Jakob Netskar

Machine Learning and Cross-Sectional Returns

An empirical analysis of machine learning for return prediction in the Norwegian equities market

Master's thesis in Economics and Business Administration
Supervisor: Christian Ewald

May 2023

Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Economics and Management
NTNU Business School

NTNU
Norwegian University of
Science and Technology

Jakob Netskar

Machine Learning and Cross-Sectional Returns

An empirical analysis of machine learning for return prediction in the Norwegian equities market

Master's thesis in Economics and Business Administration
Supervisor: Christian Ewald
May 2023

Norwegian University of Science and Technology
Faculty of Economics and Management
NTNU Business School

NTNU
Norwegian University of Science and Technology

Abstract

This study uses tree-based machine learning models for prediction of future stock returns of the constituents of the Oslo Stock Exchange All Share Index (OSEAX). Random Forest and Gradient Boosted Trees are measured against the benchmark as well as Logistic Regression, a less complex machine learning model extensively used in traditional econometric research. Long-short portfolios rebalanced both daily and monthly are constructed based on the predictions produced by the machine learning models. A diverse feature space is used for the monthly predictions, including established capital market anomalies found in the literature. The features for the daily predictions solely consist of pure momentum factors, utilizing lagged returns with varying intervals from the past trading year in addition to a few technical indicators. The result of the empirical research presents a nuanced picture of the usefulness machine learning models exhibit in return prediction. Backtesting the portfolios show that the daily portfolio is not able to outperform the OSEAX index after accounting for transaction costs, yielding a Sharpe ratio of 0.54, equal to that of the index. The monthly portfolio does however yield consistent excess returns, producing a mean annual Sharpe ratio of 0.68. This suggests that machine learning models utilizing a diverse feature set with a longer prediction horizon can capture information not incorporated in the stock price, thus challenging the views imposed by the efficient market hypotheses.

Contents

1. INTRODUCTION 1

1.1 PROBLEM DEFINITION 2

1.2 LITERATURE REVIEW 3

1.3 THESIS STRUCTURE 7

2. THEORETICAL FRAMEWORK 8

2.1 EFFICIENT MARKET HYPOTHESES 8

2.2 MACHINE LEARNING 9

2.2.1 Tree-Based Methods 10

2.2.2 Random Forest 11

2.2.3 Gradient Boosted Trees 12

2.3 MACHINE LEARNING AND RETURN PREDICTION 12

2.3.1 Challenges Applying Machine Learning for Return Prediction 14

3. DATA 16

3.1 DATA PREPROCESSING 17

4. METHODOLOGY 19

4.1 FEATURES 21

4.1.1 Fundamental features 22

4.1.2 Technical features 23

4.1.3 Feature Selection 24

4.2 MODELS 25

4.2.2 Regularization and hyperparameters 26

4.3 PORTFOLIO CONSTRUCTION AND TRADING SYSTEM 28

4.3.1 Trading Signals and Execution 28

4.3.2 Study Periods 29

4.3.3 Transaction Costs 30

4.4 PERFORMANCE EVALUATION 30

4.4.1 Model Performance 30

4.4.2 Financial Performance 32

5. RESULTS AND ANALYSIS 33

5.1 PREDICTION RESULTS 33

5.2 PORTFOLIO ANALYSIS 34
5.2.1 Portfolio Size and Initial Results *35*
5.2.2 After Transaction Costs *38*
5.2.3 Long/Short Analysis *42*
5.2.4 Sub-period Analysis *45*
5.2.5 Feature Importance *48*
5.2.6 Stock Analysis *52*

6. DISCUSSION 56

6.1 MACHINE LEARNING FOR RETURN PREDICTION 56

6.2 RESULTS IN RELATION TO THE EFFICIENT MARKET HYPOTHESES 57

6.3 LIMITATIONS AND FURTHER RESEARCH 58

7. CONCLUSION 59

BIBLIOGRAPHY 61

8. APPENDIX 69

List of Tables

Table 1.1: Overview over previous literature ... 6
Table 3.1: Summary statistics ... 16
Table 4.1: Overview of all models that produce predictions used for portfolio construction.. 26
Table 4.2: Overview of hyperparameters ... 27
Table 5.1: Results for daily predictions ... 33
Table 8.1: Full overview of predictor variables ... 69
Table 8.2: Statistics for daily portfolio ... 71
Table 8.3: Statistics for monthly portfolio ... 72
Table 8.4: Winners and losers for the long and short portfolios ... 73

List of Figures

Figure 1.1: Published articles on machine learning in financial journals ... 4
Figure 3.1: Index constituents by sector ... 17
Figure 5.1: Mean values for monthly and daily portfolios ... 35
Figure 5.2: Portfolio values for daily and monthly portfolios before transaction costs ... 37
Figure 5.3: Cumulative return before and after transaction costs ... 40
Figure 5.4: Net Asset Value for long and short portfolios ... 43
Figure 5.5: Feature importance for the daily portfolio ... 49
Figure 5.6: Feature importance for the monthly portfolio ... 50
Figure 5.7: Most held stocks ... 52

1. Introduction

Accurately predicting future stock returns have long been a sought-after feat for practitioners and academics alike. Financial time-series are however notoriously difficult to predict with precision. The assertions of the efficient market hypothesis also contribute to shedding doubt about whether predicting future returns can ever be done with confidence, stating that all available information already is incorporated in stock prices, thus making any attempt at predicting future stock prices fruitless. This view is however increasingly being challenged by empirical evidence suggesting that there exist anomalies that might be exploited for generating consecutive excess returns.

Recent years have seen a dramatic increase in both computational power and data availability, fostering the development of complex machine learning models as tools for stock price modeling and investment decision-making. The ability of machine learning models to capture complex and non-linear relationships in high-dimensional data justifies the investigation of the utility of these models to further improve academic empirical research and practical portfolio management. Earlier literature in the field of empirical asset pricing and return prediction have been dominated by classical econometric models such as the Capital Asset Pricing Model (CAPM), or different factor models such as the Fama-French five factor model. Time-series prediction have also been characterized by a few model configurations, such as the autoregressive integrated moving average (ARIMA) model. These models do however assume linearity in the relationship between stock returns and predictor variables, which poses constraints on available information that is able to be captured. The extension of traditional linear methods to machine learning methods might open for promising new research, improving the current understanding of stock returns and its drivers.

The contribution of this thesis to existing literature is threefold. First, different prediction horizons are analyzed with deployment of both daily and monthly predictions, thus assessing the impact time have on predictive power. Second, machine learning models are assessed against classical models previously employed in traditional econometric research. Finally, the impact of market efficiency on return predictability are investigating through conducting empirical analysis the Norwegian equity markets, which can be considered less efficient than the U.S. markets which dominates previous research.

1.1 Problem Definition

This thesis explores the utility of machine learning models for predicting future stock returns in the Norwegian equity market, considering constituents of the Oslo Stock Exchange All Share Index (OSEAX). The scope encompasses three main research questions.

1. *Can a machine learning driven investment strategy deliver consistent excess returns compared to the OSEAX index?*

2. *How does prediction horizon impact the performance of machine learning models and the subsequent financial performance?*

3. *How can non-linear machine learning models contribute to the existing status quo for return prediction that is mainly concerned with linear models?*

These research questions are deemed to give a holistic perspective of the added value that deployment of machine learning models can give to both practitioners and academics concerned with both quantitative portfolio management and empirical asset pricing. The objective of the thesis is to assess these questions through empirical analysis, exploring the role machine learning can play in future endeavors of stock return research. It also effectively serves as a test of the efficient market hypotheses in the Norwegian equities market, with an emphasis on the difference in market efficiency between more liquid stock markets and the Norwegian market.

1.2 Literature Review

The recent popularity spike for machine learning would suggest that research on such methods applied to stock returns is relatively new. However, research in this field have been around for several years. One of the earliest examples is (Schöneburg, 1990), who analyzed the possibility of predicting intra-day stock prices for German stocks using neural networks fed with price data. (Kryzanowski et al., 1993) extended the use of neural networks by classifying one-year-ahead returns to be positive or negative using macroeconomic and stock specific financial data. (K. Kim, 2003) introduced support vector machines to the problem of stock price prediction, attempting to predict the future direction of the Korea composite stock price index using a variety of technical indicators as features.

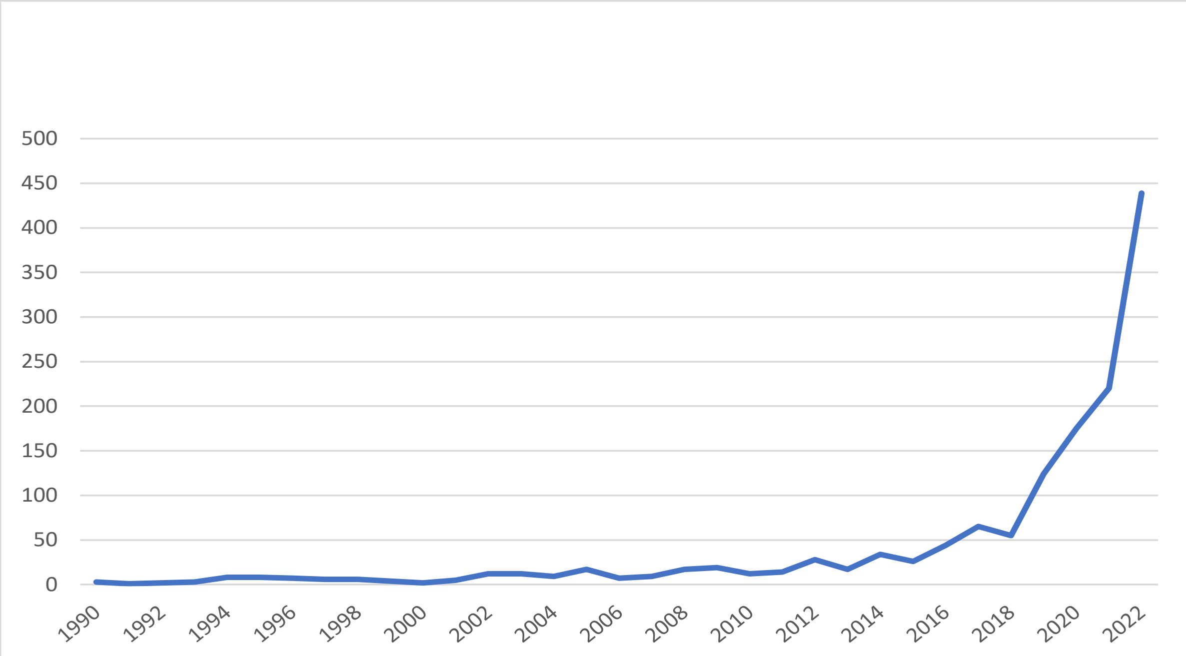
Clearly, this has been a research area of interest for several years. However, earlier studies in this field have been heavily constrained by computational power and data availability. But as computational capacity is made more available for every year that passes, these limiting factors become less of an obstacle. This, as well as the increase of availability and focus on alternative data such as in (Rechenthin et al., 2013), (Junqué de Fortuny et al., 2014) and (Kumar & Ravi, 2016) has given an exponential rise to the number of papers published on the topic. This is illustrated in figure 1.1, highlighting the number of papers published in a selected number of financial journals[1] over the years.

(Gu et al., 2020) studies the use of machine learning for return prediction using a large set of fundamental and macroeconomic variables, with an emphasis on the significance of their results relative to current empirical asset pricing literature. They compare multiple models, with neural networks and regression trees having the best performance. They also find that shallow neural networks perform better than deep neural networks and argue that can be due to the low signal-to-noise ratio found in asset pricing problems. Their study investigates a machine learning approach whilst considering classical models and find that ML-models vastly outperforms ordinary least squares (OLS) models for prediction. They also discuss how machine learning problems can be useful for inference as well as prediction and emphasize this by showing that

[1] Journal of Financial Economics, The Review of Financial Studies, The Journal of Finance, Finance Research Letters, Journal of Corporate Finance, Journal of Banking & Finance, Research in International Business and Finance, Journal of Financial and Quantitative Analysis, International Review of Financial Analysis, Journal of Accounting and Economics, Journal of Financial Data Science, Journal of Portfolio Management.

the ML- models all agreed on a small set of important predictor variables that are well established in financial research, namely momentum, liquidity, volatility, and valuation ratios.

Figure 1.1: Published articles on machine learning in financial journals



Note: Number of studies published in a selected set of financial journals with the keywords “artificial intelligence”, “deep learning”, or “machine learning” in the title.

(Moritz & Zimmermann, 2016) also investigates the usefulness of machine learning methods for developing established financial literature. They expand on techniques like portfolio sorts and Fama-MacBeth regressions by developing a strategy based on “deep conditional portfolio sorts”. Their features consist of decile ranks based on one-month return for all 24 months prior to portfolio construction. A random forest model is then used for predicting the year-ahead returns of the stocks in the portfolio. Regressed on a four-factor model from (Carhart, 1997), this strategy yields an excess monthly return of 2%.

(Bao et al., 2017) studies the possibility for an LSTM model, which is a type of recurrent neural network with the unique ability of capturing long-term dependencies in sequential data, to predict the next day’s closing price for stocks using price history and macroeconomic variables. Their results shows that predictive performance is better in less developed markets and find that there is less predictability in the S&P500 index compared to the Indian Nifty 50 index. (Hao et al., 2023) also investigates predictability for different markets. They use a classification framework by predicting whether index constituents in the US, the UK, China, Canada, and Japan outperform the cross-sectional median return of their respective indices. Out of the models they deployed the portfolio based on predictions from the deep neural network yield the best results with an annualized return ranging from 27.9% and 255.17% before transaction costs between the different markets. They find that the financial performance is vastly better in the Chinese market compared with the US and UK.

(Krauss et al., 2017) use deep neural networks, random forest, and gradient boosted trees for predicting directional movements of future stock returns. They use lagged returns as input features and a survivorship-bias free sample of constituents of the S&P500. They find that an ensemble consisting of one set of prediction from each model performs the best, with out-of-sample daily returns of 0.45% before transaction costs. They also report that the models show better performance during times with high market volatility, but that predictability, and hence returns, seemed to decrease after the year of 2000. (Fischer & Krauss, 2018) directly builds upon this study by deploying an LSTM model. They report a daily return of 0.46% and a Sharpe ratio of 5.8 prior to transaction costs. They also report findings of declining results in recent times, with a clear shift from 2010 and outwards. In analyzing this phenomenon, they find that stocks with high volatility and short-term reversal of returns are common pattern of stocks selected for trading.

Other studies have had a focus on feature selection and its consequences for performance. (Huck, 2019a) includes over 600 predictors in their models and concludes that, despite the machine learning method's flexibility and ability to generalize to high dimensional data, adding more predictors is not necessarily positive for model performance due to increased noise and risk of overfitting. (Carta et al., 2022a) has a significant focus on feature selection, in the end training 432 models with different feature sets based on combination of features, as well as stock- and sector specific features. They find that the extensive feature selection process improves both predictive and financial performance.

Several other studies have applied machine learning to return- or stock-price prediction. Table 1.1 below highlight some notable newer works.

Table 1.1: Overview over previous literature

Author	Universe	Models	Horizon	Features	Prediction
Hao et al., 2023	US, UK, Asia	DNN, RF, GBT	Daily	LR	Classification
Carta et al., 2022	S&P500	RF, GBT, SVM	Daily	LR, TI	Classification
H. Wang et al., 2021	US	RF, GBM, SVR, RNN	10-year	Macro	Regression
Flori & Regoli, 2021	S&P500	LSTM	1, 5, 10 days	LR, TI	Classification
Gu et al., 2020	US	OLS, PLS, GLM, RF, NN	Yearly	F, LR, Macro	Regression
Nikou et al., 2019	UK	NN, SVM, RF, LSTM	Daily	LR	Classification
Basak et al., 2019	US	RF, GBT	Multiple	TI	Classification
Rasekhschaffe et al., 2019	International	GBT, NN, SVM	Monthly	F	Classification
Huck, 2019	S&P500	NN, RF, Elastic Net	1 and 5 days	LR, Macro	Classification
Weng et al., 2018	US	GBT, SVR, RF, NN	1-10 days	TI, AD	Regression
Henrique et al., 2018	Brazil, US, Asia	SVR	Daily	TI	Regression
Fischer et al, 2018	S&P500	LSTM, RF, LR	Daily	LR	Classification
Chen & Hao, 2017	China	SVM, KNN	1, 5, 10, 15, 20, 30 days	TI	Both
Bao et al., 2017	International	LSTM	Daily	TI, Macro	Regression
Krauss et al., 2017a	S&P500	DNN, GBT, RF	Daily	LR	Classification
Moritz et al., 2016	US	RF	Daily	LR, F	Regression

Note: DNN = Deep Neural Network, RF = Random Forest, GBT = Gradient Boosted Trees, SVM = Support Vector Machines, RNN = Recurrent Neural Network, LSTM = Long Short Term Memory, OLS = Ordinary Least Squares, PLS = Partial Least Squares, GLM = Generalized Linear Model, NN = Neural Network, SVR = Support Vector Regression, LR = Logistic Regression, KNN = k-Nearest Neighbour. LR = Lagged Returns, TI = Technical Indicators, F = Fundamentals

1.3 Thesis Structure

The thesis is structured as follows: Chapter 2 will review the theoretical foundation needed for the analysis and subsequent discussion. Chapter 3 presents the data and how it was collected and preprocessed. Chapter 4 contains a detailed analysis of the methodology applied in this study. In chapter 5 the results and analysis are presented, first analyzing predictive performance before studying the financial performance of the constructed portfolios. Chapter 6 provide discussions regarding the results in relation to the efficient market hypothesis, as well as limitations in backtesting.

2. Theoretical Framework

2.1 Efficient Market Hypotheses

The Efficient Market Hypothesis (EMH) has been a fundamental concept in economic theory ever since Eugene Fama's influential paper "Efficient Capital Markets" was published in 1970 (Fama, 1970). The theory outlines the ability for financial markets to reflect all currently available, and relevant, information in security prices. This suggests that markets are "efficient" and that all prices can be considered fair. According to the EMH only new information can move prices. Since it is unknown when this new information is available it is therefore unpredictable, and thus will lead to unpredictable and random fluctuations in stock prices. This consequentially leads to that both technical- and fundamental analysis, two main frameworks for stock price prediction, is considered fruitless (Bodie et al., 2018).

The EFM includes three forms of market efficiency: weak, semi-strong and strong. The weak form suggests that security prices reflect all available market information such as historical prices and trading volume, thus rendering technical analysis useless for prediction. The semi-strong form builds on the weak form and adds that all publicly available information is reflected in the price. This includes stock specific information such as financial ratios or macroeconomic variables. This form of market efficiency suggests that neither technical nor fundamental analysis bear any meaningful merit. The last and most extreme form of market efficiency suggests that stock prices reflect all available information, including insider information (Bodie et al., 2018).

Fama published a revised paper on the efficient market hypotheses in 1991, this time defining weak market efficiency as the concept of prices reflecting information up to the point where marginal benefits of acting on that information do not exceed the marginal costs (Fama, 1991). At this time the dominance of the original theory presented in 1970 was far less prominent, as many financial researchers began publishing studies on the predictability of stock returns relating to psychological and behavioral elements, as well as stock specific anomalies such as financial ratios (Malkiel, 2003). Numerous other studies challenge the notion of the efficient market hypotheses. (Campbell et al., 1998) found that there exists short-term momentum in stock prices, rejecting the hypothesis that stock prices behave as true random walks. (Fama & French, 1988a) and (Poterba & Summers, 1988) provides evidence of long term mean reversion in stock markets, suggesting that variation in returns over long holding periods can be predicted by considering the negative correlation with past returns. Predictability of future returns have also been explored using fundamental information, such as in (Fama & French, 1988b) where it is suggested that a large part of the variance in future returns can be explained by dividend yields.

2.2 Machine Learning

Machine learning is a tool in the statistical learning toolbox generally used to provide accurate predictions by utilizing historical data. Machine learning is based on *algorithms* that are effective at handling big data sets and that can produce accurate predictions (Foundations of Machine Learning, 2018). In general, machine learning algorithms are concerned with estimating a relationship, or mapping a function, f between a quantitative response variable Y and a vector of p predictors $\mathbf{X} = (X_1, X_2, \ldots, X_p)$. This relationship can then be written as:

$$Y = f(x) + e \tag{1}$$

Where f is an unknown function of $\mathbf{X}$ and ϵ is a random error term. This error term has a mean of zero, and is uncorrelated with $\mathbf{X}$. There are two main areas of benefit that can be achieved by estimating f, namely *inference* and *prediction* (James et al., 2021).

Prediction tasks estimates Y based on a set of predictors $\mathbf{X}$ as such:

$$\hat{y} = \hat{f}(x) \tag{2}$$

where $\hat{f}$ is the estimate of the true relationship between Y and $\mathbf{X}$, and $\hat{y}$ is the prediction based on this estimate. The estimated function f often yields a black box, where one is not too concerned with the form of the function if it provides accurate predictions for Y. This accuracy is dependent on two numbers, namely the reducible and irreducible error. Since the true relationship is only estimated, some error is bound to exist. This error is reducible since there are steps that can be taken to reduce this quantity, for example choosing the correct model. However, even if $\hat{f}$ were to estimate f perfectly, there would still be some error present because Y is also a function of ϵ, which cannot be predicted using **X**. This quantity represents the irreducible error. Consider that both $\hat{f}$ and **X** are fixed, so that the only source of variability is produced from ϵ. Then:

$$E[(Y - \hat{Y})^2] = E[(f(X) + e - \hat{f}(X))]^2 = [f(x) - \hat{f}(x)]^2 + Var(e) \quad (3)$$

Where $E[(Y - \hat{Y})^2]$ is the expected value of the squared difference between the actual value of Y and the estimated value of Y. The $Var(e)$ represents the irreducible error, while $f(x) - \hat{f}(x)$ represents the reducible error. Machine learning models generally estimate f with the goal of minimizing the reducible error (James et al., 2021)

Inference is mainly concerned with dissecting the relationship between Y and the predictor variables **X,** while not necessarily trying to produce a prediction for Y. The estimated function $\hat{f}$ cannot be treated as a black box in this case, as its form is clinical for the task of inference. Applying machine learning methods for inference purposes can shed light on important questions, such as which variables provide predictive power over Y, or if the relationship between Y and **X** can be explained using linear models, or if more advanced models are needed (James et al., 2021).

Machine learning can be split into two domains: supervised learning and unsupervised learning. During supervised learning the model is trained on a labelled dataset, so that the Y is known for each observation. This contrasts with unsupervised learning where there is no Y label being utilized in training, leaving the model to figure out relationships and structure in the data on its own (James et al., 2021). Supervised learning can further be split into two categories, namely regression and classification. Regression is concerned with predicting an exact quantitative number, while classification is used when Y is categorical rather than numerical, thus the prediction task becomes assigning an observation to a category (James et al., 2021).

2.2.1 Tree-Based Methods

Tree-based methods are machine learning algorithms with their main trait being that they segment the predictor space into several simpler sub-groups. Usually, the mean or mode response value for the training observations in the sub-group that a given observation belongs to is used for producing predictions. The rules that constitute the process of splitting the predictor space can explained with a tree- structure and are thus also known as decision trees. Decision trees can be used in both regression and classification, with the only difference being the output, taking a numerical or categorical value respectively. For each observation in a classification tree the prediction is based upon the most commonly occurring class of training observations in the sub-group that particular observation belongs. The classification tree is built by using binary splitting, guided by *entropy* or the *Gini index*. Entropy can be defined as:

$$D = -\sum_{k=1}^{K} \hat{p}_{mk} \log \hat{p}_{mk} \tag{4}$$

Where $\hat{p}_{mk}$ represents the proportion of the training observations in the mth sub-group that are from the kth class. The entropy will adopt a small value if the mth node is small, because it can be shown that if the different values of $\hat{p}_{mk}$ are near 0 or 1, the entropy will converge to 0. This measure is used to guide the splitting, evaluating the performance of each split (James et al., 2021).

2.2.2 Random Forest

Decision trees generally experience high variance, meaning that applying a decision tree to two randomly split parts of a training set can yield very different results. This is an unfavorable trait, that can be mitigated through the process of *bagging*. This process is based on taking repeated samples from the training set, fitting a model to each sample, and then averaging the results as such:

$$\hat{f}_{\text{bag}}(x) = \frac{1}{B} \sum_{b=1}^{B} \hat{f}^{*b}(x) \tag{5}$$

While bagged trees increase the performance from decision trees, the random forest algorithm introduces *decorrelation* of the trees, further increasing performance. Random forest builds decision trees like you would do in bagging, but during the building process of these decision trees the random forest algorithm takes a random sample of m predictors from the full set of p

predictors at each split. Assuming a dataset with a few strong predictors while the rest show moderate predicting power, all the trees in the bagged collection of trees would utilize the strong predictors in the top split, which results in all trees being similar. By only considering a subset of the variables at each split, the trees will contain different sets of variables, thus decorrelating the trees. This leads to the average produced by the different trees to exhibit less variance, making the result more robust.

2.2.3 Gradient Boosted Trees

Boosting is another technique for improving the performance of tree-based models. While bagging was based on taking repeated samples of the training set, fitting different decision trees and then combining all the trees, boosting grows the tree sequentially. This means that the trees are constructed one after the other, with each tree utilizing information from past trees. Boosted trees does not use the bagging technique, but rather fit each decision tree on a modified training set. Gradient Boosted Trees uses the *gradient descent* to minimize the loss function each time a new tree is constructed, hence the name Gradient Boosted Trees.

2.3 Machine Learning and Return Prediction

Research on stock price behavior have been a fundamental part of the financial science for many decades. (Fama F, 1965) famously studied the random walk characteristics of price behavior, followed up by the introduction of the efficient market hypothesis in 1970 (Malkiel & Fama, 1970). This research, that has been considered a cornerstone of financial theory since its publication, suggests that forecasting or predicting future asset prices is an impossible task. However, contrary to the efficient market hypothesis, numerous studies have shown that stock markets are predictable to some extent. Time series prediction have seen prominent use of parametric statistical models, such as the autoregressive moving average (ARMA) and the autoregressive integrated moving average (ARIMA) (Box et al., 2015). Several studies have been published showing that the autoregressive framework is plausible for short-term prediction tasks, such as in (Ariyo et al., 2014) and (Virtanen & Yli-Olli, 1987b).

Literature on stock level return prediction is traditionally concerned with models that explains differences in returns between stocks based on certain characteristics of these stocks. This approach usually involves predicting future returns based on few historical stock characteristics using regression methods, like in (Lewellen, 2014). Traditional approaches like these are characterized by constraints that ML-methods can potentially mitigate, with the most important constraint being the ability to handle the vast amount of predictor variables that have shown up in the literature over the years (Gu et al., 2020). Asset pricing literature suggests a plethora of economic variables that affects returns, with the number of stock-level and macroeconomic characteristics reaching the hundreds (Harvey et al., 2016). Machine learning can add additional value and perform better than linear models in situations where high-dimensional data is present.

Linear models are prone to overfitting when the number of variables is large in relation to the number of observations and perform poorly when the variables are closely correlated (Gu et al., 2020). The practitioner is therefore forced to only include a small subset of the variables potentially excluding vast amounts of conditional information and relying on the practitioner's domain expertise to construct the functional form and identity relevant variables. This is unlike ML- methods that are much more robust to overfitting and do not require a pre-specified functional form, which allows for flexibility that can discover true but complex relationships that would be ignored in a linear model (H. Wang et al., 2021). Linear models on the other hand assume a linear relationship between variables, which often is not the case (Hoang & Wiegratz, 2022).

While prediction problems are prominent in econometric research, it generally solves a different problem than machine learning. In economic analysis, the values of interest are usually the parameters β that explains relationships between independent and dependent variables, making econometrics more concerned with *inference* rather than *prediction*, which is the main difference between machine learning and econometrics (Bzdok et al., 2018; Gu et al., 2020). While ML- methods can generate these parameters, they are usually inconsistent. Using a method meant for producing a prediction $\hat{y}$, and inferring information from their $\hat{\beta}$ as you would in econometric models, can in many cases lead to misleading conclusions. Thus, machine learning is more useful in applications concerned with $\hat{y}$, rather than $\hat{\beta}$ (Mullainathan & Spiess, 2017).

The importance of inference in econometric research is exemplified in empirical asset pricing where the main goal is to deduce the behavior of risk premiums. This is because even if future returns were completely observable, the behavior would still need to be explained by financial theory and tested empirically. Return variation is however often highly influenced by unforecastable news due to market efficiency, which makes it incredibly difficult to measure (Gu et al., 2020). ML- methods might excel traditional approaches in this context, due to both its predicting power and ability to extract valuable information about the predictor variables. This can aid in the problem of measuring risk premiums, and by extension enable further investigation into the cross section of asset pricing and economic mechanisms. Measuring risk premium involves approximating the equation:

$$E(r_{i,t+1}|F_t) \tag{6}$$

Where:

$r_{i,t+1}$: $Return\ in\ excess\ of\ risk - free\ rate\ for\ asset\ i$

$\mathcal{F}_t$: $True\ and\ unobservable\ information\ set\ of\ market\ participants$

ML- models might perform approve upon this approximation, thus better explaining expected return behavior compared to linear models, through penalization and dimension reduction, which makes ML-methods useful nonetheless for understanding relationships (Gu et al., 2020).

2.3.1 Challenges Applying Machine Learning for Return Prediction

One would believe that the many incredible tasks machine learning have mastered in different fields should imply the possibility for it to succeed in financial applications like return prediction. There are however several characteristics of financial data makes it notoriously challenging to work with (Israel et al., 2020). The first challenge of applying machine learning in the field of finance is the lack of *big data*. Israel et al. consider the regression in equation (7) to highlight the issue of *small data* for return prediction tasks.

$$y_t = \sum_{i=1}^{N} \beta_i x_{i,t-1}, \quad t = 1, ..., T \tag{7}$$

This simple regression is trying to predict future returns y_t with the help of $(x_{1,t-1},\ldots,x_{N,t-1})$ predictors. The authors argue that the limiting factor for return prediction never have been the number of predictor variables *N*, but rather the number of independent return observations, *T*. Big data is often thought about in conjunction with the number of predictors, and subsequent inclusion of variables that span from price data to satellite images. However, without a big enough sample of observations, the model is effectively constrained to be small. Assuming a monthly holding period for a few thousand equities that have data ranging from a few months to a few decades, the total number of observations for cross-sectional returns amount to a few hundred thousand. Relative to many other fields, this is a small number, and is even smaller when accounting for cross-sectional correlations. The only way to increase the size of this dataset, without generating synthetic data, is to wait, which differs from other fields where big data is readily available or can be generated synthetically through experiments (Israel et al., 2020).

The second challenge for predicting returns is the low signal-to-noise ratio. This ratio is important for the success of machine learning applications, and essentially describes the level of predictability that resides within a system. The low signal-to-noise ratio of the financial markets is due to the constant pressure of the economic forces of profit maximization, which serves as the basis for the efficient market hypothesis (Israel et al., 2020).

Practitioners that seek to predict future returns is also concerned with understanding how this prediction came to be. Machine learning models are infamous for acting as black boxes, but for many asset managers it is important to be able to interpret the model that is being used (Gu et al., 2020). It is even stated in the EU General Data Protection Regulation (GDPR) that decision-making systems that are automated should retain meaningful information regarding the logic of the system, as well as the impacts from these decisions (Goddard, 2017) . This means that practitioners of machine learning driven systems should be able to derive meaningful information about the underlying logic of decision- making. Practical use of machine learning is also constrained by the concern that the model commits fatal logical mistakes, such as loading up on concentrated sources of risk. ML- driven investment strategies are therefore less useful if they are not interpretable, and generally not suited for problems where it is necessary with an understanding of the economic mechanisms (Li et al., 2022). This concern has given rise to development in research regarding so called “explainable artificial intelligence”, as seen in (Adadi & Berrada, 2018; Gunning et al., 2019; Molnar et al., 2020) for the general domain, and

(Bussmann et al., 2021; Carta et al., 2022a) for financial applications. This subfield of AI and ML research seeks to control, discover, justify, and improve machine learning tasks, and is an important initiative for the widespread use of ML in finance (Carta et al., 2022a).

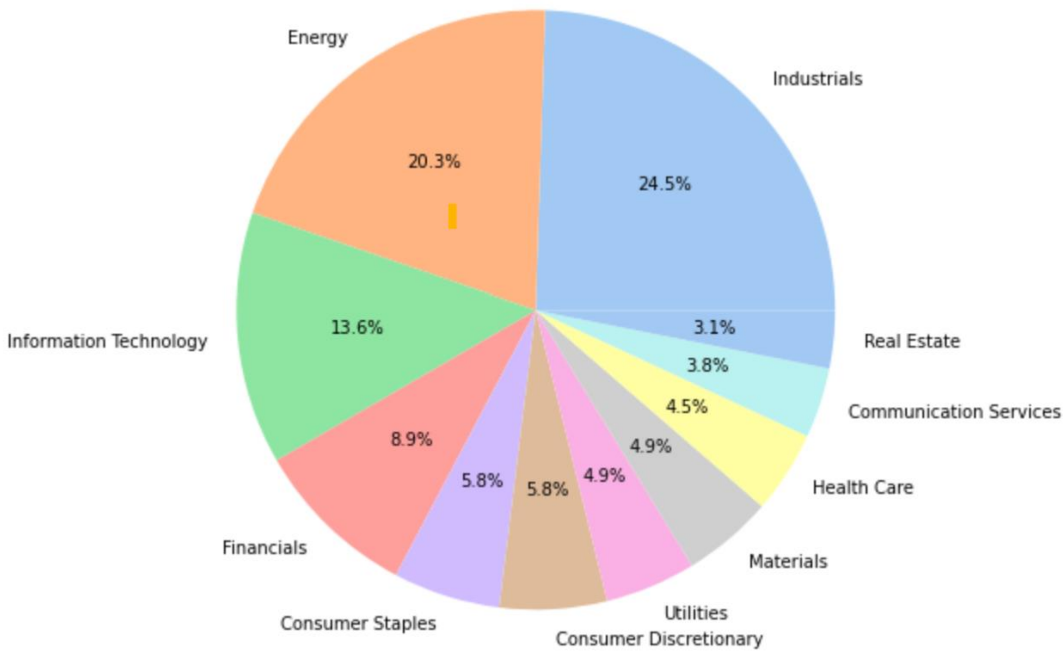
3. Data

This thesis uses equity data from constituents of the Oslo Stock Exchange All Share Index (OSEAX) from April 1999 to 31 December 2022. The noisy environment and continuity of financial time series demands the need to facilitate for as much historical memory as possible (Chourmouziadis & Chatzoglou, 2016), and ideally data should be collected from as far back as possible. This study collects daily and monthly data from 1999 since data before this year show clear signs of diminishing quality and missing values. The dataset includes daily observations of prices, trading volume, and shares outstanding, and monthly stock specific financial data from Thomson Reuters Datastream. Macroeconomic data like oil price, consumer price index and exchange rates are sourced from Bloomberg. The risk- free rate is collected from Norges Bank. Preprocessing of data and fitting of machine learning models are done with the Python programming language and the scikit-learn library.

Table 3.1: Summary statistics

Summary Statistics Investment Universe	
Min. number of constituents	153
Max. number of constituents	251
Total number of constituents	414
Annualized return	13.49%
Annualized standard deviation	18.61%

Figure 3.1: Index constituents by sector



Note: Based on the Global Industry Classification Standard categorization for each stock, collected from Bloomberg.

3.1 Data Preprocessing

Common challenges working with financial data sets includes cleaning the data set in a consistent and robust way, handling missing values, and structuring the data such that predictive performance is maximized. This thesis utilizes cross-sectional times-series data, and one should therefore check for irregularities such as outliers and corporate actions (Kelliher, 2022). Corporate actions such as mergers and acquisitions, dividends, and stock splits might have an abnormal impact on the return of a stock. A stock might for example double its return overnight due to a reverse stock split, or dividends might play a big role in inflating returns such that the underlying attractiveness of the company is not fully reflected (Kelliher, 2022). Stock prices are therefore adjusted for corporate events, as this makes them more suitable for prediction tasks and return calculations. Further, in line with (Hou et al., 2011), monthly returns that are reported to be over 300% and reversed the next month are removed. Since Datastream repeats the last valid entry, all stocks that reports a monthly return of 0% twice in a row for the monthly predictions are removed, while the threshold is set to four consecutive trading days of 0% returns for the daily predictions.

Constituents of an index will evolve over time, and companies are included or discarded for different reasons such as bankruptcy or mergers. Considering only the current constituents of the OSEAX would lead to so-called survivorship bias, which is based on the notion that current constituents are likely to have been historically successful. Excluding companies that have been removed from the investment universe might lead to inflated returns. To mitigate this, data is downloaded for all stocks that have ever been part of the OSEAX.

Another bias that can affect backtesting performance is look-ahead bias, which means using information that would not yet be available at that point in time. Stock specific financial information are thus lagged by a period corresponding to its reporting frequency. For example, accounting ratios that are reported annually are lagged by one year. Furthermore, since the publication of macroeconomic data also is associated with a lag, this thesis follows (Qi & Maddala, 1999) by lagging macroeconomic variables by two months.

Following (Tobek & Hronec, 2021a), liquidity filters are applied to reduce trading friction and to avoid high transactions costs, as well as to avoid micro-cap stocks that would be impossible to short. Considering the difference in liquidity and efficiency between the OSEAX and the S&P500, where most studies are conducted, this is deemed an important step and will help remove the impact of market microstructure on the results. We exclude stocks that fall below a market capitalization threshold, which is set to be the 20th percentile of the mean market capitalization across all stocks for that specific period. We further exclude stocks that exhibits low trading activity by following the same approach as with market capitalization, this time sorting by turnover. A stricter volume threshold is applied for the short portion of the portfolio to ensure lending availability and thus replicating a real-life scenario more accurately. Lastly, a stock is excluded if the trade price falls below 10 NOK. These filters aid in creating a liquid universe of stocks, which further mitigates challenges regarding short-sale constraints, non-viable trades, and high transaction costs. By constraining the universe to the most traded companies the desired characteristics of a less developed market are kept, while also retaining a viable level of market friction that can serve as a foundation for empirical research.

The models deployed in this study are not able to handle missing values, and thus stocks with missing values must either be excluded from the set of tradable stocks, or the values must be filled. Since one would like to preserve as much data as possible for the models, missing values are generally filled by forward interpolation. The primary motivation for using interpolation is its simplicity, as well as its suitability for filling small gaps in the data (quant python). When larger periods of missing values are present, in this case 10 consecutive observations, the stock is removed entirely. Estimates used for filling gaps only considers historical data, so that look-ahead bias is avoided.

Lastly, the observations are standardized as follows:

$$x_t^i = \frac{(x_t^i - \mu_{\text{train}}^i)}{\sigma_{\text{train}}^i}, \quad i \in \{1, 2, ..., n\} \tag{8}$$

Where *i* is a particular stock, *n* is the number of elements in the set of tradeable stocks, and *train* denotes that only training data is used to avoid look-ahead bias. The data consists of features that vary greatly in numerical size, both within the feature itself and with other features. This difference in scale can have a large impact on the prediction and might lead to a few features having a dominating influence on the results. By standardizing the variables to have a mean of 0 and a standard deviation of 1, the variables will be on a comparable scale, ensuring adequate robustness of the predictions.

4. Methodology

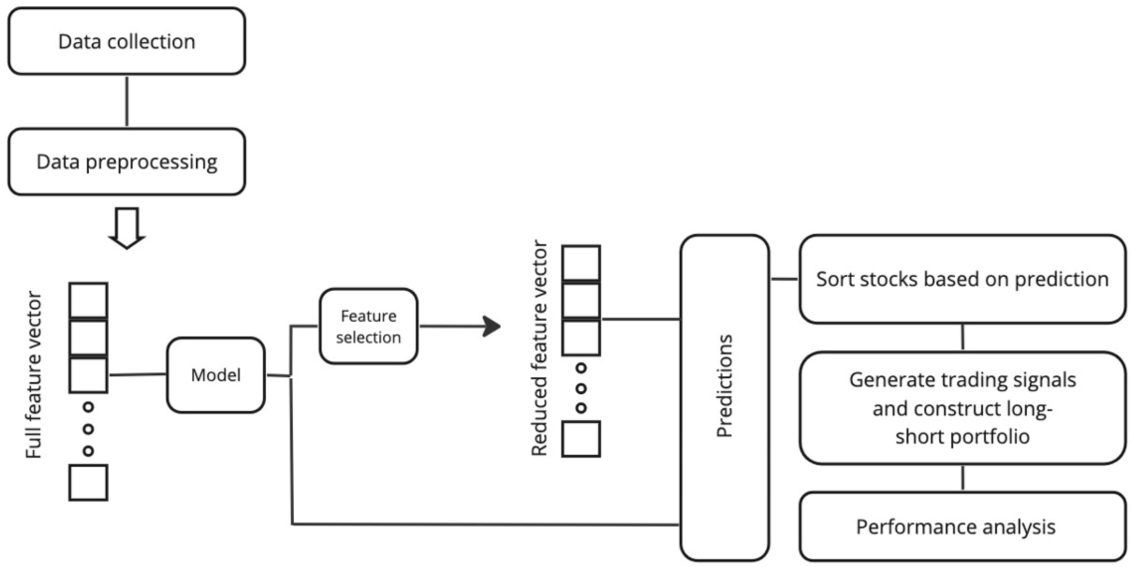
Research on machine learning driven trading strategies in the Norwegian market is scarce. Most empirical applications are done on highly liquid and efficient markets such as the constituents of the S&P500 index. The term “academic home bias puzzle” coined by (Andrew Karolyi, 2016) denotes that only 16% of financial research on asset pricing literature in the main finance research journals is conducted in markets outside of the US. While being a highly developed market, the Norwegian market is less mature than other markets such as in the US and UK. The Norwegian market is also affected by other factors than other highly efficient markets (Grobys & Huhta-Halkola, 2019) and is therefore an interesting entity to apply machine learning methods to.

This thesis also investigates the implication of different prediction horizons, with both daily and monthly predictions. Two different portfolios are thus studied, one that rebalances every month and one that rebalances every day. (Lewellen, 2014) states that the accuracy of predictions generally will diminish the further the forecasting horizon. This can be attributed to the efficient market hypotheses, which states that stock prices need time for including all available information (Malkiel & Fama, 1970). This thesis applies two different sets of variables, one for daily predictions and one for monthly predictions. The daily predictions use lagged returns and technical indicators, while the monthly predictions use a range of features from fundamentals and momentum to macroeconomic variables.

In short, the procedure for prediction and portfolio construction follows four steps. The primary step is to split the full dataset into training- and trading sets for different study periods, following a walk-forward approach. The training set is used for in-sample training and tuning of the different models, while the trading set is used for out-of-sample predictions. The trained models then produce trading signals and long-short portfolios are constructed. (Fischer & Krauss, 2018; Huck, 2019; Krauss et al., 2017) approach the prediction task as a classification problem. Additional support from (Enke & Thawornwong, 2005) and (Leung et al., 2000) has served as motivation for this thesis to also adopt this approach. The binary response variable is constructed by taking a value of 1 if a specific stock is predicted to outperform the OSEAX index at the rebalancing date, or 0 otherwise.

This method is repeated for both daily and monthly predictions, except for the reduced feature vector step showed in figure 4.1, which only applies for monthly predictions. This is because the monthly feature vector has a length of 48, while the daily feature vector has a length of 13, making feature space reduction redundant. The following sub-chapters will go into greater depth of the features and models used for prediction, and will explain the trading setup, portfolio construction and performance analysis in more detail.

Figure 4.1: Overview of the methodology



4.1 Features

Choosing the correct type of features, as well as the number of features, is highly influential for predictive performance. The recent increase in available data with some degree of relevancy to the financial markets has resulted in a notably large potential feature space. Selecting features that are grounded in economic theory can increase the signal-to-noise ratio, thus only features that are related to future stock returns should be included in the model, as this will reduce overfitting and minimize runtime (Rasekhschaffe & Jones, 2019). In most previous studies using ML for return prediction the feature space consists of factors related to price data such as lagged returns and technical indicators (Fischer & Krauss, 2018; Huck, 2019a). Other studies include factors related to company specific financial ratios and the macroeconomic environment (Moritz & Zimmermann, 2016; H. Wang et al., 2021), or a combination of all (Gu et al., 2020). Some studies also include alternative data such as textual data (Junqué de Fortuny et al., 2014). This wide array of possible features to include might seem daunting, but a great advantage to ML-models is their ability to handle collinear data. Thus, many features can still be included without specifically knowing which ones have the greatest predictive power.

The different features making up the feature space in this thesis can be categorized into three categories. The first is data based on historical stock price movements and trading volume. The other category consists of fundamental features. These are data points related to stock specific financial information, as well as macroeconomic measures. The last category are other variables such as day-of-the-week effect and an industry dummy classifier. In addition to what features have shown predictive power in previous literature, the decision of what features to include was also dependent on data availability.

4.1.1 Fundamental features

Research on asset pricing models have over the years identified different several hundred patterns in average stock returns (Feng et al., 2020). (Harvey et al., 2016) outlines 316 of them. It can therefore be a challenge to choose which features to include in the models. Information related to the financial status quo of a company, like profitability and cash-related measures, are often used to calculate the intrinsic equity value of a company, as well as to predict the cross section of average returns (Foerster et al., 2017). Some notable stock specific characteristics that seem to exhibit a relationship with returns are size and value. (Banz, 1981) show the relationship between market cap and returns, (Basu, 1983) finds that book-to-market equity exhibits a relationship with average returns, while (Lakonishok et al., 1994; Rosenberg et al., 1985) discovers that earnings over price, cash flow over price and past sales growth is related to average return. (Fama & French, 2006) reveals that profitable companies tend to have higher returns, while (Fama & French, 2018; Green et al., 2013) provides further insights into factors that may provide valuable information for future cross-sectional returns.

This study includes 24 stock specific characteristics that are well documented in financial literature and that have performed well in newer studies such as in (Feng et al., 2020), as well as important features extracted using feature importance techniques, such as in (Gu et al., 2020). Variables representing stock specific information such as size, growth, valuation, and other financial metrics are included.

The relationship between returns and the macroeconomic environment has also been extensively studied in the literature, e.g., in (Fama, 1981; Fama & Schwert, 1977; Kaul, 1987; Lee, 1992). The holistic behavior of the stock market is significantly influenced by macroeconomic variables, as they impact future investment opportunities and consumption patterns (S.-S. Chen, 2009). (Ratanapakorn & Sharma, 2007) finds that stock prices have an inverse relationship to long-term interest rates, but a positive relation to short-term interest rates and industrial production, while (S.-S. Chen, 2009) finds that yield- curve spreads and inflation rates have predictive properties for predicting recessions in the U.S. stock market. A study performed by (Gjerde & Sættem, 1999) suggests that Norwegian markets responds to oil prices and real interest rates. The subspace of this study's feature space consisting of macroeconomic variables include 10 variables outlined in (Gjerde & Sættem, 1999; Welch & Goyal, 2008), such as inflation, short- term interest rates, and oil price.

The fundamental features are used for the monthly return predictions. They are not used for the daily predictions, as it is believed that these features hold very little predictive power in short-term forecasting. They are thus excluded from the short-term model to reduce noise.

4.1.2 Technical features

Technical features are in this thesis regarded as information related to historical price movements, as well as trading volume. A much-debated topic in financial literature is the usefulness and predictable power of historical price data and technical indicators. Momentum factors are perhaps the features with the most supporting empirical evidence. Such factors are presented in (De Bondt & Thaler, 1985; Jegadeesh & Titman, 1993). The former finds that stocks with low long-term past returns exhibit higher future returns, while the latter argues that short-term trend in return tend to continue and that stocks with higher return in the past year show higher future returns. Empirical evidence of the effect of momentum on stock prices are also presented in (Carhart, 1997) and (Asness et al., 2013). (Fischer & Krauss, 2018) creates a momentum space consisting of lagged returns for the previous 20 trading days, and then the previous 11 months with a 20-day interval. This was used as input features in a machine learning driven trading system and was shown to significantly improve annual returns.

Technical indicators are measures made by applying transformations to historical price and volume data to create more informative variables (Hsu et al., 2016). Their usefulness and predictive power for forecasting returns are heavily debated in the literature. (Lo et al., 2000) finds some technical indicators with predictive power, while (Brock et al., 1992) finds that a strategy based on technical indicators can produce excess returns. Studies by (Malkiel & Fama, 1970) and (Lesmond et al., 2004) dispute this, and argue that technical indicators bear little predictive power. Despite this, many papers concerned with return forecasting using machine learning rely on either raw price data, technical indicators, or both. In contrast to traditional linear models, ML- models have the ability of dissecting nonlinear relationships and detecting complex variable interactions. Since ML-models are not common in financial literature, the predictive power of technical information might be underestimated (Hsu et al., 2016). Thus, this study includes 10 variables related to momentum, volatility, and volume.

Following (Fischer & Krauss, 2018), a lagged return feature space is also included for the short- term model. We define lagged return as:

$$LR_{t-j} = \frac{(close_{t-1} - open_{t-j})}{open_{t-j}}, \quad j \in \{1, 2, 3, 4, 5, 21, 63, 126, 252\} \tag{9}$$

Where j denotes the number of days of which return is calculated. The purpose of this is to provide additional information about the return behavior in the recent past, i.e., 1-5 trading days, and the medium and long- term behavior, with {21, 61} and {126, 252} trading days respectively. In addition to lagged returns, 4 technical indicators are included. The full set of features and which prediction horizon they are used for are provided in table 8.1 in the appendix.

4.1.3 Feature Selection

The use of feature space reduction techniques in conjunction with machine learning models are popular in previous studies. These techniques have as their main goal to reduce the dimension of an $n \times p$ data matrix X, where n is the sample size and p is the number of features (James et al., 2021), thus removing highly correlated features and reducing the number of features down to the ones that has the highest explanatory power. It also helps with computational cost, reduces the possibility of overfitting and mitigates the curse of dimensionality (Cai et al., 2012; Y. Kim, 2006; Marcos Lopez de Prado, 2018). This can be done manually through hand-selection of features, or automatically through techniques. These automated techniques can be categorized under feature selection and feature extraction. Feature selection chooses a subset of existing features and uses this subset as input for a model, while feature extraction create a subset of new features generated from the data (Htun et al., 2023). Automated methods for dimension reduction can be advantageous in when applying machine learning to financial time-series for two reasons: First, separating between relevant and irrelevant features manually is infamously difficult due to the unclear dependencies that exists in the data (Carta et al., 2022b). Second, it can reduce the concerns of the machine learning models acting as black boxes through enhancing the model interpretation (Zhao et al., 2019). However, one should be careful that the introduction of these techniques does not excessively remove relevant information. (Smolander et al., 2019).

Several studies on return prediction incorporates different feature selection techniques. (Enke & Thawornwong, 2005) use decision tree algorithm with integrated information gain analysis to select a feature subset. (W. Wang et al., 2020) use recursive feature selection, while (Long et al., 2019) use a convolutional filter in conjunction with a neural network. (Huck, 2019) combines random forest, and its integrated feature selection method. Most of these studies report better predictive performance with a smaller subset of important features. Hence, his study employs the *Random Forest Feature Importance* method[2].

The feature importance ranking from the random forest will be used in a two- step methodology combining the new subset of features with the machine learning models. The models will first be fed the full set of features, and then the reduced subset produced by the feature importance method. Note that this only applies for the monthly predictions, as previously discussed.

4.2 Models

Among the many statistical models available for predictive modelling this thesis considers three of them. The two most complex models are the Gradient Boosted Trees (GBT) and Random Forest (RF) algorithms. A simpler model, Logistic Regression, is included to enable the comparison between the predictive capabilities of models of varying complexity, and their ability to capture the information embedded in financial time-series.

[2] After training, the feature importance is calculated by averaging the information gain achieved by each feature over all the decision trees. The importance is thus assessed by looking at how much it improved the model's performance.

Random forest models have been applied in numerous studies and have been shown to achieve a high out-of-sample classification accuracy, while also being relatively easy to implement and configure (Fischer & Krauss, 2018; Krauss et al., 2017; Moritz & Zimmermann, 2016). Logistic regression is a simpler method well suited for predicting binary variables. It has also seen extensive use in the field of econometrics, which makes it interesting for comparison. It is therefore used as a baseline for comparing predictive performance to the other more complex methods.

Table 4.1: Overview of all models that produce predictions used for portfolio construction

Model group	**LR**	**RF**		**GBT**	
Short term	LR	RF	-	GBT	-
Long term	LR	RF	RF_r	GBT	GBT_r

4.2.2 Regularization and hyperparameters

A main advantage of machine learning models is their ability to regularize and tune its hyperparameters such that overfitting is reduced, and predictive ability is enhanced. This section explains the steps taken for tuning hyperparameters for the different models.

The random forest algorithm is built on the notion of combining many weak classifiers, i.e., decision trees, into one strong learner. While random forest does not require much hyperparameter tuning, three main parameters need to be considered when fitting the model. These are the number of decision trees, the maximum number of features to be evaluated when partitioning and the maximum depth of the decision tree. To find the optimal parameters for the model a grid search[3] is performed with the number of trees varying between 250 and 1000, and the depth between 5 and 30. Following previous studies (Fischer & Krauss, 2018; Huck, 2019) this study set the number of trees to 1000, while the depth is set to 20. As recommended

[3] A technique used to fine the optimal combination of hyperparameters for the Random Forest algorithm. It involves systematically assessing the performance for every possible combination of hyperparameters.

by (James et al., 2021) the maximum number of features to be evaluated is $\sqrt{n}$ where n is the total number of features.

The gradient boosted trees algorithm is more susceptible to overfitting than the random forest. The hyperparameters set for this model is therefore stricter than for the random forest, decreasing the number of trees and a predefined low learning rate. The low learning rate means that it converges slower but can lead to better generalization. The maximum depth is set to vary between 3 and 10, not allowing it to be as deep as the random forest. This is also done to reduce the risk of overfitting.

The logistic regression is fitted using L2, or ridge, regularization to encourage more evenly distributed coefficients. The regularization parameter, C, controls the regularization strength of the model, and is varied between 0.01 and 10 to prevent overfitting. The solver model is given four different options of solvers, or kernels, to maximize performance.

Table 4.2: Overview of hyperparameters

	GBT	**RF**	**LG**
Learning rate	0.05	-	-
Number of trees	50 ~ 400	250 ~ 1000	-
Maximum depth	3 ~ 10	5 ~ 30	-
Subsample ratio	n/2	-	-
Maximum features	-	$\sqrt{n}$	-
Criterion	-	Entropy	-
C	-	-	0.01~10
Kernel	-	-	[liblinear, newton-og, sag, lbfgs]
Regularization			L2
Maximum iterations			100

4.3 Portfolio Construction and Trading System

4.3.1 Trading Signals and Execution

The prediction phase of the proposed strategy starts by constructing a binary response variable Y_{t+1}^{s} for each stock s at time t for each study period. This takes the value of 1 if its one-period return R_{t+1}^{s} outperforms the benchmark at $t+1$, and 0 otherwise. Simple return for each stock s with price P is calculated as:

$$R_t^s = \frac{(P_t^s - P_{t-h}^s)}{P_{t-h}^s} \tag{9}$$

Thus, the response variable is defined as follows:

$$\begin{aligned} Y_{t+h}^{1,s} &= 1 \quad \text{if } R_{t+h}^{1,s} > \breve{R}_{t+h} \\ Y_{t+h}^{1,s} &= 0 \quad \text{if } R_{t+h}^{1,s} < R_{t+h} \end{aligned} \tag{10}$$

Where R_{t+h} is the return of the benchmark at time $t + h$, and h is the prediction horizon. The value of h is either 1 day or 1 month. Using binary cross-entropy, a loss function used for binary classification problems that measures the distance between the true binary class labels and the predicted probability estimates for those labels, a probability of stock s being classified as 1 can be obtained. Stocks are then sorted by this probability.

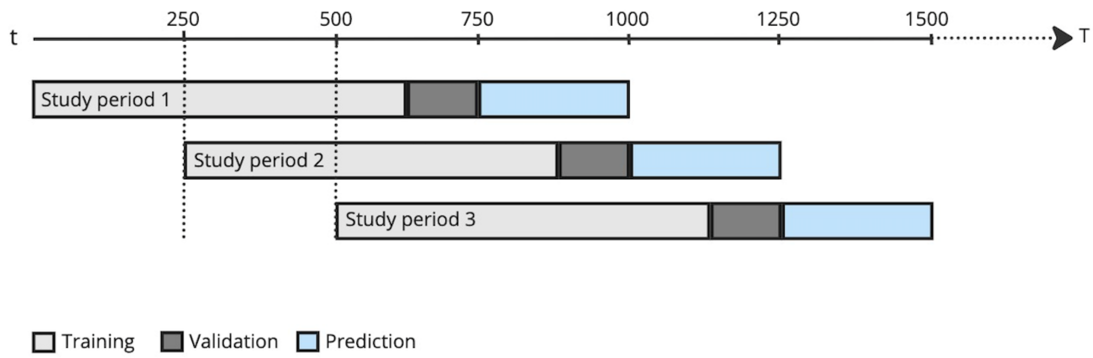
Portfolios are created by going long the k stocks with the highest probability and short the k stocks with the lowest probability at the date of rebalancing. The middle-ranked stocks exhibit the highest directional uncertainty, a risk which is mitigated through only considering the top and bottom rankings. Portfolios constructed using $k \in \{5, 10\}$ are considered. Thus, for each probability forecast a portfolio is created consisting of k long positions and k short positions. The positions in the portfolio are equal-weighted, and an equal amount of capital is allocated the long and short portfolios. Equal- weighted portfolios are deemed favorable to value-weighted portfolios due to the large discrepancy in market capitalizations, where a few stocks with high market capitalizations would dominate the returns. The return of a value-weighted portfolio could thus rely almost entirely on one or two stocks.

Similar for all models, a holding of 100 000 NOK of cash is initiated at the beginning of the trading period for both the long and short portfolios. It is assumed that all stocks can be traded for the opening price on the date of rebalancing. Further, if a stock already present in the portfolio is predicted among the *k* best performers when rebalancing, the position in that stock is adjusted so that the portfolio is equal- weighted. The same process is done for short positions.

4.3.2 Study Periods

A study period is a full completion of the trading strategy pipeline. In other words, from training and prediction to trading. A training period and trading period is set to be 750 and 250 trading days respectively. These periods are selected in line with (Huck, 2019) and is equal to approximately 3 years of training, split between training and validation, and 1 year of trading assuming there are 21 trading days a month. This creates non-overlapping batches which are shifted 250 days to the right for every study period. This method for backtesting is called the *walk-forward* validation approach and is commonly found in the time-series prediction literature. By shifting forward the training and test sets with the length of the test period, continuous out-of-sample predictions for the entire 22-year period can be obtained, except for the first 750 days which are used for the initial training, as shown in figure 4.2. It also ensures that the model is trained on relevant market conditions, as these change over time. For each walk a vector of features are created, dependent on the prediction horizon. Furthermore, the investment universe is filtered through the liquidity filters explained in section 3.1, and thus the number of investable stocks varies throughout the study periods. Missing data also impact the size of the investment universe for a given study period.

Figure 4.2: Illustration of the walk-forward approach



4.3.3 Transaction Costs

Transactions costs are notoriously difficult to replicate in backtests and return numbers can vary greatly based on the incorporated transaction costs. This thesis employs an observable transaction fee provided by Nordnet as a proxy. They charge a fee of 0.49% per transaction. A transaction cost of 0.5% have been used in earlier studies, e.g., in (Krauss et al., 2017). Applying a fee of 0.49% per share trade will therefore allow this study to compare its results with other studies, while also conforming to actual transaction costs seen in the Norwegian market.

4.4 Performance Evaluation

Both the performance of the predictions and the financial performance are analyzed using common performance metrics. This is necessary both for assessing the robustness of the models, as well as for comparability with implemented benchmarks and other studies.

4.4.1 Model Performance

Popular classification metrics are assessed for evaluating the predictive performance, including the log loss value which indicates the reliability of probabilistic predictions. The first metric that we use to analyze the predictive performance is the accuracy, which can be calculated as follows:

$$\text{Accuracy} = \frac{\text{correct predictions}}{\text{total predictions}} \tag{11}$$

We then assess the precision and recall. Precision, also known as positive predictive value, measures the ratio between correct positive predictions and total positive predictions. It thus provides information about how many predicted positive values were actually positive, indicating how confident the model is in its positive predictions. Recall, sensitivity or the true positive rate, measures the share of true positives that the model has been able to classify correctly. The F1 score is also assessed. The F1 score is the harmonic mean of precision and recall, providing a more balanced measure the model's performance. These metrics can be calculated as follows:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \tag{12}$$

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \tag{13}$$

Finally, the log loss value, or cross-entropy loss, is presented. This is a binary classification metric that quantifies the accuracy of the model by penalizing false classifications. This metric allows the performance analysis to incorporate the probability of each classification, such that it is possible to assess the confidence of the predictions. This is an important feature for this study's purpose as the portfolios are constructed based on predicted probabilities. A low log loss value is desirable, as this means the predicted probability is closer to the target. It can be calculated by the formula:

$$\text{Log loss} = -\frac{1}{N}\sum_{n=0}^{N}(Y_n \log(P_n) + (1 - Y_n)\log(1 - P_n)) \tag{14}$$

Where P_n is the predicted probability of the target being 1, Y_n is the actual target and N is the number of observations.

4.4.2 Financial Performance

Financial performance is analyzed through the classical risk-adjusted performance metrics Sharpe ratio and Sortino Ratio. Pure return and volatility numbers are also studies through return and standard deviation values. These are annualized to allow for better comparison between the monthly and daily portfolios. To study the risk profile of the portfolios, Value at Risk (5%) and Maximum Drawdown are also calculated. Formulas of the different measures are presented in table 4.3.

Table 4.3: Performance metrics

Metric	Formula	
Annualized Return	$R_p = \left(\frac{\text{Ending Value}}{\text{Beginning Value}}\right)^{1/n} - 1$	(15)
Standard Deviation	$\sigma_p = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j Cov(r_i, r_j)}$	(16)
Sharpe Ratio	$Sharpe = \frac{R_p - R_f}{\sigma_p}$	(17)
Sortino Ratio	$Sortino = \frac{R_p - R_f}{\sigma_d}$	(18)
VaR(5%)	$VaR_{5\%} = \mu - Z_{5\%}\sigma$	(19)
Maximum Drawdown	$MDD = \frac{\text{Through Value} - \text{Peak Value}}{\text{Peak Value}}$	(20)

These measure gives a holistic view of the performance of the different portfolios and ensure that favorable return numbers are not due to excessive risk taking.

5. Results and Analysis

This section presents the results obtained from the predictions and the subsequent backtests. We start by analyzing the predictions results obtained from the different machine learning algorithms, before conducting a thorough analysis of financial performance.

5.1 Prediction Results

Table 5.1 shows an overview of the classification metrics outline in section 4.4.1 for Logistic Regression (LR), Random Forest (RF) and Gradient Boosted Trees (GBT) for both daily and monthly predictions. Performance for two additional monthly models, the RF and GBT fitted using a reduced feature set, is also presented.

For short term predictions RF shows superior results, with an accuracy of 54.2%, indicating that it correctly predicts future outcomes more frequently than LR and GBT. Its precision value outperforms the other two models, implying that the positive predictions made are more likely to be actual positives. RF also yields the highest recall, which means that it performs better at identifying a higher proportion of true positives than the other two models. GBT and LR show very low recall values, indicating that the models missed many stocks that did in fact outperform the benchmark as defined by the independent variable. The ROC AUC- score measure the overall performance of the classification models, signifying that RF have better discriminative power than LR and GBT. The RF has the lowest log loss value, indicating that it provides better probabilistic outcomes, which is what we are most concerned with as these probabilities in the end decide what stocks are included in the portfolio.

Table 5.1: Results for daily predictions

	Daily predictions					
	Accuracy	Precision	Recall	F1-Score	ROC AUC - Score	Log Loss Value
LR	0.51	0.51	0.12	0.19	0.51	**0.76**
RF	**0.54**	**0.58**	**0.37**	**0.45**	**0.59**	**0.69**
GBT	0.53	0.54	0.19	0.27	0.52	**0.71**

Table 5.2: Results for monthly predictions

Monthly predictions						
	Accuracy	Precision	Recall	F1-Score	ROC AUC - Score	Log Loss Value
LR	0.51	0.51	0.51	0.49	0.53	0.73
RF	0.53	0.51	0.52	0.49	**0.54**	**0.70**
GBT	0.53	0.51	0.52	0.49	0.53	0.70
RF_r	**0.55**	**0.52**	**0.53**	**0.52**	0.53	**0.69**
GBT_r	0.53	**0.52**	0.52	0.49	**0.54**	0.70

For the monthly predictions RF and GBT initially performs similarly with an equal accuracy, precision and recall. The reduced RF however show superior performance to the original RF and GBT models, as well as the reduced GBT. There is however less of a discrepancy between the models for the monthly predictions than for the daily predictions, with the log loss value sitting around 0.70 for all models. LR show a higher log loss value at 0.73, repeating the pattern from the daily predictions.

While the accuracies shown are not very large in terms of classical machine learning performance evaluation, this is to be expected. All models do however show a accuracy higher than 50%, indicating that it is better than picking stocks at random.

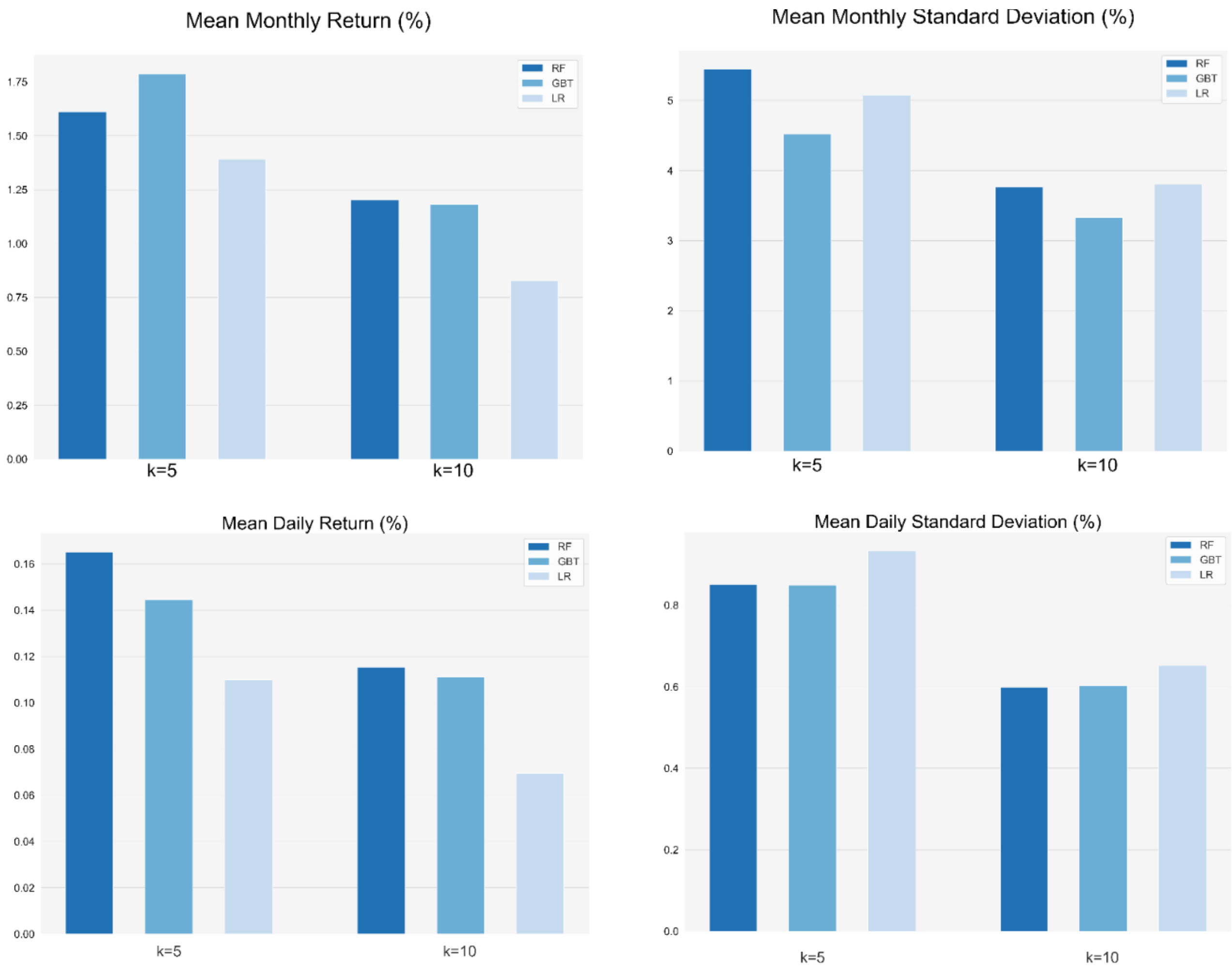
5.2 Portfolio Analysis

This section will analyze the results produced from backtesting the machine learning driven investment strategies. The following sections will first analyze the effect of different portfolio sizes for both horizons and all models. Following, different performance metrics related to return and risk while considering the effect of transaction costs are assessed, before taking a closer look at the long and short portions of the portfolios. Then an analysis of performance during sub- periods follows, before finally we analyze the specific stock positions taken. This study also dedicates a section to analyzing feature importance, with performance metrics before and after feature selection.

5.2.1 Portfolio Size and Initial Results

Monthly and daily long-short portfolios with varying holding sizes are considered, one with five stocks both legs and one with ten stocks in both legs. We start by analyzing the effect of the different holding sizes, visualized in figure X, as well as the performance by the different models. Figure X shows the mean return and mean standard deviation per rebalancing period, i.e., daily mean for the daily portfolio and monthly mean for the monthly portfolio.

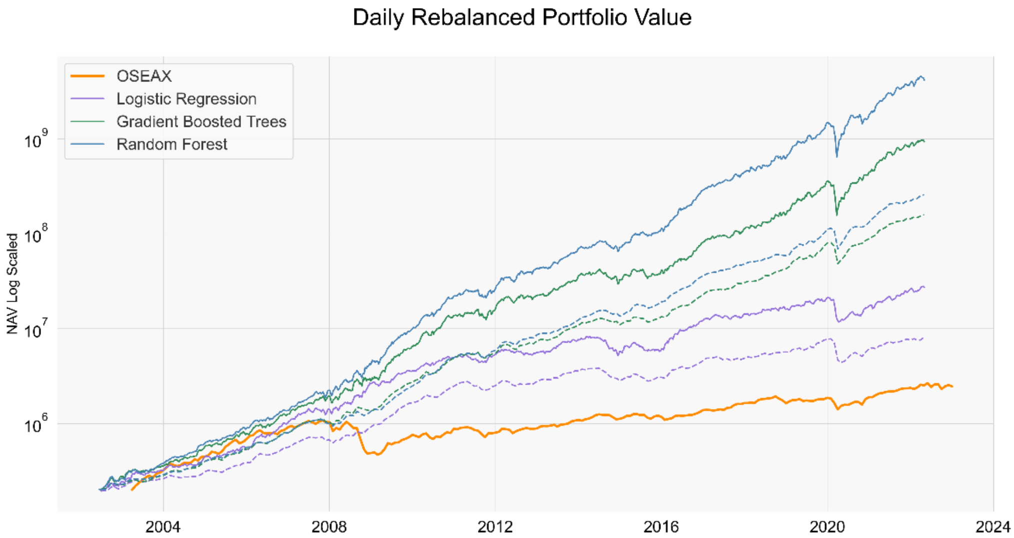
Figure 5.1: Mean values for monthly and daily portfolios

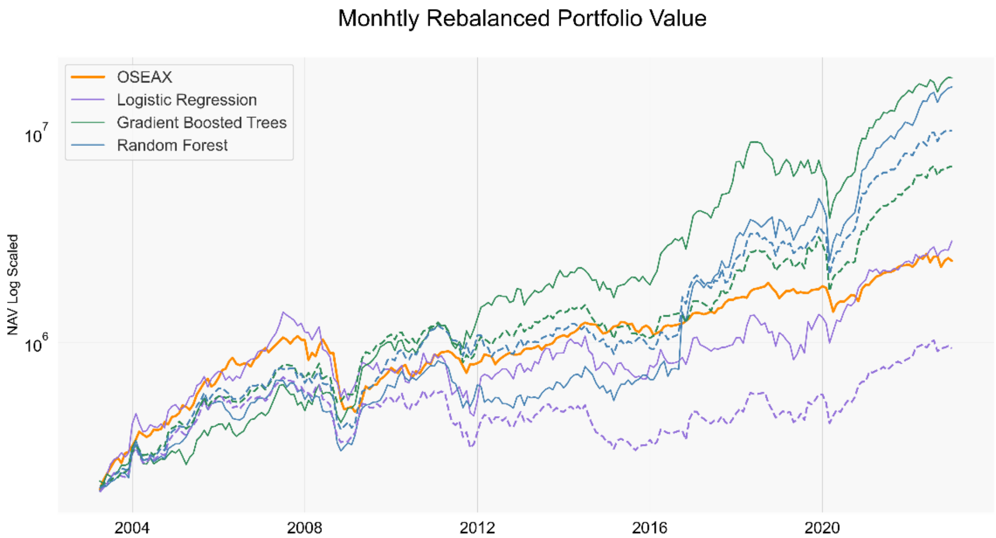


Both the daily and monthly portfolios show better performance with a holding size of 5 stocks each for the long and short portfolios. The best performing daily portfolio yields a mean daily return of 0.166% while the portfolio keeping 10 long and short positions yields a mean daily return 0.115%. Like the daily portfolio, the monthly portfolio also reports better performance for the portfolio with $k = 5$, with a mean monthly return of 1.78% as opposed to 1.20% for *k = 10*. Increasing *k* will decrease the directional certainty of the stocks that are included in the portfolio. This can explain the difference in return, as including 10 stocks in both the long and short positions would lead to holding more stocks that have a less certain probability of outperforming the benchmark. Another factor that could explain the higher return for *k = 5* is the volatility, as all portfolios exhibits a higher mean standard deviation than its *k = 10* counterpart. This behavior is expected, as established portfolio theory states that increasing the diversification of a portfolio, which in this scenario is done by adding more assets, reduces the volatility (Bodie et al., 2018).

The portfolio based on logistic regression shows the worst performance across the board, with a higher standard deviation and lower return in all portfolios, except for the monthly portfolio with *k = 10* where the random forest model reports a higher standard deviation. This is in line with the results from (Gu et al., 2020) and conforms to what was expected beforehand, as it is the simplest model of the three which consequently should result in inferior performance. It is however interesting to note that the short portfolios based on predictions from logistic regression perform much better than the long portfolios. This can be seen in tables 8.2 and 8.3 in the appendix, which showcase full return statistics for all portfolios. For both the daily and monthly portfolios the mean return from the short positions is close to the other more complex models, even outperforming both random forest and gradient boosted trees for the monthly portfolio with *k = 10.* The notable performance difference between logistic regression and the other two models for the long positions and the lack of for the short positions might serve as evidence that the short predictions are more spurious and less accurate.

Figure 5.2: Portfolio values for daily and monthly portfolios before transaction costs





Note: Full lines represent portfolios with k = 5, while dotted lines represent k = 10

As can be seen in figure 5.2, the portfolios based on predictions from the random forest- and gradient boosted trees algorithms have similar performance. Both models seem highly correlated with each other, especially for the daily portfolio, which is to be expected since they trade on the same relatively small universe of stocks. For *k = 5*, the daily portfolio based on random forest slightly outperforms gradient boosted trees, while the opposite is true for the monthly portfolio. This difference is negligible for *k = 10*, with a 0.02% difference in mean daily return and a 0.0004% difference in mean monthly returns in favor of random forest. All models produce statistically significant positive returns when testing for the null hypothesis:

$$\begin{aligned} H_0 &: \bar{X} = 0 \\ H_A &: \bar{X} \neq 0 \end{aligned} \tag{21}$$

Where $\bar{X}$ is the mean return and $\alpha = 0.05$[4]. This is true for both daily and monthly holding periods, and all holding size variations. When accounting for transaction costs however, only the random forest with *k = 5* produce statistically significant positive returns for the daily holding period. All portfolios with a daily holding period are greatly affected by transaction costs, exemplified by the dramatic drop in the t-statistic values before and after transaction costs for all portfolios with *k = 10.* From table 8.2 in the appendix we can see that mean daily returns decreased from 0.07%, 0.12% and 0.11% to -0.07%, -0.07% and -0.08% for logistic regression, random forest, and gradient boosted trees respectively after accounting for transaction costs. For the monthly holding period all portfolios with *k = 5* produce statistically significant positive returns both after and before transaction costs, while for *k = 10* none of the portfolios exhibit a statistically significant return above 0.

The further analysis will focus on the daily portfolio based on random forest predictions with *k = 5,* and the monthly portfolio based on predictions from gradient boosted trees, also with *k = 5.*

5.2.2 After Transaction Costs

While the machine learning driven strategies at first glance might yield abnormally high returns, they quickly diminish when considering transaction costs, since all strategies have a relatively high turnover. For the portfolio with daily rebalancing the high turnover is logical, as the models driving the trading activity makes decisions based on previous price history and tries to exploit any existing irregularities. This also applies for the monthly portfolio as a large portion of its model's variables are related to price history. Table 5.3 presents various risk- return metrics that provides a holistic view of portfolio performance both before and after transaction costs

[4] α (Alpha) denotes the significance level. Critical value at α = 0.05 is 1.96.

for the best performing model in both holding segments. These metrics are compared against the OSEAX index.

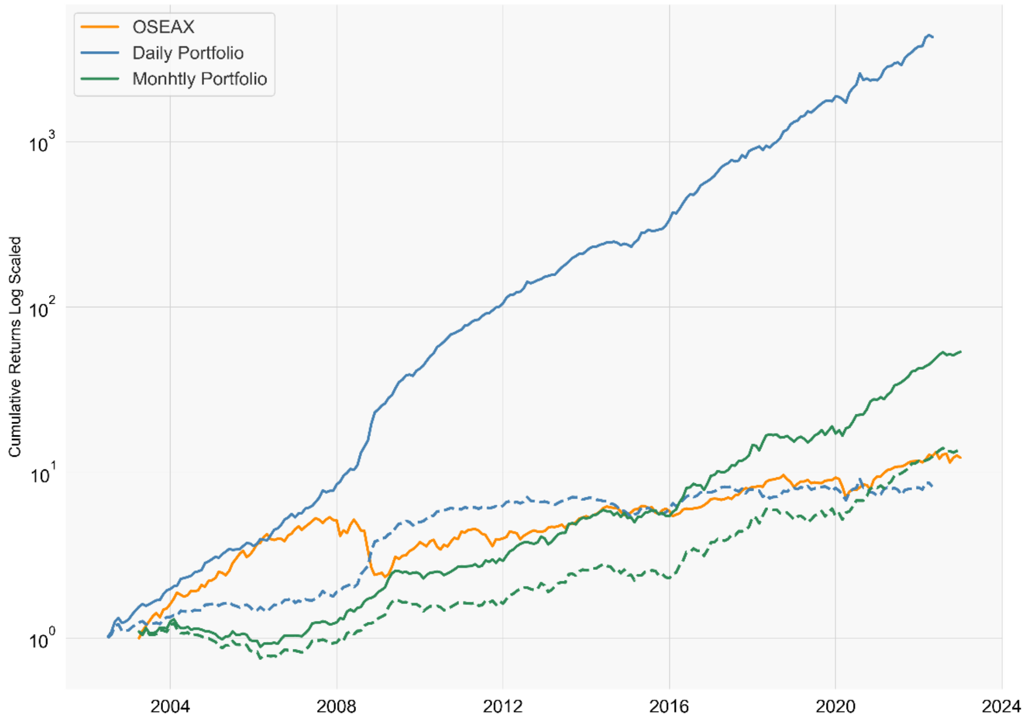
Table 5.3: Performance metrics for daily and monthly portfolio

	Daily		**Monthly**		**OSEAX**
Annualized	*Before*	*After*	*Before*	*After*	-
Return	0.5014	0.1078	0.2224	0.1413	0.1349
Std dev	0.1351	0.1353	0.1565	0.1557	0.1861
Sharpe	3.45	0.54	1.20	0.68	0.54
Sortino	5.21	0.82	2.25	1.20	0.63
MDD	0.2533	0.3252	0.2382	0.3093	0.5259
VaR 5%	-1.4570	-0.2752	-0.1668	-0.2163	-0.2504

Note: Portfolios based on random forest and gradient boosted trees for the daily and monthly holding periods respectively. Risk-free rate used in calculation of ratios is proxied by the yield on the 3-year Norwegian Government Bond.

The daily portfolio exhibits an impressive, annualized return before transaction costs of 50.14%, significantly higher than the 22.24% return of the monthly rebalanced portfolio. This indicates the potential for the daily rebalanced portfolio to generate significant return numbers and might suggest that the daily predictions are able to capture short-term price irregularities. However, after accounting for transaction costs annual return decreases significantly to 10.78%. This showcases the impact of transaction costs on a portfolio with a high turnover, and its sensitivity to transaction costs. Annualized returns from the monthly rebalanced portfolio are less sensitive to transaction costs, decreasing from 22.24% to 14.13%, thus outperforming the daily portfolio. The monthly portfolio also outperforms the benchmark slightly, yielding an excess annualized return of 0.64%. The daily portfolio however struggles to outperform the benchmark after considering transaction costs. The cumulative returns for the portfolios are shown in figure 5.3, where the eroding effect on returns from frequent transaction costs are clearly visualized.

Figure 5.3: Cumulative return before and after transaction costs



Note: Dotted lines represents the cumulative returns after accounting for transaction costs. Full line represents cumulative returns prior.

The daily portfolio does on the other hand exhibit favorable standard deviation values, which is a measure of portfolio volatility, at 13.53% after transaction costs. This might be attributed to its frequent adjustment of its holding which offsets price volatility, the fact that daily rebalancing allows the portfolio to frequently realign with its target allocation, thus reducing drift from desired risk levels, and more frequent adjustments to market movements. The annualized standard deviation for the monthly portfolio is slightly higher than the daily portfolio at 15.65% but is still smaller than the standard deviation for the benchmark.

Before transaction costs the daily rebalanced portfolio exhibits a notably higher Sharpe ratio at 3.45 compared to the monthly rebalanced portfolio at 1.20, which implies a superior risk-adjusted performance. It also delivers a higher Sortino ratio at 5.21, suggesting a better downside risk-adjusted performance. However, the transaction costs again lead to severe reductions in performance, with the Sharpe ratio decreasing to 0.54, and the Sortino ratio decreasing to 0.85. This is on similar levels with the benchmark despite that the daily portfolio delivers inferior returns, which can be attributed to the exhibited volatility as discussed above. The monthly portfolio again shows a stronger robustness to transaction costs when considering risk-adjusted metrics, outperforming both the daily portfolio and the benchmark with Sharpe- and Sortino ratios of 0.68 and 1.20 respectively after transaction costs, a relatively small reduction from the 1.20 and 2.25 values before transaction costs.

Maximum Drawdown (MDD) and Value at Risk (VaR) are two important risk metrics in portfolio analysis, each providing information on potential losses a portfolio might endure during adverse market conditions. MDD quantifies the largest observed loss from a peak-to-through of a portfolio, capturing the worst potential loss an investor could have experienced during a specific period. The daily rebalanced portfolio experiences a similar MDD to that of the monthly rebalanced portfolio after transaction costs with 32.35% to 30.93%, both outperforming the benchmark with quite a large margin. This suggests that the machine learning driven investing strategies fare better at managing exposure to large peak-to-trough losses. The Value at Risk metric is also used to measure risk within specific time frame and estimates the maximum potential loss given a level of confidence, which in this case is 5%. Both the daily and monthly portfolios show a higher VaR after transaction costs, implying that transaction costs not only reduce returns but also increase the risk of extreme losses. The daily portfolio shows a VaR of -145.7% before transaction costs, which means that, with 95% confidence, the portfolio will not lose more than 145.7% of its portfolio value. The VaR decreases to -27.52% after accounting for transaction costs which means that the maximum expected loss decreases when introducing transaction costs. While this might seem counterintuitive, a possible explanation might be that transaction costs have a dampening effect on the portfolio's return distribution by reducing the number of extreme positive and negative returns. The daily portfolio relies on a strategy that exploits sharp, short-term price movements which might not be profitable after transaction costs. This leads to a more symmetrical return distribution, which decreases the VaR.

The monthly portfolio exhibits an opposite behavior with the VaR increasing from 16.68% to 21.63% after transaction costs, which implies that it gets more sensitive to extreme losses. This increase can possibly be explained by increased exposure to adverse price movements and market volatility compared to the daily portfolio due to its longer rebalancing period, which is further exacerbated by the transaction costs. Since rebalancing is only done once a month, the portfolio might be exposed to higher risk sectors for longer, causing the portfolio to drift away from its target allocation, which are stocks with the highest probability of outperforming the benchmark. In other words, stocks that are predicted to beat the benchmark at time t might at time $T - \frac{t}{2}$ no longer exhibit the traits that prompted the machine learning model to give it a high probability in the first place. This might also lead to missed opportunities mid-holding period to capitalize on potential profitable stocks but is less of a concern as the monthly portfolio yields a higher return and Sharpe ratio than the daily portfolio.

5.2.3 Long/Short Analysis

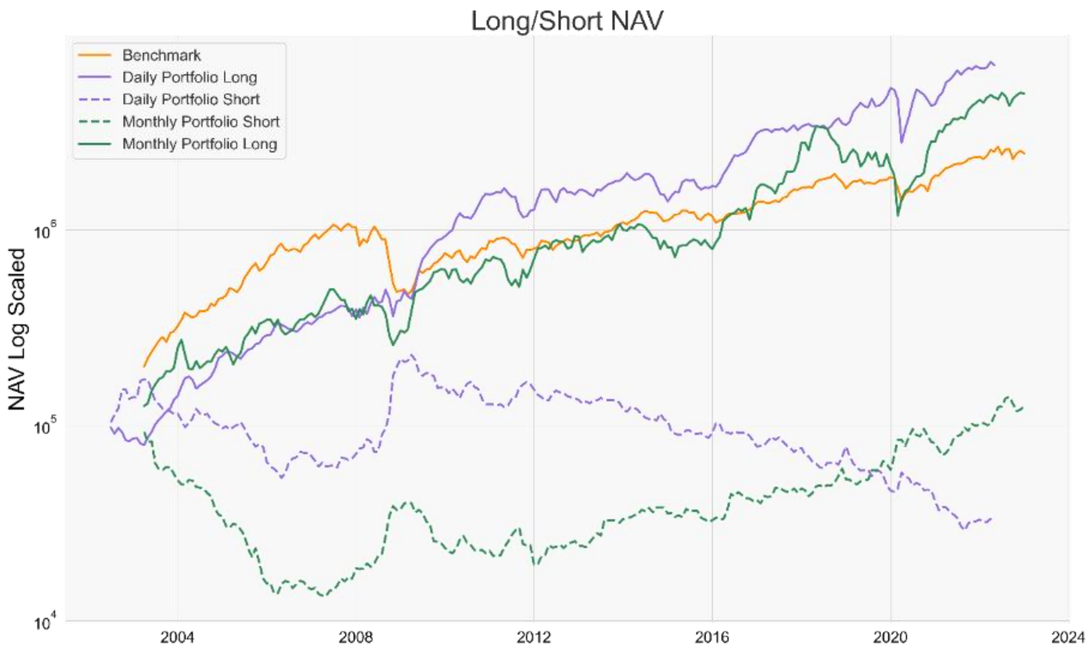
We have up until now only considered the total daily and monthly portfolios. It is however interesting to assess the constituents of the total portfolios and dissect the long and short portions. Table 5.4 presents the same risk-return metrics as the previous chapter for the long and short positions after transaction costs.

Table 5.4: Performance metrics for long and short portfolios

	Daily		**Monthly**		**OSEAX**
Annualized	*Long*	*Short*	*Long*	*Short*	-
Return	0.2251	-0.0517	0.2176	0.0116	0.1349
Std dev	0.2908	0.2481	0.3154	0.2456	0.1861
Sharpe	0.65	-0.35	0.58	-0.09	0.54
Sortino	0.91	-0.50	0.89	-0.15	0.63
MDD	0.2357	-0.8483	-0.6489	-0.8553	0.5259
Var 5%	0.3296	-0.7629	-0.4922	-0.5125	-0.2504

The most obvious observation from the table is the difference in return between the long and short portions of both portfolios. The long positions comfortably outperform the short positions with regards to annualized return, with the daily short portfolio yielding a negative annualized return and the monthly short portfolio yielding an annualized return close to zero. The risk-adjusted performance ratios for the short portfolios also deliver disappointing results, with both the Sharpe- and Sortino ratio dropping to the negatives for both holding periods. The short portfolios also experience higher risk, as can be seen by the MDD and VaR values, especially for the short portion of the daily portfolio. The superior performance of the long positions might be due to the general upward trend in the OSEAX index since the beginning of trading, resulting in fewer and less certain shorting opportunities. Furthermore, analyzing the long and short portfolios individually without considering the total effect neglects an important feature of a long-short portfolio, namely market- neutrality and reduction of volatility. As seen in table 5.4, the long portfolios exhibit higher standard deviations than the benchmark. In the previous subchapter it was shown that both the daily and monthly total portfolios had less standard deviation than the benchmark, thus experiencing less volatility. Figure 5.4 visualize the Net Asset Value (NAV) for the long and short positions, which serves as evidence that the short portfolio offsets some of the market volatility in times with high turmoil.

Figure 5.4: Net Asset Value for long and short portfolios



Note: Net Asset Value calculated by multiplying portfolio value at time t=0 with cumulative returns at time t.

The short positions effectively offset the largest drawdowns in the broad market, as seen in for example the large positive trend around the financial crisis in 2008. The short portfolios seem to be effective at exhibiting an inverse relationship with the broad market, thus reducing the overall volatility of the total portfolios. Table 5.5 provides beta values for the different portfolios. Beta is a measure of the systematic risk exposure of a portfolio in relation to the broad market, thus quantifying the portfolio's sensitivity to market movements. Since a long-short portfolio generally aims to be market-neutral, a beta value of 0 is deemed desirable as this means that the portfolio value is not at all correlated with the market and is therefore theoretically less volatile than the market. A beta value of 1 means the opposite and suggests the value of a portfolio will move in tandem with the market. Both the monthly and daily long positions exhibit a positive correlation to the OSEAX index, but the total portfolio correlations are offset by the short positions.

Table 5.5: Beta values for long and short portfolios

	Daily			***Monthly***			***OSEAX***
	Long	Short	Total	Long	Short	Total	-
β	0.044	0.010	0.027	0.172	-0.276	-0.052	1

Note: Beta values obtained by regressing the return series of the daily and monthly portfolios onto the return series of the OSEAX using Ordinary Least Squares.

Interestingly, the daily short portfolio starts a downwards trend after 2008, while the opposite is true for the monthly portfolio which regains traction around 2010 and continues to produce relatively constant positive returns. From figure 5.4 we can see that both short portfolios are highly correlated from the beginning of trading until 2012, before they start to exhibit an inverse relationship. As stated in (Lewellen, 2014), the predictability of stock returns seems to gradually vanish as we get closer to today. The large discrepancy between the two short portfolios, which are based on predictions from models of equal complexity, suggests that this holds true also for Norwegian markets when exploiting short term price history for predicting shorting candidates. Another explanation might be that predicting candidates for shorting with a relatively high precision is a much harder task than predicting candidates for long positions. Equity markets, including the OSEAX index, usually experience an inherent upward bias over the long term which means that stocks in general tend to appreciate over time. This again leads to more successful long candidates than short candidates, which can explain the difficulty in identifying shorting opportunities that generate positive returns.

We see however that the monthly short portfolio has been performing better than the daily short portfolio, indicating that there is some predictability to be captured. The monthly portfolio utilizes variables such as financial ratios and macroeconomic statistics. These factors are seemingly better at explaining negative performance for individual stocks. This can be because negative price reactions often follow poor earning calls (Fama & French, 2006), and since individual stocks are shown to exhibit momentum effects (De Bondt & Thaler, 1985; Jegadeesh & Titman, 1993), even though the financial information are lagged by its reporting frequency, the model might still capture the long- term repercussions of the negative earning calls. Another contributing factor to the better performance of the monthly short portfolio can be its inclusion of macroeconomic variables. The OSEAX index is heavily tilted towards sectors that directly reacts to the macroeconomic environment, such as energy and industrials, which constitutes that the portfolios hold stocks from these sectors, leading to improved predictability.

5.2.4 Sub-period Analysis

The trading performance of the monthly and daily portfolios are further analyzed during three different time periods. The first period is from 2002-2009, representing times of turmoil with the aftermath of the dot-com bubble and the global financial crisis of 2008. The second period is 2009-2019, characterized by recovery from the 2008 financial crisis, quantitative easing from central banks and low interest rates. In the last period, 2019-2022, financial markets were in large degree dominated by the Covid-19 pandemic and its subsequent attempt to recover, resulting in significant market volatility, as well as an interplay between escalating inflation and interest rate hikes from central banks.

Table 5.6: Performance metrics during selected time-periods.

		2002-2009		2009-2019		2019-2022	
	Annualized	*Portfolio*	*OSEAX*	*Portfolio*	*OSEAX*	*Portfolio*	*OSEAX*
Daily	Return	0.2287	0.2156	0.0471	0.0940	0.0072	0.0969
	Std dev	0.1417	0.2425	0.1163	0.1291	0.1817	0.1992
	Sharpe ratio	1.37	0.75	0.10	0.46	-0.15	0.31
	Sortino ratio	2.16	0.83	0.16	0.67	-0.21	0.36
	MDD	-0.1556	-0.5259	-0.2645	-0.0904	-0.2949	-0.2306
	VaR 5%	-0.3029	-0.3584	-0.1401	-0.0668	-0.1746	-0.1082
Monthly	Return	0.0668	0.2156	0.1335	0.0940	0.3614	0.0969
	Std dev	0.1463	0.2425	0.1642	0.1291	0.1390	0.1992
	Sharpe ratio	0.22	0.75	0.60	0.46	2.34	0.31
	Sortino ratio	0.39	0.83	1.05	0.67	3.39	0.36
	MDD	-0.3093	-0.5259	-0.1030	-0.0904	-0.0819	-0.2306
	VaR 5%	-0.2585	-0.3584	-0.2376	-0.0668	0.0850	-0.1082

Some interesting findings emerge when considering the portfolios during different time periods. Earlier studies (Fischer & Krauss, 2018; Huck, 2019) show that returns from machine learning driven investment strategies deteriorate with time, likely because of increased market efficiency. These results conform to the previous finding to some degree, with the daily portfolio showing decreasing returns. However, the monthly portfolio shows a strong positive uptrend, with a Sharpe ratio of 0.22, 0.6 and 2.34 in the periods 2002-2009, 2009-2019 and 2019-2022 respectively. The monthly portfolio shows an inverse relationship with the OSEAX index, suggesting that the increase in performance is due to other factors than positive trends in the market. It also displays a risk profile that is highly favorable relative to the OSEAX index with a MDD of only 8.19% and a VaR of 8.5% in 2019-2022, compared to 23.06% and 10.82% for the index.

The daily portfolio performs well in earlier years, gaining an excess return of 1.31% annually between 2002 and 2009 over the OSEAX index. This number decreases to -4.69% and -8.97% for the periods 2009-2019 and 2019-2022 respectively, following the same trends as reported in earlier literature. The somewhat disappointing performance between 2019-2022 might well be attributed to increased market efficiency. There are however other factors to consider, such as the length of the period and the state of the financial markets during these years. This period only spans 3 years, making it hard to tell if the trend is actual or if it was fueled in large by the Covid-19 pandemic. The variables included in the daily return predictions cannot be expected to bear any predictive power of the effect the pandemic had on stock prices, which in turn affects the return of the selected stocks from the model in an unpredictable fashion.

The different development of performance for the daily and monthly portfolio indicates that predicting stock prices using a wide array of fundamental variables, thus creating a feature space with a holistic view of individual stocks, and macroeconomic variables is beneficial. The results suggests that the market has become more efficient, but that there still is predictability to be captured in company anomalies reported in literature. This might especially be true for Norwegian stocks, as it is deemed less of an efficient market than the U.S., both because of less trading activity and competition, but also because it reacts slowly to anomalies published in financial literature. It can as a result contain opportunities for medium term return prediction, utilizing said anomalies as predictors.

Table 5.7 shows an overview over the worst drawdown periods. While one might initially expect that the portfolios would perform worse during times of financial turmoil, this seems to only be partially true. This is also logical considering the low beta values for the daily and monthly portfolios. The table shows that the worst drawdown periods for the monthly portfolio was between February 2004 and November 2008 spanning 1736 days and yielding a total drawdown of 38.71%. The daily portfolio saw its worst period start in December 2008 before recovering in July 2020, just after the initial effects of the Covid-19 pandemic. Its third worst drawdown period started just 16 days later, suggesting that the exit from the first drawdown period was in large due to positive index returns between the months of April and July[5]. Notably, none of the portfolios experienced a serious drawdown in relation to the financial crisis in 2008, but rather in much less volatile periods. This suggests that the model is better at exploiting mispricing during times of high volatility, while less so when the markets are not experiencing irregularities.

[5] The return for the OSEAX Index was 9.02%, 2.79%, -0.3% and 3.75% in the months of April, May, June and July respectively in 2020.

Table 5.7 Drawdown statistics

Monthly			
Started	*Recovered*	*Drawdown*	*Days*
29/02/2004	30/11/2008	-38.71%	1736
30/06/2014	30/04/2016	-20.18%	670
30/06/2018	30/11/2019	-17.12%	518
Avg. Drawdown		*Avg. Drawdown Days*	
-7.35%		225	
Daily			
Started	Recovered	Drawdown	Days
11/12/2018	14/07/2020	-32.53%	581
11/10/2013	20/06/2016	-26.46%	983
30/07/2020	27/04/2022	-23.68%	636
Avg. Drawdown		*Avg. Drawdown Days*	
-2.53%		39	

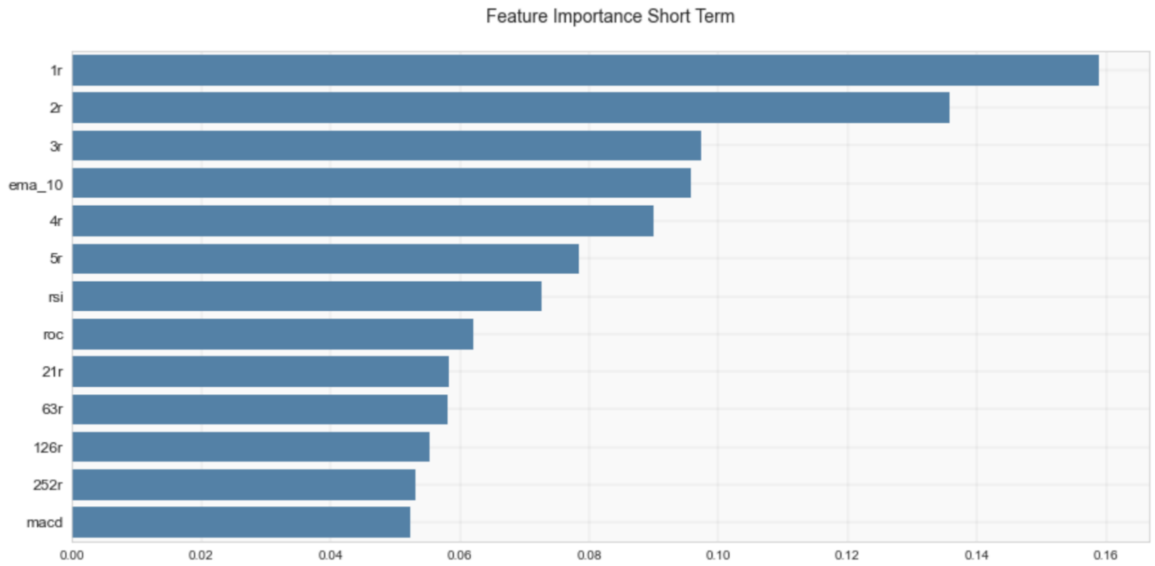
Note: The OSEAX index had an average drawdown of 8.39% with average drawdown days of 239. Daily drawdown statistics for the daily portfolio, monthly for the monthly portfolio.

5.2.5 Feature Importance

Through the integrated feature importance method in the random forest algorithm the predictive power of the different features is quantified. We start by analyzing the feature importance for the daily rebalanced portfolio, before moving on to the monthly rebalanced portfolio where we assess both the feature importance and two new portfolios utilizing a reduced feature set.

The results from the feature importance extraction are what was expected. Like in (Huck, 2019) the most recent returns are the most important features and the returns from the previous two trading days make up approximately 28% of the predictability that exists in the variables. The 10-day estimated moving average is the technical indicator that is the most important, while the rest of the indicators seem to not contribute as much as reported in (Lesmond et al., 2004). Figure 5.5 indicates that returns become less important the more distant they are.

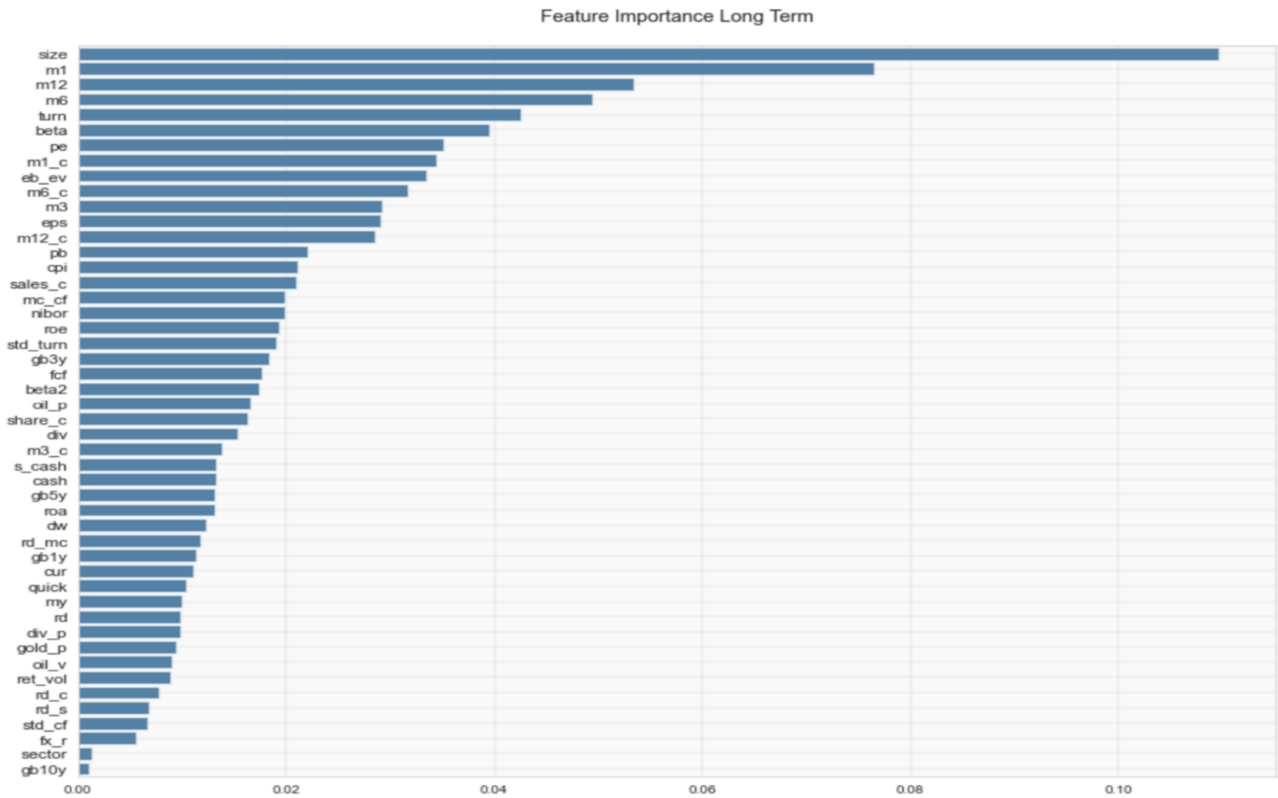
Figure 5.5: Feature importance for the daily portfolio



Next, we investigate the importance of the individual variables included in the monthly portfolio, reported in figure 5.6. The most important variable seems to be, by a great margin, size. Then follows short-term reversal (m1), also containing a great deal of the available predicting power in the feature set. The next features are also related to momentum, with stock momentum the last year (m12) and the last 6 months (m6) being among the most important variables. These four variables have in previous literature experienced much attention for its predictive power for future stock returns, making these results not surprising.

The most important variables after the mentioned four are all related to stock specific financial ratios, liquidity, and price history. Turnover (turn), beta, price-to-earnings (pe), change in short-term reversal, ebitda-to-ev (eb_ev), change in six-month momentum (m6), medium-term reversal (m3), earnings per share (eps), change in long-term momentum (m12) and price-to-book (pb) are all ranked higher than the first macroeconomic variable in the feature set. The Consumer Price Index (cpi) is the most important macroeconomic variable, followed by the 3-month NIBOR rate. Variables related to oil seem to play little part in explaining future returns, with oil trading volume (oil_v) being among the weakest predictors. The oil price (oil_p) falls behind the 3-year Norwegian Government bond yield (gb3y), while other macroeconomic variables such as the NOK/USD exchange rate, gold price (gold_p) and the 10-year Norwegian Government bond are among the least important variables.

Figure 5.6: Feature importance for the monthly portfolio



These results are like those found in (Gu et al., 2020) where features related to momentum and liquidity, as well as size, are the most important for U.S. stocks. Our results also conform with several other studies on market anomalies, suggesting that these are pervasive across different markets and geographies. The size factor however seems to be much more prominent for Norwegian stocks than for U.S. stocks. The size effect is well documented in financial literature and implies that smaller companies usually outperform larger companies. The importance of the size variable in our analysis compared to the studies performed on U.S. stocks might be caused by the Norwegian equity market containing more small cap stocks. Norwegian markets can be assumed to also be less competitive than the U.S. market, leading to a less effective reaction to published anomalies. The U.S., however, is characterized by quick responses to financial research, resulting in greater arbitrage activity and thus declining returns to stock specific variables.

Another reasons might be that the Norwegian market is likely to be less efficient than the U.S., thus opening for the size effect to persist. The U.S. market is however highly efficient, which means that capitalizing on anomalies such as size are quickly arbitraged away, which might explain that the size variable is not as dominant in studies conducted on U.S. stocks. This is consistent with results presented in (Lesmond et al., 2004) and (Hou et al., 2005) where predictive power of firm characteristics mainly exists among stocks with arbitrage opportunities or high transaction friction. Beta, another important variable, is related to the market risk of a stock and is found in classical factor models such as the Capital Asset Pricing Model (CAPM) and seems to exhibit predictive power over future returns. Further, the price-to-earnings (P/E) variable suggests that there exists a value effect where lower P/E stocks tend to outperform higher P/E stocks.

Next, the analysis of feature importance is taken one step further, fitting two new models using only the ten most important features. We then compare the financial performance for monthly rebalanced portfolios produced by the new models with the portfolios produced by models using the full feature set.

Table 5.8: Performance measures for reduced models

	Random Forest		**Gradient Boosted Trees**	
Annualized	*RF*	*RF_r*	*GBT*	*GBT_r*
Return	0.1178	0.1180	0.1413	0.1650
Std dev	0.1884	0.1947	0.1557	0.1761
Sharpe ratio	0.43	0.43	0.68	0.74
Sortino ratio	0.86	0.90	1.20	1.64
MDD	-0.2536	-0.3059	-0.3093	-0.2811
VaR 5%	-0.2366	-0.3353	-0.2163	-0.2225

Table 5.8 reports the usual performance metrics. The portfolios constructed based on predictions from both GBT and RF perform better after including only the most important variables, with the GBT showing the most improvement yielding an extra 2.37% annualized return. While only including the set of variables with the highest predictive power increase returns, it also results in higher volatility with the standard deviation increasing with 0.63% and 2.04% annually for RF and GBT respectively. Despite this, both models produce a higher or equal Sharpe and Sortino ratio. The MDD and VaR metrics show poorer performance for the

reduced Random Forest model, while the reduced Gradient Boosted Trees models show an improvement of MDD but a less favorable VaR. The reduced GBT portfolio does however exhibit a general increase in performance, while the RF does not. This might be due to the resistant nature of the RF model to overfitting. The GBT model is more prone to overfitting, which means there is more potential for increasing performance when reducing the feature space. Discarding variables with little predictive power reduces noise in the data, resulting in the model being less complex and thus less prone to overfitting. Reducing the feature space increases the probability of the model to generalize better, which might explain the increase in results for the reduced GBT portfolio.

5.2.6 Stock Analysis

Analysis on holding size, model choice, and transaction costs have all been conducted with an emphasis on its effect on portfolio performance. This sub-chapter will take a closer look at the core driver of return, namely the individual stocks that have been included in the different portfolios. As discussed previously, machine learning driver investment strategies tend to act as black boxes basing decisions on reasons unknown to the practitioner. This section is an attempt delve into this black box and shed light on the trading decisions made by the machine learning algorithms.

Figure 5.7: Most held stocks

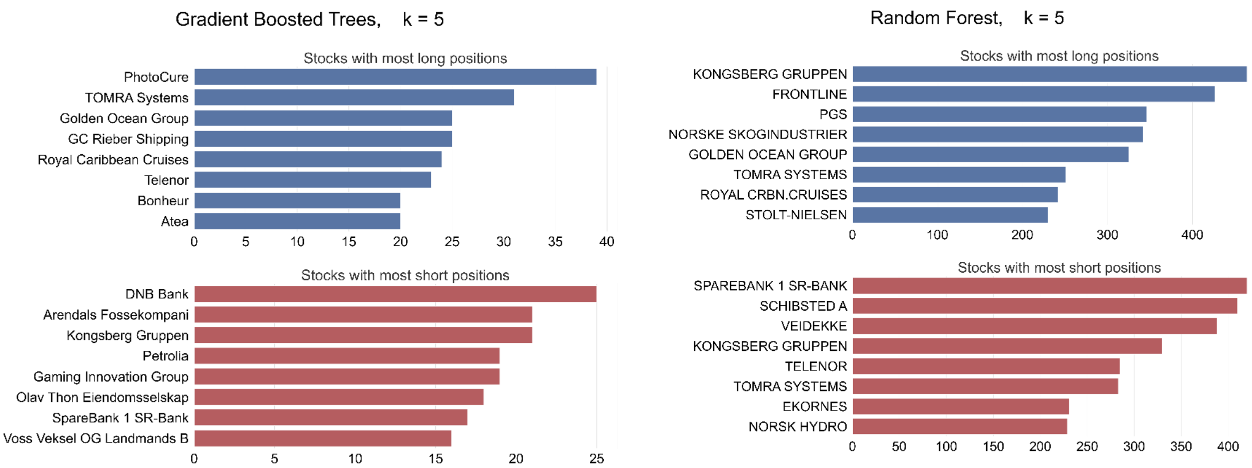


Figure 5.7 shows an overview of stocks that were held for the most trading periods, i.e. trading days for the daily portfolio and months for the monthly portfolio. They are split into two categories; one for the long and short portion of the portfolios. Some stocks are frequently held

both for the monthyl and daily portfolios. Tomra Systems, Golden Ocean Group and Royal Caribbean Cruises are all recurrent stocks for the long portion of the portfolios, while Kongsberg Gruppen and SpareBank 1 SR-Bank frequently are chosen for the short portfolios. Interestingly, the daily portfolio has the most long positions in Kongsberg Gruppen, while it also being a favorite for the short portfolio. Tomra Systems also seem to be a favorite in both portions of the daily portfolio. Furthermore, while Telenor is a repeating stock in holding for the long portion of the monthly portfolio, it is among the top held stocks for the short portion of the daily portfolio. The monthly portfolio also does not have any stocks in both the top long and top short stocks, unlike the daily portfolio. To further gain insights into what has driven the machine learning algorithms we start by dissecting the mean values of all the most important features of both the monthly and daily portfolios just prior to prediction. This will give indicators as to what values the different features had that made the machine learning algorithms choose said stocks.

We analyze the monthly portfolio first. Table 5.9 show mean feature values per stock prior to prediction for the five most important features of the monthly portfolio, size, 1- month momentum, 6-month momentum, 12-month momentum and turnover. Analyzing the table below reveals some noteworthy observations about mean characteristics of the most traded stocks. First, the mean size value for the long positions is at 34950 million NOK much larger than that for the short positions, with approximately two thirds of the value at 21032 million. This value is however skewed by a few companies, namely Telenor and Royal Caribbean Cruises for the long portfolio, and DNB and Kongsberg Gruppen for the short portfolio. The median value gives a fairer measurement, with the long positions returning a median size of 6423 million NOK compared to 5949 million NOK for the short positions. While this is not a very large discrepancy, it shows that the long portfolio prefers smaller companies providing some evidence that the machine learning model have unraveled the size effect from the data. The momentum features show that the short portfolio prefers stocks that have exhibited big long term positive returns, showing a mean 12-month momentum at 17.3% in comparison to the long portfolio's 8.6%. The 6-month (3.7%) and 1-month (0.8%) momentum variables are also lower for the long portfolio. The lower long-term momentum for the long portfolio is in line with the reversal effect found in (De Bondt & Thaler, 1985), where it is reported that stocks with low long-term past returns exhibit higher future returns. The model also seems to prefer stocks with positive short-term momentum, which is in line with short-term momentum effect reported in (Jegadeesh & Titman, 1993). The short portfolio however seems to also prefer stocks with a positive short-term momentum, suggesting that the model did not capture such a relationship between future returns and short-term momentum.

Table 5.9: Overview of mean feature values prior to prediction – monthly portfolio

	Stock	Size	M1	M6	M12	turn
Long						
	PhotoCure	768.2	0.001	0.01	-0.011	28
	Tomra Systems	6115.4	0.005	-0.011	-0.075	1416
	Golden Ocean Group	4991.4	0.012	0.082	0.221	961
	GC Rieber Shipping	941.7	-0.021	-0.118	-0.212	0.66
	Royal Caribbean Cruises	51424	0.017	0.024	0.093	9166
	Telenor	199750	-0.003	0.050	0.053	7933
	Bonheur	6730.6	0.050	0.148	0.599	143
	Atea	8878.8	0.061	0.114	0.024	487
	Mean	**34950**	**0.015**	**0.037**	**0.086**	**2516**
	Median	**6423**	**0.008**	**0.037**	**0.038**	**724**
Short						
	DNB	130812	0.029	0.174	0.375	5483
	Arendals Fossekompani	4193.7	0.019	0.112	0.237	21
	Kongsberg Gruppen	12714.5	0.025	0.084	0.167	119
	Petrolia	235.8	-0.03	-0.087	-0.123	48
	Gaming Innovation Group	1150.9	0.028	0.256	0.233	51
	Olav Thon Eiendom	11245.2	-0.046	-0.091	-0.095	24
	Sparebank 1 SR-Bank	7704.4	0.033	0.217	0.337	124
	Voss Veksel- og Landmandsbank	206	0.01	0.135	0.258	0.72
	Mean	**21032.8**	**0.0085**	**0.100**	**0.173**	**733**
	Median	**5949.0**	**0.022**	**0.123**	**0.235**	**49.5**

Another interesting observation is that, for the short portfolio, stocks that show negative mean short- and long-term momentum prior to prediction also exhibit a low turnover. This indicates that the model has captured an interaction effect between turnover and momentum, where low turnover and negative momentum might result in negative future returns. There is however contradicting evidence, as GC Rieber Shipping and Olav Thon Eiendom both have high negative returns and low turnover while being in two separate legs of the portfolio, which makes it difficult to conclude anything. These stocks can though serve as evidence that the model capture other effects than size and momentum, such as value, where it is assumed that stocks that have underperformed in the past are undervalued and therefore have potential for positive future performance. Finally, the long portfolio tilt more towards stocks with a higher turnover than the short portfolio, indicating that good liquidity is an attractive characteristic for possible future returns.

Table 5.10: Overview of mean feature values prior to prediction – daily portfolio

	Stock	R1	R2	R3
Long				
	Kongsberg Gruppen	0.0039	0.0062	0.0043
	Frontline	-0.0038	-0.0060	0.0145
	PGS	0.0021	0.0018	0.0071
	Norske Skogindustrier	-0.0008	-0.0038	0.0328
	Golden Ocean Group	0.0006	-0.0041	0.0048
	Tomra Systems	0.0007	0.0025	0.0018
	Royal Caribbean Cruises	-0.0008	0.0002	-0.0051
	Stolt-Nielsen	0.0020	0.0039	0.0047
	Mean	**0.0005**	**0.0000**	**0.0081**
	Median	**0.0006**	**0.0010**	**0.0047**
Short				
	SpareBank 1 SR-Bank	0.0010	0.0021	0.0009
	Schibsted	0.0026	0.0058	0.0059
	Veidekke	0.0050	0.0080	0.0048
	Kongsberg Gruppen	0.0041	0.0066	0.0055
	Telenor	0.0036	0.0048	0.0043
	Tomra Systems	0.0019	0.0039	0.0029
	Ekornes	0.0024	0.0071	0.0103
	Norsk Hydro	-0.0003	-0.0021	0.0006
	Mean	**0.0025**	**0.0045**	**0.0044**
	Median	**0.0025**	**0.0053**	**0.0045**

Table 5.10 is based on the same logic as the previous table, this time presenting data for lagged returns for 1, 2, and 3 trading days. The mean values for the long portfolio are lower than that of the short portfolio for R1 and R2, while higher for R3. This indicates that the model does not necessarily favor stocks that have exhibited high past returns, but rather apply a contrarian strategy expecting stocks that have low or negative returns to exhibit positive returns in the near future (De Bondt and Thaler, 1985). The high R1 and R2 for the short portfolio compared to the near-zero R1 and R2 values for the long portfolio indicates that the model expects stock to exhibit mean-reverting behavior. This is however difficult to conclude, as identifying mean-

reversion usually includes deeper analysis the relation between past and future returns. For a slightly longer period back in time the model seems to anticipate that stocks that have a high return will demonstrate stronger return continuation, as the mean R3 for the long portfolio is almost double that of the short portfolio, showing evidence of a momentum effect for slightly longer periods.

Table 8.4 in the appendix display additional stock specific information, including the biggest winners and losers for both the long and short portions of the daily and monthly portfolios.

6. Discussion

6.1 Machine Learning for Return Prediction

While research on return prediction using machine learning is an important future endeavor of financial research, there are some challenges that exists that should be discussed. This is especially true for less liquid markets, where it can be challenging to perform a proper and robust backtest without inflating, or deflating, results. The daily rebalanced portfolio is especially sensitive to trading friction and transaction costs, as many trades conducted in this study might very well never have been possible to complete due to a plethora of factors. The high sensitivity to transaction costs exhibited by the daily portfolio was made clear when analyzing performance before and after considering transaction costs, and even a basis point additional transaction cost could alter the annual performance metrics considerably. It is important to note that this study conducts empirical research assuming conditions that might be unrealistic in a real-life scenario. Factors such as lending availability, implementation shortfall and price impact will in practice make the machine learning driven investing strategy much harder to implement. While this study deploys an observable transaction cost of 0.49% per trade, it does not account for the bid-ask-spread which also might affect the reported return numbers, especially for the daily portfolio. It is however less of a prominent problem for the monthly portfolio. Furthermore, the earlier years of the deployed strategy would also be characterized by less computing power and a less developed electronic trading infrastructure, which might have increased the transaction costs. Finally, it is also important to note that while the portfolios generated results greater or equal to the benchmark, it cannot be fully excluded that this performance is due to pure luck and randomness.

While there clearly exists challenges that accompanies deploying machine learning driven strategies, the unique ability to capture non-linear relationships and complex dynamics in financial time-series still justifies its place in the financial literature. Short-term fluctuations and price irregularities are often complex and can only partly be described by high-dimensional data. Development in empirical research utilizing non-linear models can help close the gap between the traditional linear models that have dominated financial research and the complex reality of financial markets. Furthermore, as discussed previously, the quantification of an asset's risk premium is inherently a problem of prediction. The proposed set of variables that can be used for this prediction task has in recent years grown to an unmanageable size for linear models. Machine learning models can complement the established linear models, contributing to further understanding of the drivers of returns by capturing subtle patterns and relationships in large datasets. This study utilized a feature importance method embedded in the Random Forest algorithm, extracting valuable information about the predictive power of the individual variables that would be difficult to replicate with linear models. The feature importance analysis identified that established anomalies such as size, momentum and value possess predicting power for future stock returns. In addition to providing evidence for the predictive abilities of these features in the Norwegian equities market, they were also utilized to produce superior risk-adjusted performance relative to the OSEAX index. The ability of machine learning models to both identify potential anomalies and capture non-linear relationships in financial data showcases its usefulness for both practitioners and academics.

6.2 Results in relation to the Efficient Market Hypotheses

The results generated from portfolios constructed from predictions using machine learning methods seem to partially challenge the assertions of the EMH. The weak form of the EMH suggests that all available information regarding historical market data is fully incorporated in the current stock price, and therefore it is not possible to achieve superior consistent excess returns deriving relationships and patterns from past price movements. This seems to hold true for the Norwegian equity market, as the financial performance from the daily portfolio delivered a risk-adjusted performance after transaction costs on par with that of the OSEAX index, implying that there is not enough predictive power in the historical price movements of stocks when considered in isolation.

The monthly portfolio, however, challenges the semi-strong form of the EMH. This form of efficient markets asserts that all publicly available information, like financial statement information and macroeconomic events, is incorporated into the current stock prices. Yielding an annual Sharpe ratio 0.68, the monthly portfolio outperforms the Sharpe ratio of the benchmark at 0.54. The machine learning driven investment strategy appears to contradict the notion of the semi-strong form of the EMH, and the result from the monthly portfolio suggests that the Gradient Boosted Machine was able to successfully identify and exploit information that were not fully reflected in the stock price. The risk-adjusted performance of the monthly portfolio suggests that the market may not fully reflect all publicly available information, opening for quantitative strategies to exploit inefficiencies and generate consistent excess returns, disputing the semi-strong form of the EMH in the Norwegian equity markets.

6.3 Limitations and Further Research

While the results obtained from constructing machine learning driven portfolios were decent, even more complex models could push the performance even further. Deep learning has gained traction in the empirical literature for return prediction in recent times, with the Long Short Term Memory (LSTM) model especially yielding promising results. The LSTM model was planned to be included in this study but was cut due to lacking computational capacity, which showcases one of the forementioned challenges of applying complex non-linear models. Other machine learning methods can also be investigated and compared to the ones used in this study, such as the Support Vector Machine, Naïve Bayes, k-Nearest Neighbors or ensemble methods.

The set of variables used for prediction have also played a central part in this study. While at first it was planned to utilize most of the predictor variables outlined in (Gu et al., 2020) for the monthly portfolio, it quickly became apparent that lack of available data made this unfeasible. In addition to this, for the variables that there were available data the quality often was subpar, especially in the earlier years of the study period. This reduced the scope of the predictor set considerably, but with data becoming more available for every year that passes future research might deploy many more variables, thus exploiting the full capacity of machine learning models to handle many variables. This study also assessed the individual importance of the used variables using the integrated feature importance technique from the random forest algorithm. Future research could assess other methods for extracting variable importance and compare these with the RF method. This would check the robustness of the feature importance produced by the RF. Additionally, clustering methods such as Principal Compent Analysis can be utilized to extract features automatically from the data.

In terms of backtesting, there are further work that could both enhance performance and increase realistic feasibility. Due to the strict constraints placed on required liquidity, many stocks are excluded from the investment universe of tradeable stocks every study period. This leads to all models being subject to a relatively small set of stocks, which further decrease the difference in performance they could have exhibited if the number of tradeable stocks was bigger. Further, this study does not deploy any advanced portfolio optimization techniques, which could potentially enhance performance. Future work could look at portfolio optimization in conjunction with machine learning, utilizing for example mean-variance optimal portfolios. Further, as discussed, further work can be done to increase the accuracy of the backtesting process, more closely replicating a realistic trading environment.

7. Conclusion

This study has analyzed three main problems in relation to applying machine learning for return prediction. First, two different prediction horizons have been investigated. Second, the empirical extension of linear models to non-linear models in financial research. Lastly, this study investigated the difference between machine learning applied to more efficient markets such as the U.S. and less efficient markets such as in Norway, effectively testing whether the EMH holds in the Norwegian equity markets.

Three different machine learning models of varying complexity produced a total of ten long-short portfolios, five each for the daily and monthly prediction. Two different long-short portfolios were constructed per model, differing in holding size. Results show that holding ten stocks in each the long and short portion increases diversification, and thus volatility, but are not able to outperform a holding size of five stocks per portfolio when assessing the risk-return relationship. This holds true for both daily and monthly rebalancing. The portfolios produced by the logistic regression model perform poorly compared to random forest and gradient boosted trees, suggesting that the more complex models can capture non-linear relationships that the logistic regression cannot. This suggests that non-linear models might play a more central part in the empirical financial literature in the future. For the daily predictions the random forest performed the best, while for the monthly portfolios the gradient boosted trees model produced the most impressive results.

The daily rebalanced portfolios generate an impressive mean return of 0.22% per day, indicating that it can exploit information gained from short-term price history. While this might sound promising, the returns are quickly eroded once accounting for transaction costs, exemplified by the annual Sharpe ratio decreasing from 3.45 before transaction costs to 0.54 after transaction costs, thus no longer outperforming the OSEAX index. The monthly rebalanced portfolio initially shows less impressive results with a Sharpe of 1.20 before transaction costs, but its sensitivity to transaction costs is nowhere near that of the daily portfolios. While the Sharpe ratio is close to half of its value after accounting for transaction costs, it still yields an annual Sharpe of 0.68, thus outperforming the index.

Finally, this study finds that the most important variables for predicting future stock returns in large degree conform to what has previously been reported in the financial literature. Size and momentum seem to be the most dominant factors for the monthly portfolio. Size is the most important variable, which differs slightly from Gu et al. where momentum is deemed the most important variable. It is discussed that this might be due to the difference in efficiency between the Norwegian and U.S. stock markets. For the daily portfolio the most recent returns seem to be the most important predictors, outperforming technical indicators and returns from further back. This is in line with what has been reported in krauss, huck.

In summary, this study shows that applying machine learning for return prediction can generate value both for the practitioner, delivering consistent excess returns, and for the academic, analyzing anomalies and drivers of return. This study suggests that there are opportunities in the Norwegian equity markets for exploiting market- and stock specific anomalies that are not fully incorporated in the stock price, effectively disputing the semi-strong form of the efficient market hypothesis.

Bibliography

Adadi, A., & Berrada, M. (2018). Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access*, *6*, 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052

Alex, K., Alan, M., & Zvi, B. (2018). *Investments* (10th ed.). McGraw Hill Higher Education.

Andrew Karolyi, G. (2016). Home Bias, an Academic Puzzle. *Review of Finance*, *20*(6), 2049–2078. https://doi.org/10.1093/rof/rfw007

Ariyo, A. A., Adewumi, A. O., & Ayo, C. K. (2014). Stock Price Prediction Using the ARIMA Model. *2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation*, 106–112. https://doi.org/10.1109/UKSim.2014.67

ASNESS, C. S., MOSKOWITZ, T. J., & PEDERSEN, L. H. (2013). Value and Momentum Everywhere. *The Journal of Finance*, *68*(3), 929–985. https://doi.org/10.1111/jofi.12021

Banz, R. W. (1981). The relationship between return and market value of common stocks. *Journal of Financial Economics*, *9*(1), 3–18. https://doi.org/10.1016/0304-405X(81)90018-0

Bao, W., Yue, J., & Rao, Y. (2017). A deep learning framework for financial time series using stacked autoencoders and long-short term memory. *PLOS ONE*, *12*(7), e0180944. https://doi.org/10.1371/journal.pone.0180944

Basak, S., Kar, S., Saha, S., Khaidem, L., & Dey, S. R. (2019). Predicting the direction of stock market prices using tree-based classifiers. *The North American Journal of Economics and Finance*, *47*, 552–567. https://doi.org/10.1016/j.najef.2018.06.013

Basu, S. (1983). The relationship between earnings' yield, market value and return for NYSE common stocks. *Journal of Financial Economics*, *12*(1), 129–156. https://doi.org/10.1016/0304-405X(83)90031-4

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control*. John Wiley & Sons.

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

BROCK, W., LAKONISHOK, J., & LeBARON, B. (1992). Simple Technical Trading Rules and the Stochastic Properties of Stock Returns. *The Journal of Finance*, *47*(5), 1731–1764. https://doi.org/10.1111/j.1540-6261.1992.tb04681.x

Brownstone, D. (1996). Using percentage accuracy to measure neural network predictions in Stock Market movements. *Neurocomputing*, *10*(3), 237–250. https://doi.org/10.1016/0925-2312(95)00052-6

Bussmann, N., Giudici, P., Marinelli, D., & Papenbrock, J. (2021). Explainable Machine Learning in Credit Risk Management. *Computational Economics*, *57*(1), 203–216. https://doi.org/10.1007/s10614-020-10042-0

Bzdok, D., Altman, N., & Krzywinski, M. (2018). Statistics versus machine learning. *Nature Methods*, *15*(4), 233–234. https://doi.org/10.1038/nmeth.4642

Cai, X., Hu, S., & Lin, X. (2012). Feature extraction using Restricted Boltzmann Machine for stock price prediction. *2012 IEEE International Conference on Computer Science and Automation Engineering (CSAE)*, 80–83. https://doi.org/10.1109/CSAE.2012.6272913

Campbell, J. Y., Lo, A. W., MacKinlay, A. C., & Whitelaw, R. F. (1998). The Econometrics of Financial Markets. *Macroeconomic Dynamics*, *2*(4), 559–562. https://doi.org/10.1017/S1365100598009092

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

Carta, S., Consoli, S., Podda, A. S., Recupero, D. R., & Stanciu, M. M. (2022a). Statistical arbitrage powered by Explainable Artificial Intelligence. *Expert Systems with Applications*, *206*, 117763. https://doi.org/10.1016/j.eswa.2022.117763

Carta, S., Consoli, S., Podda, A. S., Recupero, D. R., & Stanciu, M. M. (2022b). Statistical arbitrage powered by Explainable Artificial Intelligence. *Expert Systems with Applications*, *206*, 117763. https://doi.org/10.1016/j.eswa.2022.117763

Chen, S.-S. (2009). Predicting the bear stock market: Macroeconomic variables as leading indicators. *Journal of Banking & Finance*, *33*(2), 211–223. https://doi.org/10.1016/j.jbankfin.2008.07.013

Chen, Y., & Hao, Y. (2017). A feature weighted support vector machine and K-nearest neighbor algorithm for stock market indices prediction. *Expert Systems with Applications*, *80*, 340–355. https://doi.org/10.1016/j.eswa.2017.02.044

Chourmouziadis, K., & Chatzoglou, P. D. (2016). An intelligent short term stock trading fuzzy system for assisting investors in portfolio management. *Expert Systems with Applications*, *43*, 298–311. https://doi.org/10.1016/j.eswa.2015.07.063

De Bondt, W. F. M., & Thaler, R. (1985). Does the Stock Market Overreact? *The Journal of Finance*, *40*(3), 793–805. https://doi.org/10.1111/j.1540-6261.1985.tb05004.x

Enke, D., & Thawornwong, S. (2005). The use of data mining and neural networks for forecasting stock market returns. *Expert Systems with Applications*, *29*(4), 927–940. https://doi.org/10.1016/j.eswa.2005.06.024

Fama, E. F. (1981). Stock Returns, Real Activity, Inflation, and Money. *The American Economic Review*, *71*(4), 545–565.

Fama, E. F. (1991). Efficient Capital Markets: II. *The Journal of Finance*, *46*(5), 1575–1617. https://doi.org/10.1111/j.1540-6261.1991.tb04636.x

Fama, E. F., & French, K. R. (1988a). Permanent and Temporary Components of Stock Prices. *Journal of Political Economy*, *96*(2), 246–273. https://www.jstor.org/stable/1833108

Fama, E. F., & French, K. R. (1988b). Dividend yields and expected stock returns. *Journal of Financial Economics*, *22*(1), 3–25. https://doi.org/10.1016/0304-405X(88)90020-7

Fama, E. F., & French, K. R. (2006). Profitability, investment and average returns. *Journal of Financial Economics*, *82*(3), 491–518. https://doi.org/10.1016/j.jfineco.2005.09.009

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

Fama, E. F., & Schwert, G. W. (1977). Asset returns and inflation. *Journal of Financial Economics*, *5*(2), 115–146. https://doi.org/10.1016/0304-405X(77)90014-9

Fama F, E. (1965). The Behaviour of Stock-Market Prices. *The Journal of Business*, *38*(1), 34–105.

Feng, G., Giglio, S., & XIU, D. (2020). Taming the Factor Zoo: A Test of New Factors. *The Journal of Finance*, *75*(3), 1327–1370. https://doi.org/10.1111/jofi.12883

Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, *270*(2), 654–669. https://doi.org/10.1016/j.ejor.2017.11.054

Flori, A., & Regoli, D. (2021). Revealing Pairs-trading opportunities with long short-term memory networks. *European Journal of Operational Research*, *295*(2), 772–791. https://doi.org/10.1016/j.ejor.2021.03.009

Foerster, S., Tsagarelis, J., & Wang, G. (2017). Are Cash Flows Better Stock Return Predictors Than Profits? *Financial Analysts Journal*, *73*(1), 73–99. https://doi.org/10.2469/faj.v73.n1.2

Gjerde, Ø., & Sættem, F. (1999). Causal relations among stock returns and macroeconomic variables in a small, open economy. *Journal of International Financial Markets, Institutions and Money*, *9*(1), 61–74. https://doi.org/10.1016/S1042-4431(98)00036-5

Goddard, M. (2017). The EU General Data Protection Regulation (GDPR): European Regulation that has a Global Impact. *International Journal of Market Research*, *59*(6), 703–705. https://doi.org/10.2501/IJMR-2017-050

Green, J., Hand, J. R. M., & Zhang, X. F. (2013). The supraview of return predictive signals. *Review of Accounting Studies*, *18*(3), 692–730. https://doi.org/10.1007/s11142-013-9231-1

Greenwood, M. (2022). Quantitative finance with Python: A practical Guide to Investment Management, Trading and Financial Engineering, *Chapman & Hall/CRC*. https://doi.org/10.1080/14697688.2023.2179939

Grobys, K., & Huhta-Halkola, T. (2019). Combining value and momentum: evidence from the Nordic equity market. *Applied Economics*, *51*(26), 2872–2884. https://doi.org/10.1080/00036846.2018.1558364

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

Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., & Yang, G.-Z. (2019). XAI—Explainable artificial intelligence. *Science Robotics*, *4*(37). https://doi.org/10.1126/scirobotics.aay7120

Hao, J., He, F., Ma, F., Zhang, S., & Zhang, X. (2023). Machine learning vs deep learning in stock market investment: an international evidence. *Annals of Operations Research*. https://doi.org/10.1007/s10479-023-05286-6

Harvey, C. R., Liu, Y., & Zhu, H. (2016). … and the Cross-Section of Expected Returns. *Review of Financial Studies*, *29*(1), 5–68. https://doi.org/10.1093/rfs/hhv059

Henrique, B. M., Sobreiro, V. A., & Kimura, H. (2018). Stock price prediction using support vector regression on daily and up to the minute prices. *The Journal of Finance and Data Science*, *4*(3), 183–201. https://doi.org/10.1016/j.jfds.2018.04.003

Hoang, D., & Wiegratz, K. (2022). Machine learning methods in finance: Recent applications and prospects. *European Financial Management*. https://doi.org/10.1111/eufm.12408

Hou, K., Karolyi, G. A., & Kho, B.-C. (2011). What Factors Drive Global Stock Returns? *Review of Financial Studies*, *24*(8), 2527–2574. https://doi.org/10.1093/rfs/hhr013

Hou, K., & Moskowitz T.-J. (2005). Market Frictions Price Delay, and the Cross-Section of Expected Returns. *Review of Financial Studies*, *18*(3), 981-1020. https://doi.org/10.1093/rfs/hhi023

Hsu, M.-W., Lessmann, S., Sung, M.-C., Ma, T., & Johnson, J. E. V. (2016). Bridging the divide in financial market forecasting: machine learners vs. financial economists. *Expert Systems with Applications*, *61*, 215–234. https://doi.org/10.1016/j.eswa.2016.05.033

Htun, H. H., Biehl, M., & Petkov, N. (2023). Survey of feature selection and extraction techniques for stock market prediction. *Financial Innovation*, *9*(1), 26. https://doi.org/10.1186/s40854-022-00441-7

Huck, N. (2019a). Large data sets and machine learning: Applications to statistical arbitrage. *European Journal of Operational Research*, *278*(1), 330–342. https://doi.org/10.1016/j.ejor.2019.04.013

Huck, N. (2019b). Large data sets and machine learning: Applications to statistical arbitrage. *European Journal of Operational Research*, *278*(1), 330–342. https://doi.org/10.1016/j.ejor.2019.04.013

Israel, R., Kelly, B. T., & Moskowitz, T. J. (2020). Can Machines “Learn” Finance? *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3624052

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An Introduction to Statistical Learning*. Springer US. https://doi.org/10.1007/978-1-0716-1418-1

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

Junqué de Fortuny, E., De Smedt, T., Martens, D., & Daelemans, W. (2014). Evaluating and understanding text-based stock price prediction models. *Information Processing & Management*, *50*(2), 426–441. https://doi.org/10.1016/j.ipm.2013.12.002

Kaul, G. (1987). Stock returns and inflation. *Journal of Financial Economics*, *18*(2), 253–276. https://doi.org/10.1016/0304-405X(87)90041-9

Kim, K. (2003). Financial time series forecasting using support vector machines. *Neurocomputing*, *55*(1–2), 307–319. https://doi.org/10.1016/S0925-2312(03)00372-2

Kim, Y. (2006). Toward a successful CRM: variable selection, sampling, and ensemble. *Decision Support Systems*, *41*(2), 542–553. https://doi.org/10.1016/j.dss.2004.09.008

Krauss, C., Do, X. A., & Huck, N. (2017a). Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. *European Journal of Operational Research*, *259*(2), 689–702. https://doi.org/10.1016/j.ejor.2016.10.031

Kryzanowski, L., Galler, M., & Wright, D. W. (1993). Using Artificial Neural Networks to Pick Stocks. *Financial Analysts Journal*, *49*(4), 21–27. https://doi.org/10.2469/faj.v49.n4.21

Lakonishok, J., Shleifer, A., & Vishny, R. W. (1994). Contrarian Investment, Extrapolation, and Risk. *The Journal of Finance*, *49*(5), 1541–1578. https://doi.org/10.1111/j.1540-6261.1994.tb04772.x

LEE, B.-S. (1992). Causal Relations Among Stock Returns, Interest Rates, Real Activity, and Inflation. *The Journal of Finance*, *47*(4), 1591–1603. https://doi.org/10.1111/j.1540-6261.1992.tb04673.x

Lesmond, D. A., Schill, M. J., & Zhou, C. (2004). The illusory nature of momentum profits. *Journal of Financial Economics*, *71*(2), 349–380. https://doi.org/10.1016/S0304-405X(03)00206-X

Leung, M. T., Daouk, H., & Chen, A.-S. (2000). Forecasting stock indices: a comparison of classification and level estimation models. *International Journal of Forecasting*, *16*(2), 173–190. https://doi.org/10.1016/S0169-2070(99)00048-5

Lewellen, J. (2014). The Cross Section of Expected Stock Returns. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2511246

Li, Y., Simon, Z., & Turkington, D. (2022). Investable and Interpretable Machine Learning for Equities. *The Journal of Financial Data Science*, *4*(1), 54–74. https://doi.org/10.3905/jfds.2021.1.084

Lo, A. W., Mamaysky, H., & Wang, J. (2000). Foundations of Technical Analysis: Computational Algorithms, Statistical Inference, and Empirical Implementation. *The Journal of Finance*, *55*(4), 1705–1765. https://doi.org/10.1111/0022-1082.00265

Long, W., Lu, Z., & Cui, L. (2019). Deep learning-based feature engineering for stock price movement prediction. *Knowledge-Based Systems*, *164*, 163–173. https://doi.org/10.1016/j.knosys.2018.10.034

Malkiel, B. G. (2003). The Efficient Market Hypothesis and Its Critics. *Journal of Economic Perspectives*, *17*(1), 59–82. https://doi.org/10.1257/089533003321164958

Malkiel, B. G., & Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, *25*(2), 383–417. https://doi.org/10.1111/j.1540-6261.1970.tb00518.x

Marcos Lopez de Prado. (2018). *Advances in Financial Machine Learning*. O’Reilly.

Molnar, C., Casalicchio, G., & Bischl, B. (2020). *Interpretable Machine Learning – A Brief History, State-of-the-Art and Challenges* (pp. 417–431). https://doi.org/10.1007/978-3-030-65965-3_28

Moritz, B., & Zimmermann, T. (2016). Tree-Based Conditional Portfolio Sorts: The Relation between Past and Future Stock Returns. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2740751

Mullainathan, S., & Spiess, J. (2017). Machine Learning: An Applied Econometric Approach. *Journal of Economic Perspectives*, *31*(2), 87–106. https://doi.org/10.1257/jep.31.2.87

Nikou, M., Mansourfar, G., & Bagherzadeh, J. (2019). Stock price prediction using DEEP learning algorithm and its comparison with machine learning algorithms. *Intelligent Systems in Accounting, Finance and Management*, *26*(4), 164–174. https://doi.org/10.1002/isaf.1459

Poterba, J. M., & Summers, L. H. (1988). Mean reversion in stock prices. *Journal of Financial Economics*, *22*(1), 27–59. https://doi.org/10.1016/0304-405X(88)90021-9

Qi, M., & Maddala, G. S. (1999). Economic factors and the stock market: a new perspective. *Journal of Forecasting*, *18*(3), 151–166. https://doi.org/10.1002/(SICI)1099-131X(199905)18:3<151::AID-FOR716>3.0.CO;2-V

Rasekhschaffe, K. C., & Jones, R. C. (2019). Machine Learning for Stock Selection. *Financial Analysts Journal*, *75*(3), 70–88. https://doi.org/10.1080/0015198X.2019.1596678

Ratanapakorn, O., & Sharma, S. C. (2007). Dynamic analysis between the US stock returns and the macroeconomic variables. *Applied Financial Economics*, *17*(5), 369–377. https://doi.org/10.1080/09603100600638944

Rechenthin, M., Street, W. N., & Srinivasan, P. (2013). Stock chatter: Using stock sentiment to predict price direction. *Algorithmic Finance*, *2*(3–4), 169–196. https://doi.org/10.3233/AF-13025

Rosenberg, B., Reid, K., & Lanstein, R. (1985). Persuasive evidence of market inefficiency. *The Journal of Portfolio Management*, *11*(3), 9–16. https://doi.org/10.3905/jpm.1985.409007

Schöneburg, E. (1990). Stock price prediction using neural networks: A project report. *Neurocomputing*, *2*(1), 17–27. https://doi.org/10.1016/0925-2312(90)90013-H

Smolander, J., Dehmer, M., & Emmert-Streib, F. (2019). Comparing deep belief networks with support vector machines for classifying gene expression data from complex disorders. *FEBS Open Bio*, *9*(7), 1232–1248. https://doi.org/10.1002/2211-5463.12652

Tobek, O., & Hronec, M. (2021a). Does it pay to follow anomalies research? Machine learning approach with international evidence. *Journal of Financial Markets*, *56*, 100588. https://doi.org/10.1016/j.finmar.2020.100588

Virtanen, I., & Yli-Olli, P. (1987a). Forecasting stock market prices in a thin security market. *Omega*, *15*(2), 145–155. https://doi.org/10.1016/0305-0483(87)90029-6

Wang, H., Ahluwalia, H. S., Aliaga-Díaz, R. A., & Davis, J. H. (2021). The Best of Both Worlds: Forecasting US Equity Market Returns Using a Hybrid Machine Learning–Time Series Approach. *The Journal of Financial Data Science*, *3*(2), 9–20. https://doi.org/10.3905/jfds.2021.3.2.009

Wang, W., Li, W., Zhang, N., & Liu, K. (2020). Portfolio formation with preselection using deep learning from long-term financial data. *Expert Systems with Applications*, *143*, 113042. https://doi.org/10.1016/j.eswa.2019.113042

Welch, I., & Goyal, A. (2008). A Comprehensive Look at The Empirical Performance of Equity Premium Prediction. *Review of Financial Studies*, *21*(4), 1455–1508. https://doi.org/10.1093/rfs/hhm014

Weng, B., Lu, L., Wang, X., Megahed, F. M., & Martinez, W. (2018). Predicting short-term stock prices using ensemble methods and online data sources. *Expert Systems with Applications*, *112*, 258–273. https://doi.org/10.1016/j.eswa.2018.06.016

Zhao, Z., Anand, R., & Wang, M. (2019). Maximum Relevance and Minimum Redundancy Feature Selection Methods for a Marketing Machine Learning Platform. *2019 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, 442–452. https://doi.org/10.1109/DSAA.2019.00059

8. Appendix

Table 8.1: Full overview of predictor variables

	Feature	Explanation	Found in
Daily Predictions			
	Returns	Lagged returns over different intervals the last trading year	(Krauss et al., 2017)
	EMA10	10-period exponential moving average	-
	MACD	Moving average convergence divergence	-
	ROC	Rate of change	-
	RSI	Relative strength index	-
Monthly Predictions			
		Fundamentals	
	size	Stock price times shares outstanding	(Banz, 1981)
	eps	Earnings per share	-
	ps	Price to sales	(Barbee et al., 1996)
	pe	Price to earnings	(Basu, 1977)
	pb	Price to book	(Brennan et al.,1998)
	ev_ebitda	ev to ebitda	-
	roe	Return on equity	(Chen et al., 2011)
	roa	Return on assets	(Balakrishnan et al.,2010)
	beta	Beta relative to index	(Fama & Macbeth, 1973)
	beta2	Beta squared	(Fama & Macbeth, 1973)
	cash	Cash holdings	(Palazzo, 2012)
	cf	Free cash flow	-
	stdcf	Cash flow volatility	(Huang, 2009)
	sale_cash	Sales to cash	(Ou & Penman, 1989)
	sale_g	Sales growth	(Lakonishok, 1994)
	cr	Current ratio	(Ou & Penman, 1989)
	qr	Quick ratio	(Ou & Penman, 1989)
	rd_mc	R&D to market capitalization	(Guo et al., 2006)
	rd_s	R&D to sales	(Guo et al., 2006)
	rd_i	Increase in R&D	(Eberhart et al., 2004)
	mc_cf	Market cap to free cash flow	-
	shar_c	Change in shares outstanding	(Pontiff et al., 2008)
	div_p	Dividend to price	(Litzenberger et al.,1979)
	turn	Turnover	(Datar et al., 1998)
	turn_std	Liquidity volatility	(Chordia et al., 2001)
		Price	
	m12	12- month momentum	(Jegadeesh, 1990)
	m6	6- month momentum	(Jegadeesh et al., 1993)
	m3	3- month momentum	(Jegadeesh et al., 1993)
	m1	1- Month momentum	(Jegadeesh et al., 1993)
	mmr	Max monthly return	(Bali et al., 2011)
	m12_c	Change in 12-month momentum	(Gettleman et al., 2006)
	m6_c	Change in 6-month momentum	(Gettleman et al., 2006)

m3_c	Change in 3-month momentum	(Gettleman et al., 2006)
m1_c	Change in 1-month momentum	(Gettleman et al., 2006)
m_ind	Industry momentum	(Moskowitz et al., 1999)
ret_vol	Return volatility	(Ang et al., 2006)
	Macroeconomics	
nr	Three-month NIBOR rate	(Gjerde & Sættem, 1999)
gov_10y	10-year government bond yields	(Welch & Goyal, 2008)
gov_5y	5-year government bond yields	(Welch & Goyal, 2008)
gov_3y	3-year government bond yields	(Welch & Goyal, 2008)
gov_1y	1-year government bond yields	(Welch & Goyal, 2008)
nok_usd	NOK to USD currency rate	-
gp	Gold price	(Jones and Kaul, 1996)
infl	Inflation measured by the Consumer Price Index (CPI)	(Welch & Goyal, 2008)
oil	Spot prices on Arabian Light crude oil	(Gjerde & Sættem, 1999)
oil_v	Oil trading volume	-
	Other	
ind	Industry classifier	(Hong et al., 2007)
dw	Day of the week	(Gibbons et al., 1981)
my	Month of the year	(Ariel, 1990)

Table 8.2: Statistics for daily portfolio

Daily Portfolio

		Before Transaction Costs			After Transaction Costs		
		LR	RF	GBT	LR	RF	GBT
$k = 5$	Mean return (long)	0.0012	**0.0022**	0.00187	0.0002	**0.0010**	0.0006
	Mean return (short)	0.0010	**0.0011**	0.00094	**0.0001**	0.0000	-0.0002
	Mean return	0.0011	**0.0017**	0.0014	0.0001	**0.0004**	0.0002
	Minimum	-0.0679	-0.0664	**-0.04092**	-0.0697	-0.0678	**-0.0429**
	Max	0.0738	**0.0914**	0.07165	0.0734	**0.0909**	0.0712
	Standard deviation	0.0093	**0.0085**	**0.0085**	0.0093	**0.0085**	**0.0085**
	Standard error	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
	t-statistic (NW)	8.47	**13.96**	12.24	0.79	**3.73**	1.40
	Skewness	0.04	0.37	0.46	0.02	0.38	0.46
	Kurtosis	4.35	6.79	4.38	4.42	6.88	4.44
$k = 10$	Mean return (long)	0.00097	**0.0016**	0.0015	**-0.0003**	**-0.0003**	-0.0005
	Mean return (short)	0.00042	0.0007	**0.0007**	**-0.0011**	-0.0012	-0.0012
	Mean return	0.00069	**0.0013**	0.0011	**-0.0007**	**-0.0007**	-0.0008
	Minimum	-0.04822	-0.0401	**-0.0277**	-0.0504	-0.0434	**-0.0292**
	Max	0.06112	0.0484	**0.0646**	0.0606	0.0479	**0.0641**
	Standard deviation	0.00652	**0.0060**	0.0060	0.0065	**0.0060**	0.0061
	Standard error	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
	t-statistic (NW)	7.67	**13.87**	13.29	**-7.90**	-9.01	-10.38
	Skewness	0.36	0.35	0.63	0.31	0.36	0.64
	Kurtosis	4.80	3.83	5.19	4.89	3.85	5.25

Table 8.3: Statistics for monthly portfolio

Monthly Portfolio

		Before Transaction Costs			After Transaction Costs		
		LR	**RF**	**GBT**	**LR**	**RF**	**GBT**
$k = 5$	Mean return (long)	0.0189	**0.0271**	0.0261	0.01379	**0.0220**	0.0207
	Mean return (short)	0.0089	0.0050	**0.0096**	**0.0034**	-0.0006	**0.0034**
	Mean return	0.0139	0.0161	**0.0178**	0.0086	0.0107	**0.0130**
	Minimum	-0.1334	-0.1356	**-0.0949**	-0.1399	-0.1458	**-0.1009**
	Max	0.2319	**0.4551**	0.1557	0.2247	**0.4515**	0.1497
	Standard deviation	0.0506	0.0544	**0.0451**	0.0504	0.0544	**0.0449**
	Standard Error	0.0032	0.0035	0.0029	0.0032	0.0035	0.0029
	t-statistic (NW)	4.24	4.56	**6.10**	2.63	3.03	**4.14**
	Skewness	0.59	2.34	0.22	0.58	2.34	0.20
	Kurtosis	2.04	17.62	0.14	2.07	17.96	0.18
$k = 10$	Mean return (long)	0.0121	**0.0234**	0.0208	0.0040	**0.0144**	0.0122
	Mean return (short)	**0.0044**	0.0006	0.0027	**-0.0043**	-0.0085	-0.0069
	Mean return	0.0083	**0.0120**	0.0118	-0.0001	**0.0029**	0.0026
	Minimum	-0.0835	-0.0898	**-0.0750**	-0.0880	-0.1098	**-0.0816**
	Max	0.1643	**0.2509**	0.1360	0.1530	**0.2419**	0.1280
	Standard deviation	0.0380	0.0376	**0.0332**	0.0381	0.0380	**0.0335**
	Standard error	0.0024	0.0024	0.0021	0.0024	0.0024	**0.0021**
	t-statistic (NW)	3.37	4.92	**5.47**	-0.07	1.19	**1.21**
	Skewness	0.58	1.17	0.32	0.53	1.09	0.24
	Kurtosis	1.30	6.85	1.17	1.23	6.82	1.15

Table 8.4: Winners and losers for the long and short portfolios

	Monthly		Daily	
	Stock	*Return*	*Stock*	*Return*
Long Portfolio				
Winners	Akastor	175.08%	Kongsberg Gruppen	362.48%
	TGS	173.50%	Ekornes	302.34%
	NRC Group	172.25%	SpareBank 1 SR-Bank	299.90%
Losers	Ensurge Micropower	-111.80%	Norwegian Air Shuttle	-132.73%
	XXL	-91.04%	Magnora	-68.83%
	Aker Solutions	-65.88%	PCI Biotech	-48.47%
Short Portfolio				
Winners	StrongPoint	131.78%	Atea	250.53%
	Norwegian Air Shuttle	122.19%	Petrolia	193.74%
	Arribatec Group	86.33%	Equinor	95.35%
Losers	Petrolia	-254.54%	Royal Crbn. Cruises	-131.40%
	Aker BP	-181.10%	Tandberg	-96.27%
	Wilh Wilhelmsen	-81.57%	Opticom	-75.02%

Note: These percentages are acquired by taking the sum of return contribution.

NTNU

Norwegian University of
Science and Technology