Business Plan –Why and How

7. Business Plan

A Business plan has been created to maximize, not only success rate but also to reduce costs.

With the Business Plan, the two main questions of this work will be answered.

- o *Why the "black box"model is predicting a client over another?*

- o *How can the results help build rapport with the client?*

<< Sales targets and leads can proiferate but they are of no use if they just create a morass of information for the sales person to go through without insights which makes Sales reps job more complicated>> Big Data ,Analytics, and the Future of Marketing and Sales[9]

To be able to carry out the analysis, I am going to analyse the predictions based on:

- - Communication channels.
- - Cost per contact.
- - Percentage of customers who open & read the communication/message.
- - Individual Important Features as oppossed to Models Global Feature Importance. Approaching customers specially from the Sales reprsentatives perspective depends on the local feature important selection.

Since new customers are not available to be used for the model, the validation file customers results will be used as an extrapolation to 100k customers.

As a result of these calculations, only 21,9% of the customers identified by the model with more than 60% probability (21.900 clients) will be approached.

7.1 Initial approach and basic information *(Why is important, from the business perspective, to avoid getting lost in translation with Sales representatives)*

Basic assumptions are created to be able to build a Business case.

Assumptions		
average loan	6.000 €	
margen per Loan	4%	240 €
loan term	4 years	
total number of clients	100k	
Success prediced cust	22%	21.900

Channels Basic information		
Channels	Unit cost	% open the communication
web	1 €	5%
email	3 €	3%
whatsApp	5 €	15%
mail	15 €	20%
phone	30 €	35%
mail + phone	45 €	55%
branch	100 €	70%
email + branch	103 €	73%
mail + branch	115 €	90%

To Break even, a mínimum number of customers is required based on the opening and success rates).Previous cost information was also used.

Minimun number of clientes by Open & Success rate								
Success Rate	Web	email	whatsApp	mail	phone	mail + phone	branch	email + branch
25%	80	133	27	20	11	7	5	4
35%	57	95	19	14	8	5	4	3
45%	44	74	15	11	6	4	3	2
55%	36	61	12	9	5	3	2	2
65%	31	51	10	8	4	3	2	2
75%	27	44	9	7	4	2	2	1
85%	24	39	8	6	3	2	2	1
95%	21	35	7	5	3	2	1	1
100%	20	33	7	5	3	2	1	1

Taking the previous analysis results to our dataset, the following table shows us the best channel approach that should be used in order to maximize our results:

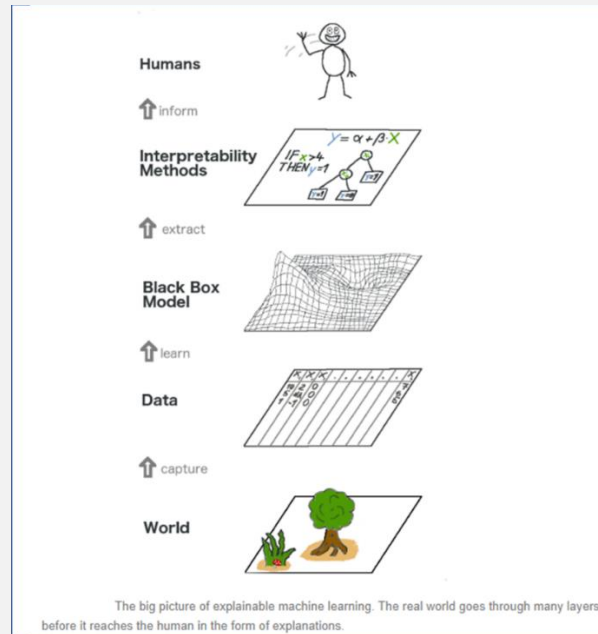|	Unit cost	1	3	5	15	30	45	100	103	115	
	% open rate	5%	3%	15%	20%	35%	55%	70%	73%	90%	
Maximun profitability matrix (based on Unit Cost,Success Rate & Open Rate)											
Success Rate	number of clients	Web	email	whatsApp	mail	phone	mail + phone	branch	email + branch	mail + branch	Most profitable Channel
25%	40	81	-48	161	-121	-363	-484	-2.339	-2.388	-2.460	whatsApp
33%	40	121	-24	282	40	-81	-40	-1.775	-1.799	-1.734	whatsApp
43%	121	501	10	1.262	674	726	1.400	-3.388	-3.377	-2.714	mail + phone
50%	484	2.420	290	6.292	4.356	5.808	10.164	-7.744	-7.453	-3.388	mail + phone
55%	444	2.484	426	6.566	5.058	7.187	12.245	-3.372	-2.946	1.686	mail + phone
57%	323	1.890	360	5.024	4.010	5.808	9.818	-1.291	-931	2.719	mail + phone
60%	1.089	6.751	1.437	18.077	15.028	22.215	37.242	871	2.309	15.899	mail + phone
61%	444	2.810	621	7.542	6.359	9.464	15.823	1.183	1.804	7.542	mail + phone
62%	645	4.120	923	11.069	9.382	13.998	23.380	2.184	3.107	11.566	mail + phone
64%	282	1.874	447	5.056	4.389	6.622	11.010	1.951	2.397	6.339	mail + phone
64%	363	2.437	591	6.586	5.756	8.712	14.467	2.904	3.495	8.660	mail + phone
65%	444	3.001	736	8.116	7.124	10.804	17.929	3.862	4.598	10.987	mail + phone
67%	1.129	7.905	2.033	21.456	19.198	29.361	48.559	13.551	15.584	32.749	mail + phone
69%	444	3.216	865	8.762	7.986	12.311	20.297	6.877	7.742	14.862	mail + phone
69%	363	2.653	720	7.232	6.617	10.219	16.837	5.919	6.640	12.537	mail + phone
70%	282	2.089	576	5.703	5.251	8.131	13.382	4.969	5.545	10.220	mail + phone
71%	605	4.581	1.296	12.532	11.667	18.149	29.816	12.099	13.396	23.767	mail + phone
72%	1.049	8.039	2.307	22.021	20.623	32.158	52.780	22.371	24.677	42.993	mail + phone
75%	726	5.808	1.742	15.971	15.245	23.957	39.202	18.875	20.617	34.120	mail + phone
76%	645	5.255	1.604	14.473	13.920	21.940	35.860	18.069	19.673	31.989	mail + phone
77%	403	3.320	1.024	9.152	8.842	13.961	22.803	11.789	12.813	20.631	mail + phone
78%	282	2.353	734	6.493	6.305	9.975	16.280	8.658	9.392	14.963	mail + phone
79%	1.210	10.253	3.248	28.338	27.701	43.940	71.641	39.482	42.730	67.184	mail + phone
80%	645	5.550	1.781	15.358	15.100	24.005	39.105	22.198	23.979	37.299	mail + phone
85%	686	6.308	2.139	17.552	17.689	28.385	46.075	29.345	31.484	47.035	mail + branch
87%	524	4.929	1.699	13.737	13.947	22.440	36.387	23.909	25.607	37.855	mail + branch
88%	565	5.364	1.863	14.963	15.245	24.562	39.807	26.538	28.401	41.783	mail + branch
89%	323	3.119	1.097	8.712	8.927	14.412	23.338	15.917	17.015	24.844	mail + branch
90%	726	7.114	2.526	19.891	20.472	33.104	53.576	37.170	39.696	57.642	mail + branch
92%	484	4.877	1.765	13.663	14.184	23.008	37.192	26.656	28.421	40.840	mail + branch
93%	524	5.318	1.932	14.905	15.505	25.167	40.671	29.361	31.294	44.866	mail + branch
93%	565	5.759	2.100	16.149	16.826	27.329	44.155	32.072	34.172	48.898	mail + branch
97%	1.129	11.955	4.463	33.606	35.397	57.710	93.107	70.249	74.712	105.646	mail + branch
100%	3.872	42.590	16.262	120.027	127.770	209.078	336.849	263.284	279.546	391.054	mail + branch

In summary:

A potential outcome of 21% ROI (~350k €) per year could result if these customers are addressed:

Summary						
Channels	% Success Rate	number of clients	Profitability	*Income*	*Cost*	*~ ROI per year*
whatsApp	<40%	80	444 €	*847 €*	*403 €*	28%
mail + phone	40%-80%	12.422	560.043 €	*1.119.037 €*	*558.994 €*	25%
mail + branch	>80%	9.397	840.464 €	*1.921.146 €*	*1.080.682 €*	19%
Total		21.899	1.400.951 €	*3.041.030 €*	*1.640.079 €*	**21%**

This Business Case has been developed by channel which would need to be completed with a comunication, marketing and product plan which is not included in this exercise. Further actions would be required and put into place(i.e triggers for customers for simulation purposes or through any other channel that provides further information). The following section will cover how to help sales representatives to approach the clients.
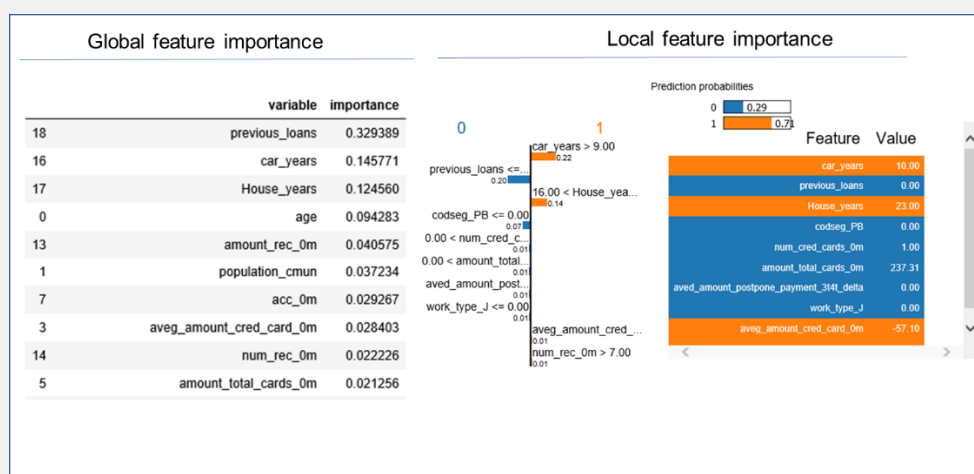
7.2 How can 'Lost in translation' be avoided with Sales representatives.

o After defining the number of customers (9.397) that will be managed by Sales Representative, explaining to them why this model has selected them and help them find the best manner to present it to the client is crucial.

o To do this comparing Global Feature Importance with Local Features selected for a client. For the Local Features selection, the model used was LIME(Local Interpretable Model-agnostic Explanations).

o LIME is a Python library for model explainability. LIME builds sparse linear models around each prediction to explain how the 'black box' model works in that local vicinity. According to <<Why should I Trust You?>> original paper[10], LIME is a subset of SHAP. Is used because is faster than Shap and it uses the Chi-Squared Test, a distribution-based approximation as a good approximation not as accurate as the Fisher Test used by Shap.[10]

o <<We capture the world by collecting data, and abstract it further by learning to predict the data (for the task) with a machine learning model. Interpretability is just another layer on top that helps humans understand>> Interpretable Machine Learning by Christoph Molnar [11]

The big picture of explainable machine learning. The real world goes through many layers before it reaches the human in the form of explanations.

o   A tool tradicionally used to detect bias in machine learning models, can also be used to give sales representatives tips on how best approaching a client. [12]

o   Steps to follow(*Interpretable Machine Learning by Christoph Molnar*):

-   Select instance of interest for which you want to have an explanation of its black box prediction.
-   Perturb dataset and get the black box predictions for these new points.
-   Weight the new samples according to their proximity to the instance of interest.
-   Train a weighted, interpretable model on the dataset with the variations.
-   Explain the prediction by interpreting the local model.
-   Plotting results

Random Forest Classifier

	variable	importance
18	previous_loans	0.329389
16	car_years	0.145771
17	House_years	0.124560
0	age	0.094283
13	amount_rec_0m	0.040575
1	population_cmun	0.037234
7	acc_0m	0.029267
3	aveg_amount_cred_card_0m	0.028403
14	num_rec_0m	0.022226
5	amount_total_cards_0m	0.021256

Insights:

- The model rates all the features but does not give you how each affects the prediction. By using LIME,one can see that this customer has a 8-years-old car and his house is 23. Sales Representatives can use this information to offer clients the possibility to change their car without having to pay all upfront or to ask if hi/she is thinking on reforming any part of his/her house.
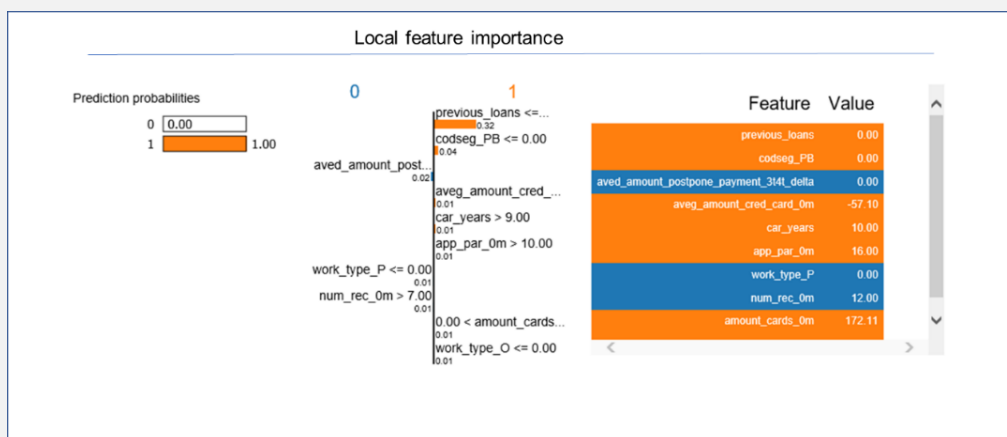
XGBoost
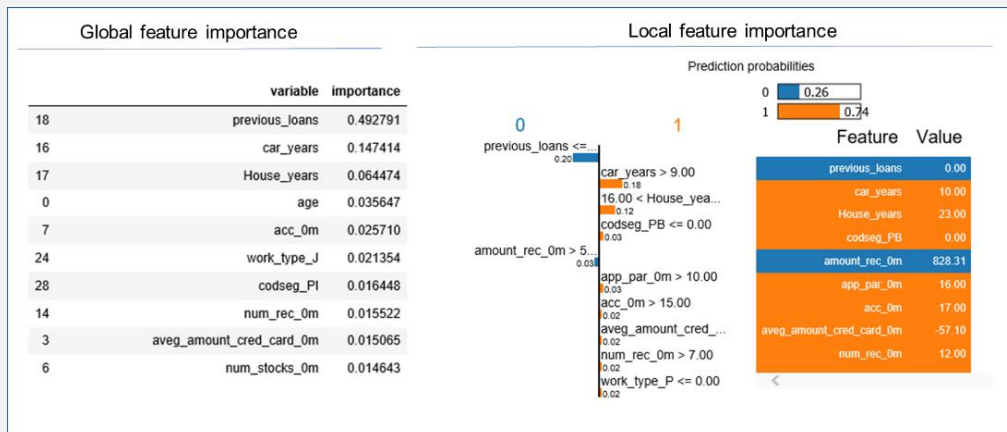
Insights:

- Similar results. As seen above, this model includes more features than the Random Forest Model. Features like no belonging to a private banking sector for instance.

Stacking model

Insights:

- The Global feature importance is the average of the global features of the trees used.
- The same results were obtained with the Stacking Method than with the previous models. It is interesting to see how this model has selected additional features, like how many times he/she checks her bank details through the app or the amount of money expent on cards.

[9]Big Data ,Analytics, and the Future of Marketing and Sales by Mckinsey Chef Marketing & Sales Officer Forum.
[10]Why should I Trust You? Explaining the Predictions of Any Classifier by Marco Tulio Ribeiro
[11]Interpretable Machine Learning: https://christophm.github.io/interpretable-ml-book/storytime.html
[12]LIME: https://blog.dominodatalab.com/shap-lime-python-libraries-part-1-great-explainers-pros-cons/

Although the initial information included in the dataset is real, in order to present this work and to comply with current regulations, all data has been anonymized.