Modern Finance

Modern Finance Institute

Article

Machine learning stock selection: Evidence from the South African factor zoo

Daniel Page [1], **Yudhvir Seetharam** [2,*] **and Christo Auret** [3]

[1] School of Economics and Finance, University of the Witwatersrand; daniel.page@wits.ac.za
[2,*] Correspondence: School of Economics and Finance, University of the Witwatersrand; Yudhvir.seetharam@wits.ac.za.
[3] School of Economics and Finance, University of the Witwatersrand; christo.auret@wits.ac.za

Abstract: This study examines whether machine learning can predict Johannesburg Stock Exchange stock returns using South African factor zoo features. Six models are tested in an expanding-window, walk-forward design, with portfolio performance evaluated across alternative weighting schemes and factor-spanning regressions. Ensemble tree-based models, particularly Random Forest, XGBoost and LightGBM, deliver the strongest out-of-sample performance, especially under more intensive training and linear rank weighting. Market-capitalisation weighting weakens alpha. Although machine-learning portfolios generate meaningful alphas under Fama–French models, the inclusion of momentum materially reduces alpha, highlighting momentum's dominance in South African equity returns.

Keywords: Machine learning asset pricing; Stock return prediction; Factor zoo; Johannesburg Stock Exchange; Emerging equity markets; Portfolio weighting; Momentum factor

JEL Codes: C45, G11, G12, G15

1. Introduction

The inevitable amalgamation of machine learning ("ML") with asset pricing and the 'factor zoo' per Cochrane (2011) has yielded a recent plethora of studies that focus on leveraging the power of ML to improve cross-sectional performance prediction. The literature typically assesses predictability across several ML models in two forms. 'Explanatory' studies focus on out-of-sample R^2 while 'trading strategy' studies concentrate on out-of-sample alpha. Both streams assess predictive accuracy yet focus on different outcomes. The consensus across both is that, in general, linear predictive models are inferior to ML models that can exploit non-linear relationships present in both time-series and cross-sectional financial data. This study focuses on out-of-sample alpha, where we apply a large sub-set of South African 'factor zoo' proxies as features to a liquid universe of single share excess performance measured over the following quarter. We opt for the trading strategy approach for two reasons. First, increased out-of-sample R^2 does not necessarily translate to superior out-of-sample performance. Second, and more importantly, in the context of stock selection, we are less concerned with the accuracy of 'actual' return predictions but rather precision of the prediction signal. Given the inherent noise of time-series and cross-sectional financial return data, whether a ML model is better at predicting the returns determined via R^2 is moot when compared to the informational content of the signal i.e. does the model prediction result in consistent and superior alpha?

Using several ML models, we find evidence consistent with the international literature as out-of-sample performance (and therefore out-of-sample accuracy) improves specifically with ensemble decision tree models that are adept to exploiting non-linear relationships between the feature set and forward performance, even when applied to a

Citation: Page, D., Seetharam, Y., & Auret, C. (2026). Machine learning stock selection: Evidence from the South African factor zoo. *Modern Finance*, *4*(3), 1-25.

Accepting Editor: Adam Zaremba

Received: 11 April 2026
Accepted: 1 July 2026
Published: 10 July 2026

relatively less liquid, smaller emerging market bourse such as the Johannesburg Stock Exchange ('JSE'). Importantly, our analysis covers several less considered empirical design implications, namely the impact of more versus less training, weighting method applied in back-tests as well as assessing model performance using factor spanning regressions. Our results indicate that first, notwithstanding the computational expense, more training results in better out-of-sample generalization through improved out-of-sample performance. Second, linear weighting followed by equal weighting produces the highest out-of-sample long-short and long-only portfolio performance across ML models, while market capitalization weighting generally negates any alpha potential. Last, we find that in factor spanning tests, the inclusion of the momentum premium (per Jegadeesh and Titman (1993,2001)) dramatically suppresses alpha, irrespective of the weighting method applied. The latter result reinforces evidence presented by Page and Auret (2019) and Muller and Ward (2013) regarding the presence and dominance of price momentum premium on the JSE.

The remainder of the study is structured as follows; section 2 provides a brief synopsis of the recent ML based asset pricing literature; section 3 describes the research methodology; section 4 details the results of the empirical analysis and section 5 concludes.

2. Literature Review

The section that follows considers the most recent pertinent literature that explores the intersection of asset pricing, factor premiums and forecasting via ML. Gu, Kelly and Xiu (2020) conducted a comparative analysis of several linear and ML based prediction models to measure asset risk premiums of both the aggregate market and individual shares. The study considered a large cross-section of individual shares sourced from the CRSP database over March 1957 - December 2016, using a total of 900 raw signals and further applied feature engineering in the form interaction and non-linear transformations. The authors found that ML predictive models performed best when estimating out-of-sample R^2 over the following month, improving from 0.16% when applying cross-sectional linear regression to 0.33% and 0.4% when applying tree-based ML models and shallow-learning neural nets. Importantly, the study concludes that the benefits of ML based predictions centers around the ability to exploit non-linear interrelationships between features and future returns, which is best displayed when applied to large liquid shares and portfolios.

Leung, Lohre, Mishlich, Shea and Stroh (2021) compare the predictive ability of gradient booster (GB) ML algorithms to linear models on a large cross-section of global large and mid-cap shares associated with the MSCI, FTSE, S&P and STOXX indices over the period January 1991 - December 2018. The study applied a 'relatively' small set of 20 firm level characteristics that form part of four general factor pricing proxies, specifically momentum, quality, size and value to predict individual 1 and 6-month future outperformance based on industry or region. Using both linear and GB models, predictions were converted to relative ranks and equally weighted decile portfolios were sorted monthly. The study found that ML techniques proved superior when compared to linear prediction models, with the 1-month GB long-short portfolio producing an average monthly return of 1.34% and an annualized information ratio of 2.5, compared to the equally weighted signal (naïve aggregation) portfolio which achieved an information ratio of 1.99.

Leippold, Wang and Zhou (2022) conduct an equivalent assessment to Gu et al. (2020) on a cross-section of Chinese listed shares over the sample period January 2000 - June 2020. The authors apply a total of 1160 features (using interaction and non-linear transformations) encompassing stock level, macro-economic and industry dummy variables to predict out-of-sample individual share returns over the following month. The authors employ 11 ML models (linear and non-linear) where linear models are of the least squares variety as well as popular non-linear models such as decision trees and neural

networks. Like, Gu et al. (2020), the study found that ML models more adept to identifying non-linear relationships between features and labels delivered superior out-of-sample R^2, while portfolio level back-tests confirm the findings with neural networks and variable subsample aggregation delivering superior out-of-sample nominal returns and Sharpe ratios.

Tsai, Gao and Yuan (2023) apply four ML models to forecast share performance on the cross-section of Taiwanese listed shares over the period March 2013 - June 2020. Using Random Forests (RF), Feed-Forward Neural Nets (FNN), Gated Recurrent Unit (GRU) and Financial Graph Attention Networks (FinGAT), the authors applied 18 features, encompassing liquidity, leverage, asset efficiency, relative valuation and profitability ratios to predict next quarter performance in excess of the benchmark. The authors found that when applying ML predictions to define the top 10 and 20 best shares in each sorting period, the RF and FinGAT models produced the highest out-of-sample excess returns above the benchmark, returning average excess returns of 9.8% and 9.7% per annum respectively.

Wolff and Echterling (2024) applied ML to select shares from the universe of S&P500 constituents over the period January 1999 - March 2021. In contrast to several similar studies, the study focused on weekly data for the estimation of features and applied weekly forward returns as labels. Further, the empirical design of the study applied a classification framework, where share performance was converted to binary variable based on under or outperformance relative to the benchmark. Lastly, the study considered a fixed historical window of 4 years, where 3 years (156 weeks) were applied for training and 1 year for testing (52 weeks). In contrast to Gu et al (2020), the feature set considered 47 firm level characteristics that covered well-documented style based proxies, sector dummies as well as technical indicators. The authors found that over the sample period, ML-based strategies achieved significant regression alphas of between 8% to 12% per annum when applying the Fama-French 6 factor model, implying that common risk-factors fail to explain ML driven investment strategies. Notably, the authors also found that all ML models displayed very high levels of portfolio turnover, ranging from 9.9 – 31.1 times per year on average.

Notwithstanding this increasingly popular branch of literature, the proliferation of ML based stock selection studies has been significantly less explored in emerging markets. From a South African perspective, virtually no studies have been conducted in this regard, barring the recent study by Page, McClelland and Auret (2024) where the authors apply machine learning to style rotation. The study specifically compared naïve style rotation based on style momentum to ML based approaches and found that tree based methods (specifically gradient boosters) achieved highest out-of-sample performance. Importantly, Page et al. (2024) was limited to style-level portfolio analysis as opposed to individual share prediction applying style proxies as features. In comparison, this study differs by applying a methodological framework consistent with several global studies that utilize style proxies as features in predicting individual share performance on the JSE. In doing so, this study adds to the body of SA specific and emerging market corpus of literature by being the first to consider the JSE factor zoo within a machine learning framework for the purpose of stock selection.

3. Data and Methods

3.1. Data

The data used in this study covers the cross-section of JSE listed shares over the period January 2002 - August 2025 and is sourced primarily from Bloomberg. The data is limited to listed operating companies and therefore excludes ETF's, ETN's, cash shells and special purpose investment trusts. Price data (and therefore return data) are adjusted for corporate actions including share splits, unbundlings, consolidations, normal and special dividends as well as name changes to mitigate the impact of structural breaks. Daily data

is sourced for price, volume, market capitalization as well as trailing dividend yield, earnings yield, price-to-book and price-to-cash flow ratios. Accounting data is typically reported semi-annually (as per the JSE listing requirements) and is captured for each firm based on their respective mid-year and financial year end and is lagged 3 months to account for look-ahead bias.

To mitigate the impact of survivorship bias, the database covers a cross-section of 426 shares, which includes both new listings and delistings. In the event of delisting, shares are assigned a 'delisted' code, rendering them no longer eligible for portfolio inclusion. Similarly, shares that list over the sample period are included from the point of listing but only become eligible for portfolio inclusion after being listed for a full trading year (252 days). Lastly, if a share delists due to a scheme of arrangements, the last share price is maintained in the data set and are no longer eligible for inclusion in portfolio sorts, while delisting due to corporate failure results in shares being assigned a -100% return[1]. Over and above share price data, the market benchmark (FTSE JSE All-Share Total Return Index), broad sector indices and the USDZAR exchange rate are also included in the data over the sample period and the 91-day SA Government T-bill is applied as the risk-free proxy.

3.2. Methodology

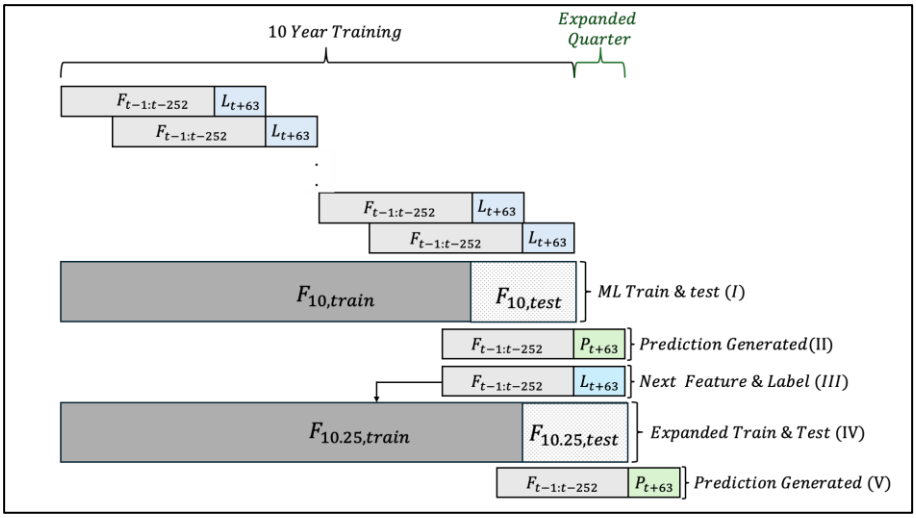
The methodology applied broadly conforms to the current body of international studies, however, there are several notable differences. First, daily data is applied to estimate the historical feature set over rolling 252-day (trading year) windows and is mapped to forward 63-day (trading quarter) individual share returns in excess of the benchmark. We opt for quarterly predictions for several reasons. First, the FTSE-JSE applies quarterly rebalancing for all major indices, resulting in back-tested portfolio returns being consistent with the market benchmark rebalance framework and reporting frequency. Second, in comparison to developed market bourses, emerging markets typically display lower volume and in-turn, greater bid-ask spreads. Limiting rebalancing frequency to quarterly therefore provides a more consistent depiction of out-of-sample performance, less impacted by round-trip transaction costs (which are not explicitly considered however portfolio turnover analysis is presented in Appendix C).

In terms of features applied, we consider 107 features which are detailed in Appendix A that cover broad set of stylised proxies that span technical indicators (momentum, moving averages), historical risk (volatility, value-at-risk, market beta, currency beta, sector betas, VIX betas), liquidity, market capitalization as well as 19 fundamental style proxies based on accounting and financial ratio data. As described in the literature review, a consistent conclusion across studies with respect to superior predictive accuracy of ML over linear models lies in the latter's ability to capture non-linear relationships between historical features and future performance. To this end, we apply simple feature engineering where each of the 107 features is squared, resulting in a total of 214 features. In contrast to several studies covered, the application of 214 features is considered relatively parsimonious, however, given that there has not been a comprehensive study of the factor zoo in a South African context, we have attempted to apply a feature set which is well covered in the local literature (see van Rensburg (2001); van Rensburg and Robertson (2003); Basiewicz and Auret (2009); Strugnell, Gilbert and Kruger (2011); Hodnett, Hsieh and van Rensburg (2012); Muller and Ward (2013); Page, Britten and Auret (2015); Page and Auret (2019); Cox and Britten (2019); van Rensburg and van Vuuren (2020); Page, McClelland and Auret (2022)).

1. Corporate failure is broadly used to encompass share suspension due to violations of JSE/FTSE listing rules. Generally, prior to being removed from the JSE trading boards, such shares retain a residual value (i.e. the last traded price). Since such shares are suspended and can therefore not be traded, we apply the punitive penalty and exclude the share from future portfolio sorts.

Consistent with global literature, such as Gu et al. (2020) and Leippold et al. (2022), we apply an expanding feature window that is described in Figure 1 that follows:

Figure 1. Expanding feature window that incorporates prior training feature set and labels after each prediction period



As mentioned, the time-period considered runs from January 2002 until August 2025 and we reserve the first 10 year period for initial training. Using the expanding feature set window per figure 1; (I) the initial 10 years of features and labels (F10) are used for model training applying a 70:30 train-test-split after which fitted parameters are applied to the most recent historical feature window (II) to predict out-of-sample share performance. After each portfolio formation, the most recent set of features and actual label returns are added to the training data (III and IV) and the process is repeated.

Second, we consider six ML models, namely K Nearest Neighbors (KNN), Ridge Regression (RDG), Multi-layer Perceptron neural network (NN), Random Forest (RF), XGBoost (XGB) and LightGBM (LGB). Importantly, the purpose of the study is not to test an exhaustive set of ML models but rather includes models that are well-established in the literature and include those that have been found able to exploit the associated benefit of identifying non-linear relationships between factor related features and future share returns. Consistent with Gu et al. (2020), the objective of each ML model is to minimize out-of-sample prediction error, where predictions follow the general form:

$$P\left(e[r_{i,t+63}]\right) = g\left(z_{i,t}\right) \tag{1}$$

And

$$e[r_{i,t+63}] = P\left(e[r_{i,t+63}]\right) + \varepsilon_{i,t+63} \tag{2}$$

Equation 1 describes the generalised prediction of share i's excess return over the benchmark associated with each ML model g based on the vector of features z. Equation 2 depicts the generalized form of the objective function where the actual excess return of share i over the subsequent quarter is assumed to be a combination of the predicted excess return plus prediction error. To minimize data-leakage, we apply a walk-forward back-test method where at each portfolio formation date, models are trained on historical data assuming a 70:30 train-test-split. Using the training segment of the data, a randomized search cross-validation of model specific hyperparameters are applied with the objective of minimizing the forecast root-mean-squared error (RMSE). Once optimal hyperparameters are defined, tuned models are assessed on unseen test data to evaluate out-of-sample generalization and overfitting. Models are then deployed on held-out historical validation data (not included in the train-test split) and quarterly forward predictions are estimated (model specific parameter grids are detailed in Appendix B).

In training the respective models via randomized parameter search, we apply two broad sets of assumptions related to the maximum iterations of hyperparameter combinations and cross-validation k-folds that consider 'more' and 'less' training. For the former, we set the maximum iterations to 50 and cross-validate applying 5 k-folds, where predictions and subsequent portfolio returns are suffixed with MT ('more training'). For the latter, we set the number of maximum iterations to 10 and cross-validate using 3 k-folds and suffix subsequent predictions and portfolio returns LT ('less training'). The purpose of the methodological nuance is to determine whether computational cost has a material relationship with out-of-sample portfolio performance.

A final difference in methodology relates to the main objective of the study which differs from the broader literature as our test of predictive accuracy does not consider the explanatory ability of ML in the form of out-of-sample R^2. As mentioned, this study focuses on quarterly predictions which conforms to the rebalance frequency of the underlying market benchmark and minimizes the potential impact of trading costs that would naturally increase if sorting was conducting monthly. Therefore, we direct the focus of this study to determining the profitability potential (and feasibility) of applying ML to a cross-section of developing market counters that suffer from the expected differences in depth and breadth when compared to developed market economies.

Lastly, to reduce the impact of illiquid penny-stocks skewing out-of-sample returns, we limit the eligible universe to the top 100 shares at each quarterly portfolio sort date, using the most recent data, based on an average rank of latest trading volume and market capitalization[2]. Shares are then ranked on predicted performance and two portfolios per ML model are constructed, where the 'top' portfolio represents the best 30 shares based on predicted excess return while the 'bottom' portfolio, the worst 30. This implies that only historical data is applied for training, testing, signal prediction and sorting. Out-of-sample daily portfolio returns are measured on a buy-and-hold basis over the following 63-days assuming equal, market capitalization and rank-based linear weighting. We apply differing weighting methodologies to determine the impact of portfolio construction on out-of-sample performance as several individual and multi-style based studies have found that weighting method can significantly impact on performance.

To evaluate the profitability of the respective ML model strategies, long-only and long-short performance of the ML test portfolios is assessed using t-tests, Sharpe ratios and Probabilistic Sharpe ratios ("PSR") per Bailey and Lopez de Prado (2012). In specific reference to the PSR tests conducted, each portfolios annualized Sharpe ratio is compared to the market benchmark Sharpe ratio over the assessment period and provides a probabilistic measure which accounts for portfolio specific distributional properties overlooked by the conventional Sharpe ratio. Mathematically,

$$PSR(SR^*) = P\left(\widehat{SR} > SR^*\right) = Z\left(\frac{(\widehat{SR}-SR^*)\sqrt{n-1}}{\sqrt{1-\hat{\gamma}_3\widehat{SR}+\frac{\hat{\gamma}_4-1}{4}\widehat{SR}^2}}\right) \tag{3}$$

Where $\widehat{SR}$ represents the Sharpe ratio of the portfolio, SR^* the benchmark Sharpe ratio, $\hat{\gamma}_3$ the observed test portfolio skewness, $\hat{\gamma}_4$ the observed test portfolio kurtosis and Z is the cumulative distribution function of the Standard Normal distribution.

Last, risk-adjusted performance is assessed using factor spanning regressions using four established asset pricing models, namely; the Fama-French 3 factor model (FF3) per Fama and French (1993), the Carhart (1997) 4 factor model (CH4), the Fama-French 5 factor model (FF5) per Fama and French (2015) as well as the Fama-French 6 factor model (FF6) per Fama and French (2018). To ensure methodologically consistency, the underlying style proxy and factor premium construction matches the respective test portfolio weighting

2. Limiting the investable universe to the top 100 counters at each portfolio sort date reveals the level of concentration of the JSE. Over the full sample period, the quarterly defined top 100 covers 97% of the full index market capitalization on average, varying between 94% in 2004 and increasing to 99% in 2025.

methodology. As an example, when testing alphas and factor loadings of the equally weighted ML test portfolios, the zero-cost factors applied in spanning tests are also estimated assuming equal weighting. The sections that follow describes nominal and risk-adjusted performance of the ML driven strategies.

4. Results and Discussion

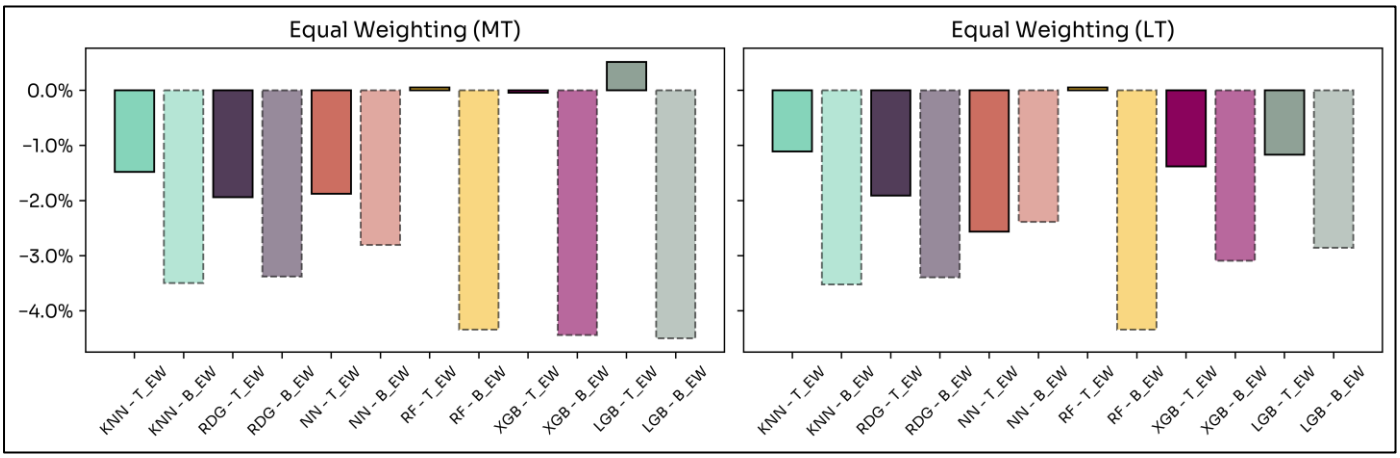
4.1. Nominal Performance

Table 1. Annualised performance of the equally weighted ML test portfolios

		More Training (50 Iterations, CV=5)			Less Training (10 Iterations, CV=3)		
Model (I)	Portfolio (II)	Ann. Return (III)	Sharpe (IV)	PSR (V)	Ann. Return (VI)	Sharpe (VII)	PSR (VIII)
KNN	Top	10.09%***	0.43	45.02%	10.54%***	0.46	49.59%
	Bottom	7.57%*	0.26	22.96%	7.48%*	0.25	22.20%
RDG	Top	9.52%**	0.39	39.44%	9.55%**	0.39	39.58%
	Bottom	7.70%*	0.27	23.67%	7.71%*	0.27	23.80%
NN	Top	9.40%**	0.37	36.26%	8.69%**	0.33	31.62%
	Bottom	8.56%**	0.33	30.69%	8.94%**	0.34	32.75%
RF	Top	11.55%***	0.49	53.67%	11.53%***	0.49	53.41%
	Bottom	6.88%*	0.23	19.45%	6.86%*	0.21	18.16%
XGB	Top	11.51%***	0.49	54.40%	9.92%**	0.40	40.37%
	Bottom	6.71%*	0.23	18.37%	8.27%**	0.31	28.96%
LGB	Top	12.07%***	0.53	57.62%	10.14%**	0.41	41.54%
	Bottom	6.66%*	0.23	18.19%	8.50%**	0.33	30.63%

Note. Table 1 presents the annualised performance of the equally weighted ML test portfolios over the period January 2012 – August 2025 across ML models where columns III to IV describe portfolio performance based on 'more training' and columns V to IX (less training). Returns are annualised and tested via two-tailed 1-sample t-tests, while Sharpe ratios and their respective PSR's are also presented.

Figure 2. Equally weighted ML test portfolio excess average performance over the period January 2012 to August 2025 applying more ('MT') and less ('LT') training.



Note. Returns are annualised and expressed in excess of the JSE ALSI where darker bars represent performance of the respective ML models top 30 portfolio and lighter shaded bars represent performance of the respective ML models bottom 30 portfolio.

Table 1 describes the annualized gross performance of the ML based portfolios under 'more training' and 'less training' assuming equal weighting. Focusing on 'more training' (i.e. columns III-V), 'top' portfolios outperform their respective 'bottom' portfolios, with

only RF, XGB and LGB producing long-only annualized returns north of 11% that are significant at the 1% level. Similarly, the same models produce Sharpe ratios above the (unreported) annualized benchmark Sharpe ratio of 0.46 with PSR's in excess of 50%. Focusing on the 'less training' portfolio results, the computationally cheaper hyper-parameter tuning negatively impacts NN and the gradient boosters, yet favors the KNN model, improving the 'top' portfolio annualized return to 10.5% (significant at the 1% level) and PSR to 50%. In terms of consistency, RF is superior, achieving virtually identical results when compared to 'more training' with an annualized return of 11.5% (significant at the 1% level) and a PSR of 53.4%.

The tabulated results are confirmed on examination of figure 2. Figure 2 (right hand sub-figure) describes the average annualized performance in excess of the market benchmark across the ML models assuming 'more' training and shows that each model's 'top' portfolio outperforms their respective 'bottom' but only RF, XGB and LGB produce long-only returns that match and outperform the benchmark. Similarly, the RF, XGB and LGB discrepancy between 'top' and 'bottom' is starker when compared to the KNN, RDG and NN. The results are largely consistent with the body of literature, indicating that ML models that are more adept to identifying non-linearities achieve relative outperformance, notwithstanding the relatively poor results of NN. Conversely, the left-hand sub-figure describes the cumulative performance results when applying 'less training'. 'Less training' seems to favor the less sophisticated ML models with improvements in the long-only 'top' KNN portfolio, while the RF 'top' portfolio performs in line with the benchmark. Importantly, the impact of 'less training' negatively impacts the performance of the NN and gradient booster models through lower 'top' portfolio performance and improved 'bottom' portfolio performance.

Table 2. Hypothetical zero-cost long-short annualised excess returns under 'more' and 'less' training assuming equal weighting

Equal Weighting		
Model	Excess Return - More Training	Excess Return - Less Training
KNN (Top – Bottom)	1.51%	1.87%
RDG (Top – Bottom)	0.82%	0.86%
NN (Top – Bottom)	0.28%	-0.84%
RF (Top – Bottom)	3.97%	3.91%
XGB (Top – Bottom)	4.04%*	1.18%
LGB (Top – Bottom)	4.74%**	1.23%

Table 2 describes the equal-weighted "top" minus "bottom" long-short annualised excess returns across the respective ML models assuming "more" and "less" training. Consistent with table 1 and figure 2, "more training" benefits RF, LGB and XGB, all producing annualized excess returns at and above 4%. Notably, the gradient boosters achieve the highest excess returns, with XGB and LGB producing annualized alphas of 4.04% and 4.74% that are significant at the 10% and 5% level respectively. "Less training" leads to economic improvements in excess returns for the less sophisticated ML models (KNN and RDG), however, the results lack statistical significance. More importantly, the best performing models under "more training" (gradient boosters) produce the 3rd and 4th worst excess returns when applying less training, dropping by 71% and 74% for XGB and LGB respectively.

Table 3 and Figure 3 present the annualized gross returns of the respective 'top' and 'bottom' portfolios across 'more' and 'less' training assuming market capitalization weighting. Under 'more training', the largest notable difference can be seen in five of the six 'top' portfolios underperforming their respective 'bottom' portfolios across the ML models employed. Moreover, the economic size and statistical significance of the respective model's gross performance is degraded, with only LGB producing a 'top' portfolio return of 12.5% that is significant at the 1% level. The results are confirmed by

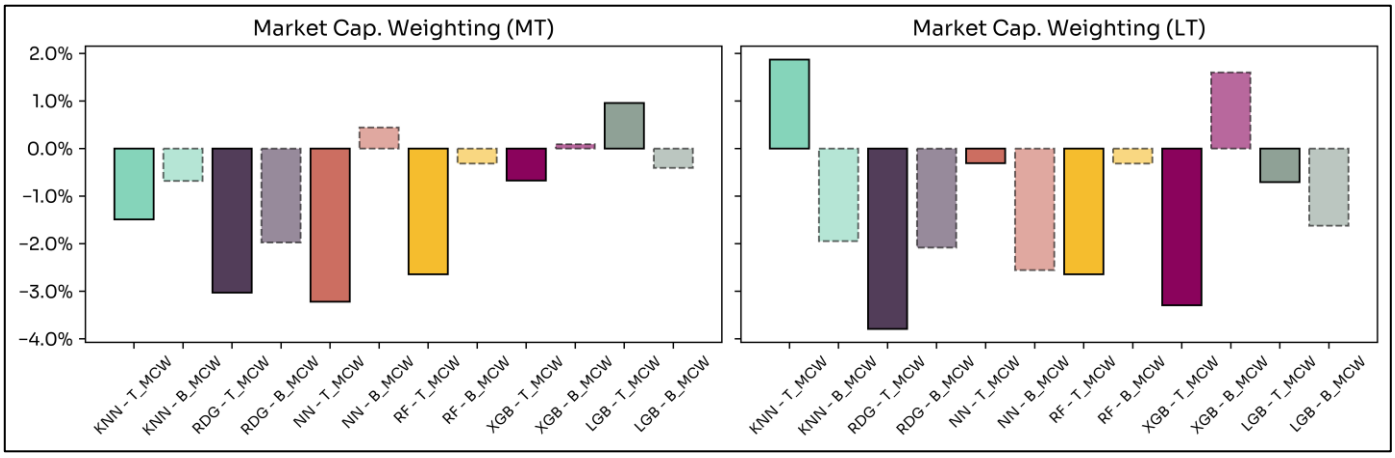
both portfolio Sharpe ratios and PSR, where only the LGB model produces an annualized Sharpe ratio above 0.5 and a PSR of more than 50%.

Table 3. Annualised performance of the market capitalization weighted ML test portfolios

		More Training (50 Iterations, CV=5)			Less Training (10 Iterations, CV=3)		
Model (I)	Portfolio (II)	Ann. Return (III)	Sharpe (IV)	PSR (V)	Ann. Return (VI)	Sharpe (VII)	PSR (VIII)
KNN	Top	10.11%**	0.40	41.28%	13.94%***	0.62	71.05%
	Bottom	10.41%**	0.38	38.58%	8.87%*	0.31	28.40%
RDG	Top	8.07%*	0.28	25.08%	7.20%*	0.24	20.24%
	Bottom	9.09%**	0.32	30.24%	9.05%**	0.32	30.30%
NN	Top	7.67%*	0.26	22.47%	11.09%**	0.43	45.72%
	Bottom	12.14%***	0.49	53.60%	8.35%*	0.29	25.86%
RF	Top	8.42%*	0.29	26.80%	8.42%*	0.29	26.80%
	Bottom	11.15%**	0.44	45.96%	11.15%**	0.44	45.96%
XGB	Top	10.71%**	0.41	42.55%	7.78%*	0.27	23.46%
	Bottom	11.43%**	0.44	46.80%	13.15%***	0.53	58.82%
LGB	Top	12.48%***	0.50	55.73%	10.63%**	0.41	42.49%
	Bottom	11.03%*	0.43	45.01%	9.53%**	0.35	33.73%

Note. Table 3 presents the annualised performance of the market capitalisation weighted ML test portfolios over the period January 2012 – August 2025 across ML models where columns III to IV describe portfolio performance based on 'more training' and columns V to IX (less training). Returns are annualised and tested via two-tailed 1-sample t-tests, while Sharpe ratios and their respective PSR's are also presented.

Figure 3. Market capitalisation weighted ML test portfolio annualised excess average performance over the period January 2012 to August 2025 applying more ('MT') and less ('LT') training.



Note. Returns are annualized and expressed in excess of the JSE ALSI where darker bars represent performance of the respective ML models top 30 portfolio and lighter shaded bars represent performance of the respective ML models bottom 30 portfolio.

Consistent with the equally weighted results, the application of 'less' training again favors KNN and is exacerbated by market capitalization weighting. KNN produces the best 'top' portfolio performance, achieving 13.9% and is significant at the 1% level. Moreover, KNN produces the highest Sharpe ratio (0.62) and a PSR of 71%. Conversely, tree based models, namely RF, XGB and LGB produce lower long-only 'top' returns, while XGB's 'bottom' portfolio achieved the 3rd highest performance, returning an annualized 13.15% that is significant at the 1% level as well as a Sharpe ratio in excess if 0.5 and a PSR of 59%. In contrast to the equally weighted results, NN actually shows an economic improvement under 'less training', with the top portfolios return increasing to 11.1% and

'bottom' portfolio reducing to 8.4%, with the former being statistically significant at the 5%.

Table 4. Hypothetical zero-cost long-short annualised excess returns under 'more' and 'less' training assuming market capitalisation weighting

Market Cap. Weight		
Model	Excess Return - More Training	Excess Return - Less Training
KNN (Top – Bottom)	-2.28%	2.29%*
RDG (Top – Bottom)	-2.88%	-3.52%
NN (Top – Bottom)	-5.38%	0.60%
RF (Top – Bottom)	-3.78%	-3.78%
XGB (Top – Bottom)	-2.24%	-6.12%
LGB (Top – Bottom)	-0.06%	-0.46%

Table 4 confirms the long-only results, showing that irrespective of the ML model applied, 'more' training results in statistically insignificant negative excess returns when applying market capitalization weighting. Of the more sophisticated models, LGB, which was the best performing model under equal weighting, performed best under market capitalization weighting, delivering a neutral -0.06%. Under 'less' training, KNN produces the largest annual excess return of 2.3% and is significant at the 10% level.

Table 5. Annualised performance of the linear weighted ML test portfolios over the period January 2012 to August 2025

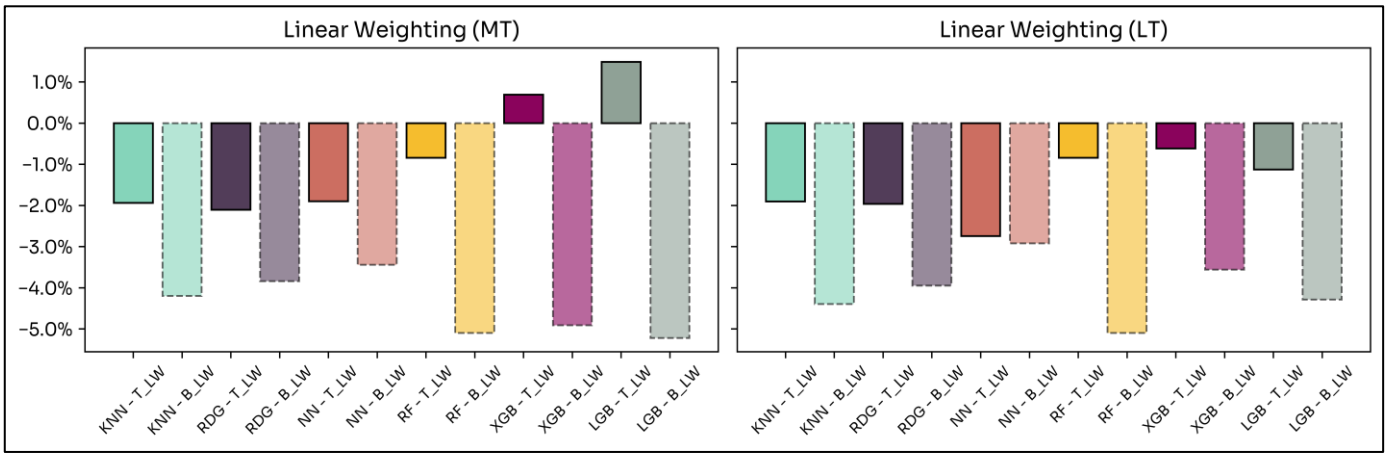
		More Training (50 Iterations, CV=5)			Less Training (10 Iterations, CV=3)		
Model (I)	Portfolio (II)	Ann. Return (III)	Sharpe (IV)	PSR (V)	Ann. Return (VI)	Sharpe (VII)	PSR (VIII)
KNN	Top	9.60%**	0.40	40.33%	9.66%**	0.40	41.13%
	Bottom	6.71%	0.21	17.70%	6.43%	0.20	16.38%
RDG	Top	9.32%**	0.37	36.77%	9.48%**	0.38	38.12%
	Bottom	7.13%*	0.23	19.79%	7.02%*	0.23	19.27%
NN	Top	9.39%**	0.36	35.32%	8.47%**	0.31	29.35%
	Bottom	7.82%*	0.28	24.70%	8.30%**	0.30	27.21%
RF	Top	10.56%**	0.43	44.77%	10.52%**	0.41	43.24%
	Bottom	6.03%	0.18	14.63%	6.87%	0.18	15.71%
XGB	Top	12.32%***	0.53	59.47%	10.72%**	0.43	45.01%
	Bottom	6.14%	0.18	15.17%	7.77%*	0.28	24.81%
LGB	Top	13.17%***	0.57	64.94%	10.21%**	0.41	41.68%
	Bottom	5.88%	0.17	13.91%	6.90%*	0.23	19.36%

Table 5 presents the annualized performance of the linear weighted ML test portfolios over the period January 2012 – August 2025 across ML models where columns III to IV describe portfolio performance based on 'more training' and columns V to IX (less training). Returns are annualized and tested via two-tailed 1-sample t-tests, while Sharpe ratios and their respective PSR's are also presented.

Both Table 5 and Figure 4 describe the long-only and excess annualised performance of respective 'top' and 'bottom' portfolios across ML models applying linear rank weighting. Linear rank weighting involves assigning linearly degrading weights based on predicted rank and is directly related to prediction strength. Under 'more' training (left-hand sub-figure of Figure 4 and Table 5, columns III-IV), linear weighting has the most

positive impact on the gradient boosters, with both LGB and XGB producing long-only, statistically significant 'top' annualized returns in excess of 12%. XGB's 'top' portfolio achieves 12.3% (significant at the 1% level) as well as a PSR of 60%. LGB is once again the best performing model, achieving an annualized return of 13.2% (significant at the 1% level) and delivers a PSR relative to the benchmark of 65%. Under 'less' training, most models produce economically lower 'top' portfolio returns and simultaneously higher 'bottom' portfolio performance, with RF once again being the most consistent, albeit with an economically lower return which is still significant at the 5% level.

Figure 4. Linear weighted ML test portfolio annualized excess average performance over the period January 2012 to August 2025 applying more ('MT') and less ('LT') training.



Note. Returns are annualised and expressed in excess of the JSE ALSI where darker bars represent performance of the respective ML models top 30 portfolio and lighter shaded bars represent performance of the respective ML models bottom 30 portfolio.

Table 6. Hypothetical zero-cost long-short annualized excess returns under 'more' and 'less' training assuming linear rank weighting

	Linear Weight	
Model	Excess Return - More Training	Excess Return - Less Training
KNN (Top – Bottom)	1.38%	1.59%
RDG (Top – Bottom)	0.59%	0.84%
NN (Top – Bottom)	0.60%	-0.84%
RF (Top – Bottom)	3.52%	3.33%
XGB (Top – Bottom)	5.02%*	2.15%
LGB (Top – Bottom)	6.32%**	2.58%

Table 6 presents the long-short zero cost performance of the respective models top minus bottom average annual performance under linear weighting. As noted above, gradient boosters outperform when applying linear weighting using 'more training', with XGB and LGB producing statistically significant excess returns of 5% and 6.3% respectively. Importantly, RF produces consistent performance independent of 'more' and 'less' training, achieving 3.52% and 3.33% respectively, making it the 3rd best model under 'more' training and best model under 'less' training.

Evaluating the results across weighting methodologies, it appears that linear weighting achieves the best results on a long-only and excess return basis under 'more' training, especially for the gradient boosters. The performance improvement under linear weighting strengthens the case for gradient boosters as the nature of the weighting methodology is directly linked to the economic magnitude of model predictions. Placing more weight on model prediction signals is an inadvertent test of model accuracy, as

prediction strength is evaluated through portfolio position size and subsequent returns. Conversely, market capitalization weighting, which naturally makes portfolios more benchmark cognizant, favors the less complex ML models as well as NN, especially when applying 'less' training.

Considering model performance, there is a clear difference between ensemble and gradient boosting models, with RF, XGB and LGB producing the best aggregate performance under more 'training', with LGB standing out as the best performing ML model. The results are therefore mostly consistent with the body of literature as RF, XGB and LGB are regularly found to be superior predictors given their ability in capturing and exploiting non-linear relationships between historical features and future individual share excess performance. Importantly, a notable difference emerges between the results presented and the body of literature related to the performance of NN. Across most of the literature surveyed, NN's are typically found to produce superior performance through their ability to capture non-linearities. The results above indicate the contrary, with NN producing the worst performance and only improving under market capitalization weighting in conjunction with less training. A credible reason for this could be related to empirical design, specifically related to hyperparameter tuning and maximum iterations associated with the randomized model specific parameter search applied. Several studies find that NN requires significantly more training and generally underperform gradient boosters when applied to structured tabulated data (see Grinsztajn et al. (2022)).

4.2. Risk-adjusted Performance

The section that follows details the risk-adjusted performance of the respective ML model portfolios through the application of factor-spanning regressions. Factor spanning regressions are run using four asset pricing models, namely; the Fama-French 3 factor model (FF3), Carhart 4 factor model (CH4), Fama-French 5 factor model (FF5) and the Fama-French 6 factor model (FF6). For each of the abovementioned models, we match the underlying ML portfolio weighting methodology with that of the individual style and factor premiums applied in regression analysis to ensure that weighting does not conflate regression parameter estimates. For brevity purposes, we limit our analysis to the 'more' training test portfolio returns and present both long-only regression parameter estimates as well as long-short portfolio alphas.

Table 7. Factor spanning regression results of long-only, equally weighted 'top' and 'bottom' ML model test portfolios using.

Panel A: Fama-French 3 factor model $R_{ML,i,t} - r_f = \alpha_i + \beta_{MKT} EMRP_t + \beta_{SMB} SMB_{EW,t} + \beta_{VMG} VMG_{EW,t} + \varepsilon_{i,t}$.

Model/Portfolio	α	β_{EMRP}	β_{SMB}	β_{VMG}	Adj R^2
KNN - Top 30	5.00%***	0.94***	0.30***	-0.14***	82.80%
KNN - Bottom 30	1.85%	0.99***	0.25***	0.17***	81.59%
RDG - Top 30	4.40%***	0.96***	0.28***	-0.16***	83.14%
RDG - Bottom 30	1.99%	1.00***	0.27***	0.19***	80.97%
NN - Top 30	3.75%**	0.99***	0.22***	-0.01	81.29%
NN - Bottom 30	3.38%	0.96***	0.32***	0.03	78.58%
RF - Top 30	5.71%***	1.00***	0.26***	-0.03	83.82%
RF - Bottom 30	1.81%	0.93***	0.27***	0.04**	77.79%
XGB - Top 30	5.81%***	0.98***	0.25***	-0.05***	82.87%
XGB - Bottom 30	1.51%	0.96***	0.30***	0.07***	80.73%
LGB - Top 30	6.16%***	1.00***	0.25***	-0.01	82.87%
LGB - Bottom 30	1.52%	0.95***	0.29***	0.03**	80.84%

Panel B: Carhart 4 factor model $R_{ML,i,t} - r_f = \alpha_i + \beta_{MKT} EMRP_t + \beta_{SMB} SMB_{EW,t} + \beta_{VMG} VMG_{EW,t} + \beta_{WML} WML_{EW,t} + \varepsilon_{i,t}$.

Model	α	β_{EMRP}	β_{SMB}	β_{VMG}	β_{WML}	Adj R^2
KNN - Top 30	3.83%**	0.95***	0.31***	-0.09***	0.10***	83.72%
KNN - Bottom 30	4.15%**	0.97***	0.22***	0.07***	-0.20***	84.41%
RDG - Top 30	2.89%	0.97***	0.31***	-0.09***	0.13***	84.61%
RDG - Bottom 30	4.61%**	0.98***	0.23***	0.08***	-0.23***	84.58%
NN - Top 30	3.80%**	0.99***	0.22***	-0.01	0.00	81.28%
NN - Bottom 30	4.48%**	0.95***	0.31***	-0.02	-0.10***	79.31%
RF - Top 30	4.31%***	1.02***	0.28***	0.03	0.12***	84.95%
RF - Bottom 30	4.35%**	0.91***	0.24***	-0.07***	-0.22***	81.82%
XGB - Top 30	4.87%***	0.98***	0.26***	-0.01	0.08***	83.40%
XGB - Bottom 30	3.64%**	0.94***	0.27***	-0.02	-0.18***	83.50%
LGB - Top 30	5.34%***	1.00***	0.26***	0.03	0.07***	83.26%
LGB - Bottom 30	3.41%**	0.93***	0.26***	-0.05***	-0.16***	83.09%

Panel C: Fama-French 5 factor model $R_{ML,i,t} - r_f = \alpha_i + \beta_{MKT} EMRP_t + \beta_{SMB} SMB_{EW,t} + \beta_{VMG} VMG_{EW,t} + \beta_{RMW} RMW_{EW,t} + \beta_{CMA} CMA_{EW,t} + \varepsilon_{i,t}$.

Model	α	β_{EMRP}	β_{SMB}	β_{VMG}	β_{RMW}	β_{CMA}	Adj R^2
KNN - Top 30	4.48%***	0.94***	0.30***	-0.07***	0.08***	-0.07***	83.43%
KNN - Bottom 30	1.66%	0.99***	0.26***	0.26***	0.16***	0.08***	82.56%
RDG - Top 30	3.80%**	0.95***	0.29***	-0.05***	0.15***	-0.03	84.44%
RDG - Bottom 30	1.81%	1.00***	0.27***	0.25***	0.08***	0.02	81.23%
NN - Top 30	3.02%	0.98***	0.23***	0.10***	0.15***	-0.06***	82.60%
NN - Bottom 30	3.33%	0.96***	0.33***	0.08***	0.09***	0.06***	78.95%
RF - Top 30	4.94%***	0.99***	0.26***	0.06***	0.09***	-0.12***	84.91%
RF - Bottom 30	1.82%	0.93***	0.28***	0.11***	0.13***	0.10***	78.70%
XGB - Top 30	4.96%***	0.97***	0.26***	0.08***	0.17***	-0.07***	84.78%
XGB - Bottom 30	1.62%	0.96***	0.31***	0.10***	0.07***	0.08***	81.03%
LGB - Top 30	5.58%***	0.99***	0.26***	0.07***	0.09***	-0.07***	83.57%
LGB - Bottom 30	1.33%	0.95***	0.29***	0.11***	0.14***	0.06***	81.75%

Panel D: Fama-French 6 factor model $R_{ML,i,t} - r_f = \alpha_i + \beta_{MKT} EMRP_t + \beta_{SMB} SMB_{EW,t} + \beta_{VMG} VMG_{EW,t} + \beta_{WML} WML_{EW,t} + \beta_{RMW} RMW_{EW,t} + \beta_{CMA} CMA_{EW,t} + \varepsilon_{i,t}$

Model	α	β_{EMRP}	β_{SMB}	β_{VMG}	β_{WML}	β_{RMW}	β_{CMA}	Adj R^2
KNN - Top 30	3.44%**	0.95***	0.32***	-0.02	0.10***	0.09***	-0.03	84.27%
KNN - Bottom 30	3.68%**	0.97***	0.22***	0.15***	-0.20***	0.12***	-0.01	85.08%
RDG - Top 30	2.33%	0.97***	0.32***	0.03	0.14***	0.18***	0.03**	86.07%
RDG - Bottom 30	4.29%**	0.97***	0.23***	0.12***	-0.24***	0.04**	-0.08***	84.91%
NN - Top 30	3.13%	0.98***	0.23***	0.10***	-0.01	0.15***	-0.07***	82.60%
NN - Bottom 30	4.25%**	0.95***	0.31***	0.03	-0.09***	0.08***	0.02	79.55%
RF - Top 30	3.76%**	1.01***	0.28***	0.12***	0.11***	0.11***	-0.07***	85.83%
RF - Bottom 30	4.03%**	0.90***	0.24***	0.00	-0.21***	0.10***	0.01	82.23%
XGB - Top 30	4.11%***	0.98***	0.27***	0.13***	0.08***	0.19***	-0.04	85.29%
XGB - Bottom 30	3.51%**	0.94***	0.28***	0.01	-0.18***	0.04	0.00	83.55%
LGB - Top 30	4.87%***	1.00***	0.27***	0.11***	0.07***	0.10***	-0.04**	83.89%
LGB - Bottom 30	2.99%	0.93***	0.27***	0.03**	-0.16***	0.11***	-0.01	83.74%

Note. The WML naming convention is used for consistency as opposed to UMD. All regressions are run assuming Newey-West consistent standard errors and alphas as well as factor loadings are emboldened to indicate statistical significance where ***,**,* represents significance at the 1%, 5% and 10% level.

Table 7 presents the equally weighted long-only regression results. Focusing first on the FF3 regressions per Table 7, Panel A, irrespective of the underlying ML model used, 'top' portfolios produce consistently larger alphas then their respective 'bottom' counterparts. As with gross performance described in section 4.1, RF, XGB and LGB produce the economically largest 'top' portfolio alphas, all of which are significant at the 1% level. In terms of factor loadings, unlike the KNN and RDG, advanced ML models (NN and tree-based models) 'top' portfolios tend to load more on the market, less on small and mid-caps and have higher exposures to growth relative to value.

Per Table 7, Panel B, when applying the CH4 model, the introduction of a momentum premium broadly reduces 'top' portfolio alphas while simultaneously increasing those of the 'bottom' portfolios. Of the six ML models employed, only XGB and LGB maintain their 'top' portfolio dominance, achieving alphas of 4.9% and 5.3% against 'bottom' alphas of 3.6% and 3.4%. Interestingly, all ML model 'top' portfolios tend to load positively on the momentum premium (barring NN) while 'bottom' portfolios load negatively. Like the FF3 results, the FF5 model per Panel C (which excludes momentum) produces superior 'top' portfolio alphas. Once again, RF, XGB and LGB produce the economically largest alphas, however, exploration of the additional factor premiums (RMW and CMA) reveals differences across ML model portfolios. All portfolios load positively on RMW yet differ in terms of the economic size of the factor loadings. KNN, RF and XGB 'bottom' portfolios tend to have larger RMW factor loadings (i.e. larger exposure to 'robust' companies based on ROE) relative to their 'top' portfolios, while the inverse is true for RDG, NN and LGB. With respect to CMA, 'top' and 'bottom' models are consistent where the former load negatively on CMA, while the latter load positively.

This implies that ML driven 'top' portfolios favor companies with more aggressive asset growth and 'bottom' portfolios favor conservative asset growth. The results of the factor spanning regressions run using the FF6 model per Table 7, Panel D (which includes the momentum premium) deliver alphas consistent with the CH4 results with 'top' portfolio alphas reducing in both economic size and significance. The sign and magnitude of the momentum factor loadings across 'top' and 'bottom' portfolios is maintained, however, 'bottom' portfolio alphas supersede 'top' portfolio alphas for KNN, RDG, NN and RF. Therefore, only XGB and LGB produce 'top' portfolio alphas greater than their 'bottom' portfolio counterparts.

Figure 5 describes the equal weighted long-short factor spanning regression alphas of the ML model portfolios across the four asset pricing models applied, where each bar denotes the annualized long-short excess alpha, p-value in square brackets and color coded to indicate statistical significance (green entails significant while grey, insignificant). Consistent with the results of Panel A, under the FF3 model, the best performing ML models are RF, XGB and LGB, with the latter two delivering annualized excess alphas of 4.3% and 4.6% that are statistically significant at the 10% level respectively. The introduction of momentum reduces annualized long-short alphas when applying the CH4 model, resulting in only XGB and LGB producing economically positive alphas, yet neither are statistically significant. The results for the FF5 model show that the incorporation of quality proxies RMW and CMA reduces the economic size of alphas, however, XGB and LGB still produce the largest, with the latter still being significant at the 10% level. Lastly, consistent with the CH4 results, the inclusion of the momentum premium within FF6 renders all alphas statistically insignificant, with only XGB and LGB producing a positive annualised excess returns of 0.6% and 1.9%, yet neither are statistically significant.

Tables 8 describe the long-only factor spanning regression results across the respective market capitalization weighted ML 'top' and 'bottom' portfolios. As shown in the gross return analysis, the impact of market capitalization weighting dramatically impacts results. Per Panel A, under the FF3, only four alphas are significant, three of which are 'bottom' portfolios per the NN, RF and XGB models. LGB produces the only 'top' portfolio with a positive and statistically significant alpha. In terms of factor loadings,

market betas are marginally lower compared to the equal weighted results with 'bottom' portfolios tending to load more on the market risk premium, barring NN. As expected, all portfolios load negatively on the size premium due to market capitalization weighting, while 'bottom' portfolios typically load positively on the value premium, with only the NN 'top' portfolio producing a positive value premium factor loading larger than its 'bottom' portfolio.

Figure 5. Equally weighted, long-short annualized regression alphas

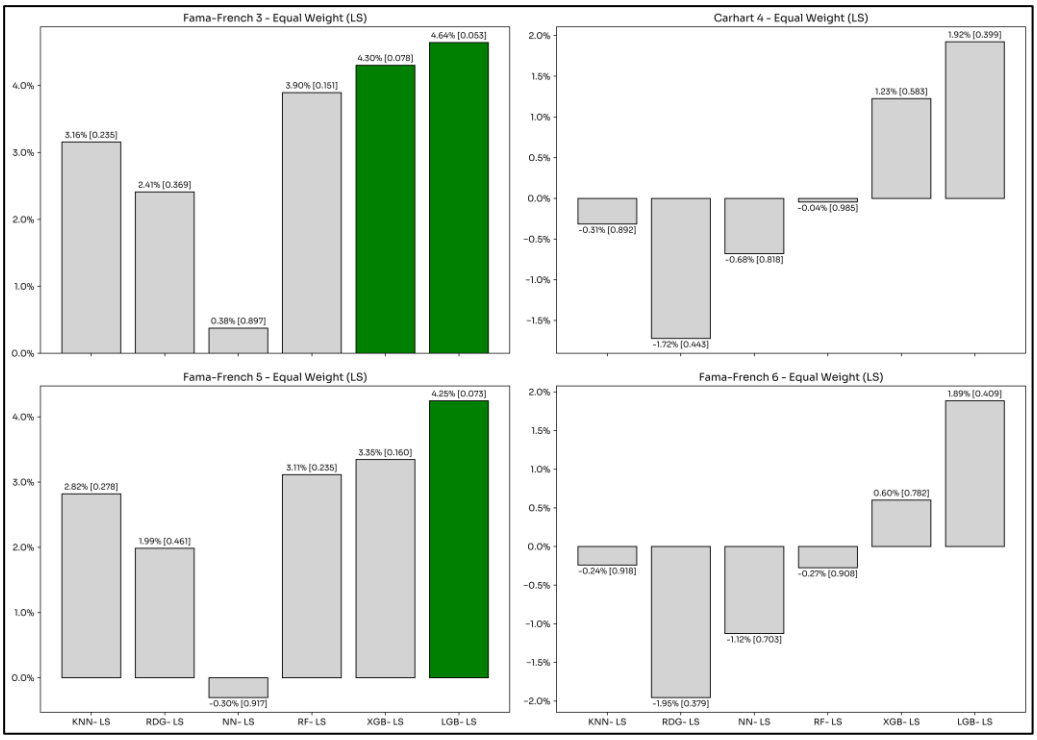


Table 8. Factor spanning regression results of long-only, market capitalization weighted 'top' and 'bottom' ML model test portfolios

Panel A: Fama-F`rench 3 factor model $R_{ML,i,t} - r_f = \alpha_i + \beta_{MKT} EMRP_t + \beta_{SMB} SMB_{MCW,t} + \beta_{VMG} VMG_{MCW,t} + \varepsilon_{i,t}$.

Model	α	β_{EMRP}	β_{SMB}	β_{VMG}	Adj R^2
KNN - Top 30	4.82%	0.76***	-0.13***	-0.02	63.26%
KNN - Bottom 30	4.26%	0.96***	-0.15***	0.11***	72.26%
RDG - Top 30	2.32%	0.89***	-0.07***	-0.10***	68.22%
RDG - Bottom 30	3.54%	0.91***	-0.19***	0.24***	70.66%
NN - Top 30	1.84%	0.94***	-0.14***	0.07***	69.98%
NN - Bottom 30	6.65%**	0.81***	-0.10***	0.03	59.70%
RF - Top 30	2.56%	0.90***	-0.16***	-0.04	68.07%
RF - Bottom 30	5.52%**	0.85***	-0.16***	0.12***	66.83%
XGB - Top 30	4.79%	0.86***	-0.17***	-0.03	68.04%
XGB - Bottom 30	5.52%**	0.91***	-0.13***	0.13***	70.43%
LGB - Top 30	6.33%**	0.88***	-0.15***	0.04**	71.13%
LGB - Bottom 30	5.30%	0.85***	-0.18***	0.09***	67.26%

Panel B: Carhart 4 factor model $R_{ML,i,t} - r_f = \alpha_i + \beta_{MKT} EMRP_t + \beta_{SMB} SMB_{MCW,t} + \beta_{VMG} VMG_{MCW,t} + \beta_{WML} WML_{MCW,t} + \varepsilon_{i,t}$.

Model	α	β_{EMRP}	β_{SMB}	β_{VMG}	β_{WML}	Adj R^2
KNN - Top 30	4.08%	0.78***	-0.14***	0.01	0.11***	65.38%
KNN - Bottom 30	4.83%	0.95***	-0.15***	0.09***	-0.08***	73.15%
RDG - Top 30	1.34%	0.91***	-0.08***	-0.06***	0.14***	71.21%
RDG - Bottom 30	4.48%	0.90***	-0.19***	0.20***	-0.13***	73.06%
NN - Top 30	2.12%	0.93***	-0.14***	0.06***	-0.04**	70.19%
NN - Bottom 30	6.30%**	0.82***	-0.10***	0.05**	0.05**	60.07%
RF - Top 30	1.82%	0.91***	-0.17***	-0.01	0.11***	69.71%
RF - Bottom 30	6.26%**	0.84***	-0.15***	0.09***	-0.10***	68.56%
XGB - Top 30	4.05%	0.87***	-0.17***	0.00	0.10***	69.74%
XGB - Bottom 30	6.07%**	0.90***	-0.12***	0.11***	-0.08***	71.33%
LGB - Top 30	5.66%**	0.89***	-0.15***	0.07***	0.09***	72.58%
LGB - Bottom 30	5.75%**	0.84***	-0.18***	0.08***	-0.06***	67.89%

Panel C: Fama-French 5 factor model $R_{ML,i,t} - r_f = \alpha_i + \beta_{MKT} EMRP_t + \beta_{SMB} SMB_{MCW,t} + \beta_{VMG} VMG_{MCW,t} + \beta_{RMW} RMW_{MCW,t} + \beta_{CMA} CMA_{MCW,t} + \varepsilon_{i,t}$.

Model	α	β_{EMRP}	β_{SMB}	β_{VMG}	β_{RMW}	β_{CMA}	Adj R^2
KNN - Top 30	4.91%	0.76***	-0.14***	0.00	0.04	0.01	63.38%
KNN - Bottom 30	4.23%	0.97***	-0.14***	0.13***	0.07***	0.06**	72.75%
RDG - Top 30	2.22%	0.91***	-0.06***	-0.07***	0.09***	0.09***	69.39%
RDG - Bottom 30	3.85%	0.89***	-0.22***	0.24***	-0.02	-0.08***	71.21%
NN - Top 30	2.41%	0.92***	-0.17***	0.13***	0.11***	-0.04	71.27%
NN - Bottom 30	6.37%**	0.84***	-0.08***	0.04	0.05**	0.10***	60.71%
RF - Top 30	2.79%	0.89***	-0.17***	-0.01	0.07***	0.00	68.42%
RF - Bottom 30	5.91%**	0.83***	-0.18***	0.15***	0.04**	-0.06***	67.38%
XGB - Top 30	4.91%	0.87***	-0.17***	0.01	0.09***	0.05	68.85%
XGB - Bottom 30	5.60%**	0.91***	-0.13***	0.14***	0.01	-0.01	70.44%
LGB - Top 30	6.27%**	0.90***	-0.14***	0.05**	0.05**	0.05***	71.49%
LGB - Bottom 30	5.58%**	0.85***	-0.19***	0.13***	0.08***	0.00	67.84%

Panel D: Fama-French 6 factor model $R_{ML,i,t} - r_f = \alpha_i + \beta_{MKT} EMRP_t + \beta_{SMB} SMB_{MCW,t} + \beta_{VMG} VMG_{MCW,t} + \beta_{WML} WML_{MCW,t} + \beta_{RMW} RMW_{MCW,t} + \beta_{CMA} CMA_{MCW,t} + \varepsilon_{i,t}$

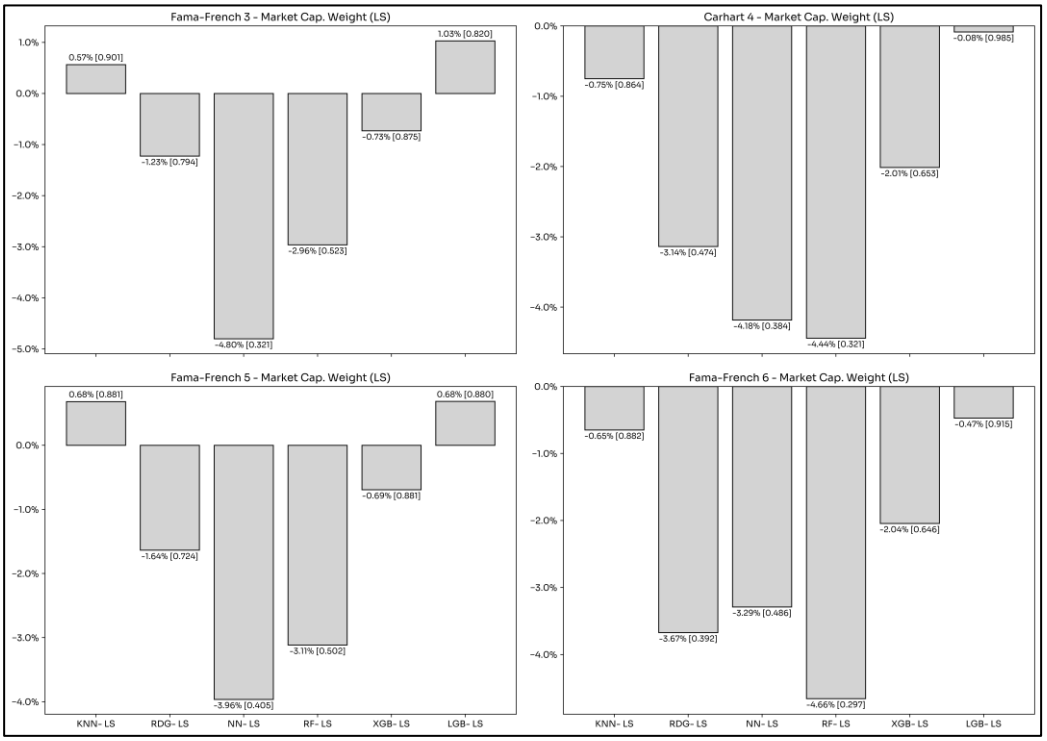
Model	α	β_{EMRP}	β_{SMB}	β_{VMG}	β_{WML}	β_{RMW}	β_{CMA}	Adj R^2
KNN - Top 30	4.14%	0.78***	-0.14***	0.03	0.11***	0.04**	0.02	65.57%
KNN - Bottom 30	4.79%	0.96***	-0.14***	0.11***	-0.08***	0.06***	0.05	73.55%
RDG - Top 30	1.17%	0.93***	-0.06***	-0.03	0.15***	0.10***	0.11***	72.67%
RDG - Bottom 30	4.84%	0.87***	-0.21***	0.21***	-0.14***	-0.02	-0.09***	73.78%
NN - Top 30	2.69%	0.91***	-0.17***	0.12***	-0.04**	0.10***	-0.05**	71.48%
NN - Bottom 30	5.98%**	0.85***	-0.08***	0.05**	0.05***	0.05**	0.10***	61.18%
RF - Top 30	2.02%	0.91***	-0.18***	0.02	0.11***	0.07***	0.01	70.12%
RF - Bottom 30	6.67%**	0.82***	-0.18***	0.12***	-0.11***	0.03	-0.06***	69.17%
XGB - Top 30	4.13%	0.88***	-0.17***	0.04	0.11***	0.10***	0.06**	70.69%
XGB - Bottom 30	6.17%**	0.90***	-0.13***	0.12***	-0.08***	0.00	-0.02	71.35%
LGB - Top 30	5.56%**	0.91***	-0.14***	0.08***	0.10***	0.05***	0.06***	73.05%
LGB - Bottom 30	6.03%**	0.84***	-0.19***	0.11***	-0.06***	0.08***	0.00	68.44%

Note. The WML naming convention is used for consistency as opposed to UMD. All regressions are run assuming Newey-West consistent standard errors and alphas as well as factor loadings are emboldened to indicate statistical significance where ***,**,* represents significance at the 1%, 5% and 10% level.

Per Table 8, Panel B, applying the CH4 spanning model results in five of the twelve long-only portfolios achieving statistically significant alphas, four of which relate to 'bottom' portfolios, while once again, only the LGB 'top' portfolio achieves a statistically significant alpha. Like the equal weighted results, 'top' portfolios load positively on the momentum premium while 'bottom' portfolios load negatively, barring NN, which produces the opposite loading signs. As shown in Panels C and D, both the FF5 and FF6 spanning regression results are virtually identical in terms of alphas compared to the FF3 and CH4 results, with only five portfolios producing significant alphas, however, the factor loadings of the quality proxies (RMW and CMA) are disparate and fail to show any pattern across ML model 'top' and bottom' portfolios.

Figure 6 describes the annualised long-short excess return alphas between market capitalization weighted 'top' and 'bottom' portfolios across the ML models considered. Excluding the momentum in factor spanning regressions (i.e. per the FF3 and FF5 models) produces virtually identical patterns with KNN and LGB producing positive, yet statistically insignificant alphas. Conversely, RDG, NN, RF and XGB produce negative excess alphas, with NN being the most economically negative. When including the momentum premium (per CH4 and FF6), all alphas are negative and insignificant, however, LGB still produces the highest excess alphas of -0.1% and -0.5%.

Figure 6. Market capitalisation weighted, long-short annualised regression alphas



The final set of factor spanning regressions are presented in Table 9 and Figure 7. As mentioned, linear rank weighting is akin to factor weighting, where at each portfolio formation date, model predictions are ranked and assigned linear descending weights that sum to unity. The resulting portfolios positions are therefore directly linked to the relative prediction size, increasing the connectedness between portfolio returns beyond share selection, which is a limitation of both equal and market capitalization weighting. Consistent with the equal weighted results presented in Table 7, Panel A of Table 9 shows that most 'top' portfolios produce statistically significant alphas that are consistently larger than their 'bottom' counterparts, with only NN being statistically insignificant. More importantly, the gradient booster alphas increase in magnitude when compared to equally weighted FF3 results, with XGB and LGB 'top' portfolios producing annualized excess alphas of 6.6% and 7.3% respectively. Considering factor loadings, all portfolios

produce market betas close to unity and load positively on the size premium. In terms of the value premium, once again 'top' portfolios tend to load on growth while 'bottom' portfolios load on value.

Table 9. Factor spanning regression results of long-only, linear weighted 'top' and 'bottom' ML model test portfolios

Panel A: Fama-French 3 factor model $R_{ML,i,t} - r_f = \alpha_i + \beta_{MKT} EMRP_t + \beta_{SMB} SMB_{LW,t} + \beta_{VMG} VMG_{LW,t} + \varepsilon_{i,t}$.

Model	α	β_{EMRP}	β_{SMB}	β_{VMG}	Adj R^2
KNN - Top 30	4.78%***	0.96***	0.33***	-0.15***	81.80%
KNN - Bottom 30	0.94%	1.02***	0.26***	0.21***	76.63%
RDG - Top 30	4.44%**	0.98***	0.30***	-0.20***	81.23%
RDG - Bottom 30	1.37%	1.01***	0.28***	0.26***	75.42%
NN - Top 30	3.90%	0.99***	0.23***	-0.03	76.92%
NN - Bottom 30	2.69%	0.97***	0.33***	0.08***	75.16%
RF - Top 30	4.99%***	1.01***	0.28***	-0.04***	80.66%
RF - Bottom 30	1.10%	0.92***	0.27***	0.08***	71.99%
XGB - Top 30	6.64%***	0.99***	0.25***	-0.06***	80.11%
XGB - Bottom 30	0.93%	0.96***	0.29***	0.12***	76.46%
LGB - Top 30	7.26%***	1.00***	0.25***	-0.02	80.22%
LGB - Bottom 30	0.92%	0.95***	0.30***	0.05**	77.22%

Panel B: Carhart 4 factor model $R_{ML,i,t} - r_f = \alpha_i + \beta_{MKT} EMRP_t + \beta_{SMB} SMB_{LW,t} + \beta_{VMG} VMG_{LW,t} + \beta_{WML} WML_{LW,t} + \varepsilon_{i,t}$.

Model	α	β_{EMRP}	β_{SMB}	β_{VMG}	β_{WML}	Adj R^2
KNN - Top 30	3.34%**	0.96***	0.34***	-0.09***	0.09***	82.96%
KNN - Bottom 30	4.38%	1.01***	0.24***	0.07***	-0.22***	81.22%
RDG - Top 30	2.00%	0.99***	0.32***	-0.10***	0.16***	84.33%
RDG - Bottom 30	5.07%**	1.00***	0.26***	0.11***	-0.24***	80.58%
NN - Top 30	3.82%	0.99***	0.23***	-0.02	0.01	76.91%
NN - Bottom 30	3.83%	0.97***	0.33***	0.03	-0.07***	75.76%
RF - Top 30	2.38%	1.02***	0.29***	0.06***	0.17***	83.81%
RF - Bottom 30	4.67%**	0.91***	0.25***	-0.07***	-0.23***	78.15%
XGB - Top 30	4.94%***	0.99***	0.26***	0.01	0.11***	81.51%
XGB - Bottom 30	3.93%**	0.95***	0.27***	-0.01	-0.20***	80.66%
LGB - Top 30	5.38%***	1.00***	0.26***	0.06***	0.12***	81.89%
LGB - Bottom 30	3.60%	0.94***	0.29***	-0.07***	-0.17***	80.87%

Panel C: Fama-French 5 factor model $R_{ML,i,t} - r_f = \alpha_i + \beta_{MKT} EMRP_t + \beta_{SMB} SMB_{LW,t} + \beta_{VMG} VMG_{LW,t} + \beta_{RMW} RMW_{LW,t} + \beta_{CMA} CMA_{LW,t} + \varepsilon_{i,t}$.

Model	α	β_{EMRP}	β_{SMB}	β_{VMG}	β_{RMW}	β_{CMA}	Adj R^2
KNN - Top 30	4.35%***	0.95***	0.33***	-0.10***	0.07***	-0.02	82.15%
KNN - Bottom 30	0.90%	1.02***	0.27***	0.26***	0.10***	0.11***	77.18%
RDG - Top 30	3.82%**	0.97***	0.31***	-0.11***	0.12***	0.00	82.06%
RDG - Bottom 30	1.61%	1.02***	0.28***	0.24***	0.00	0.05**	75.48%
NN - Top 30	3.09%	0.98***	0.23***	0.08***	0.12***	-0.04**	77.88%
NN - Bottom 30	2.91%	0.98***	0.34***	0.08***	0.03	0.08***	75.42%
RF - Top 30	4.24%**	1.00***	0.27***	0.03	0.06***	-0.09***	81.40%
RF - Bottom 30	1.54%	0.93***	0.28***	0.07***	0.05**	0.15***	72.86%
XGB - Top 30	5.75%***	0.97***	0.25***	0.05***	0.12***	-0.06***	81.33%
XGB - Bottom 30	1.41%	0.97***	0.29***	0.09***	0.01	0.11***	76.96%
LGB - Top 30	6.77%***	0.99***	0.25***	0.03	0.04***	-0.06***	80.55%
LGB - Bottom 30	1.03%	0.95***	0.31***	0.08***	0.09***	0.13***	78.02%

Panel D: Fama-French 6 factor model $R_{ML,i,t} - r_f = \alpha_i + \beta_{MKT} EMRP_t + \beta_{SMB} SMB_{LW,t} + \beta_{VMG} VMG_{LW,t} + \beta_{WML} RMW_{LW,t} + \beta_{RMW} RMW_{LW,t} + \beta_{CMA} CMA_{LW,t} + \varepsilon_{i,t}$

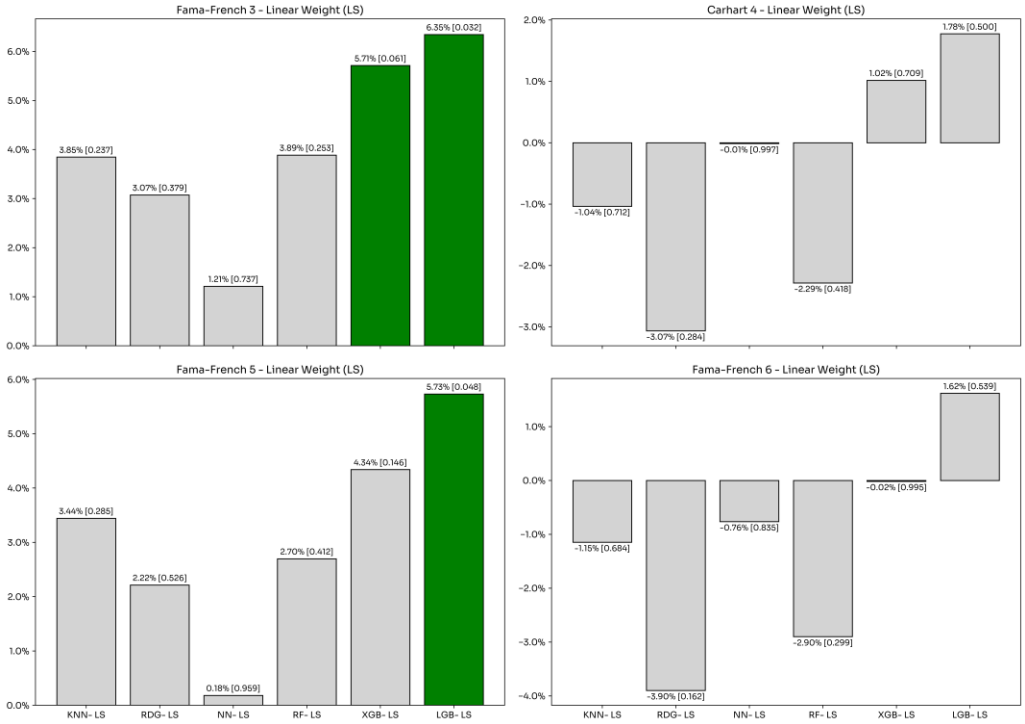
Model	α	β_{EMRP}	β_{SMB}	β_{VMG}	β_{WML}	β_{RMW}	β_{CMA}	Adj R^2
KNN - Top 30	2.91%	0.96***	0.34***	-0.03**	0.10***	0.09***	0.03	83.40%
KNN - Bottom 30	4.06%	1.00***	0.24***	0.11***	-0.22***	0.05***	-0.01	81.36%
RDG - Top 30	1.32%	0.99***	0.32***	0.01	0.18***	0.16***	0.09***	85.59%
RDG - Bottom 30	5.22%**	1.00***	0.25***	0.08***	-0.26***	-0.06***	-0.09***	80.81%
NN - Top 30	3.07%	0.98***	0.23***	0.08***	0.00	0.12***	-0.04**	77.87%
NN - Bottom 30	3.83%	0.97***	0.33***	0.04	-0.07***	0.02	0.05**	75.84%
RF - Top 30	1.83%	1.01***	0.29***	0.14***	0.17***	0.10***	-0.01	84.33%
RF - Bottom 30	4.73%**	0.91***	0.26***	-0.07***	-0.23***	0.00	0.03	78.18%
XGB - Top 30	4.13%**	0.98***	0.26***	0.13***	0.12***	0.15***	0.00	82.71%
XGB - Bottom 30	4.14%**	0.95***	0.27***	-0.03	-0.19***	-0.03	0.01	80.73%
LGB - Top 30	4.99%***	1.00***	0.27***	0.11***	0.13***	0.07***	0.01	82.14%
LGB - Bottom 30	3.37%	0.94***	0.29***	-0.03	-0.17***	0.06***	0.04**	81.02%

Note. The WML naming convention is used for consistency as opposed to UMD. All regressions are run assuming Newey-West consistent standard errors and alphas as well as factor loadings are emboldened to indicate statistical significance where ***,**,* represents significance at the 1%, 5% and 10% level.

Per Table 9, Panel B, CH4-spanning regressions show that incorporating the momentum premium degrades 'top' portfolio alphas while improving 'bottom' portfolio alphas. Now, six portfolios produce statistically significant alphas, with only half being 'top' portfolios. Consistent with the other weighting methods, 'top' portfolios load positively on the momentum premium, which explains the majority of 'top' portfolios producing inferior alphas relative to their 'bottom' counterparts. The only portfolios that maintain 'top' portfolio dominance is once again limited to the gradient boosters, with XGB and LGB producing annualized alphas of 4.9% and 5.4% that are significant at the 1% level. Like the FF3, the FF5 results per Panel C show that excluding momentum positively impacts 'top' portfolio alphas, with all (excluding NN) being statistically significant and economically larger than their respective 'bottom' counterparts. The results further point to the included quality proxies (RMW and CMA) proving unable to explain performance across the respective ML 'top' and 'bottom' portfolios. However, the results of the FF6 factor spanning regressions per Panel D mirror those of Panel B. Once again, the inclusion of the momentum premium results in only five portfolios generating statistically significant alphas, with only two being produced by the 'top' XGB and LGB portfolios.

Figure 7 describes factor spanning regression annualized excess alphas run on long-short linear weighted portfolio returns across the ML models applied. Like the equal weighted results presented in figure 5, XGB and LGB produce the economically largest annualized alphas across asset pricing models employed. Furthermore, long-short alphas are dramatically impaired when including the momentum premium. Under FF3, all alphas are positive, yet only those produced by XGB and LGB are statistically significant at the 10% and 5% level. A similar pattern emerges under FF5, however, only LGB's alpha (5.7%) is statistically significant. Like the long-only regression analysis per Table 9, Panel B, the inclusion of momentum dramatically reduces excess alphas where, under CH4, only XGB and LGB are positive, delivering 1% and 1.8% respectively, yet lack statistical significance. The degradation of factor spanning alphas under FF6 is more pronounced, with only LGB producing a positive alpha of 1.6% but is statistically insignificant.

Figure 7. Linear weighted long-short annualised regression alphas



5. Discussion and Conclusion

This study is the first to examine the predictability of share returns on the cross-section of listed equities on the JSE via ML and the South African factor zoo. Using a combination of raw and engineered features (style proxies), our study produces novel findings on the application of ML based investment strategies specific to the South African market. The growing body of international literature focuses on developed market bourses and typically finds that the key benefit of ML over econometric models is the ability to identify and capitalize on non-linear relationships between features and future returns. The broader research question is whether the benefits of ML, in terms of predictive accuracy, extend to developing markets with smaller equity cross-sections and less liquidity. To this end, we consider the predictive accuracy of a subset of ML models through portfolio based tests and extend the experiment to consider the impact of 'more' and 'less' training, portfolio weighting method as well as matching the prediction period to the underlying benchmark rebalance frequency. Our findings show that 'more' training tends to favor tree-based ensemble ML models that are best adept at capturing non-linear dependencies, specifically RF and the gradient boosters, XGB and LGB. Conversely, 'less' training favors the less sophisticated ML models, namely KNN and RDG. In contrast to the body of literature, our findings show that under the empirical design applied, NN is the worst performing model. The result emphasizes that contrary to the international findings, NN fails to exploit non-linear relationships when applied to South African data, however, this could be largely attributable to the noted difference in training and validation required between NN and tree-based models. Second, in both nominal and risk-adjusted performance tests, market capitalisation weighting has a negative effect on portfolio performance and therefore predictability. The impact is almost bi-directional where both equal and linear weighting typically result in 'top' portfolios outperforming their 'bottom' counterparts, while market capitalization weighting conversely improves 'bottom' portfolios and reduces performance of 'top' portfolios.

More importantly, the findings both conform and contradict the global literature. Most of our results show that the incorporation of non-linear features benefits ensemble and gradient booster decision-tree models, with LGB achieving the consistently best

portfolio performance. Across both gross return and risk-adjusted tests, LGB, XGB and RF produce the highest annualized long-only and long-short performance. In portfolio tests, performance can be treated as synonymous predictability, entailing that the findings are broadly consistent with the current body of literature. However, our results depart from the literature in factor spanning tests as the inclusion of the momentum premium dramatically reduces long-only annualized alphas and makes long-short, annualized alphas insignificantly different from zero. A reasonable explanation for this outcome is the dominance of price momentum on JSE, especially when compared to other South African factors. As seen in factor spanning regressions, the incorporation of the momentum premium absorbs the majority of alpha associated with ML based strategies. Moreover, the fact that all 'top' portfolios consistently load positively on the momentum premium strengthens this hypothesis.

Lastly, notwithstanding the contribution to emerging market literature, this study provides several avenues of future research. An important expansion would be broadening the number of ML models applied, specifically the incorporation of additional deep learning neural nets and autoencoder models, which have been found to perform well in capturing non-linear dynamics between historical features and future returns. Second, both developed and developing market studies should consider the practical implications of applying ML based strategies, specifically considering performance across market conditions, implied portfolio turnover and the consideration of transaction costs. As shown in Appendix C, long-only ML based portfolios typically display average stock turnover ratios of between 37% and 55%, with the best performing ML models, namely XGB and LGB turning over 43% and 45% of their stocks at each rebalance on average, which would certainly reduce alpha when including trading costs and buy-side securities transfer tax.

On this basis, we emphasise that the study is not an exhaustive test of ML models in predicting returns and that the results may differ considerably when applying transaction costs. Nevertheless, our findings broadly show that ML based stock selection strategies are both feasible and profitable when applied in an emerging market context, albeit to a lesser degree when compared to developed market studies, especially when considering alpha degradation when including momentum. Lastly, our findings emphasize the importance of hyperparameter tuning assumptions and the effects of feature engineering to capture non-linearities, especially when applied to ML models that have been proven to capture non-linear relationships between historical stylized proxy based features and future returns.

Author Contributions: Conceptualisation: D.P; Methodology: D.P and Y.S; Formal Analysis: D.P and Y.S.; Validation: C.A; Writing – original draft: D.P; Writing – review and editing: Y.S and C.A.

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments: We thank the Editor and Reviewers for their constructive comments in improving our submission.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Feature Set

Feature Family	Feature	Description	# Features
Returns based	Price Momentum, no skip	Historical momentum calculated over historical 252-day period applying t-z historical windows where z initially equals 21 trading days and increases by 21-days until 252	12
Returns based	Price Momentum, skip most recent 21-days	Historical momentum calculated over historical 252-day period, skipping the most recent 21-day period applying t-z historical windows where z initially equals 21 trading days and increases by 21-days until 252	11
Returns based	Price scaled by historical moving average	Latest price scaled by historical moving average measured over t-z window periods where z initially equals 21 trading days and increases by 21-days until 252	12
Returns based	Standard deviation	Historical standard deviation measured over historical 252-day window period applying t-z historical windows where z initially equals 21 trading days and increases by 21-days until 252	12
Returns based	95% Value-at-Risk	Historical VaR measured over historical 252-day window period applying t-z historical windows where z initially equals 63 trading days and increases by 21-days until 252	10
Returns based	Sharpe Ratio	Historical excess return scaled by standard deviation calculated over historical 252-day period applying t-z historical windows where z initially equals 21 trading days and increases by 21-days until 252	12
Returns based	Trading volume	Historical total trading volume calculated over historical 252-day period applying t-z historical windows where z initially equals 21 trading days and increases by 21-days until 252	12
Returns based	Market Beta	Historical market beta measured via OLS calculated over historical 252-day period	1
Returns based	Broad Sector Beta	Historical sector betas measured via OLS against broad JSE Financial, Industrial and Mining Indices calculated over historical 252-day period	3
Returns based	Exchange rate beta	Historical exchange rate beta measured against changes in the USDZAR exchnage rate over a historical 252-day period	1
Returns based	VIX Beta	Historical VIX beta meaured against changes in the CBOT VIX index over a historical 252-day period	1
Fundamental	Earnings Yield	Lagged earnings yield measured as lagged EPS scaled by latest share price	1
Fundamental	Dividend Yield	Trailing annual dividends scaled by latest share price	1
Fundamental	Book-to-Market ratio	Lagged BVPS scaled by latest share price	1
Fundamental	Cashflow per share	Lagged CFPS	1
Fundamental	Cashflow to price	Lagged CFPS scaled by latest share price	1
Fundamental	Debt-to-assets	Lagged balance sheet interest-bearing debt scaled by total assets	1
Fundamental	NOPAT-to-assets	Lagged NOPAT scaled by total assets	1
Fundamental	Net Profit-to-assets	Lagged Net Profit scaled by total assets	1
Fundamental	Earnings growth	Lagged percentage change in reported EPS	1
Fundamental	Total Asset growth	Lagged percentage change in reported Total Assets	1
Fundamental	Sales growth	Lagged percentage change in reported Sales	1
Fundamental	Book value per share growth	Lagged percentage change in reported BVPS	1
Fundamental	Net Profit growth	Lagged percentage change in reported Net Profit	1
Fundamental	Cashflow per share growth	Lagged percentage change in reported CFPS	1

Fundamental	Debt-to-equity	Lagged balance sheet interest-bearing debt scaled by total equity	1
Fundamental	Return on equity	Lagged Net Profit scaled by total equity	1
Fundamental	Return on Assets	Lagged Net Profit scaled by total assets	1
Fundamental	Return on Invested Capital	Lagged NOPAT scaled by average invested capital	1
Fundamental	Net Profit Margin	Lagged Net Profit scaled by total Sales	1
Fundamental	Log Market Capitalization	Natural log of latest market capitalization	1
Total Features			107

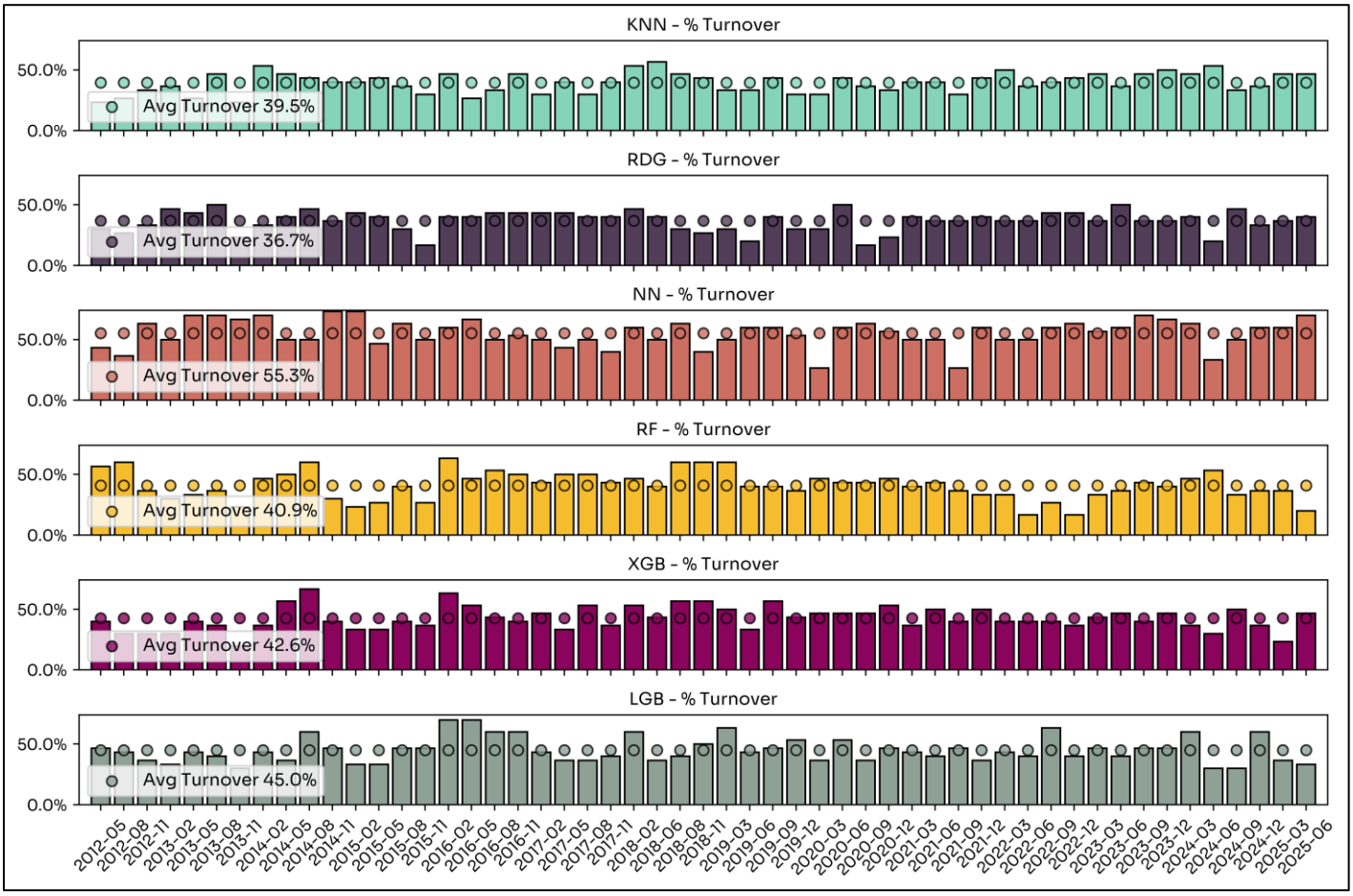
Appendix B

Table B1. Parameters

Model	Model Python Name	Parameter Grid
KNN	KNeighborsRegressor	n_neighbors: list(range(5, 305)); weights: uniform, distance; p: 1, 2
RDG	Ridge	alpha: np.linspace(0.01, 100, 1000); fit_intercept: True, False; max_iter: 100, 250, 500, 1000, 5000
NN	MLPRegressor	hidden_layer_sizes: (32,), (50,), (100,), (32, 16), (50, 25), (100, 50); activation: identity, logistic, tanh, relu; solver: lbfgs, sgd, adam; alpha: 1e-7, 1e-5, 1e-3, 1e-1, 0.025, 0.05; learning_rate_init: 1e-8, 1e-6, 1e-4, 1e-2, 0.01
RF	RandomForestRegressor	n_estimators: 10, 50, 75, 100, 200; max_features: sqrt, log2, None; max_depth: 3, 6, 9, 12; max_leaf_nodes: 3, 6, 9, 12; min_samples_split: 2, 4, 6, 8
XGB	XGBRegressor	n_estimators: 10, 50, 75, 100, 200, 500; max_depth: 3, 4, 5, 6, 7, 8, 9, 10; subsample: 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9; colsample_bytree: 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9; min_child_weight: 1, 2, 3, 4, 5, 6, 8, 9, 10; learning_rate: 1e-5, 1e-4, 1e-3, 1e-2, 1e-1
LGB	LGBMRegressor	n_estimators: 10, 50, 75, 100, 200, 500; num_leaves: 20, 30, 40, 50; max_depth: 3, 4, 5, 6, 7, 8, 9, 10; min_data_in_leaf: 10, 20, 30, 40, 50; feature_fraction: 0.6, 0.7, 0.8, 0.9, 1.0; learning_rate: 1e-5, 1e-4, 1e-3, 1e-2, 1e-1

Appendix C

Figure C1. Turnover statistics



Note. Figure C1 describes portfolio turnover statistics associated with the long-only ML model portfolios applied at each rebalance date. Stock turnover is calculated simply as the number of new shares divided by the total number of shares in the portfolio. Stock turnover is then averaged over the full out-of-sample period and reported.

References

Bailey, D. H., & Lopez de Prado, M. (2012). The Sharpe ratio efficient frontier. *Journal of Risk, 15*(2), 13. https://doi.org/10.21314/JOR.2012.255

Basiewicz, P. G., & Auret, C. J. (2009). Another look at the cross-section of average returns on the JSE. *Investment Analyst Journal, 38*(69), 23-38. https://doi.org/10.1080/10293523.2009.11082507

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance, 52*(1), 57-82. https://doi.org/10.1111/j.1540-6261.1997.tb03808.x

Cochrane, J. H. (2011). Presidential address: Discount rates. *The Journal of Finance, 66*(4), 1047-1108. https://doi.org/10.1111/j.1540-6261.2011.01671.x

Cox, S., & Britten, J. (2019). The Fama-French five-factor model: evidence from the Johannesburg Stock Exchange. *Investment Analyst Journal, 48*(3), 240-261. https://doi.org/10.1080/10293523.2019.1647982

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics, 33*(1), 3-56. https://doi.org/10.1016/0304-405X(93)90023-5

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics, 116*(1), 1-22. https://doi.org/10.1016/j.jfineco.2014.10.010

Fama, E. F., & French, K. R. (2018). Choosing factors. *Journal of Financial Economics, 128*(2), 234-252. https://doi.org/10.1016/j.jfineco.2018.02.012

Grinsztajn, L., Oyallon, E., & Varoquaux, G. (2022). Why do tree-based models still outperform deep learning on typical tabular data? *Advances in neural information processing systems, 6*(35), 507-520.

https://proceedings.neurips.cc/paper_files/paper/2022/file/0378c7692da36807bdec87ab043cdadc-Paper-Datasets_and_Benchmarks.pdf

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies, 33*(5), 2223-2273. https://doi.org/10.1093/rfs/hhaa009

Hodnett, K., Hsieh, H. H., & van Rensburg, P. (2012). Payoffs to equity investment styles on the jse securities exchange: The case of South African equity market. *The International Business & Economics Research Journal , 11*(1), 19. https://www.researchgate.net/profile/Paul-Van-Rensburg/publication/297722786_Payoffs_To_Equity_Investment_Styles_On_The_JSE_Securities_Exchange_The_Case_Of_South_African_Equity_Market/links/57e4f2b208aee9b409fe860d/Payoffs-To-Equity-Investment-Styles-On-The-JSE-Securities-Exchange-The-Case-Of-South-African-Equity-Market.pdf

Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of finance, 48*(1), 65-91. https://doi.org/10.1111/j.1540-6261.1993.tb04702.x

Jegadeesh, N., & Titman, S. (2001). Profitability of momentum strategies: An evaluation of alternative explanations. *The Journal of finance, 56*(2), 699-720. https://doi.org/10.1111/0022-1082.00342

Leippold, M., Wang, Q., & Zhou, W. (2022). Machine learning in the Chinese stock market. *Journal of Financial Economics, 145*(2), 64-82. https://doi.org/10.1016/j.jfineco.2021.08.017

Leung, E., Lohre, H., Mischlich, D., Shea, Y., & Stroh, M. (2021). The Promises and Pitfalls of Machine Learning for Predicting Stock Returns. *The Journal of Financial Data Science, 3*(2), 21-50. https://doi.org/10.3905/jfds.2021.1.062

Muller, C., & Ward, M. (2013). Style-based effects on the Johannesburg Stock Exchange: A graphical time-series approach. *Investment Analyst Journal, 42*(77), 1-16. https://doi.org/10.1080/10293523.2013.11082552

Page, D., & Auret, C. J. (2019). Can non-momentum factor premiums explain the momentum anomaly on the JSE? An in-depth portfolio attribution analysis. *Investment Analyst Journal, 48*(1), 1-17. https://doi.org/10.1080/10293523.2018.1483792

Page, D., Britten, J., & Auret, C. J. (2015). Rand Hedge as an Investment Strategy on the JSE. *Studies in Economics and Econometrics, 39*(2), 1-19. https://doi.org/10.1080/10800379.2015.12097279

Page, D., McClelland, D., & Auret, C. J. (2022). Style rotation on the JSE. *Finance Research Letters, 46*, 102504. https://doi.org/10.1016/j.frl.2021.102504

Page, D., McClelland, D., & Auret, C. J. (2024). Machine learning style rotation–evidence from the Johannesburg Stock Exchange. *Cogent Economics & Finance, 12*(1), 2402893. https://doi.org/10.1080/23322039.2024.2402893

Strugnell, D., Gilbert, E., & Kruger, R. (2011). Beta, size and value effects on the JSE. *Investment Analyst Journal, 40*(74), 1-17. https://doi.org/10.1080/10293523.2011.11082537

Tsai, P. F., Gao, C. H., & Yuan, S. M. (2023). Stock selection using machine learning based on financial ratios. *Mathematics, 11*(23), 1-18. https://doi.org/10.3390/math11234758

van Rensburg, J., & van Vuuren, G. (2020). Evaluating investment decisions based on the business cycle: A South African sector approach. *Cogent Economics & Finance, 8*(1), 1-24. https://doi.org/10.1080/23322039.2020.1852729

van Rensburg, P. (2001). A decomposition of style-based risk on the JSE. *Investment Analyst Journal, 30*(54), 45-60. https://doi.org/10.1080/10293523.2001.11082431

van Rensburg, P., & Robertson, M. (2003). Size, price-to-earnings and beta on the JSE Securities Exchange. *Investment Analyst Journal, 32*(58), 7-16. https://doi.org/10.1080/10293523.2003.11082449

Wolff, D., & Echterling, F. (2024). Stock picking with machine learning. *Journal of Forecasting, 43*(1), 81-102. https://doi.org/10.1002/for.3021