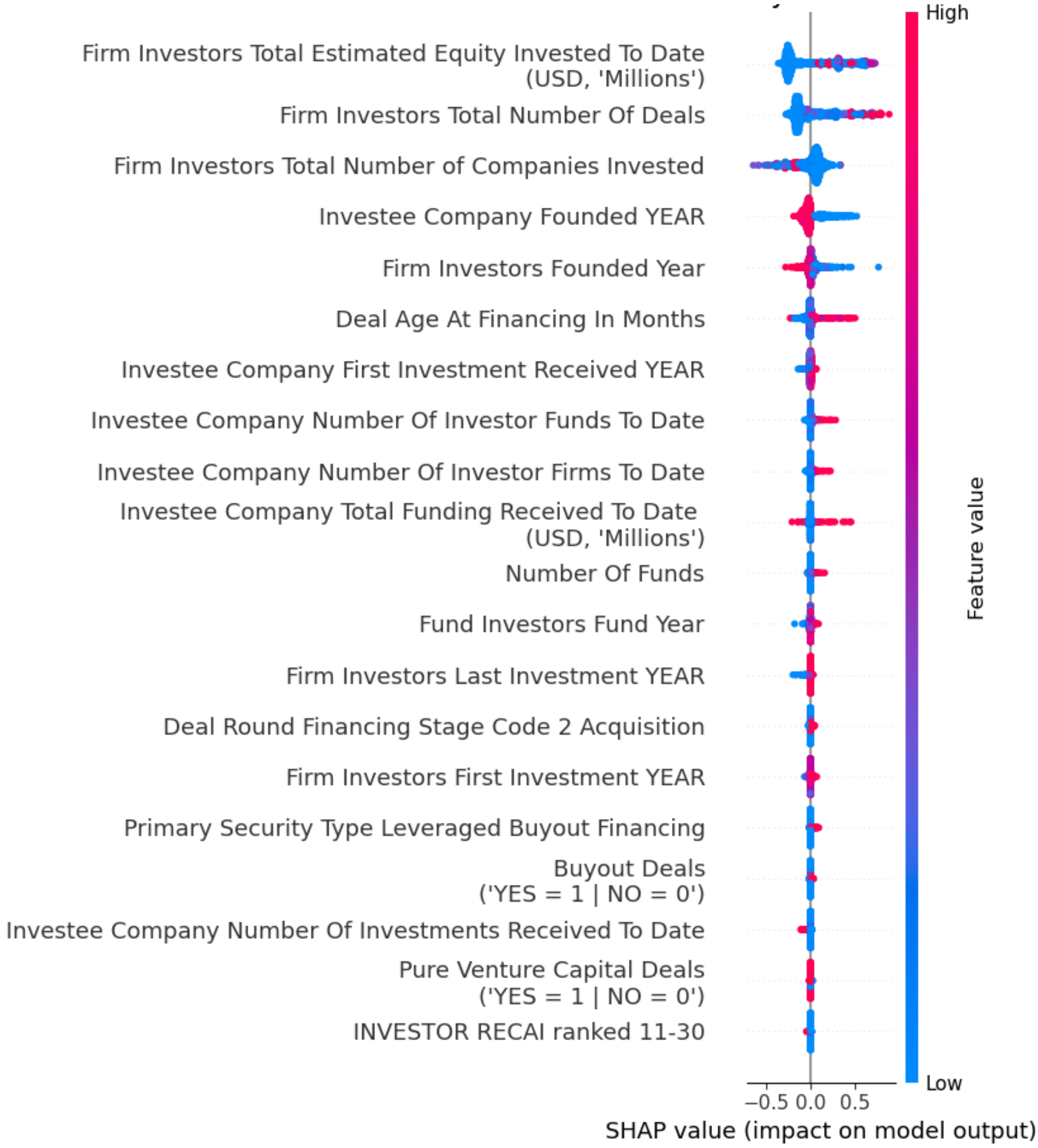


Exhibit 42: SHAP Summary Dot Plot – KNN (20 most relevant featu res, showing distribution, direction, and magnitude of SHAP values)

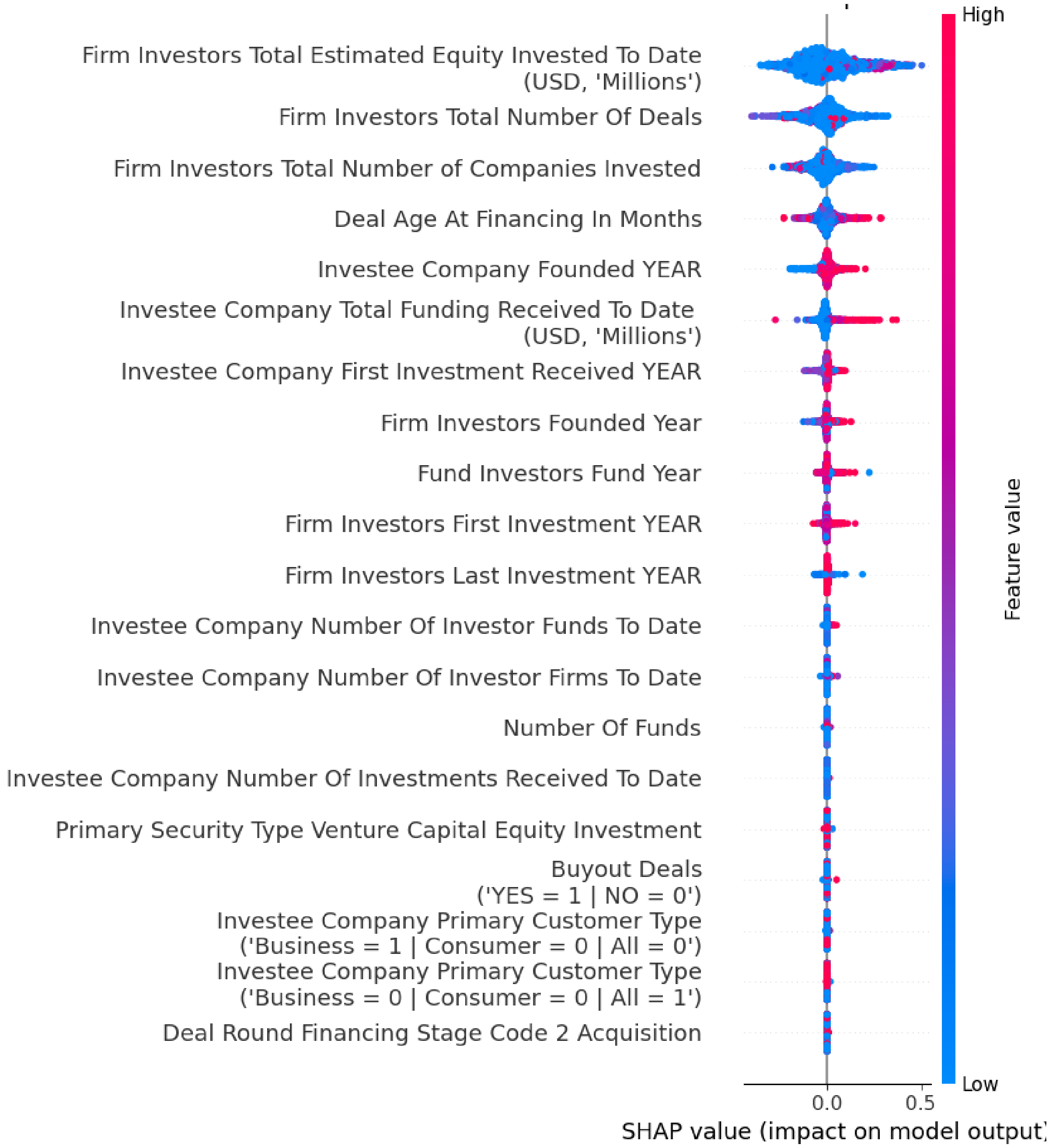
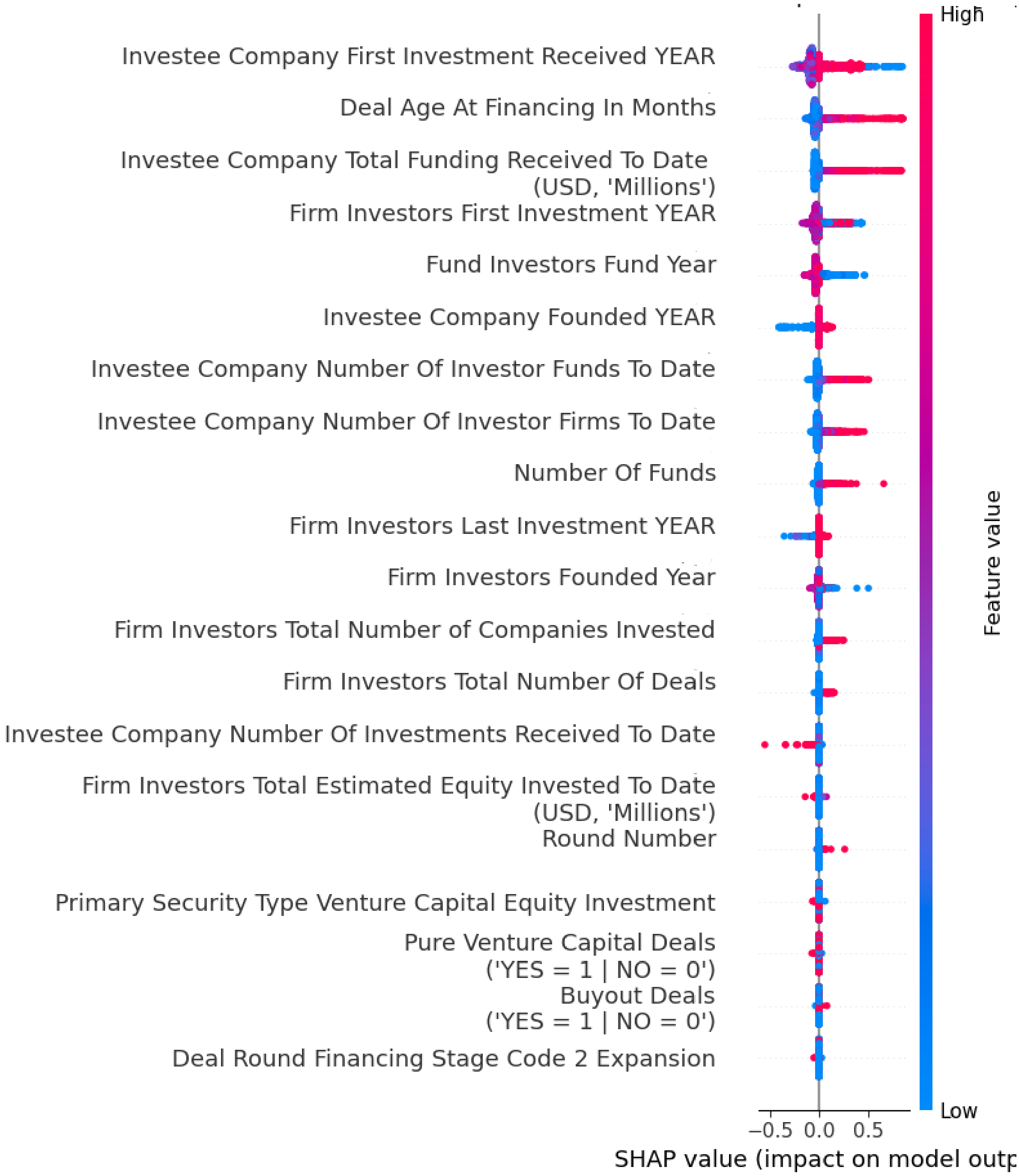


Exhibit 43: SHAP Summary Dot Plot – Naïve Bayes (20 most relevant featu res, showing distribution, direction, and magnitude of SHAP values)



10. References

Amemiya, T. (1981). Qualitative response models: A survey. *Journal of Economic Literature, 19*(4), 1483–1536.

Arroyo, J., Corea, F., Jimenez-Diaz, G., & Recio-Garcia, J. A. (2019). Assessment of machine learning performance for decision support in venture capital investments. *Technological Forecasting and Social Change, 146*, 725–739.

Bengtsson, O., & Hsu, D. H. (2015). Ethnic matching in the U.S. venture capital market. *Journal of Business Venturing, 30*(2), 338–354.

Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research, 13*(10), 281–305.

Bertoni, F., Colombo, M. G., & Quas, A. (2015). The patterns of venture capital investment in Europe. *Small Business Economics, 45*(3), 543–560.

Bertoni, F., Croce, A., & D'Adda, D. (2010). Venture capital investments and patenting activity of high-tech start-ups: A micro-econometric analysis. *Research Policy, 39*(7), 995–1006.

Beyer, K., Goldstein, J., Ramakrishnan, R., & Shaft, U. (1999). When is "nearest neighbor" meaningful? In *Database Theory – ICDT'99* (pp. 217–235). Springer.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.

Block, J. H., Colombo, M. G., Cumming, D. J., & Vismara, S. (2018). New players in entrepreneurial finance and why they are there. *Small Business Economics, 50*(2), 239–250.

Bocken, N. M. P. (2015). Sustainable venture capital – Catalyst for sustainable start-up success? *Journal of Cleaner Production, 108*(Part A), 647–658.

Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition, 30*(7), 1145–1159.

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32.

Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). *Classification and regression trees*. Wadsworth.

Brynjolfsson, E., & McElheran, K. (2016). The rapid adoption of data-driven decision-making. *American Economic Review, 106*(5), 133–139.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research, 16*, 321–357.

Chemmanur, T. J., Krishnan, K., & Nandy, D. K. (2011). How does venture capital financing improve efficiency in private firms? A look beneath the surface. *Review of Financial Studies, 24*(12), 4037–4090.

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). ACM.

Colombo, M. G., & Grilli, L. (2010). On growth drivers of high-tech start-ups: Exploring the role of founders' human capital and venture capital. *Journal of Business Venturing, 25*(6), 610–626.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273–297.

Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. IEEE Transactions on Information Theory, 13(1), 21–27.

Crespo, C., Gutiérrez, C., & Fernández, J. (2023). Predicting startup success with machine learning: Evidence from Crunchbase. *Technological Forecasting and Social Change, 189*, 122353.

Crespo, N. F., Simões, N., & Simões, A. M. (2023). Predicting the success of startups using a machine learning approach. *Journal of Business Research, 157*, 113546.

Cumming, D., & Johan, S. (2017). The problems with and promise of entrepreneurial finance. Strategic Entrepreneurship Journal, 11(3), 357–370.

Cumming, D., & Johan, S. (2009). *Venture capital and private equity contracting: An international perspective*. Elsevier.

Cumming, D., Fleming, G., & Suchard, J.-A. (2005). Venture capitalist value-added activities, fundraising and fund performance. *Journal of Banking & Finance, 29*(2), 495–528.

Da Rin, M., Hellmann, T., & Puri, M. (2013). A survey of venture capital research. In G. M. Constantinides, M. Harris, & R. M. Stulz (Eds.), *Handbook of the Economics of Finance* (Vol. 2, pp. 573–648). Elsevier.

Dalle, J. M., den Besten, M., & Menon, C. (2017). Using Crunchbase for economic and managerial research. *OECD Science, Technology and Industry Working Papers, 2017*(08).

Davis, J., & Goadrich, M. (2006). The relationship between Precision-Recall and ROC curves. In *Proceedings of the 23rd International Conference on Machine Learning* (pp. 233–240). ACM.

Dietterich, T. G. (2000). Ensemble methods in machine learning. In *Multiple Classifier Systems* (pp. 1–15). Springer.

Domingos, P., & Pazzani, M. (1997). On the optimality of the simple Bayesian classifier under zero-one loss. *Machine Learning, 29*(2), 103–130.

Ewens, M., & Rhodes-Kropf, M. (2015). Is a VC partnership greater than the sum of its partners? *Journal of Finance, 70*(3), 1081–1113.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters, 27*(8), 861–874.

Finney, D. J. (1947). *Probit analysis*. Cambridge University Press.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics, 7*(2), 179–188.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics, 29*(5), 1189–1232.

Fuerst, F., & McAllister, P. (2011). Green noise or green value? Measuring the effects of environmental certification on office values. *Real Estate Economics, 39*(1), 45–69.

Fuster, A., Plosser, M. C., Schnabl, P., & Vickery, J. I. (2018). The role of technology in mortgage lending. *FRB of New York Staff Report No. 836*.

Garreau, D., & von Luxburg, U. (2020). Explaining the explainer: A first theoretical analysis of LIME. *arXiv preprint arXiv:2001.03447*.

Ghosh, S., & Nanda, R. (2010). Venture capital investment in the clean energy sector. *Harvard Business School Working Paper, No. 11-020*.

Gompers, P., & Lerner, J. (2001). *The money of invention: How venture capital creates new wealth*. Harvard Business School Press.

Gompers, P., Kovner, J., Lerner, J., & Scharfstein, D. (2008). Venture capital investment cycles: The impact of public markets. *Journal of Financial Economics, 87*(1), 1–23.

Grilli, L., & Murtinu, S. (2014). Government, venture capital and the growth of European high-tech entrepreneurial firms. *Research Policy, 43*(9), 1523–1543.

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *Review of Financial Studies, 33*(5), 2223–2273.

Gupta, A., Lohani, M. C., & Manchanda, M. (2021). Financial fraud detection using naive Bayes algorithm in highly imbalance data set. *Journal of Discrete Mathematical Sciences and Cryptography, 24*(5), 1559–1572.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.

He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering, 21*(9), 1263–1284.

Hellmann, T., & Puri, M. (2002). Venture capital and the professionalization of start-up firms: Empirical evidence. *Journal of Finance, 57*(1), 169–197.

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd ed.). Wiley.

Huang, P., Meoli, M., & Vismara, S. (2020). The geography of venture capital and entrepreneurial finance. *Journal of Economic Geography, 20*(1), 233–265.

International Energy Agency (IEA). (2023). *World energy investment 2023*. IEA.

Intergovernmental Panel on Climate Change (IPCC). (2022). *Climate change 2022: Mitigation of climate change*. Cambridge University Press.

International Renewable Energy Agency (IRENA). (2022). *World energy transitions outlook 2022*. IRENA.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning: With applications in R*. Springer.

Jiang, C., & Liu, D. (2022). Effects of venture capital on green technology innovation in new energy vehicle industry in China. *Energy & Environment, 35*(1), 418–437.

Kaplan, S. N., & Strömberg, P. (2001). Venture capitalists as principals: Contracting, screening, and monitoring. *American Economic Review, 91*(2), 426–430.

Kaplan, S. N., & Strömberg, P. (2004). Characteristics, contracts, and actions: Evidence from venture capitalist analyses. *Journal of Finance, 59*(5), 2177–2210.

Kaplan, S. N., & Strömberg, P. (2009). Leveraged buyouts and private equity. *Journal of Economic Perspectives, 23*(1), 121–146.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., … Liu, T. Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems (NeurIPS 30)*.

Kuhn, M., & Johnson, K. (2013). Applied predictive modeling. Springer.

Kleinert, S., Volkmann, C., & Grünhagen, M. (2020). Third-party signals in equity crowdfunding: The role of prior financing. *Small Business Economics, 54*(1), 277–292.

Kuncheva, L. I. (2004). *Combining pattern classifiers: Methods and algorithms*. Wiley.

Lerner, J. (1994). The syndication of venture capital investments. *Financial Management, 23*(3), 16–27.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems (NeurIPS 30)*.

Lundberg, S. M., Erion, G. G., & Lee, S.-I. (2020). Consistent individualized feature attribution for tree ensembles. *Nature Machine Intelligence, 2*(1), 56–67.

Menard, S. (2002). *Applied logistic regression analysis* (2nd ed.). Sage.

Mienye, I. D., & Sun, Y. (2022). A survey of ensemble learning: Concepts, algorithms, applications, and prospects. *IEEE Access, 10*, 99129–99149.

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence, 267*, 1–38.

Mitchell, T. M. (1997). *Machine learning*. McGraw-Hill.

Molnar, C. (2019). *Interpretable machine learning*. Independently published.

Molnar, C. (2022). *Interpretable machine learning* (2nd ed.). Independently published.

Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives, 31*(2), 87–106.

Naidu, G., Zuva, T., & Sibanda, E. (2023). A review of evaluation metrics in machine learning algorithms. In *Artificial Intelligence Application in Networks and Systems* (pp. 15–25). Springer.

Niculescu-Mizil, A., & Caruana, R. (2005). Predicting good probabilities with supervised learning. In *Proceedings of the 22nd International Conference on Machine Learning* (pp. 625–632). ACM.

Pestov, V. (2013). Is the k-NN classifier in high dimensions affected by the curse of dimensionality? *Computers & Mathematics with Applications, 65*(10), 1427–1437.

Polikar, R. (2012). Ensemble learning. In C. Zhang & Y. Ma (Eds.), *Ensemble machine learning*. Springer.

Polzin, F., Sanders, M., & Täube, F. (2017). A diverse and resilient financial system for investments in the energy transition. Current Opinion in Environmental Sustainability.

Polzin, F. (2017). Mobilizing private finance for low-carbon innovation – A systematic review of barriers and solutions. *Renewable and Sustainable Energy Reviews, 77*, 525–535.

Powers, D. M. W. (2011). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. *Journal of Machine Learning Technologies, 2*(1), 37–63.

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: Unbiased boosting with categorical features. In *Advances in Neural Information Processing Systems (NeurIPS 31)*.

Puri, M., & Zarutskie, R. (2012). On the life cycle dynamics of venture-capital- and non-venture-capital-financed firms. *Journal of Finance, 67*(6), 2247–2293.

Quinlan, J. R. (1993). *C4.5: Programs for machine learning*. Morgan Kaufmann.

Radovanović, M., Nanopoulos, A., & Ivanović, M. (2009). Nearest neighbors in high dimensional data: The emergence and influence of hubs. In *Proceedings of the 26th International Conference on Machine Learning* (pp. 865–872).

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why Should I Trust You?": Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1135–1144). ACM.

Ross, G., Das, S., Sciro, D., & Raza, H. (2021). CapitalVX: A machine learning model for startup selection and exit prediction. *The Journal of Finance and Data Science, 7*(2), 94–114.

Rosenbusch, N., Brinckmann, J., & Müller, V. (2013). Does acquiring venture capital pay off for the funded firms? A meta-analysis on the relationship between venture capital investment and funded firm financial performance. *Journal of Business Venturing, 28*(3), 335–353.

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence, 1*(5), 206–215.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. Nature, 323(6088), 533–536.

Sagi, O., & Rokach, L. (2018). Ensemble learning: A survey. *WIREs Data Mining and Knowledge Discovery, 8*(4), e1249.

Schabek, T. (2020). The financial performance of sustainable power producers in emerging markets. *Renewable Energy, 160*, 1408–1419.

Shapley, L. S. (1953). A value for n-person games. In H. W. Kuhn & A. W. Tucker (Eds.), *Contributions to the theory of games* (Vol. 2, pp. 307–317). Princeton University Press.

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management, 45*(4), 427–437.

Sollich, P., & Krogh, A. (1995). Learning with ensembles: How overfitting can be useful. In *Proceedings of NIPS* (pp. 190–196).

Sørensen, M. (2007). How smart is smart money? A two-sided matching model of venture capital. *Journal of Finance, 62*(6), 2725–2762.

Tykvová, T. (2018). Venture capital and private equity financing: An overview of recent literature and an agenda for future research. *Journal of Business Economics, 88*(3), 325–362.

Wolpert, D. H. (1992). Stacked generalization. Neural Networks, 5(2), 241–259.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data* (2nd ed.). MIT Press.

Yeh, J. Y., & Chen, C. H. (2020). A machine learning approach to predict the success of crowdfunding fintech projects. *Electronic Commerce Research and Applications, 40*, 100935.

Zacharakis, A. L., & Meyer, G. D. (1998). The potential of actuarial decision models: Can they improve the venture capital investment decision? *Journal of Business Venturing, 13*(1), 57–76.

Zenobi, G., & Cunningham, P. (2001). Using diversity in preparing ensembles of classifiers based on different feature subsets to minimize generalization error. In L. De Raedt & P. Flach (Eds.), *Machine Learning: ECML 2001*. Lecture Notes in Computer Science (Vol. 2167). Springer.

Żbikowski, K., & Antosiuk, P. (2021). A machine learning, bias-free approach for predicting business success using Crunchbase data. *Information Processing & Management, 58*(4), 102555.