

SKLearn ML Toolbox for Making Money on the Financial Index OMXS30

Technical Report

Autor: Alexander Wendt

**Document history**

|  |  |  |  |
| --- | --- | --- | --- |
| Version | Date | Changes | Authors |
| 0.1 | 2022-01-22 | Created initial template | Alexander Wendt |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

Table of Contents

[1. Introduction 1](#_Toc91755316)

[1.1 Motivation 1](#_Toc91755317)

[1.2 Problem Statement 1](#_Toc91755318)

[1.3 Task Setting 1](#_Toc91755319)

[1.4 Methodology 1](#_Toc91755320)

[2. OMXS30 Long-Term Trend Automatically Generated 3](#_Toc91755321)

[2.1 Problem Statement 3](#_Toc91755322)

[2.2 Raw Dataset 3](#_Toc91755323)

[2.3 Generate Features and Outcomes 4](#_Toc91755324)

[2.4 Preparation and Analysis 6](#_Toc91755325)

[2.4.1 Cleaning data 6](#_Toc91755326)

[2.4.2 Correlation and Important Features 6](#_Toc91755327)

[2.5 Support Vector Machine Training and Evaluation 9](#_Toc91755328)

[2.6 XGBoost Training and Evalution 9](#_Toc91755329)

[2.7 Conclusion 10](#_Toc91755330)

[3. Usecases 11](#_Toc91755331)

[Literature 12](#_Toc91755332)

[Internet references 13](#_Toc91755333)

[Annex A 14](#_Toc91755334)

# Introduction

The purpose is to use SKLearn machine learning methods to make predictions about the current trend, tops and bottom and other structures for stock market courses. This document contains the descriptions of the process.

## Motivation

The use case document will first introduce the overall idea behind the structure of the use cases, then give an overview of all use cases and finally describe each use case in detail, including internal variations in form of usage scenarios.

## Problem Statement

## Task Setting

## Methodology

# OMXS30 Long-Term Trend Automatically Generated

## Problem Statement

How good do technical indicators describe current stock market trend?

## Raw Dataset

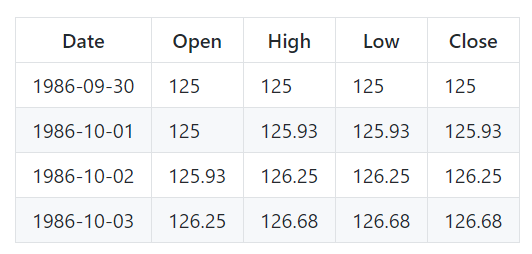


Figure 2.1: Example Figure

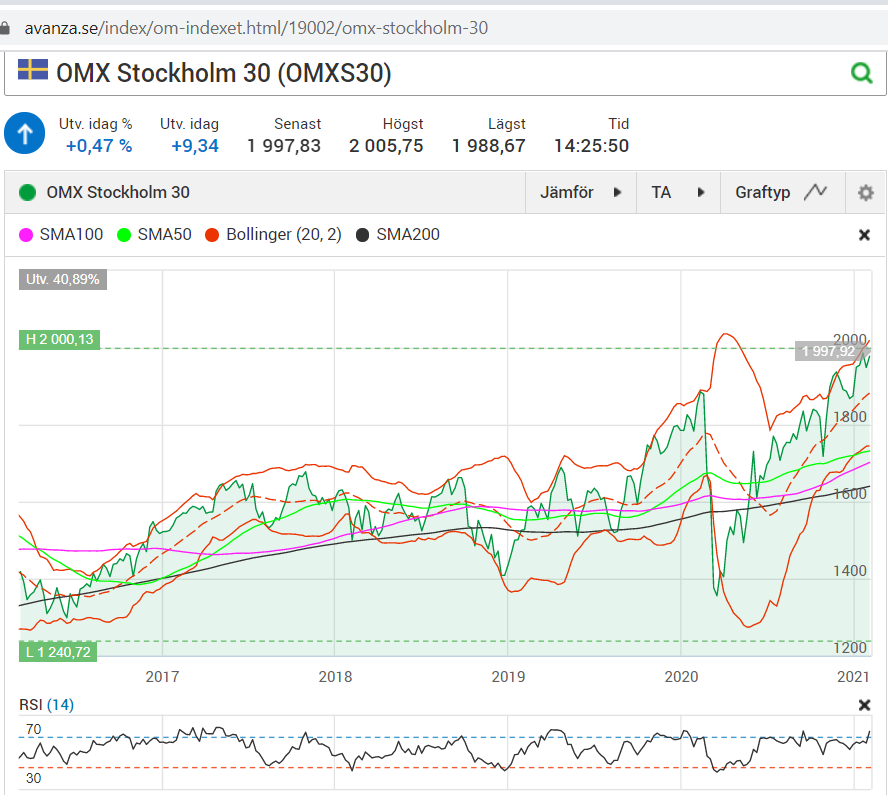


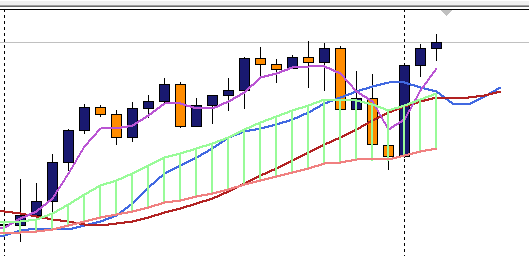
Figure 2.2: Example Figure

## Generate Features and Outcomes

The dataset ground truth was generated with Lowess.

Moving Averages

Oscillators



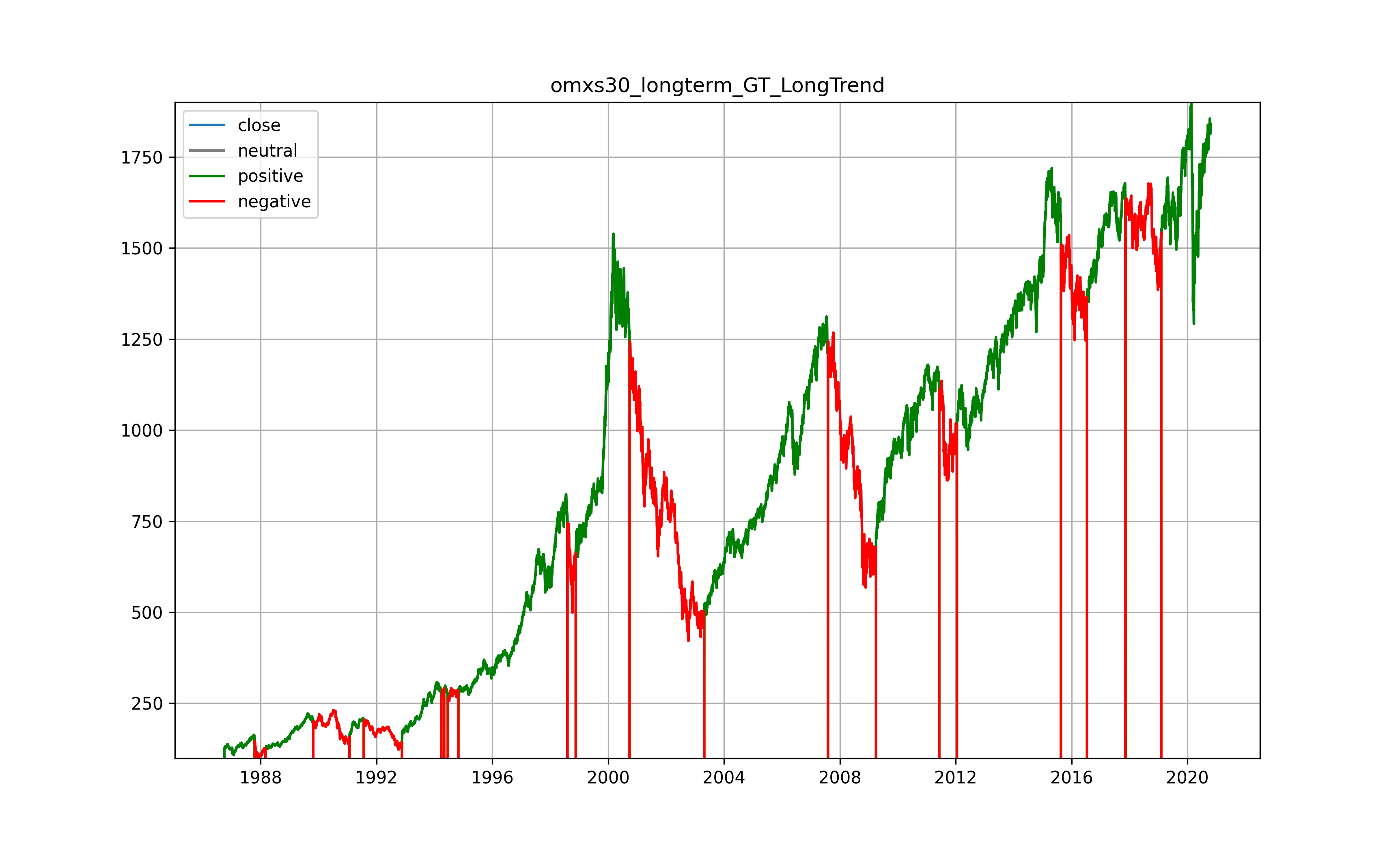


Figure 2.3: Example Figure

| Indicator | Group | Meaning |
| --- | --- | --- |
| day\_of\_month | Cycles |  |
| day\_of\_year | Cycles |  |
| day\_week\_\_0..6 | Cycles |  |
| MACD | Momentum | 12\_26\_9, 5\_35\_5 |
| MACD\_Diff | Momentum |  |
| MACD\_DiffDirChange | Momentum |  |
| month\_year\_\_1..12 | Cycles |  |
| NormKurs 5, 20, 50, 100, 200 | Momentum |  |
| NumberRise 50, 100, 200 | Momentum |  |
| RSI\_2..20 | Momentum |  |
| RSI difference 2..20 | Momentum |  |
| SMA 5..200 | Trend |  |
| SMA Difference 2..200 | Trend |  |
| Stoch\_Sd1338 | Momentum | 533, 1338, 1433, 211414, 2177 |
| Stoch\_Sk1338 | Momentum | 533, 1338, 1433, 211414, 2177 |
| week\_month\_\_1..6 | Cycles |  |
| week\_of\_year | Cycles |  |

## Preparation and Analysis

### Cleaning data

Cleaning

Column names are stripped

" " are replaced by "\_" to get unified naming

"/" are replaced by "-"

All string values "?" or empty values are replaced by numpy "NaN"

Adaptation

Binarize Labels: One vs. All

Nominals: One-hot-encoding

All to int or float

### Correlations

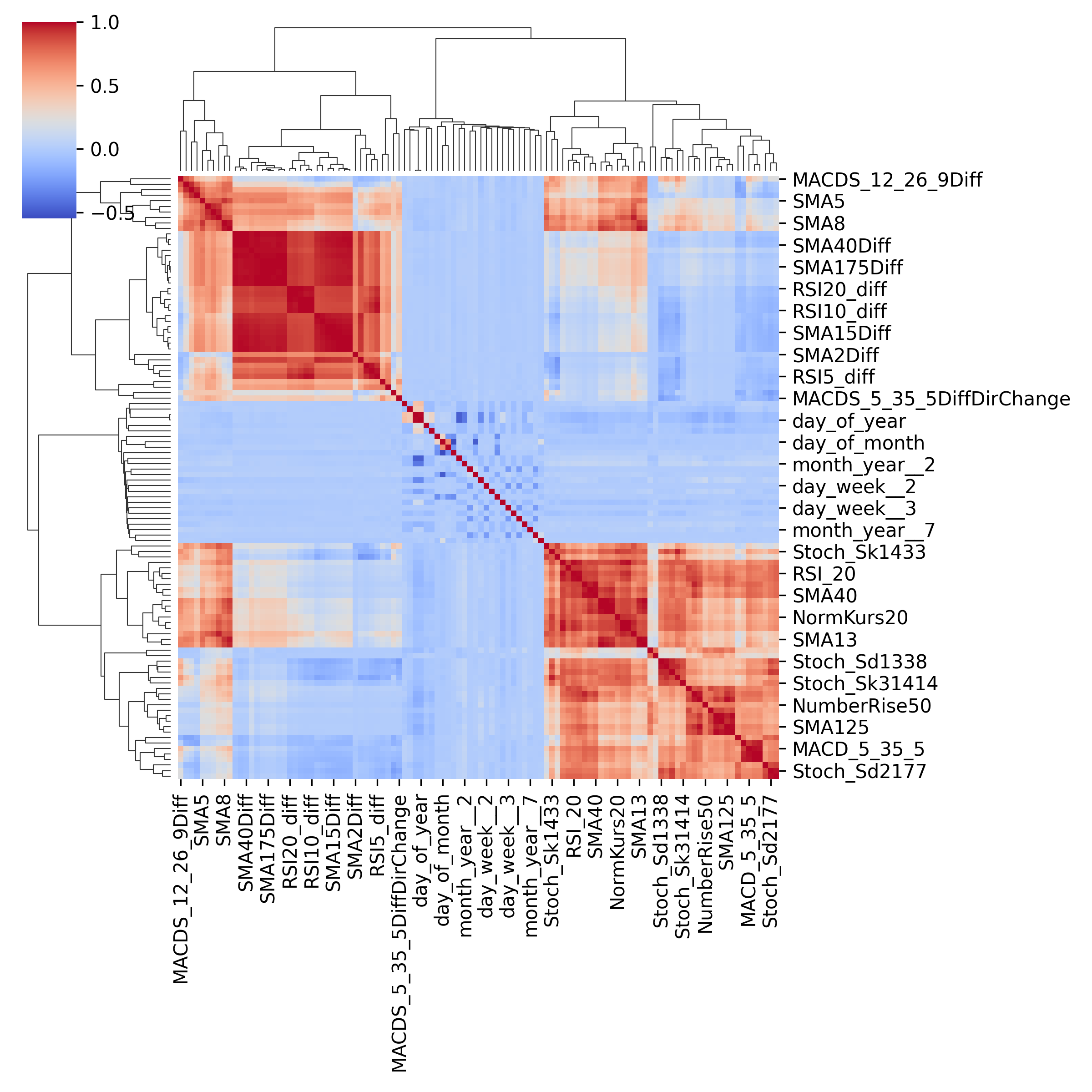


Figure 2.6: Feature Selection: Significant features for different methods, not all listed

### Feature Selection

Visual correlation analysis

See important features

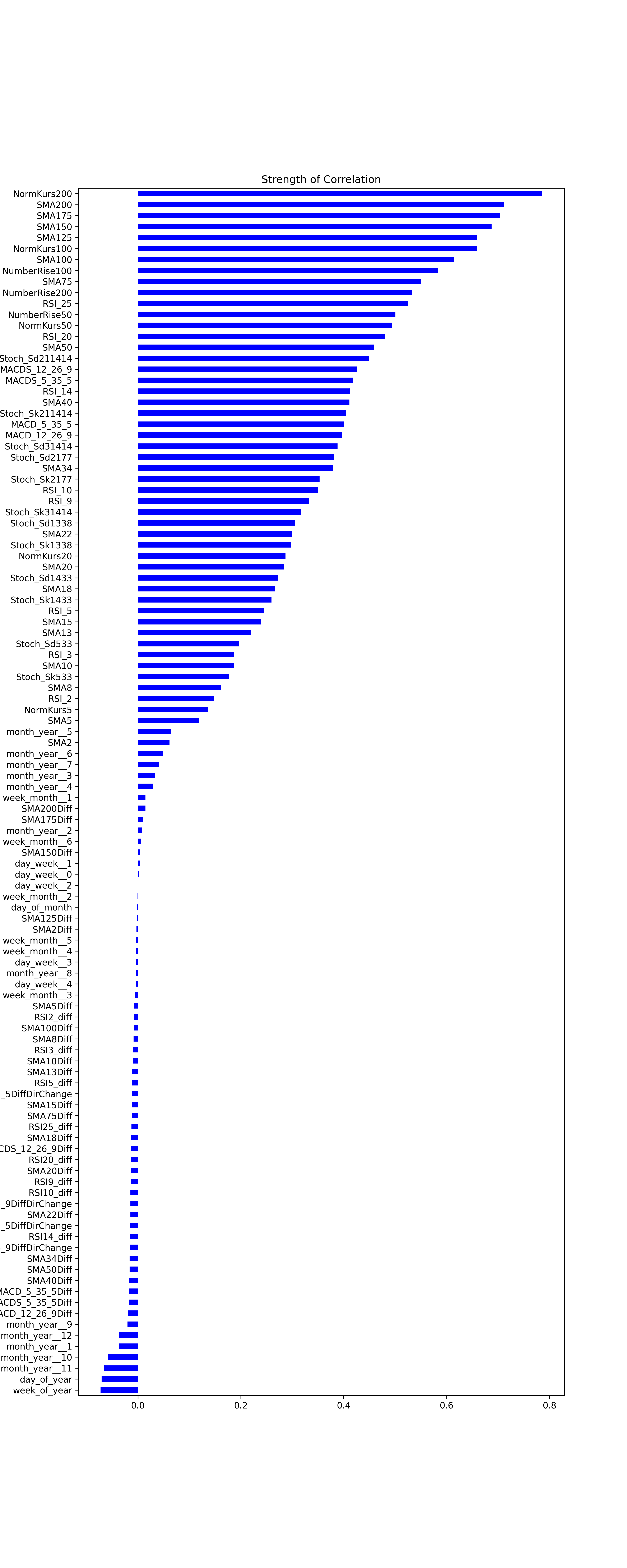


Figure 2.4: Correlation strength of the features with the outcome

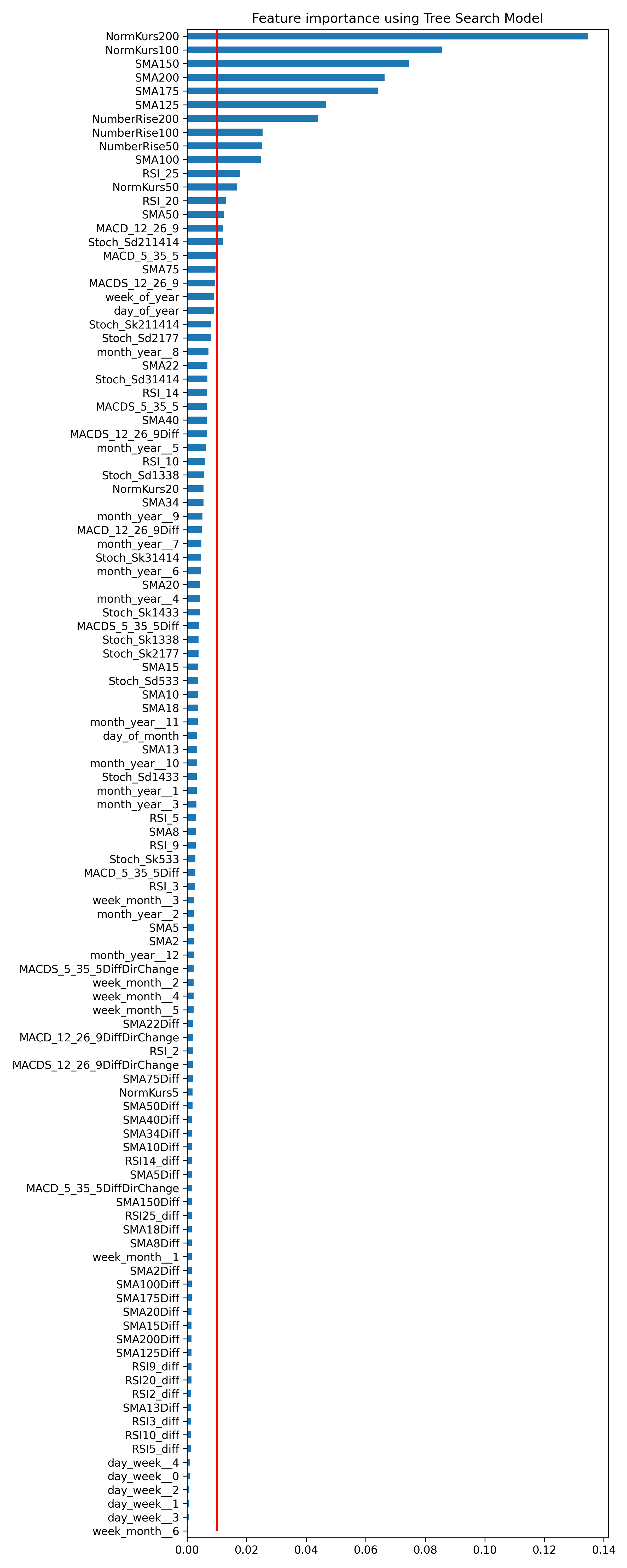
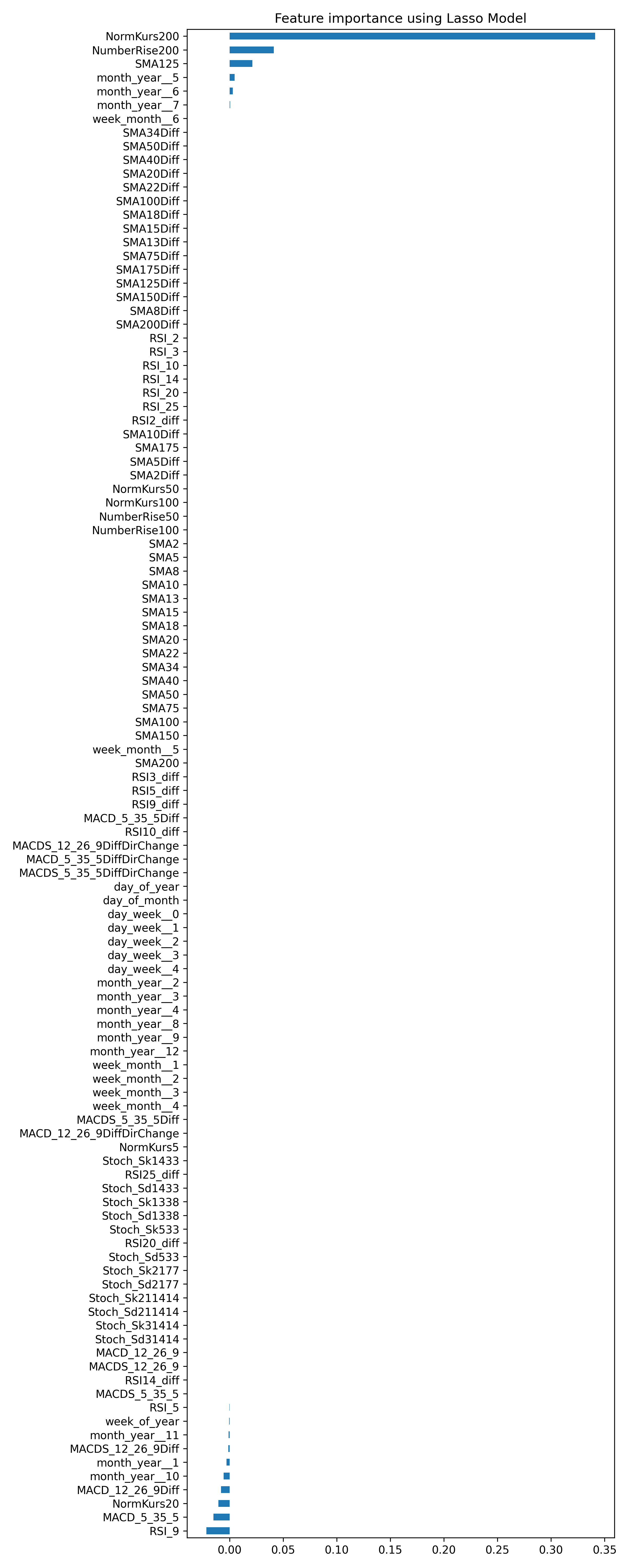


Figure 2.5: Feature Selection: Significant Features for the Lasso Regularization Model and tree-based model



Figure 2.6: Feature Selection: Significant features for different methods, not all listed

## Support Vector Machine Training and Evaluation

## XGBoost Training and Evalution

## Conclusion

Best model

Figure 2.7: Example Figure

# Usecases

Literature

[Acht04] Achterberg, T.: SCIP - a framework to integrate Constraint and Mixed Integer Programming, [ZIB report 04-19](http://www.zib.de/Publications/abstracts/ZR-04-19), Berlin 2004.

[Bar04] Barbose, G., Goldman, C., Neenan B.: A Survey of Utility Experience with Real Time Pricing”, Lawrence Berkeley National Laboratory, Paper LBNL-54238, 2004.

[Bra06] Brauner, G. et al.: Verbraucher als virtuelles Kraftwerk, Berichte aus Energie- und Umweltforschung 44/2006, BmVIT, Vienna, 2006.

[Heb02] Hebner, R., Beno, J., Walls, A.: Flywheel batteries come around again, IEEE spectrum, Volume 39, Issue 4, pp. 46 to 51, April 2002.

[Ish06] Ishida, S., Roesener, C., Ichimura, J., Nishi, H., Implementation of Internet based Demand Controlling System, Proc. of 6th IEEE Int. Conference in Industrial Informatics, Singapore, August 2006.

[Kil06] Kiliccote, S., Piette, M. A., Hansen, D.: Advanced Controls and Communications for Demand Response and Energy Efficiency in Commercial Buildings, Proc. of Second Carnegie Mellon Conference in Electric Power Systems, Pittsburgh PA. LBNL-59337, January 2006.

[Koch04] Koch, T.: Rapid Mathematical Programming, PhD Thesis, Technische Universität Berlin, 2004.

[Law03] Lawrence, D.J.; Neenan, B.F.: The status of demand response in New York, [Power Engineering Society General Meeting, 2003, IEEE](http://ieeexplore.ieee.org/xpl/RecentCon.jsp?punumber=8949), Volume 4, pp 13–17, July 2003.

[Nie06] Nieuwenhout, F. D. J., Hommelberg, M. P. F., Schaefler, G. J., Kester J. C. P., Visscher, K., Feasibility of distributed electrictity storage, International Journal of Distributed Energy Resources, Volume 2 Number 4, October – December 2006, pp 307 – 323.

[Pal01] Palenksy, P.: Distributed Reactive Energy Management, Doctoral Thesis, TU Vienna, 2001

[Sch05] Schreiber, M., Whitehead, A. H., Harrer, M., Moser, R.: The vanadium redox battery - an energy reservoir for stand-alone ITS applications along motor and expressways: Proceesings of Intelligent Transportation Systems, pp. 391 to 395, IEEE 2005

[Smi94] Smith, H. L.: DA/DSM directions. An overview of distribution automation and demand-side management with implications of future trends, [Computer Applications in Power, IEEE](http://ieeexplore.ieee.org/xpl/RecentIssue.jsp?punumber=67) Volume 7, Issue 4,  pp. 23–25, October 1994

[Stad06] Stadler, M. Firestone, R. M. Curtil, D. and Marnay, C.:On-Site Generation Simulation with EnergyPlus for Commercial Buildings, ACEEE Summer Study on Energy Efficiency in Buildings, August 13-18, 2006, Pacific Grove, California, ISBN 0-918249-56-2.

[UCTE04] UCTE Operation Handbook, Appendix 1: Load-Frequency-Control and Performance, UCTE Brussels, Belgium 2004

# Internet references

[1] [www.ieee.org](http://www.ieee.org/)

# Annex A