Spline-Rule Ensemble Classifiers with Structured Sparsity Regularization for Interpretable Customer Churn Modeling: Supplementary Materials (Appendices B-E)

Koen W. De Bock[1] and Arno De Caigny[2,3]

[1] Audencia Business School, 8 Route de la Jonelière, F-44312, Nantes, France

[2] IESEG School of Management, 3 Rue de la Digue, F-59000, Lille, France

[3] LEM-CNRS 9221, 3 Rue de la Digue, F-59000, Lille, France

E-mail addresses: kdebock@audencia.com (Koen W. De Bock), a.de-caigny@ieseg.fr (Arno De Caigny)

Declaration of Interest: none

Corresponding author: Koen W. De Bock, Audencia Business School, 8 Route de la Jonelière, F-44312, Nantes, France. E-mail address: kdebock@audencia.com, Tel.: +33 2 40 37 34 34

Appendix B: Optimized classifier hyperparameters and candidate values

Classifier	# models per algorithm[1]	Hyperparameter	Candidate values[2]
Decision tree[3] (DT)	36	Confidence threshold for pruning	0.01, 0.15, …, 0.30
		Min. leaf size	n*[0.01,0.025,0.05,0.1,0.25,0.5]
Regularized logistic regression (LR)	20	Elastic net mixing parameter $\alpha_{\text{elestic-net}}$	[0,0.05, …, 0.95,1]
Random forests (RF)	6	No. of CART trees	100
		No. of randomly sampled variables[5]	$\sqrt{v * [0.1, 0.25, 0.5, 1, 2, 4]}$
Generalized additive model (GAM)[4]	1		
Multivariate adaptive regression splines (MARS)	10	Maximum degree of interaction	[1,2, … ,9,10]
Rule ensemble (RE)	20	No. of CART trees	100
		Maximum tree depth	10
		Penalty parameter λ	20 values, dynamically determined based on data characteristics
Spline-rule ensemble (SRE)[4]	20	No. of CART trees	100
		Maximum tree depth	10
		Penalty parameter λ	20 values, dynamically determined based on data characteristics
Spline-rule ensemble with sparse group lasso regularization (SRE-SGL)[4]	20	No. of CART trees	100
		Maximum tree depth	10
		Penalty parameter λ	20 values, dynamically determined based on data characteristics
		Sparse group lasso mixing parameter α_{sgl}	[0,0.05, …, 0.95,1]

1 A classifier is trained for every unique combination of hyperparameter values. For example, DT offers two meta-parameters with both 6 candidate values, so we create 6*6= 36 classification models.
2 The quantity *n* denotes the number of observations and *v* denote the number of independent variables in the data set
3 The number of randomly sampled variables is always rounded to the next integer
4 Cubic regression splines depend on smoothness parameter ρ that is optimized directly during spline estimation using the GCV criterion [1,2]

Table B.1: Optimized classifier hyperparameters and candidate values

Appendix C: Empirical Evaluation of Spline Type Choice on SRE and SRE-SGL Performance

The choice for penalized cubic regression splines in SRE-SGL is inspired on the study in which SRE was first presented [3]. In this online appendix, we further investigate the adoption of alternative spline types as well as shrinkage-enforcing spline variants. A priori there are arguments both in favor of, and against this modification in search of increased model performance. On the one hand: SRE (and SRE-SGL) apply shrinkage in their final step: the application of a regularized regression to the pool of candidate terms. This regularization ensures model/term selection optimizing loss over the *full* set of candidate terms. Preliminary shrinkage, pursued at the time of spline estimation, effectively reduces the set of candidate terms, limiting the playing field of subsequent regularization. On the other hand, it is known that simultaneous shrinkage and curvature penalization could lead to better performance in GAMs.

Specifically, we conduct an experimental comparison of SRE and SRE-SGL implemented with penalized cubic regression splines (identified henceforth using the suffix *cr*) with four variations of SRE and SRE-SGL implemented with the following spline types:

- Penalized cubic regression splines with shrinkage (identified using the suffix *cs*)
- Penalized thin plate regression splines (identified using the suffix *tp*)
- P-splines (identified using the suffix *ps*)
- Penalized thin plate regression splines with shrinkage (identified using the suffix *ts*)

Experimental conditions and data sets are fully described in the main manuscript. Results of this additional benchmarking experiment are presented in the Tables C1-C4 below. Average ranks and adjusted p-values are reported in Tables C5-C6. Using the same statistical analysis as fully disclosed in the benchmarking experiment in the paper, we could not detect any significant differences between the penalized cubic regression splines (*cr*) on the one hand and the variations adopting alternative spline types on the other.

Data set	***SRE - cr***	***SRE - cs***	***SRE - ps***	***SRE - ts***	***SRE - tp***
Ds1	0.885 (0.003)	0.879 (0.002)	0.884 (0.002)	0.881 (0.003)	0.882 (0.003)
Ds2	0.862 (0.001)	0.861 (0.001)	0.859 (0.001)	0.854 (0.001)	0.855 (0.001)
Ds3	0.653 (0.008)	0.652 (0.007)	0.648 (0.005)	0.660 (0.006)	0.660 (0.006)
Ds4	0.843 (0.002)	0.842 (0.001)	0.844 (0.002)	0.839 (0.001)	0.840 (0.002)
Ds5	0.791 (0.002)	0.792 (0.002)	0.788 (0.001)	0.795 (0.001)	0.792 (0.001)
Ds6	0.789 (0.001)	0.781 (0.002)	0.789 (0.002)	0.785 (0.002)	0.784 (0.002)
Ds7	0.716 (0.005)	0.715 (0.004)	0.720 (0.005)	0.714 (0.005)	0.714 (0.005)
Ds8	0.831 (0.003)	0.829 (0.003)	0.821 (0.002)	0.831 (0.002)	0.830 (0.002)
Ds9	0.626 (0.005)	0.628 (0.004)	0.621 (0.005)	0.611 (0.004)	0.612 (0.004)
Ds10	0.620 (0.004)	0.618 (0.005)	0.621 (0.004)	0.609 (0.004)	0.608 (0.004)
Ds11	0.725 (0.014)	0.722 (0.012)	0.730 (0.013)	0.721 (0.012)	0.722 (0.011)
Ds12	0.816 (0.002)	0.817 (0.002)	0.814 (0.003)	0.820 (0.003)	0.819 (0.003)
Ds13	0.863 (0.004)	0.864 (0.003)	0.859 (0.004)	0.862 (0.003)	0.862 (0.003)
Ds14	0.756 (0.012)	0.754 (0.011)	0.753 (0.012)	0.758 (0.012)	0.758 (0.012)

Table C.1: Average AUC - results for SRE variants

Data set	***SRE - cr***	***SRE - cs***	***SRE - ps***	***SRE - ts***	***SRE - tp***
Ds1	5.515 (0.249)	5.560 (0.242)	5.305 (0.210)	4.988 (0.193)	5.002 (0.205)
Ds2	5.177 (0.047)	5.050 (0.052)	5.258 (0.055)	5.351 (0.044)	5.005 (0.046)
Ds3	2.053 (0.153)	2.005 (0.150)	2.152 (0.168)	2.105 (0.161)	2.053 (0.158)
Ds4	4.407 (0.037)	4.102 (0.040)	4.360 (0.045)	4.485 (0.040)	4.522 (0.039)
Ds5	3.701 (0.072)	3.801 (0.065)	3.652 (0.080)	3.789 (0.082)	3.659 (0.078)
Ds6	3.593 (0.063)	3.410 (0.065)	3.337 (0.052)	3.466 (0.063)	3.687 (0.064)
Ds7	2.333 (0.285)	2.315 (0.287)	2.250 (0.197)	2.346 (0.211)	2.346 (0.208)
Ds8	4.564 (0.122)	4.488 (0.110)	4.378 (0.140)	4.322 (0.122)	4.488 (0.118)
Ds9	1.585 (0.027)	1.655 (0.030)	1.629 (0.034)	1.499 (0.041)	1.629 (0.044)
Ds10	1.468 (0.616)	1.501 (0.430)	1.444 (0.510)	1.318 (0.055)	1.408 (0.620)
Ds11	3.322 (0.185)	3.288 (0.170)	3.322 (0.162)	3.488 (0.165)	3.440 (0.172)
Ds12	2.782 (0.058)	2.882 (0.078)	2.533 (0.066)	2.436 (0.081)	2.533 (0.077)
Ds13	5.246 (0.077)	5.273 (0.095)	5.336 (0.080)	5.011 (0.073)	5.098 (0.076)
Ds14	2.072 (0.430)	2.052 (0.377)	2.158 (0.412)	2.358 (0.442)	2.255 (0.444)

Table C.2: Average TDL - results for SRE variants

Data set	***SRE - cr***	***SRE - cs***	***SRE - ps***	***SRE - ts***	***SRE - tp***
Ds1	0.883 (0.003)	0.877 (0.003)	0.884 (0.002)	0.882 (0.003)	0.883 (0.003)
Ds2	0.862 (0.001)	0.862 (0.002)	0.858 (0.002)	0.863 (0.001)	0.862 (0.001)
Ds3	0.656 (0.006)	0.654 (0.004)	0.648 (0.005)	0.661 (0.004)	0.658 (0.005)
Ds4	0.843 (0.002)	0.843 (0.003)	0.846 (0.003)	0.842 (0.002)	0.842 (0.002)
Ds5	0.790 (0.002)	0.792 (0.002)	0.789 (0.001)	0.794 (0.001)	0.793 (0.002)
Ds6	0.789 (0.001)	0.828 (0.002)	0.822 (0.001)	0.832 (0.001)	0.831 (0.001)
Ds7	0.711 (0.006)	0.710 (0.004)	0.715 (0.005)	0.709 (0.004)	0.711 (0.004)
Ds8	0.830 (0.004)	0.785 (0.003)	0.787 (0.004)	0.785 (0.004)	0.784 (0.003)
Ds9	0.625 (0.004)	0.624 (0.002)	0.623 (0.003)	0.619 (0.003)	0.619 (0.003)
Ds10	0.623 (0.004)	0.628 (0.005)	0.620 (0.004)	0.619 (0.005)	0.619 (0.005)
Ds11	0.728 (0.017)	0.729 (0.015)	0.729 (0.012)	0.724 (0.013)	0.724 (0.013)
Ds12	0.815 (0.002)	0.816 (0.002)	0.813 (0.001)	0.814 (0.002)	0.814 (0.002)
Ds13	0.863 (0.006)	0.864 (0.004)	0.858 (0.005)	0.862 (0.004)	0.863 (0.004)
Ds14	0.761 (0.010)	0.755 (0.009)	0.758 (0.010)	0.759 (0.009)	0.759 (0.009)

Table C.3: Average AUC results for SRE-SGL variants

Data set	*SRE - cr*	*SRE - cs*	*SRE - ps*	*SRE - ts*	*SRE - tp*
Ds1	5.487 (0.272)	5.268 (0.252)	5.551 (0.271)	5.502 (0.211)	5.218 (0.252)
Ds2	5.149 (0.049)	5.306 (0.051)	5.149 (0.055)	5.009 (0.049)	5.015 (0.044)
Ds3	2.097 (0.127)	2.213 (0.123)	2.255 (0.114)	2.117 (0.114)	2.005 (0.127)
Ds4	4.404 (0.038)	4.205 (0.042)	4.008 (0.036)	4.265 (0.042)	4.489 (0.041)
Ds5	3.705 (0.085)	3.591 (0.088)	3.408 (0.085)	3.520 (0.093)	3.782 (0.085)
Ds6	3.568 (0.067)	4.655 (0.064)	4.441 (0.064)	4.707 (0.062)	4.655 (0.061)
Ds7	2.347 (0.285)	2.456 (0.279)	2.200 (0.238)	2.551 (0.240)	2.504 (0.211)
Ds8	4.560 (0.114)	3.780 (0.113)	3.588 (0.129)	3.307 (0.114)	3.442 (0.097)
Ds9	1.587 (0.027)	1.506 (0.026)	1.503 (0.025)	1.400 (0.028)	1.355 (0.028)
Ds10	1.628 (0.283)	1.488 (0.220)	1.422 (0.275)	1.467 (0.244)	1.722 (0.220)
Ds11	3.537 (0.246)	3.442 (0.236)	3.242 (0.226)	3.688 (0.244)	3.255 (0.228)
Ds12	2.777 (0.059)	2.573 (0.068)	2.805 (0.060)	2.892 (0.061)	2.777 (0.059)
Ds13	5.250 (0.091)	5.407 (0.082)	5.428 (0.086)	5.351 (0.085)	5.407 (0.088)
Ds14	2.231 (0.090)	2.123 (0.091)	2.105 (0.094)	2.057 (0.90)	2.289 (0.091)

Table C.4: Average TDL results for SRE-SGL variants

Algorithm role	Algorithm	Metric	
		AUC	**TDL**
Control	SRE - cr	**2.214**	**2.786**
Benchmarks	SRE - cs	3.143 (0.292)	2.964 (0.999)
	SRE - ps	3.107 (0.292)	3.250 (0.999)
	SRE - ts	3.250 (0.292)	3.107 (0.999)
	SRE - tp	3.286 (0.292)	2.893 (0.999)

[1] Lower average ranks indicate better performance. The best performing algorithm is indicated in bold. The adjusted p-value for Holm post-hoc test is shown between brackets.

Table C.5: Average SRE variations ranks across data sets for different performance measures and significance test results

Algorithm role	Algorithm	Metric	
		AUC	**TDL**
Control	SRE-SGL - cr	**2.357**	**2.714**
Benchmarks	SRE-SGL - cs	2.893 (0.427)	2.857 (0.999)
	SRE-SGL - ps	3.321 (0.427)	3.393 (0.999)
	SRE-SGL - ts	3.231 (0.427)	3.071 (0.999)
	SRE-SGL - tp	3.250 (0.427)	2.964 (0.999)

[1] Lower average ranks indicate better performance. The best performing algorithm is indicated in bold. The adjusted p-value for Holm post-hoc test is shown between brackets.

Table C.6: Average SRE-SGL variations ranks across data sets for different performance measures and significance test results

Appendix D: Detailed results of the benchmarking experiment

Data set	Decision tree (DT)	Regularized logistic regression (LR)	Generalized additive model (GAM)	Multivariate additive regression splines (MARS)	Random forest (RF)	Rule ensemble (RE)	Spline-rule ensemble (SRE)	Spline-rule ensemble with SGL (SRE-SGL)
Ds1	0.861 (0.007)	0.877 (0.002)	0.875 (0.002)	0.880 (0.002)	0.875 (0.002)	0.880 (0.005)	<u>0.885 (0.003)</u>	0.883 (0.003)
Ds2	0.843 (0.007)	0.859 (0.002)	0.855 (0.002)	0.859 (0.002)	0.845 (0.002)	0.861 (0.002)	0.862 (0.001)	<u>0.862 (0.001)</u>
Ds3	0.634 (0.015)	0.648 (0.010)	0.643 (0.008)	0.629 (0.011)	0.639 (0.009)	0.652 (0.008)	0.653 (0.008)	<u>0.656 (0.006)</u>
Ds4	0.816 (0.005)	0.839 (0.001)	0.835 (0.001)	0.835 (0.002)	0.837 (0.001)	0.840 (0.002)	0.843 (0.002)	<u>0.843 (0.002)</u>
Ds5	0.768 (0.006)	0.781 (0.002)	0.782 (0.002)	0.790 (0.002)	<u>0.792 (0.002)</u>	0.786 (0.002)	0.791 (0.002)	0.790 (0.002)
Ds6	0.766 (0.015)	0.779 (0.004)	0.780 (0.003)	0.786 (0.003)	0.789 (0.003)	0.787 (0.001)	0.789 (0.001)	<u>0.789 (0.001)</u>
Ds7	0.734 (0.013)	0.686 (0.004)	0.703 (0.005)	0.762 (0.008)	<u>0.780 (0.004)</u>	0.708 (0.001)	0.716 (0.005)	0.711 (0.006)
Ds8	0.804 (0.007)	0.811 (0.004)	0.828 (0.003)	0.822 (0.004)	<u>0.845 (0.003)</u>	0.826 (0.005)	0.831 (0.003)	0.830 (0.004)
Ds9	0.620 (0.006)	0.593 (0.004)	0.616 (0.003)	0.624 (0.004)	<u>0.634 (0.003)</u>	0.614 (0.004)	0.626 (0.005)	0.625 (0.004)
Ds10	0.601 (0.008)	0.615 (0.006)	0.619 (0.005)	0.614 (0.008)	0.614 (0.005)	0.617 (0.005)	0.620 (0.004)	<u>0.623 (0.004)</u>
Ds11	0.678 (0.018)	<u>0.724 (0.011)</u>	0.726 (0.009)	0.712 (0.006)	0.722 (0.009)	0.715 (0.014)	0.725 (0.014)	0.728 (0.017)
Ds12	0.792 (0.008)	0.812 (0.005)	0.808 (0.004)	0.806 (0.005)	0.812 (0.004)	0.812 (0.002)	<u>0.816 (0.002)</u>	0.815 (0.002)
Ds13	0.835 (0.005)	0.858 (0.002)	0.856 (0.002)	0.858 (0.002)	<u>0.866 (0.002)</u>	0.860 (0.008)	0.863 (0.004)	0.863 (0.006)
Ds14	0.712 (0.018)	0.754 (0.010)	0.740 (0.012)	0.729 (0.018)	0.743 (0.012)	0.755 (0.011)	0.756 (0.012)	<u>0.761 (0.010)</u>

Table D.1: Results of the benchmarking experiment using the AUC performance measure

Data set	Decision tree (DT)	Regularized logistic regression (LR)	Generalized additive model (GAM)	Multivariate additive regression splines (MARS)	Random forest (RF)	Rule ensemble (RE)	Spline-rule ensemble (SRE)	Spline-rule ensembles with SGL (SRE-SGL)
Ds1	2.619 (1.746)	5.000 (0.101)	5.167 (0.218)	5.477 (0.184)	5.175 (0.850)	5.128 (0.059)	5.515 (0.249)	5.487 (0.272)
Ds2	4.929 (0.174)	5.014 (0.045)	4.991 (0.055)	5.068 (0.152)	3.207 (1.302)	5.117 (0.082)	5.177 (0.047)	5.149 (0.049)
Ds3	0.615 (0.616)	1.934 (0.161)	1.790 (0.117)	1.903 (0.176)	1.801 (0.155)	2.094 (0.114)	2.053 (0.153)	2.097 (0.127)
Ds4	0.741 (0.295)	4.389 (0.023)	4.122 (0.030)	4.176 (0.037)	4.419 (0.045)	4.353 (0.036)	4.407 (0.037)	4.404 (0.038)
Ds5	2.482 (0.906)	3.305 (0.067)	3.466 (0.044)	3.578 (0.055)	3.884 (0.070)	3.340 (0.030)	3.701 (0.072)	3.705 (0.085)
Ds6	2.391 (1.030)	3.483 (0.073)	3.374 (0.053)	3.504 (0.043)	3.522 (0.056)	3.558 (0.067)	3.593 (0.063)	3.568 (0.067)
Ds7	1.430 (1.125)	2.245 (0.199)	2.101 (0.120)	3.348 (0.210)	3.965 (0.227)	1.264 (0.372)	2.333 (0.285)	2.347 (0.285)
Ds8	3.901 (0.599)	4. 083 (0.097)	4.120 (0.093)	4.286 (0.108)	4.845 (0.112)	4.528 (0.113)	4.564 (0.122)	4.560 (0.114)
Ds9	1.084 (0.288)	1.414 (0.030)	1.580 (0.030)	1.382 (0.037)	1.576 (0.049)	1.484 (0.035)	1.585 (0.027)	1.587 (0.027)
Ds10	0.768 (0.616)	1.814 (0.096)	1.484 (0.098)	1.358 (0.232)	1.484 (0.277)	1.606 (0.254)	1.468 (0.616)	1.628 (0.283)
Ds11	1.712 (0.614)	3.333 (0.202)	3.156 (0.161)	3.125 (0.225)	3.184 (0.184)	3.288 (0.154)	3.322 (0.185)	3.537 (0.246)
Ds12	1.109 (0.529)	2.745 (0.056)	2.504 (0.053)	2.382 (0.059)	2.652 (0.075)	2.705 (0.064)	2.782 (0.058)	2.777 (0.059)
Ds13	2.625 (0.089)	5.107 (0.079)	5.073 (0.159)	5.066 (0.199)	5.153 (0.072)	5.092 (0.265)	5.246 (0.077)	5.250 (0.091)
Ds14	0.762 (0.373)	2.139 (0.086)	2.006 (0.145)	1.915 (0.199)	2.078 (0.139)	1.550 (1.032)	2.072 (0.430)	2.231 (0.090)

Table D.2: Results of the benchmarking experiment using the TDL performance measure

Appendix E: Analysis of Variable Interactions

An alternative use of partial dependence functions is the analysis of variable interactions. The presence of multivariate rules in (S)RE models indicate the existence of such interactions, but when there are multiple interactions, or when these variable interactions are represented by several rules with the same set of defining variables, a more formal analysis is required to reveal the nature and strength of these effects.

Based on equation (12), the two interaction effects that are observed in the SRE-SGL model are confirmed: an interaction between *recchrge* and *eqpdays* ($H^2_{recchrge,eqpdays} = 0.9939$), and between *retcalls* and *eqpdays* ($H^2_{retcalls,eqpdays} = 0.5513$). Figure E.1 shows partial dependence plots $\hat{F}_{jk}\left(x_{ij}, x_{ik}\right)$ for these two interaction effects. The first interaction effect dictates that the impact of *retcalls* on churn probabilities increases for decreasing values of *eqpdays*, while the second interaction effect shows that positive effects of *eqpdays* and *recchrge* partially cancel each other out: the positive effect of *eqpdays* on the probability to churn becomes less pronounced for higher values of *recchrge*.

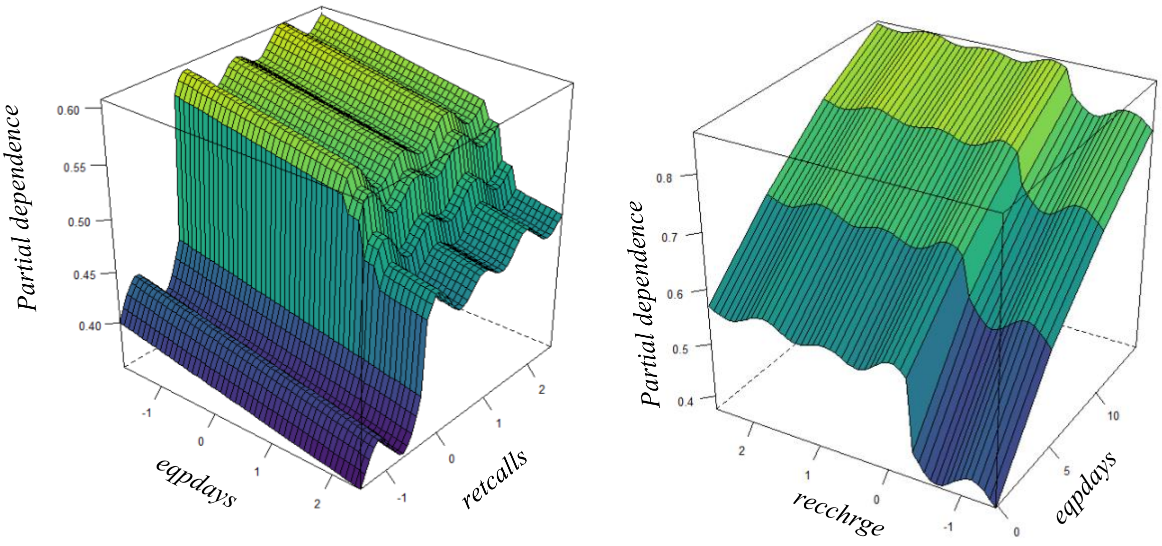

Figure E.1: Graphical representation of two interaction effects present in the SRE-SGL model through partial dependence plots

References

[1] P. Craven, G. Wahba, Smoothing noisy data with spline functions, Numer. Math. 31 (1978) 377–403.

[2] S.N. Wood, Stable and efficient multiple smoothing parameter estimation for generalized additive models, J. Am. Stat. Assoc. 99 (2004) 673–686.

[3] K.W. De Bock, The best of two worlds: Balancing model strength and comprehensibility in business failure prediction using spline-rule ensembles, Expert Syst. Appl. 90 (2017).