Executive Summary

The technical advancements in recent years have changed and shaped our everyday life in crucial ways. New methods in various scientific fields have emerged over the years. Machine learning is already a rather old technique as one of the first people to use this term was Arthur L. Samuel from IBM, who developed a game of checkers that could learn to play better. He published one of the first papers on machine learning in 1959 (Samuel (1959)). However, machine learning gained quite some momentum over the past few years. Thus, various research was done around machine learning and its potential in different fields was further explored. For this thesis, we enter the field of finance. Gu et al. (2019) show in their pioneering paper how different machine learning methods can be used to predict asset returns. Various papers which build on the findings of Gu et al. (2019) were published afterwards.

But in what kind of setups do machine learning methods perform best? Are there some methods which are superior to others in certain settings? Hence, which machine learning method should be chosen to predict asset returns? And in which particular setting should this choice be made? To answer those questions, I implement a Monte Carlo simulation which is based on the proposed data generating process of Gu et al. (2019). By deviating and adjusting various parameters in the data generating process, I test different machine learning algorithms such as OLS, LASSO, RIDGE, random forest and neural networks in different setups.

My findings suggest that from the linear methods, LASSO performs best in most cases. However, if the signal-to-noise ratio becomes very large, the other linear machine learning methods, namely

OLS and RIDGE, come very close to the predictive performance of LASSO. If we leave the linear environment, the performance of the linear models worsens greatly. Hence, we have to rely on more complex models such as random forest or neural networks. Those perform well in linear as well as non linear settings. But if the signal-to-noise ratio is very low, even the performance of the more complex machine learning methods becomes rather poor. However, the predictions for asset returns become very reliable if a lot of covariates are driving the asset returns and the signal-to-noise ratio of these covariates is large.

Thus, machine learning is a valuable tool in the field of finance to predict asset returns. There is still a great potential for research in this field, as the current literature only scratches the surface of what may be possible with machine learning methods. Further research in this field may not only be interesting for academics but also for banks, insurance companies and asset managers in general.