How to build better segmentation typing tools:

The role of classification and imbalance correction methods

Marco Vriens[1], Nathan Bosch[2] & Jason Talwar[3]

Abstract

A typing tool is a predictive model that predicts which respondents in a segmentation study fall in which segment. The ability to do that accurately is vital in segmentation, especially if the predictors are background, firm, and media usage variables, as such variables allow the marketer to reach their audiences. Typically, acceptability of a typing tool is judged by overall predictive accuracy. There are three practical challenges. One, typing tools based on passive segmentation variables often don't reach high accuracy. Two, even if overall accuracy is good, specific segments can show very low prediction accuracy, especially in unbalanced situations (when we have one segment that is much larger than the smallest segment). Three, the best overall accuracy may not equate to the best profits of the segmentation implementation. In this paper, we study how the accuracy of typing tools varies by classification methods and imbalance correction methods. We show which methods do best in terms of overall accuracy, segment-specific accuracy, and profitability. Some key insights: 1) The difference between the best & worst classifier can mean an increase of 60% in profitability, 2) highest accuracy doesn't always mean highest profits, and 3) using imbalance correction methods can sometimes results in solutions that lead to higher expected profits.

1. INTRODUCTION

A market segmentation project usually has two steps. In step one, market segments are identified based on a set of active segmentation variables (e.g. Wedel & Kamakura, 2000). In step two, predictive models are developed that predict which respondent falls in which segment using either the active segmentation variables or a set of passive segmentation variables[4]. When these predictive models are programmed (e.g., in an interactive Excel spreadsheet) so new respondents can be allocated to specific segments, we refer to this as a typing tool (TT). TTs based on passive segmentation variables enable the accessibility of the segments by giving the marketer information about how to reach these segments, either via mass marketing or via marketing to customers in their customer database.

However, developing typing tools can be challenging:

---

[1] Kwantum analytics. Marco can be reached at marco.vriens@teamkwantum.com.

[2] Kwantum Analytics.

[3] Salesforce.com.

[4] In segmentation typing tools, we can use either active or passive segmentation variables. Active segmentation variables are the variables that were used to identify the segments in the first place. Passive segmentation variables are defined as variables that were not used in the identification of segments, but are used to further profile the segments, e.g. demographic, firmo-graphic or media usage variables.

Challenge 1
In most practical situations it is hard to develop a typing tool that has a high classification accuracy, especially when passive variables are used as predictors (background, firmographic, media usage). Accuracy in such typing tools may be modest (50%-70% range); see Liu, Ram, Lusch & Brusco (2010) and Vriens et al. (2022).

Challenge 2
Not all segments are predicted with the same accuracy. Vriens et al. (2022) showed that prediction can vary substantially across segments. In some cases, the prediction may be so poor that essentially the firm cannot reach a segment, making it un-actionable.

Challenge 3
When a given segment is more important or valuable to the firm than the other segments, firms may want to prioritize accuracy for such a segment. Misclassifying respondents in a highly valuable segment can have dramatic profit implications for the segmentation strategy. This raises the question of how accuracy, both overall and at the segment-level, is related to the expected profitability of the segmentation strategy. Vriens et al. (2022) showed that imbalance correction methods can improve the prediction accuracy of specific segments and they showed that different methods lead to different expected profits.

Typing tools based on active segmentation variables will result in higher accuracy than typing tools based on passive segmentation variables. However, since the latter are used to reach the segments, this paper will evaluate typing tools based on passive segmentation variables. We will illustrate how various base classifiers yield different overall and segment-level classification accuracy, and we show that when we have approximate knowledge of the targeting cost and expected revenue from the various segments, that we derive profitability estimates that can better help us select the best typing tool.

In this paper, we build on the Vriens et al. (2022) study. We study how misclassification for specific segments can be reduced by using better classifiers and imbalance correction methods, and we show how such improvements positively impact the profitability of the segmentation strategy. This paper is structured as follows. In section two, we briefly review the key findings from the Vriens et al. (2022) study. In section three we discuss how to measure accuracy and expected profitability. In section four, we present the results of two studies. We conclude with a set of practical recommendations.

2. PREVIOUS RESEARCH

Accuracy of a typing tool can vary based on the type of predictor variables, classification method used, and the use of imbalance correction methods.

Vriens et al (2022) studied how *overall* segment prediction accuracy can vary dramatically across various base classifiers. For example, in one of their studies standard multinomial logistic regressions (LR) achieved an overall accuracy of 71% whereas the more advanced Support Vector Machines (SVM) achieved 92% (all accuracies are evaluated using10-fold cross-validation). Prediction for specific segments can also vary dramatically. In one of their studies, the minority segment was predicted with 0% accuracy using LR. By applying SVM and an imbalance correction method, they increased the prediction of this segment to 71% with only a modest decline in overall accuracy.

Specifically, they studied how applying imbalance correction methods and classification methods can impact overall and segment level accuracy. They compared six base classifiers: Logistic regression (LR), Naïve Bayes (NB), Decision Trees (DT), Random Forest (RF), Support Vector Machines (SVM), and Gradient Boosting (GB). They compared four imbalance correction methods: random under-sampling (RUS), random over-sampling (ROS), weighting, and synthetic minority over-sampling technique (SMOTE). In their study, they used an extensive simulation study and two empirical datasets for their comparisons. Several key findings came out of this study. First, overall, they find that imbalance correction methods in most cases do not improve overall accuracy and are likely to result in a small decrease (although there can be exceptions where they do improve overall accuracy). Two, imbalance correction methods were shown to dramatically improve the prediction of the minority (smallest) segment. Overall RUS and SMOTE performed best. Third, different classifiers can yield dramatic differences in overall prediction accuracy rates. Overall, SVM and GB performed best. The same held for improving the prediction of the minority segment, where again, SVM and GB performed best. Vriens et al. (2022) also studied the impact on profitability. Dramatic profit differences were found between a) a situation where we have a poor prediction of the minority segment versus b) where we have the best feasible prediction of the minority segment by applying the best base classifier and imbalance correction method.

So, we know that different classification methods can achieve different levels of predictive accuracy. We also know that some classifiers and imbalance correction methods can lead to a much better prediction of the minority segment.

Profitability

It seems reasonable to assume that, given everything else equal, higher accuracy means higher profits. However, classifiers not only differ in terms of the overall accuracy they achieve, but they also differ in terms of how well they predict specific segments and different segments can have differing importance to the firm. Segments can be attractive for a variety of reasons such as expected segment growth, relative market share, competitive intensity (e.g., a small segment could be a white space where there is no competition yet), entry barriers and purchasing power (Tonks, 2009) or expected profitability (Liu et al., 2010; Vriens et al. 2022).

If different segments differ in terms of the value for the firm, then we also need to look at the segment-specific accuracy rates, and what the cost is of reaching certain segment members and what revenue they are likely to yield. In the next section, we define the concepts of accuracy and profitability.

3. MEASURING ACCURACY AND PROFITABILITY

Typically, typing tools are evaluated by looking at their overall accuracy: i.e. how many respondents across the various segments can the tool predict in the correct segment.

Accuracy

In this paper, we look at overall performance via an unweighted hit rate[5]. Let TP (True Positives) be the true positives (the number of respondents the model correctly predicts in the correct segment). True Negatives (TN) is the number of respondents correctly predicted not to belong to segments to which they don't belong too. False Positives (FP) is the number of respondents that the model incorrectly predicts to belong to a certain segment to which they don't belong. False Negatives (FN) is the number of respondents that the model predicts not be in a certain segment whereas they do belong to that certain segment. We calculate an overall hit rate as (TP + TN)/(TP + FP + TN +FN). We calculate segment-specific classification accuracy as TP/(TP + FP) (or, the proportion of respondents classified into a segment who are truly in that segment).

Profitability

The reason differences in segment-level prediction matter is that different levels of misclassification rates for different segments can affect the profitability of a segmentation strategy.

In our analyses, we assume the perfect consumer, i.e. we assume that if we predict a consumer correctly to the right segment, this consumer will purchase what we market to them. So, there is a fixed cost in reaching them via marketing channels and there is a fixed expected revenue for each customer reached. We assume that consumers in a certain segment have a certain targeting cost and a certain expected revenue. Thus, each correctly classified consumer will generate the expected profit. We also assume that any mis-classified respondent will yield zero revenue while incurring the cost of reaching the segment in which they were classified. The formula for this scenario is:

Perfect consumer scenario:

$$P = \sum_{i}^{S} (c_i + r_i \times h_i) \times N_i$$

[5] In Vriens et al. 2022) a variety of accuracy measures was used (specifically, weighted, unweighted, and F1 score) but they all converged to the same conclusions. Hence here for simplicity's sake we just use the unweighted hit rate.

Where P is the expected profit, S are the number of segments, C is the cost, R is the revenue, H is the segment hit rate, and N are the number of consumers in each segment. This assumes that every successful segment prediction will result in a successful marketing campaign (e.g., if the segment prediction is correct then we guarantee that the consumer will spend money).

4. BASE CLASSIFIERS AND IMBALANCE CORRECTION METHODS

In this paper we use four base classifiers and two imbalance-correction methods

Base classifiers

The first is Multinomial Logistic Regression (LR). We include this method as it is the most often applied method for typing tool models and it has been shown to perform quite well[6]. Second, we use Random Forests (RF) (Breiman et al.,1984), because this method is also quite common in marketing research and is known to lead to good predictions. We added two more advanced methods, that are less common in marketing research: Support vector machines (SVM) (Bennett & Campbell, 2000 and Cortes & Vapnik, 1995), and Neural Nets (NN) (Schmidhuber, 2015). SVM came out as the best performing classification method in Vriens et al. (2022), who compared LR, decision trees (DT)[7], random forests (RF), gradient boosting (GB), and support vector machines (SVM). We added Neural Nets (NN) as these have not been used for typing tools to our knowledge. Our goal is not a comprehensive comparison of base classifiers, but our set constitutes a range from basic (LR) to more advanced (RF, SVM, NN) so that we can get insight into how different base classifiers may yield different overall accuracy. Different classifiers may also yield different profitability because even when they achieve the same level of overall accuracy, they can still differ on how they predict individual segments.

For LR, RF, SVM, and NN we used the scikit-learn Python package (Pedregosa et al., 2011). The exact implementation details can be found in Vriens et al. (2022). For NN, we used a neural network with 2 hidden layers with 16 and 8 neurons, with ReLU activation (https://en.wikipedia.org/wiki/Rectifier_(neural_networks)). We used a constant learning rate of 0.001 with the Adam optimizer (Kingman & Ba, 2014). The batch size for feeding data to the network was 16 datapoints for every update step. Training was done until convergence.

Imbalance-correction methods

There are many imbalance correction methods, including random under-sampling, random over-sampling, SMOTE (and versions of SMOTE), weighting etc. In this study, we only used the random under sampling and the SMOTE technique (Chawla et al., 2002) as they resulted in the best solutions in our previous study. In random under-sampling we randomly delete observations from the majority

[6] Linear discriminant analysis is also used quite often; this method is like LR and hence we leave it out here.
[7] CHAID is a well-known variety of a Decision Tree method. Whereas CHAID uses the Chi-squared statistic to decide the splits we used the Gini impurity measure.

segment until the size of the majority segment is the same as the size of the minority segment. Then, we move to the second largest segment, and we repeat this process. This process continues until all segments have the same size. Obviously, this approach will work best when there is sufficient sample size. Random over-sampling takes the opposite approach. We randomly select cases from the minority segment, and we add these as duplicates to that minority segment. We repeat this process until the minority segment is as large as the majority segment. Then, we move to the second smallest segment, and we again re-sample until that segment is the same size as the majority segment. SMOTE is a version of over-sampling. It selects pairs of two neighboring points in the minority segment. Then, a random value is chosen between each of the two points, and this gets added to the minority class. This process is repeated until the minority class is equal to the majority class. Then, this process is repeated for the second smallest segment and so forth.

5. METHODOLOGY AND DATA

In this section, we outline the two empirical datasets and the hypothetical scenarios for what is would cost to reach segment members successfully and what the expected revenue would be for members that were successfully reached. The dataset and cost/revenue scenarios were also used in Vriens et at. (2022) and pertained to consumer durables.

Data

The first empirical dataset comes from a commercial project we were involved in. To protect the confidentiality of our client, we cannot the name the firm nor the category other than that this was a durable consumer good. The basis for the segments was a series of MaxDiff questions (Louviere, Flynn, & Marley, 2015), analyzed using a latent class anchored MaxDiff model. An optimal three-segment solution was deemed useful and actionable by management, and a typing tool model was requested. Note, we also ran our analyses for a four-segment solution. This did not alter the general findings as presented in the results section. In this application, the firm wanted to predict segment membership based on passive[8] variables that the client also had in their database with the purpose of "scoring" the entire database on the four segments: i.e. assign a segment membership to each customer in the database. We had 13 variables for this that were part of the survey and part of the customer database.

The second empirical dataset was a strategic segmentation study, also pertaining to a consumer durable product, and was commissioned to inform the firm's brand and product strategy after a merger with a competitor. In this dataset we had a larger sample size (n=6000). A four-segment solution was deemed optimal and used for the TT development (we note that we also did our analyses on a five-segment and a six-segment solution. This did not alter the general findings as presented in the results section. Here, the TT used 21 passive segmentation variables.

[8] Just to re-iterate, all our TTs are based on passive variables, specifically ones that can be used to reach the segments.

Cost/Revenue Scenarios

We defined 12 cost/revenue scenarios for minority segment. The cost to reach a customer can take on three values: \$ 10, \$ 20, and \$ 30. The expected revenue can take on four values: \$ 50, \$ 75, \$ 150, and \$ 250. For the non-minority segment the cost is set \$ 10 and the expected revenue to \$ 50. Hence: 12 scenarios in total. For each scenario, and each solution under different classifiers we will calculate the total expected profit.

6. RESULTS

Study 1

In Tables 1a and 1b, we show the typing tool results for a 3-segment MaxDiff segmentation. Table 1a gives the results for the unbalanced data, while in Table 1b, the segments were balanced using the SMOTE technique.

Table 1a shows several interesting results. Among the base classifiers, SVM performs best, both in terms of accuracy and profitability. However, note that LR has almost the same level of accuracy but substantially lower profits. Third, the difference in profits between the worst and the best classifier is on average 20%, quite a dramatic difference. Table 1b shows something even more interesting. Although it shows that SVM yields a substantially higher accuracy, and on average yields the highest profits, it does not do so in every single scenario. In the \$10/\$250 and 20/250 scenarios, NN would be best. Also, note, that the highest profit under the SMOTE balanced scenario is higher than the highest profit under the unbalanced scenario.

Table 1.A

Study 1: A 3-segment MaxDiff segmentation (n=400)

UNBALANCED		LR	RF	SVM	NN	
ACCURACY		63%	59%	64%	55%	
COST	REVENUE	PROFIT				BEST vs WORST
$10	$50	2860	2610	**2960**	2360	25%
$10	$75	2885	2660	**3060**	2435	26%
$10	$150	2960	2810	**3360**	2660	26%
$10	$250	3060	3010	**3760**	2960	25%
$20	$50	2810	2540	**2860**	2210	13%
$20	$75	2835	2590	**2960**	2285	14%
$20	$150	2910	2740	**3260**	2510	19%
$20	$250	3010	2940	**3660**	2810	24%
$30	$50	2760	2470	**2760**	2060	12%
$30	$75	2785	2520	**2860**	2135	13%
$30	$150	2860	2670	**3160**	2360	18%
$30	$250	2960	2870	**3560**	2660	24%
AVERAGE		2891	2703	**3185**	2454	**20%**

Table 1.B

Study 1: A 3-segment MaxDiff segmentation (n=400)

SMOTE BALANCED		LR	RF	SVM	NN	
ACCURACY		57%	54%	63%	54%	
COST	REVENUE	PROFIT				BEST vs WORST
$10	$50	2460	2260	**2860**	2310	25%
$10	$75	2510	2410	**2985**	2510	24%
$10	$150	2660	2860	**3360**	3110	26%
$10	$250	2860	3460	3860	**3910**	37%
$20	$50	2350	2050	**2710**	2130	32%
$20	$75	2400	2200	**2835**	2330	29%
$20	$150	2550	2650	**3210**	2930	26%
$20	$250	2750	3250	3710	**3730**	35%
$30	$50	2240	1840	**2560**	1950	39%
$30	$75	2290	1990	**2685**	2150	35%
$30	$150	2440	2440	**3060**	2750	25%
$30	$250	2640	3040	**3560**	3550	35%
AVERAGE		2513	2538	**3116**	2780	**31%**

Study 2

In Tables 2a and 2b we show the unbalanced and balanced results for Study 2. Table 2.A shows an interesting result: all classifiers perform equally well on the unbalanced data, but since specific segments can be predicted differently, we see different profitability numbers across the various classifiers. In this study, the best profitability varies by scenario though on average RF yields the highest profits.

Table 2.A

Study 2: A 4-segment attitudinal segmentation (n=6000)

UNBALANCED		LR	RF	SVM	NN	
ACCURACY		47%	47%	47%	47%	
COST	REVENUE	PROFIT				BEST VS WORST
$10	$50	23990	23440	23640	**23990**	2%
$10	$75	**25565**	25415	23865	25540	7%
$10	$150	30290	**31340**	24540	30190	28%
$10	$250	36590	**39240**	25440	36390	54%
$20	$50	22370	21380	**23410**	22450	9%
$20	$75	23945	23355	23635	**24000**	3%
$20	$150	28670	**29280**	24310	28650	20%
$20	$250	34970	**37180**	25210	34850	47%
$30	$50	20750	19320	**23180**	20910	20%
$30	$75	22325	21295	**23405**	22460	10%
$30	$150	27050	**27220**	24080	27110	13%
$30	$250	33350	**35120**	24980	33310	41%
		27489	**27799**	24141	27488	**21%**

In Table 2.B, where we balanced the data using the random under-sampling technique, we see that NN has a slightly better overall accuracy but interestingly this results in substantially higher profits--on average, an increase of 17% in profits.

Table 2.B
Study 2: A 4-segment attitudinal segmentation (n=6000)

RANDOM UNDER-SAMPLING		LR	RF	SVM	NN	
ACCURACY		43%	45%	44%	46%	
COST	REVENUE	PROFIT				BEST VS WORST
$10	$50	20490	21390	21290	**22690**	11%
$10	$75	23690	24715	24390	**26190**	11%
$10	$150	33290	34690	33690	**36690**	10%
$10	$250	46090	47990	46090	**50690**	10%
$20	$50	16630	17330	17550	**18920**	14%
$20	$75	19830	20655	20650	**22420**	13%
$20	$150	29430	30630	29950	**32920**	12%
$20	$250	42230	43930	42350	**46920**	11%
$30	$50	12770	13270	13810	**15150**	19%
$30	$75	15970	16595	16910	**18650**	17%
$30	$150	25570	26570	26210	**29150**	14%
$30	$250	38370	39870	38610	**43150**	12%
		27030	28136	27625	**30295**	**13%**

Imperfect consumer
In our profitability calculations, we assumed what we call a perfect consumer: If we predict a consumer correctly to the segment they belong to, then we assume that will buy our product. If we mis-classify them, we assume they will buy nothing. Of course, these assumptions need not be true.

We replicated our analyses with several imperfect consumer assumptions that alleviated the first assumption: even if classified correctly, they might not buy. We can define the imperfect consumer in various ways. Here we use: higher variable variance means lower probability of success. In Table 3 below we show the profits under the 12 scenarios.

Table 3

Study 2: A 4-segment attitudinal segmentation (n=6000)

RANDOM UNDER-SAMPLING		LR	RF	SVM	NN
ACCURACY		43%	45%	44%	46%
COST	REVENUE	PROFIT			
$10	$50	-655	-233	-224	**348**
$10	$75	921	1311	1579	**2183**
$10	$150	5648	5941	6986	**7687**
$10	$250	11951	12114	14195	**15026**
$20	$50	-4995	-4303	-5384	**-4292**
$20	$75	-3419	-2759	-3581	**-2457**
$20	$150	1308	1871	1826	**3047**
$20	$250	7611	8044	9035	**10386**
$30	$50	-9335	-8373	-10544	**-8932**
$30	$75	-7759	-6829	-8741	**-7097**
$30	$150	-3032	-2199	-3334	**-1593**
$30	$250	3271	3974	3875	**5746**
		126	713	474	1671

As we can see in Table 3, some of the profit estimates turn negative. If a profit turns negative, this may mean that a segmentation strategy may not be best way to go about, or the firm should review alternative segmentation solutions or review what the profit scenario would look like if fewer than four segments would be pursued. Ideally, the imperfect consumer is defined based on expert estimates. E.g., experts provide an estimate of how likely a group may be to respond to a campaign. That will almost always be more useful than the three scenarios we provided.

None of the imperfect consumer scenarios affected our conclusions regarding classification methods or unbalanced/balanced, or how the best classification/imbalance correction combo can vary by cost/revenue scenario. The main reason why one should consider the imperfect consumer scenario is because it gives more realistic estimates of the upside of the planned segmentation strategy. Second, the analyses did show one interesting finding: In some scenarios the profits turn negative. This means a segmentation strategy should not be pursued, and alternative segmentation solutions should be reviewed.

7. DISCUSSION

This paper is a continuation of an earlier paper, Vriens et al. (2022). The key results of that study were that SVM, and Gradient Boosting were best in terms of achieving overall and minority segment prediction accuracy. Also, it was shown that prediction of minority segments can be substantially improved by applying imbalance correction methods. Random under-sampling and SMOTE were found to be the best methods. Lastly, they found that profitability differs substantially between the best and worst predictions. In this paper we investigated the performance of several classification methods (we added NN to our comparison) and the difference between unbalanced and optimally balanced.

We found that across datasets we can see substantial differences in prediction success. Overall, SVM performed best in our first study and NN did best in our second study. Prediction success also varies by whether we balance the segments. In general, overall accuracy decreases a little bit when we balance the data. In study one, average profitability is highest when overall accuracy is also highest. This means average profitability is higher under the unbalanced condition relative to the balanced condition. However, this is not the case for specific cost/revenue scenarios. In study 1, for the $10/$250 and $20/$250 scenarios the optimal profits are found under the balanced data and under the NN classifier that did not have the highest overall accuracy. In study two, under the unbalanced data, we found that all classification methods yielded the same accuracy, and that profitability really varies by cost/revenue scenario. Here too, overall profits are higher under the unbalanced condition except for scenarios $10/$150 and $10/$250 where we find much higher profits in the balanced situations.

Of course, there are other classification methods and imbalance correction methods that were not included in our study (see also Vriens et al., 2022, where of some of these methods are mentioned). In this study, we applied imbalance correction methods until there was complete balance. This may not be necessary or may not even be the best way to correct for imbalance. We could balance up to 90%, 80%, 70% etc. For example, when we apply SMOTE we could re-sample until the minority segment is 80% of the size of the majority segment. These topics are left for further research.

8. PRACTICAL RECOMMENDATIONS

Based on Vriens et al. (2022) and this study, we offer the following practical recommendations:

1. Run multiple classifiers, preferably some basic and some advanced (such as SVMs and NN).
2. Run the classification methods with and without imbalance correction & evaluate the solutions on overall and segment-level accuracy.
3. If the segments differ in value and the segment-level prediction varies by classification

method, calculate expected profitability. If you know the cost and revenue by segment, calculate profitability for that scenario. If you are unsure about the specific targeting cost and revenue, run several plausible cost/revenue scenarios and select the typing tool solution with highest average profitability

REFERENCES

Bennett, K.P. & Campbell, C. (2000). Support Vector Machines: Hype or Hallelujah. *SIGKDD Exploration*, 2, 2, 1-13.

Breiman L., Friedman J. H., Olshen R. A., and Stone, C. J. (1984). Classification and Regression Trees. Wadsworth.

Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, 321-357.

Cortes, C. & Vapnik, V. (1995). Support-Vector Networks. *Machine Learning*, 20, 273-297.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Liu, Y., Ram, S., Lusch, R.F. & Brusco, M. (2010). Multi-criterion market segmentation: A new model, implementation, and evaluation. *Marketing Science*, 29, 5, 880-984.

Louviere, J.J., Flynn, T.N., & Marley, A.A.J. (2015). Best-worst scaling: Theory, methods and applications. Cambridge University Press. Cambridge, United Kingdom.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, D., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J.T., Passos, A., Cournapeau, D., Brucher, M., OPerrot, M., Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks*, *61*, 85-117.

Tonks, D.G. (2009). Validity and the design of market segments. *Journal of Marketing Management*, 25, 3-4, 341-356.

Vriens, M., Bosch, N., Vidden, C. & Talwar, J. (2022). Prediction and profitability in market segmentation typing tools. *Journal of Marketing Analytics*. https://doi.org/10.1057/s41270-

021-00145-4

Wedel, M. and Kamakura, W. A. (2000). Market Segmentation: Conceptual and methodological Foundations, Kluwer Academic Publishers, Dordrecht, the Netherlands.