Grouped Coefficient Model

An Explainable ML Model for Fast, Accurate and Understandable Trial-to-Paid Lead Conversion

Agenda

01.
Introduction

02.
Background

03.
Model Proposal

04.
Results

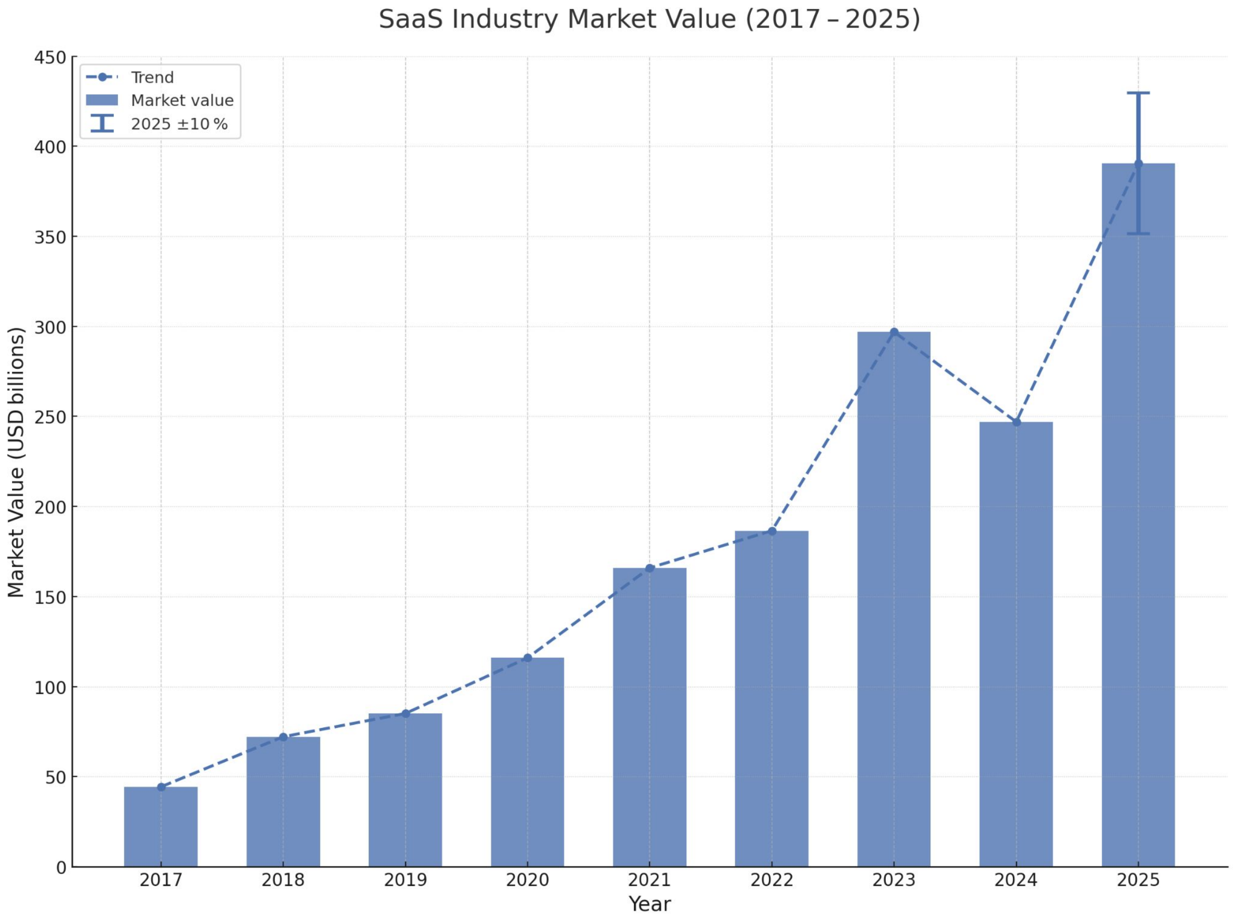
05.
Ablation Study

06.
Discussion

01 Introduction

Customer Acquisition Problem

The Software-as-a-Service (SaaS) is a Huge Market in Information Technology



- SaaS market valued $247 billion in 2024.
- Expected to be $390 billion in 2025.
- Projections say $790 billion by 2030.
- For Comparison: The market for On-Premise software is around $2.248 billion, but declining.

(Vailshery, 2025)

SaaS Is Beginning To Dominate As It Comes With Many but Mostly Financial Benefits

Company	Overall revenue (latest, USD bn)	Estimated SaaS revenue (USD bn)	SaaS share of total
Microsoft	211.9	80–100	38–47 %
Adobe	19.4	12.95	67 %
Salesforce	34.8	34.8	≈ 100 %
Intuit	15	> 7.5	> 50 %
ServiceNow	8.9	8.9	≈ 100 %
Dropbox	2.5	2.5	≈ 100 %
Amazon (AWS)	90	(subset)	n/a

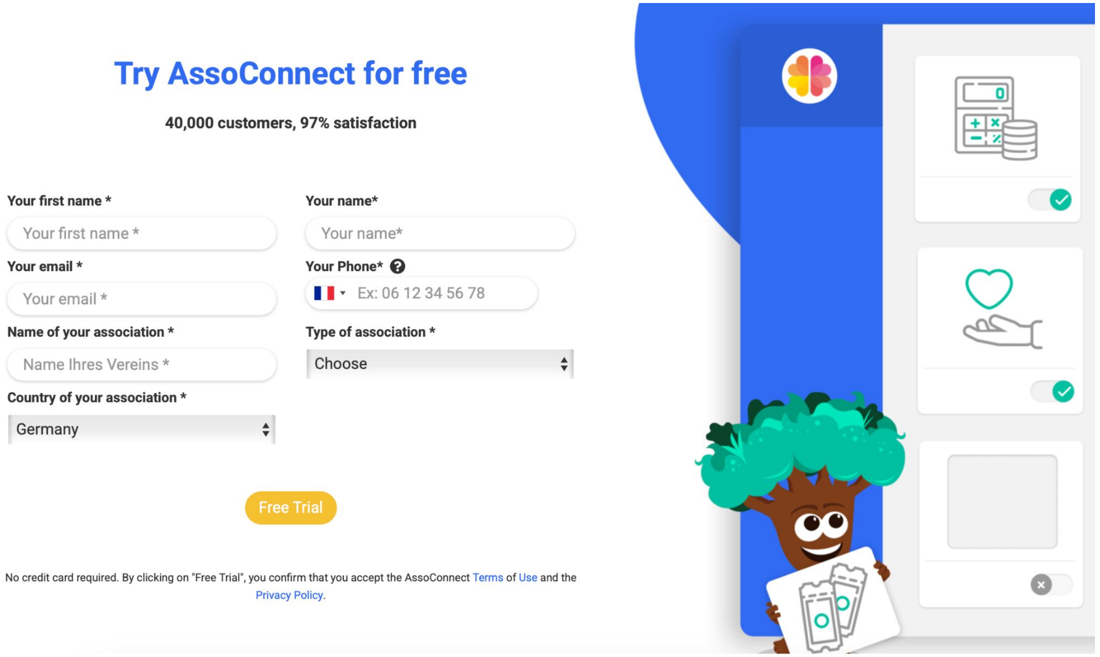
SaaS is superior in software, turning predictable subscriptions into shareholder gold while on-premise solutions fade into obsolescence.

(Howarth, 2025)

02 Background

Predicting Lead Conversion and Model Parsimony

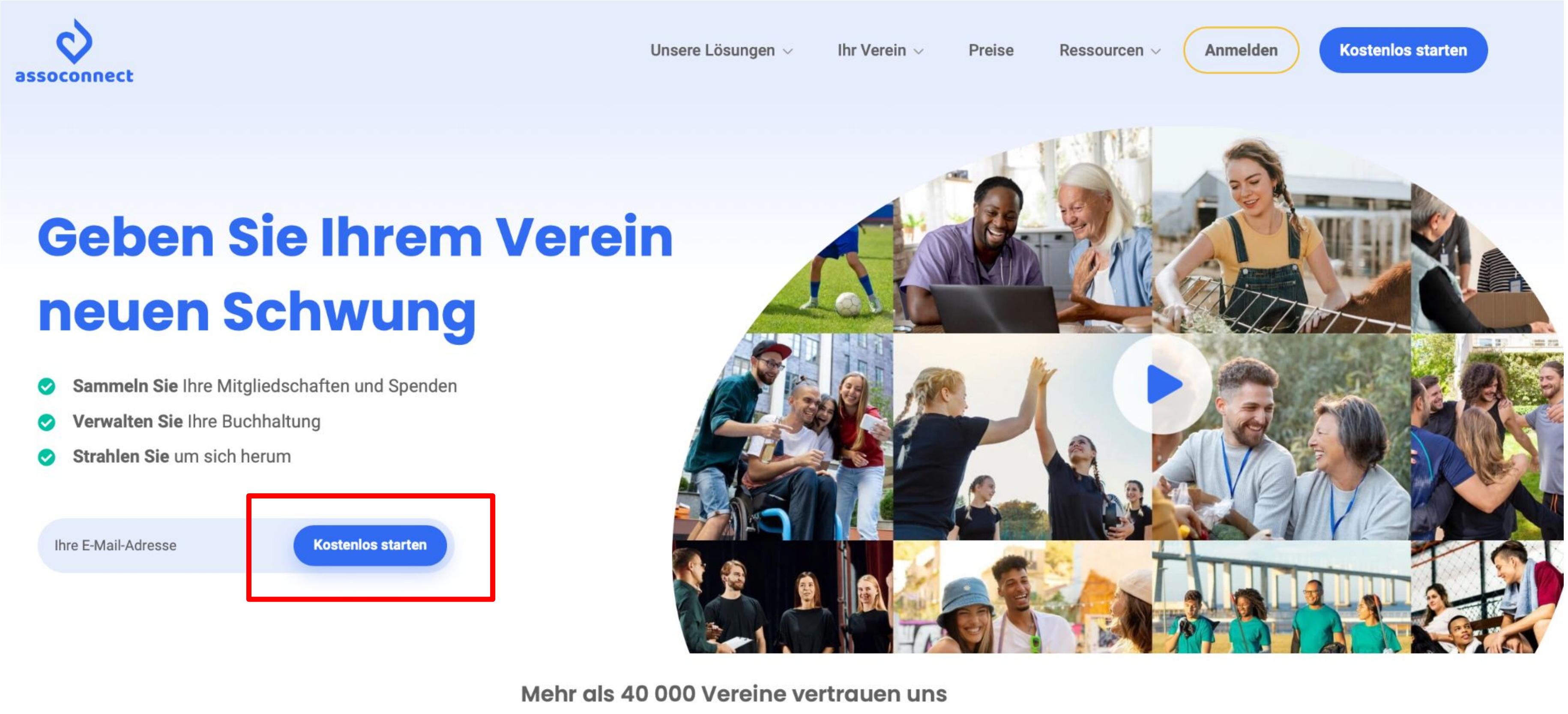
Providers of SaaS Use Simple Sign-Up Forms to Get in Touch with Their Potential Customer



- Extremely simple and fast contact form to get a call
- Often just this information is required:
    - Name
    - E-Mail
    - Phone Number
    - Software Service of Interest

However, SaaS providers know little about the behaviour of their potential customers.

At our Partnering Company Assoconnect, they use a free trial which is targeted at business customers (B2B).



Sales Agents Mostly Use Arbitrary Rules When Reaching Out to a Lead

Communication illustrations by [Storyset]

Predicting Potential Customer Behaviour Helps Support Personalization and Campaign Targeting

Idea:

- Predict Future Customer Behaviour
- Understand Customer Interests
- Support Personalization

Goal:

- Targeted Acquisition Campaigns
- Personalization
- Insights to Drivers of Customer Choices

The Sales Funnel is a CRM Concept which Explains Customer Acquisition and Aims for the Maximum Conversion.

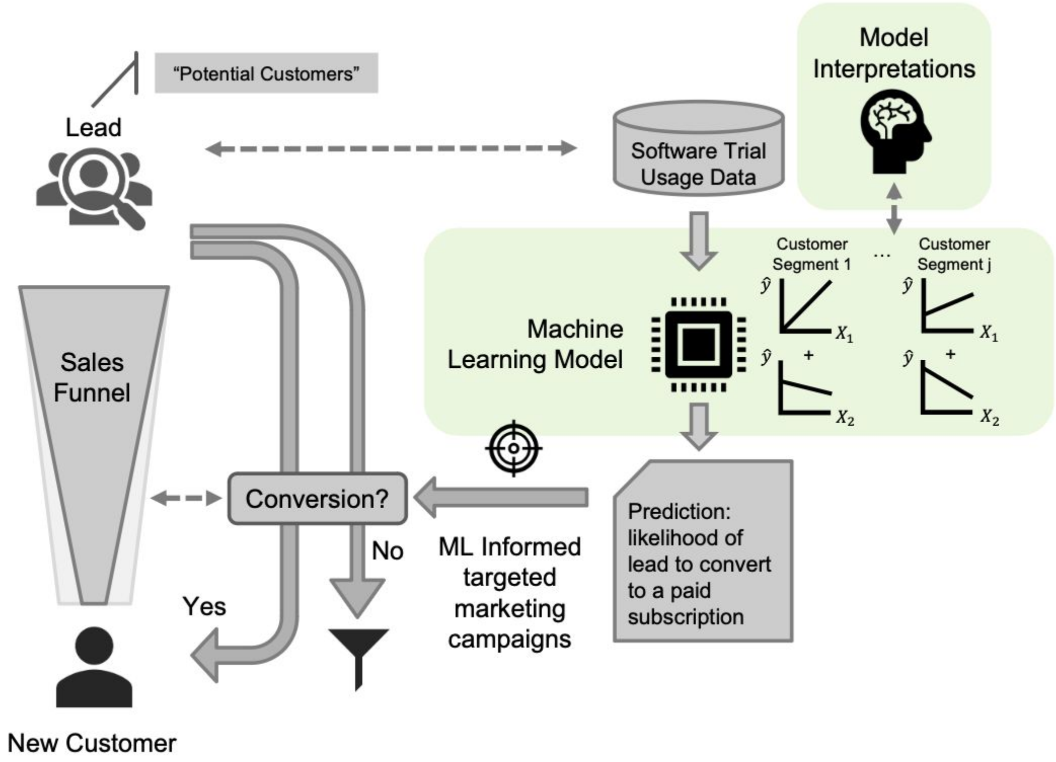


Figure 2: Machine learning enhanced sales funnel: Targeted acquisition of leads and conversion to subscribing customers.

03 Methods

Data at AssoConnect SaaS Is Collected During a 14-Day Free-Trial With a 28-Day Targeted Subscription Campaign

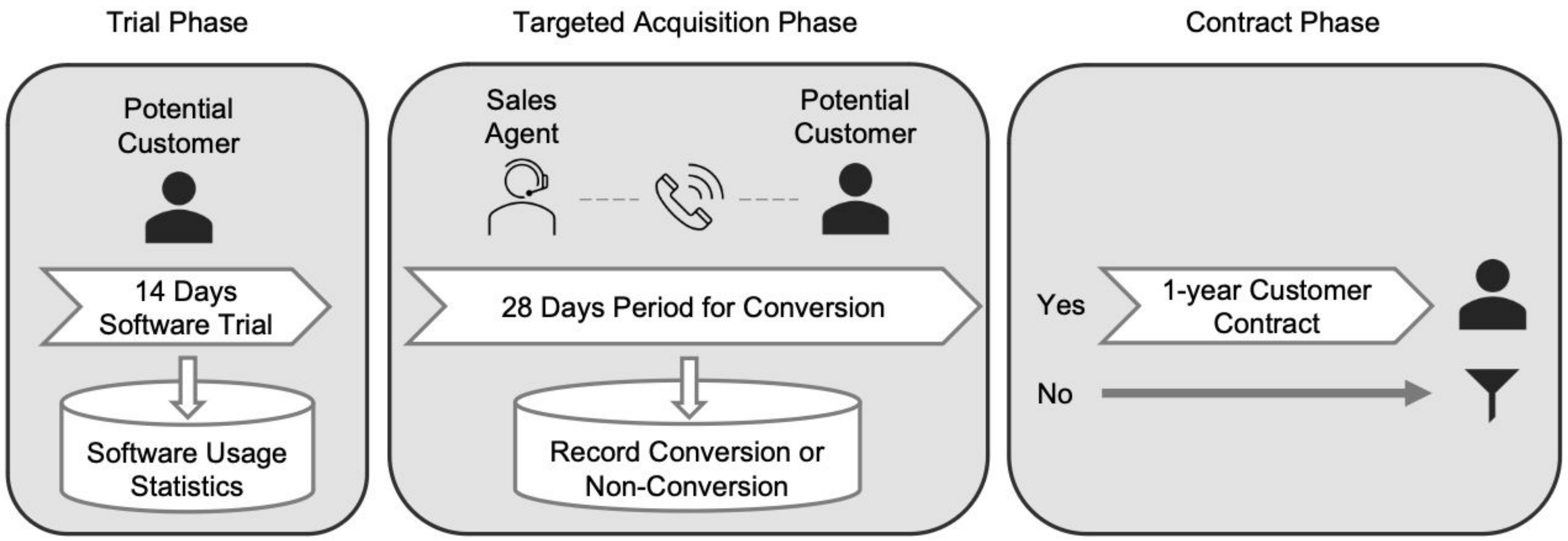


Figure 5: Data collection process. Usage statistics are tracked during a 14-day SaaS free-trial. Conversion is measured up until four weeks after the end of the free-trial.

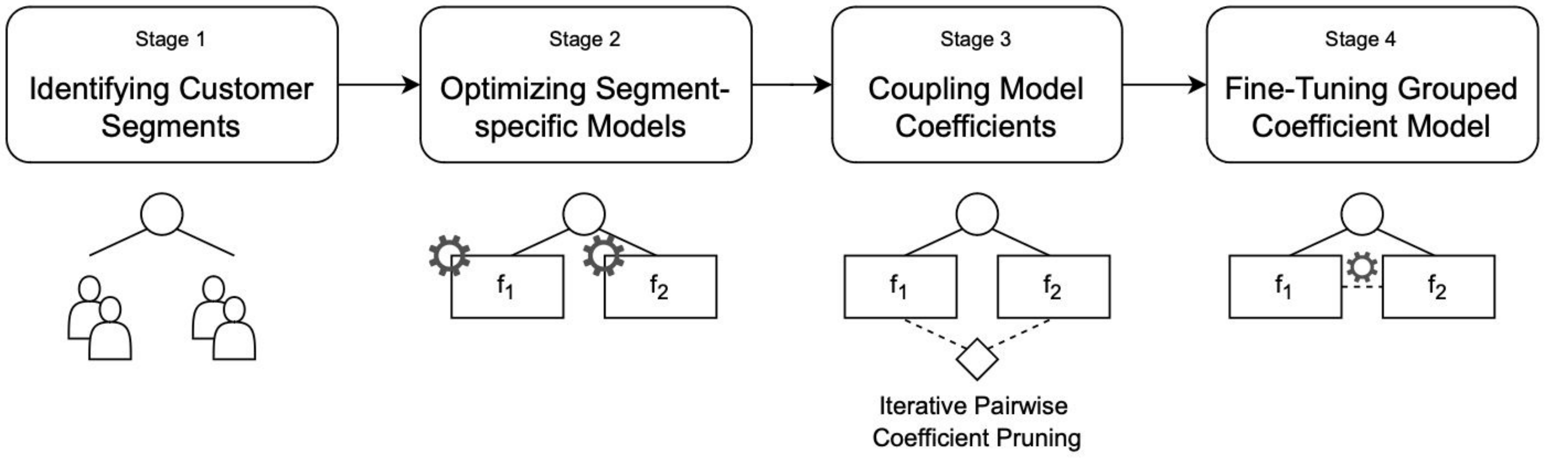
The Data We Received From Assoconnect Is Comprehensive, but Unbalanced and Not Entirely Simple With a Bunch of Missing Values

Table 1: Overview of initial datasets statistics upfront data preprocessing.

Total Samples	Number of Features	Non-Conversion	Conversion	Incidence (%)
22,439	45	21,071	1,368	6.06%

- The Dataset includes 22,439 samples. It has 45 numerical features and no categorical ones.
- 1,368 Potential customers did convert to a paid subscription during the 28-day targeted acquisition phase.
- Additionally, a lot of potential customers did not use the free-trial a lot, so some features are very sparse.

We Propose the GCM Which Models Segments of Data but Combines the Knowledge in a Few Coefficients



The GCM 4-Stage Training Algorithm:

1. First the GCM distinguishes into a few segments of data using a Decision Tree (DT).
2. The GCM trains segment specific linear regression as starting points and then couples coefficients to learn from each others segment.
3. The GCM then iteratively prunes coefficients pairwise.
4. In the End fine-tuning of the coefficients is performed on the train set.

We Developed Our GCM With the Idea of Coefficient Parsimony in Mind

Conceptual Parsimony

- Focus on simplicity in ideas and structure
- The Goal is to have clarity and interpretability
- Useful in communication, e.g. for commercial models

Parametric Parsimony

- Focus on fewest parameters possible
- Goal is to be as efficient in data representation as possible
- Measured is the number of free parameters (“order of parsimony”)
- The emphasis lies on model efficiency and avoidance of overfitting

→ We implemented this idea, but call it Coefficient Parsimony.

(Roberts, 1984)

04 Results

AUROC Results for Assoconnect Lead Conversion

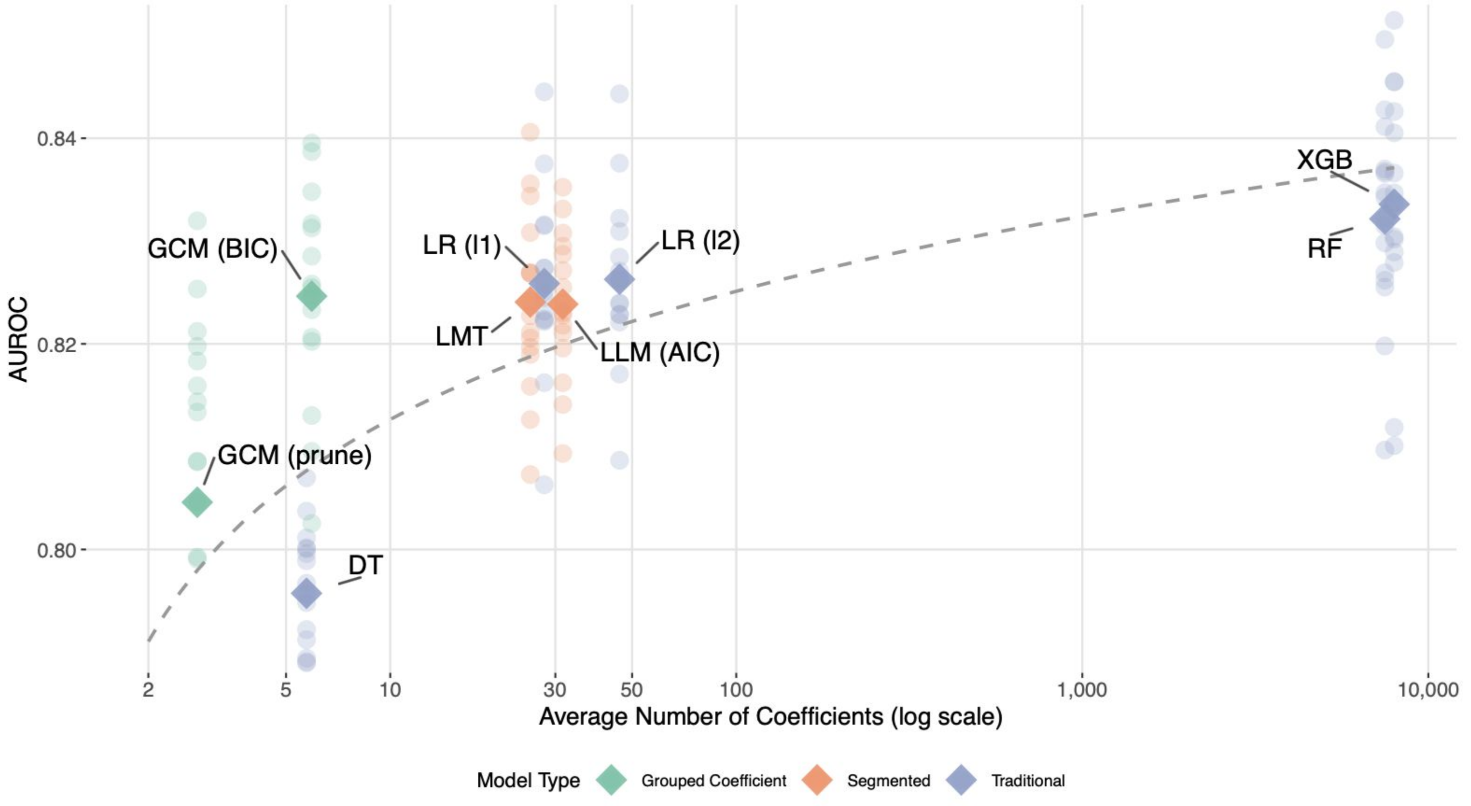


Figure 6: Results of the predictive performance in relation to number of model coefficients under AUROC performance criterion.

TDL Results for Assoconnect Lead Conversion

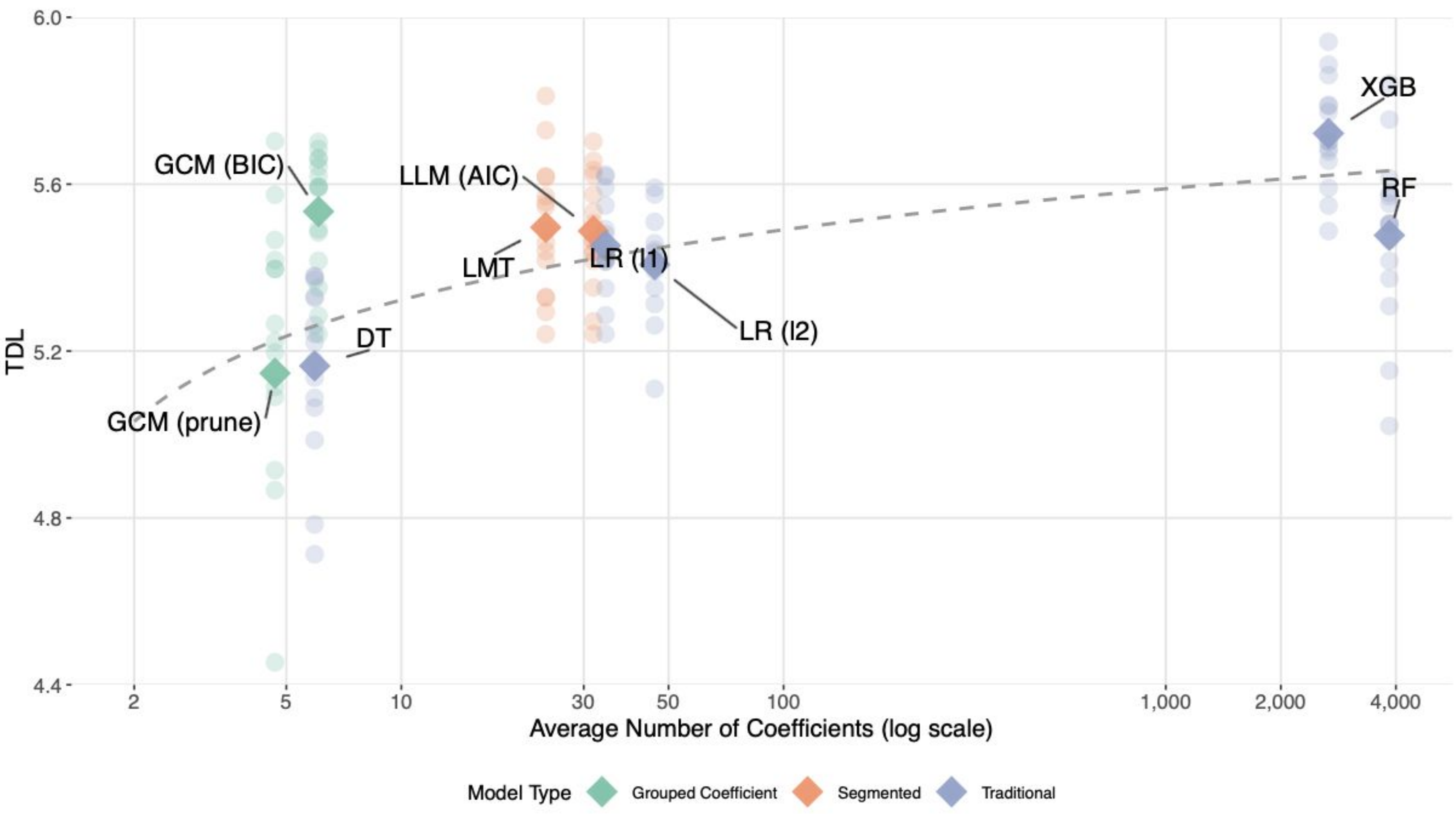


Figure 7: Results of the predictive performance in relation to number of model coefficients under TDL performance criterion.

The Statistical Results Capture the Core Idea of the Paper and the Power of the GCM Approach

Table 3: Average classifier ranks (lower is better) in terms of AUROC and TDL. Holm post-hoc tests compare each method to GCM (BIC) as the control. We find similar performance between all linear models with GCM having significantly less coefficients.

		Performance				Number of coefficients			
		Mean rank under AUROC		Mean rank under TDL		Mean coefficient count under AUROC		Mean coefficient count under TDL	
Control	GCM (BIC)	4.667		3.533		5.933		6.067	
Benchmark	LLM (AIC)	5.200	(1.000)	4.267	(0.927)	31.533**	(0.000)	31.800**	(0.000)
	LMT	5.933	(0.821)	4.267	(0.927)	25.400**	(0.000)	23.867**	(0.000)
	GCM (prune)	8.000**	(0.006)	7.733**	(0.000)	2.667**	(0.002)	3.067**	(0.005)
	LR (l1)	4.800	(1.000)	4.933	(0.646)	27.867**	(0.000)	34.200**	(0.000)
	LR (l2)	4.267	(1.000)	5.833	(0.118)	46.000**	(0.000)	46.000**	(0.000)
	DT	8.733**	(0.000)	8.367**	(0.000)	5.733	(0.842)	5.933	(0.894)
	RF	1.867*	(0.026)	4.867	(0.646)	7487.000**	(0.000)	3846.333**	(0.000)
	XGB	1.533*	(0.010)	1.200	(0.118)	7968.667**	(0.000)	2665.867**	(0.000)

Note: *Lower ranks indicate better performance; ** $p \leq 0.01$, * $p \leq 0.05$ (Holm post-hoc, family-wise error correction).

The GCM Delivers a Visualization Which Shows Segmentation First and Coefficients Groups Second

Table 4: Visualization of the GCM with a hyperparameter configuration that allows three coefficients. Consecutive feature names are grouped together, segment specific coefficients are underlined. Segment decision rules use the feature X_8 (crm_R) records the recency in days since CRM system use and the feature X_5 (acc_visiting_pageviews_M) records the number of page views.

Customer segment	Segment decision rules		Linear model			
	Rule 1	Rule 2	Intercept	Coefficients		
				−0.293	−0.032	0.106
Segment 1	$X_8 \leq$ -1.25	-	−2.802	X_{2-3}, X_{19}, X_{21}, X_{35}	$\underline{X_{4-6}}$, $\underline{X_{8-10}}$, X_{14}, X_{16}, X_{18}, $\underline{X_{22-24}}$, X_{28-29}, $\underline{X_{32-33}}$, X_{36}, $\underline{X_{40-42}}$	$\underline{X_1}$, X_7, X_{11-13}, X_{15}, X_{17}, X_{20}, X_{25-27}, X_{30-31}, X_{34}, X_{37-39}, $\underline{X_{43-45}}$
Segment 2	$X_8 > -1.25$	$X_5 \leq -0.07$	−3.451	X_{2-3}, X_{19}, X_{21}, X_{35}	$\underline{X_1}$, $\underline{X_{4-6}}$, X_{8-10}, X_{14}, X_{16}, X_{18}, X_{22-23}, X_{28-29}, X_{33}, X_{36}, X_{40}, X_{42}	X_7, X_{11-13}, X_{15}, X_{17}, X_{20}, $\underline{X_{24-27}}$, X_{30-32}, X_{34}, X_{37-39}, $\underline{X_{41}}$, $\underline{X_{43-45}}$
Segment 3	$X_8 > -1.25$	$X_5 > -0.07$	−3.200	X_{2-3}, X_{19}, X_{21}, X_{35}	$\underline{X_1}$, $\underline{X_{5-6}}$, X_8, X_{10}, X_{14}, X_{16}, X_{18}, X_{23}, X_{25}, X_{33}, X_{37}, X_{40}, $\underline{X_{43}}$	$\underline{X_4}$, X_7, $\underline{X_9}$, X_{11-13}, X_{15}, X_{17}, X_{20}, $\underline{X_{22}}$, X_{24}, X_{26-27}, $\underline{X_{28-29}}$, X_{30-32}, X_{34}, X_{36}, X_{38-39}, $\underline{X_{41-42}}$, X_{44-45}

Interpreting a GCM’s Customer Segmentation

Table 4: Visualization of the GCM with a hyperparameter configuration that allows three coefficients. Consecutive feature names are grouped together, segment specific coefficients are underlined. Segment decision rules use the feature X_8 (crm_R) records the recency in days since CRM system use and the feature X_5 (acc_visiting_pageviews_M) records the number of page views.

Customer segment	Segment decision rules		Linear model			
	Rule 1	Rule 2	Intercept	Coefficients		
				−0.293	−0.032	0.106
Segment 1	$X_8 \leq$ -1.25	-	−2.802	X_{2-3}, X_{19}, X_{21}, X_{35}	$\underline{X_{4-6}}$, $\underline{X_{8-10}}$, X_{14}, X_{16}, X_{18}, $\underline{X_{22-24}}$, X_{28-29}, $\underline{X_{32-33}}$, X_{36}, $\underline{X_{40-42}}$	$\underline{X_1}$, X_7, X_{11-13}, X_{15}, X_{17}, X_{20}, X_{25-27}, X_{30-31}, X_{34}, X_{37-39}, $\underline{X_{43-45}}$
Segment 2	$X_8 > -1.25$	$X_5 \leq -0.07$	−3.451	X_{2-3}, X_{19}, X_{21}, X_{35}	$\underline{X_1}$, $\underline{X_{4-6}}$, X_{8-10}, X_{14}, X_{16}, X_{18}, X_{22-23}, X_{28-29}, X_{33}, X_{36}, X_{40}, X_{42}	X_7, X_{11-13}, X_{15}, X_{17}, X_{20}, $\underline{X_{24-27}}$, X_{30-32}, X_{34}, X_{37-39}, $\underline{X_{41}}$, $\underline{X_{43-45}}$
Segment 3	$X_8 > -1.25$	$X_5 > -0.07$	−3.200	X_{2-3}, X_{19}, X_{21}, X_{35}	$\underline{X_1}$, $\underline{X_{5-6}}$, X_8, X_{10}, X_{14}, X_{16}, X_{18}, X_{23}, X_{25}, X_{33}, X_{37}, X_{40}, $\underline{X_{43}}$	$\underline{X_4}$, X_7, $\underline{X_9}$, X_{11-13}, X_{15}, X_{17}, X_{20}, $\underline{X_{22}}$, X_{24}, X_{26-27}, $\underline{X_{28-29}}$, X_{30-32}, X_{34}, X_{36}, X_{38-39}, $\underline{X_{41-42}}$, X_{44-45}

Interpreting Segments:

- Three customer segments based on X8 (recency since CRM system use in days) and X5 (number of page views).
- Segment 1 has Leads with very infrequent CRM system users. These Potential Customers did almost not use the Free-Trial.
- Their Recency of use was less than -1.25 of the stdv. away from the mean usage Recency.

Interpreting a GCM's Coefficient Groups using

I with a hyperparameter configuration that allows three coefficients. Consecutive feature
nt specific coefficients are underlined. Segment decision rules use the feature X_8 (crm_R)
RM system use and the feature X_5 (acc_visiting_pageviews_M) records the number of page

ules	Linear model			
2	Intercept	Coefficients		
		−0.293	−0.032	0.106
	−2.802	X_{2-3}, X_{19}, X_{21}, X_{35}	X_{4-6}, X_{8-10}, X_{14}, X_{16}, X_{18}, X_{22-24}, X_{28-29}, X_{32-33}, X_{36}, X_{40-42}	X_1, X_7, X_{11-13}, X_{15}, X_{17}, X_{20}, X_{25-27}, X_{30-31}, X_{34}, X_{37-39}, X_{43-45}
−0.07	−3.451	X_{2-3}, X_{19}, X_{21}, X_{35}	X_1, X_{4-6}, X_{8-10}, X_{14}, X_{16}, X_{18}, X_{22-23}, X_{28-29}, X_{33}, X_{36}, X_{40}, X_{42}	X_7, X_{11-13}, X_{15}, X_{17}, X_{20}, X_{24-27}, X_{30-32}, X_{34}, X_{37-39}, X_{41}, X_{43-45}
−0.07	−3.200	X_{2-3}, X_{19}, X_{21}, X_{35}	X_1, X_{5-6}, X_8, X_{10}, X_{14}, X_{16}, X_{18}, X_{23}, X_{25}, X_{33}, X_{37}, X_{40}, X_{43}	X_4, X_7, X_9, X_{11-13}, X_{15}, X_{17}, X_{20}, X_{22}, X_{24}, X_{26-27}, X_{28-29}, X_{30-32}, X_{34}, X_{36}, X_{38-39}, X_{41-42}, X_{44-45}

Interpreting the Coefficients:

Example:

- The variables X2 (frequency of deleted product collection campaign) and X3 (monetary value of deleted membership collection campaign) in segment 1 have a very **negative** impact, meaning that further increases of these feature values reduce the conversion probability.
- Other features such as X1 (number of visitors of the web system), X7 (page views on the webpage) or X11 (frequency of bookable donation collection campaigns) have a **positive** impact, indicating that increasing the value of these features will increase the conversion probability.

Interpreting a GCM’s Contrasting Feature Effects Between Coefficient Groups

I with a hyperparameter configuration that allows three coefficients. Consecutive feature
nt specific coefficients are underlined. Segment decision rules use the feature X_8 (crm_R)
RM system use and the feature X_5 (acc_visiting_pageviews_M) records the number of page

ules	Linear model			
2	Intercept	Coefficients		
		−0.293	−0.032	0.106
	−2.802	X_{2-3}, X_{19}, X_{21}, X_{35}	$\underline{X_{4-6}}$, $\underline{X_{8-10}}$, X_{14}, X_{16}, X_{18}, $\underline{X_{22-24}}$, X_{28-29}, $\underline{X_{32-33}}$, X_{36}, $\underline{X_{40-42}}$	$\underline{X_1}$, X_7, X_{11-13}, X_{15}, X_{17}, X_{20}, X_{25-27}, X_{30-31}, X_{34}, X_{37-39}, $\underline{X_{43-45}}$
−0.07	−3.451	X_{2-3}, X_{19}, X_{21}, X_{35}	$\underline{X_1}$, $\underline{X_{4-6}}$, X_{8-10}, X_{14}, X_{16}, X_{18}, X_{22-23}, X_{28-29}, X_{33}, X_{36}, X_{40}, X_{42}	X_7, X_{11-13}, X_{15}, X_{17}, X_{20}, $\underline{X_{24-27}}$, X_{30-32}, X_{34}, X_{37-39}, $\underline{X_{41}}$, $\underline{X_{43-45}}$
−0.07	−3.200	X_{2-3}, X_{19}, X_{21}, X_{35}	$\underline{X_1}$, $\underline{X_{5-6}}$, X_8, X_{10}, X_{14}, X_{16}, X_{18}, X_{23}, X_{25}, X_{33}, X_{37}, X_{40}, $\underline{X_{43}}$	$\underline{X_4}$, X_7, $\underline{X_9}$, X_{11-13}, X_{15}, X_{17}, X_{20}, $\underline{X_{22}}$, X_{24}, X_{26-27}, $\underline{X_{28-29}}$, X_{30-32}, X_{34}, X_{36}, X_{38-39}, $\underline{X_{41-42}}$, X_{44-45}

Interpreting the Coefficients:

Example (Continued):

- Features can have **contrasting effects** on the model output across different customer segments. For example feature X1 (number of visitors of the web system) shows a **positive** effect for **segment 1**, but has a **negative** effect for segments **2 and 3**.

In Comparison to the GCM, Other Segmented Models like the Logit Leaf Model Yield a Different Type of Global Interpretation and Perspective

Table 6: Visualization of Logit leaf model (LLM; AIC). Features are selected per leaf individually based on AIC. Decision rules use the feature X_{37} (web_edition_R), which records the recency of web page creation tool usage, and the feature X_8 (crm_R), which records the recency in days since CRM system use, and the feature X_{41} (web_edition_pageviews_F), which indicates the frequency of visiting the web page creation tool system.

Customer segment	Segment decision rules		Shared features									Individual features						
	Rule 1	Rule 2	Intercept	X_1	X_8	X_{12}	X_{13}	X_{14}	X_{19}	X_{23}	X_{39}							
Segment 1	$X_{37} \leq 0.17$	$X_{41} \leq 2.86$	-	Always predict class: 1 (conversion) for Segment 1.														
Segment 2	$X_{37} \leq 0.17$	$X_{41} > 2.86$	-4.053		0.26	-0.191	-0.258		-0.613	-0.141	0.017	X_2 -0.65	X_{24} 0.32	X_{29} 0.262	X_{30} -0.261	X_{31} -0.349	X_{35} 0.112	X_{42} 0.212
Segment 3	$X_{37} > 0.17$	$X_8 \leq -1.50$	-2.66	0.437	0.334			0.183	-0.219	-0.24	0.15	X_{10} -0.117	X_{17} -0.167	X_{18} 0.156	X_{26} 0.331	X_{28} 0.072		
Segment 4	$X_{37} > 0.17$	$X_8 > -1.50$	-3.414	0.544	0.342	-0.193	-0.328	0.227			0.183	X_6 -0.328	X_{16} 0.109	X_{41} 0.111	X_{44} -0.172			

05 Ablation Study

Our Ablation Study Tests the GCM on Our Common Set of 10 Classification Tasks

Table 7: Overview of selected datasets covering classification and regression tasks. Numerical (num) and categorical (cat) features are counted after preprocessing, respectively encoding.

Dataset	Samples	Features			Prediction Target
		num*	cat*	cat**	
College (Mukti, 2022)	1,000	4	6	10	Will a high school student go to college?
Water potability (Kadiwal, 2021)	3,276	9	0	0	Will the water be safe for consumption?
Stroke (Palacios, 2021)	5,110	3	7	16	Will a patient suffer from a stroke?
Customer churn (IBM, 2019)	7,043	3	16	37	Will a customer leave the company?
Recidivism (Angwin et al., 2016)	7,214	7	4	11	Will a defendant recidivate?
Credit scoring (Fair Isaac Corporation, 2018)	10,459	21	2	16	Will a client repay within 2 years?
Income adults (Kohavi, 1996)	32,561	6	7	59	Will the income exceed $50,000 per year?
Bank marketing (Moro et al., 2014)	45,211	6	9	41	Will a client subscribe to a deposit?
Airline satisfaction (Klein, 2020)	103,904	18	4	6	Will a passenger be satisfied?
Weather forecast (Bureau of Meteorology, 2010)	142,193	16	4	54	Will it rain the next day in Australia?

Note: * after data preprocessing; ** after feature encoding

The GCM Delivers Strong Performance but Incredibly Little Coefficients for the 10 Various Datasets As Well

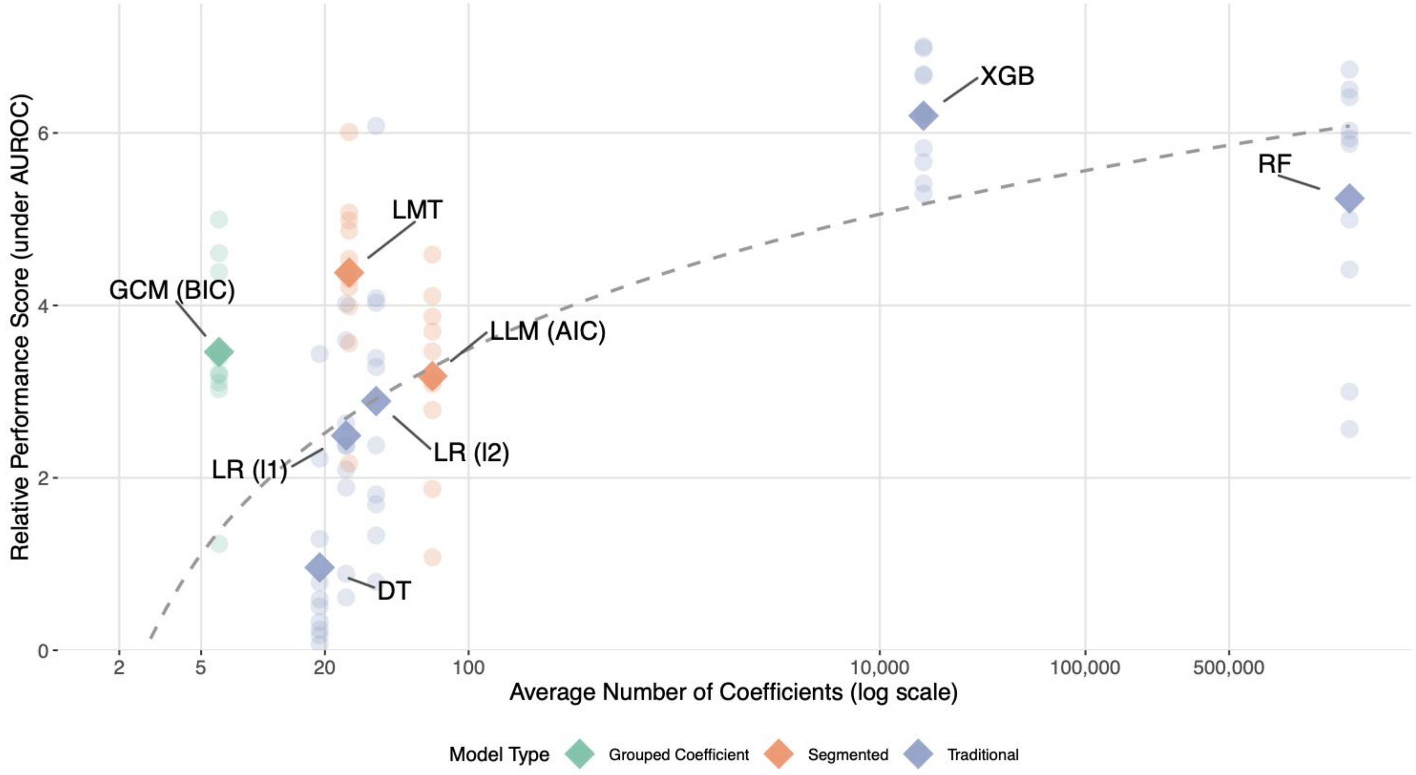


Figure 8: Ablation study on ten further operations research datasets. The relative performance score is calculated as the worst possible rank subtracted by the actual performance rank.

06 Discussion

The GCM opens up an interesting discussion on how parsimonious can decision making in interpretable AI be?

Implications and Learnings

- Incredible and Astonishing how few coefficients are enough.
- The GCM amends the current understanding of Interpretability.
- A conceptual shift to model simplification!?
- We can be very parsimonious ("sparsam") with coefficients.
- What about the Performance-Interpretability Trade-Off again? Seems there is no Trade-Off for Coefficient-Performance!

Limitations and Future Research

- Does less coefficients translate to better Interpretability??
- Visualizations of the GCM is a complex thing that would benefit from further research.
- Can users actually create a mental model out of the GCM?
- Does it really improve the fast prediction when a user knows those very few patterns by heart? Would Sales Agents love the GCM and visualization tool?

Three Key Takeaways

01. We propose a novel interpretable machine learning model called Grouped Coefficient Model (GCM).

02. GCM performs equally well compared to state-of-the-art using significantly less coefficients. There exists no significant differences in predictive performance.

03. We provide a novel way of Visualization for the novel GCM with a strong focus on coefficient parsimony.