UTN.BA
UNIVERSIDAD TECNOLÓGICA NACIONAL
FACULTAD REGIONAL BUENOS AIRES

Practical Work Presentation

Group 4

Carlos Comesaña
Mauro Giraldez
Daniel Lunati
Juan Pablo Gonzalez
Juan Ignacio Cuiule

Data Universe & Objective

238615 Records - 77 Columns

9 Months of data

August 2018 - April 2019

7 CC products

8430 clients with no CC product

26006 clients

without package & ~26% Target

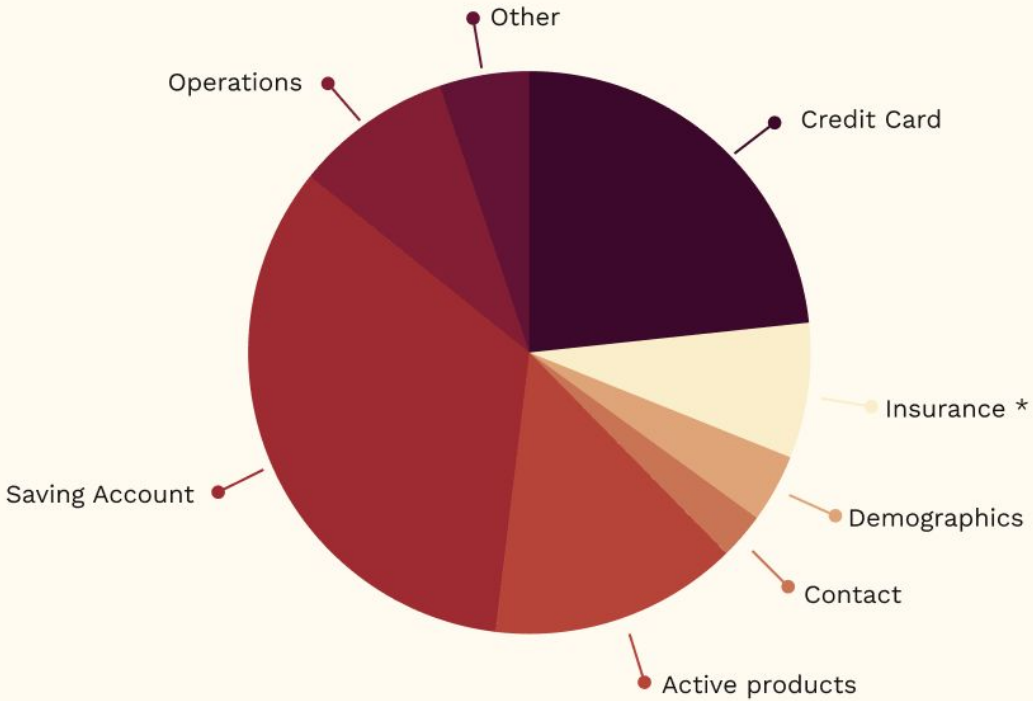
7 Regions

🔮 Predict clients that could get a package

analyzing the behaviour of users with
the “Target” label in the dataset

Features

🔍 Analysis



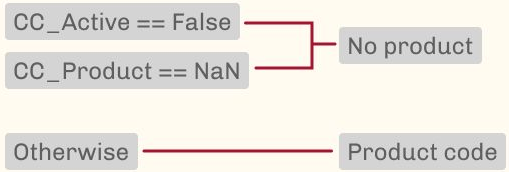
Features

🧰 Preprocessing and transformation

❌ Missing values

Credit Card

Fix the relation between “CC_Active” and “CC_Product”.



Insurance

We didn’t find any data about insurance in those columns.

Region

Filled the “NaN” region to a “NO REGION” string.

Transform categorical features

- Sex
- Credit card
- Age groups
- Regions

Aggregate features

- Credit card spending
- Days with saving account operations
- Total # of operations

Features selection

💯 Top features by model

Model 1

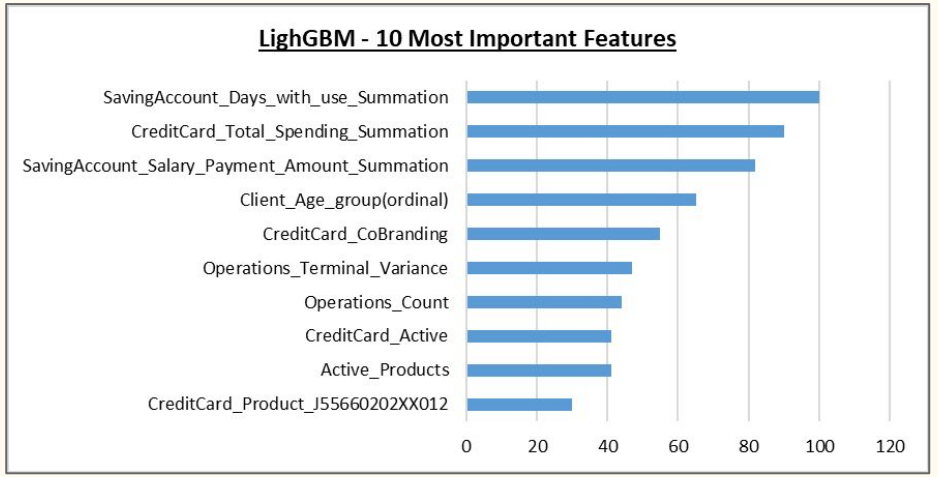
LightGBM

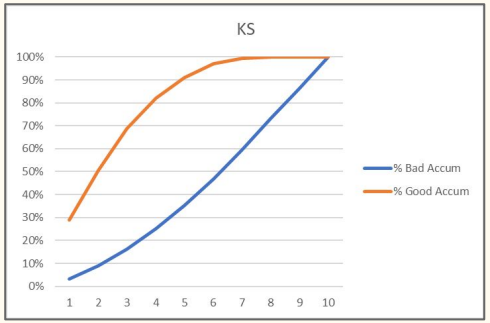
Training

Score	bad	good	Total	Bad accum	Goods accum	% bads	% goods	Lift	% bad accum	% good accum	%pob	KS	%pob accum
1	445	1.376	1.821	445	1376	24%	76%	2,88	3%	29%	10%	25%	10%
2	780	1.040	1.820	1225	2416	43%	57%	2,18	9%	51%	10%	41%	20%
3	945	874	1.819	2170	3290	52%	48%	1,83	16%	69%	10%	53%	30%
4	1.195	627	1.822	3365	3917	66%	34%	1,31	25%	82%	10%	57%	40%
5	1.386	434	1.820	4751	4351	76%	24%	0,91	35%	91%	10%	56%	50%
6	1.537	288	1.825	6288	4639	84%	16%	0,60	47%	97%	10%	50%	60%
7	1.708	107	1.815	7996	4746	94%	6%	0,22	60%	99%	10%	40%	70%
8	1.806	25	1.831	9802	4771	99%	1%	0,05	73%	100%	10%	27%	80%
9	1.802	9	1.811	11604	4780	100%	0%	0,02	86%	100%	10%	14%	90%
10	1.820	-	1.820	13424	4780	100%	0%	-	100%	100%	10%	0%	100%
	13.424	4.780	18.204				26%						

Testing

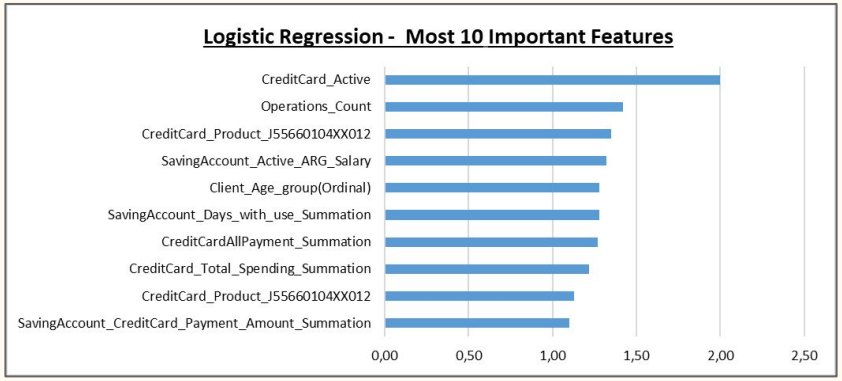
Score	bad	good	Total	Bad accum	Goods accum	% bads	% goods	Lift	% bad accum	% good accum	%pob	KS	%pob accum
1	195	636	831	195	636	23%	77%	2,91	3%	31%	11%	28%	11%
2	363	435	798	558	1.071	45%	55%	2,08	10%	52%	10%	43%	21%
3	404	339	743	962	1.410	54%	46%	1,74	17%	69%	10%	52%	30%
4	489	277	766	1.451	1.687	64%	36%	1,38	25%	82%	10%	57%	40%
5	585	205	790	2.036	1.892	74%	26%	0,99	35%	92%	10%	57%	50%
6	650	108	758	2.686	2.000	86%	14%	0,54	47%	98%	10%	51%	60%
7	697	39	736	3.383	2.039	95%	5%	0,20	59%	100%	9%	41%	69%
8	786	9	795	4.169	2.048	99%	1%	0,04	72%	100%	10%	27%	80%
9	803	1	804	4.972	2.049	100%	0%	0,00	86%	100%	10%	14%	90%
10	781	-	781	5.753	2.049	100%	0%	-	100%	100%	10%	0%	100%
	5.753	2.049	7.802				26%						

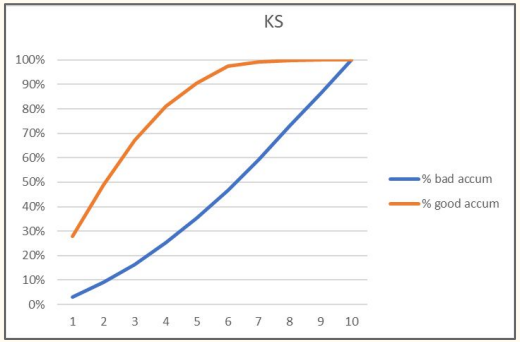




Model 2

Logistic Regression





Training

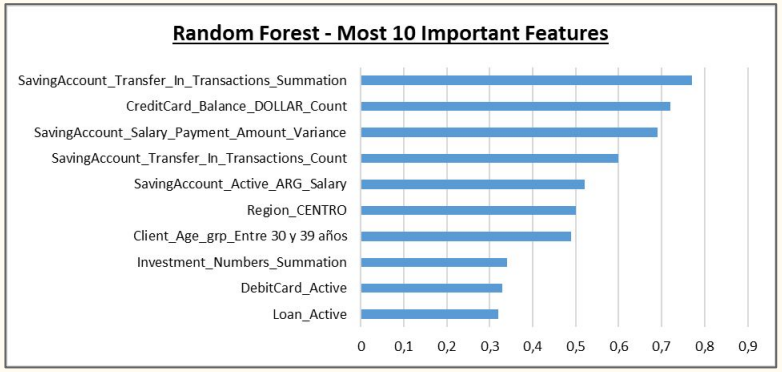
Score	bad	good	Total	Bad accum	Goods accum	% bads	% goods	Lift	% bad accum	% good accum	%pob	KS	%pob accum
1	471	1.350	1.821	471	1350	26%	74%	2,83	4%	28%	10%	25%	10%
2	749	1.071	1.820	1220	2421	41%	59%	2,25	9%	51%	10%	42%	20%
3	1.017	804	1.821	2237	3225	56%	44%	1,69	17%	68%	10%	51%	30%
4	1.158	662	1.820	3395	3887	64%	36%	1,39	25%	82%	10%	56%	40%
5	1.369	452	1.821	4764	4339	75%	25%	0,95	35%	91%	10%	56%	50%
6	1.536	284	1.820	6300	4623	84%	16%	0,60	47%	97%	10%	50%	60%
7	1.725	95	1.820	8025	4718	95%	5%	0,20	60%	99%	10%	39%	70%
8	1.795	26	1.821	9820	4744	99%	1%	0,05	73%	100%	10%	26%	80%
9	1.803	17	1.820	11623	4761	99%	1%	0,04	86%	100%	10%	13%	90%
10	1.816	4	1.820	13439	4765	100%	0%	0,01	100%	100%	10%	0%	100%
	13.439	4.765	18.204				26%						

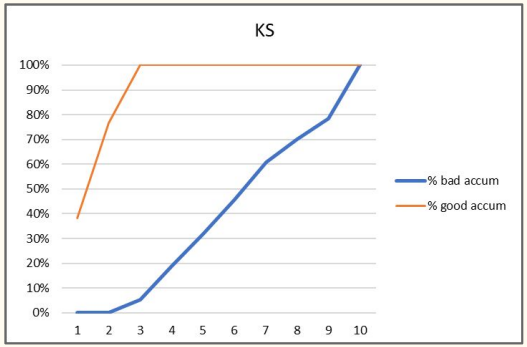
Testing

Score	bad	good	Total	Bad accum	Goods accum	% bads	% goods	Lift	% bad accum	% good accum	%pob	KS	%pob accum
1	178	575	753	178	575	24%	76%	2,89	3%	28%	10%	25%	10%
2	344	435	779	522	1.010	44%	56%	2,11	9%	49%	10%	40%	20%
3	415	378	793	937	1.388	52%	48%	1,80	16%	67%	10%	51%	30%
4	514	282	796	1.451	1.670	65%	35%	1,34	25%	81%	10%	56%	40%
5	582	201	783	2.033	1.871	74%	26%	0,97	35%	91%	10%	55%	50%
6	642	143	785	2.675	2.014	82%	18%	0,69	47%	98%	10%	51%	60%
7	734	35	769	3.409	2.049	95%	5%	0,17	59%	99%	10%	40%	70%
8	798	10	808	4.207	2.059	99%	1%	0,05	73%	100%	10%	26%	80%
9	749	4	753	4.956	2.063	99%	1%	0,02	86%	100%	10%	14%	90%
10	782	1	783	5.738	2.064	100%	0%	0,00	100%	100%	10%	0%	100%
	5.738	2.064	7.802				26%						

Model 3

Random Forest





Training

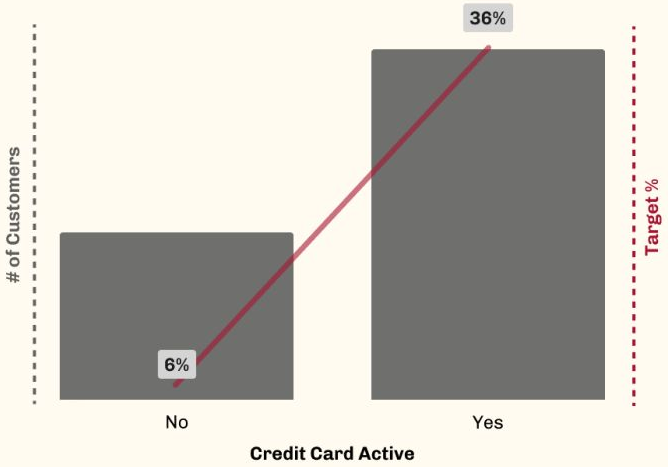
Score	bad	good	Total	Bad accum	Goods accum	% bads	% goods	Lift	% bad accum	% good accum	%pob	KS	%pob accum
1	-	1.829	1829	0	1829	0%	100%	3,82	0%	38%	10%	38%	10%
2	-	1.824	1824	0	3653	0%	100%	3,82	0%	77%	10%	77%	20%
3	695	1.112	1807	695	4765	38%	62%	2,35	5%	100%	10%	95%	30%
4	1.850	-	1850	2545	4765	100%	0%	-	19%	100%	10%	81%	40%
5	1.759	-	1759	4304	4765	100%	0%	-	32%	100%	10%	68%	50%
6	1.862	-	1862	6166	4765	100%	0%	-	46%	100%	10%	54%	60%
7	1.996	-	1996	8162	4765	100%	0%	-	61%	100%	11%	39%	71%
8	1.263	-	1263	9425	4765	100%	0%	-	70%	100%	7%	30%	78%
9	1.134	-	1134	10559	4765	100%	0%	-	79%	100%	6%	21%	84%
10	2.880	-	2880	13439	4765	100%	0%	-	100%	100%	16%	0%	100%
	13.439	4.765	18.204				26%						

Testing

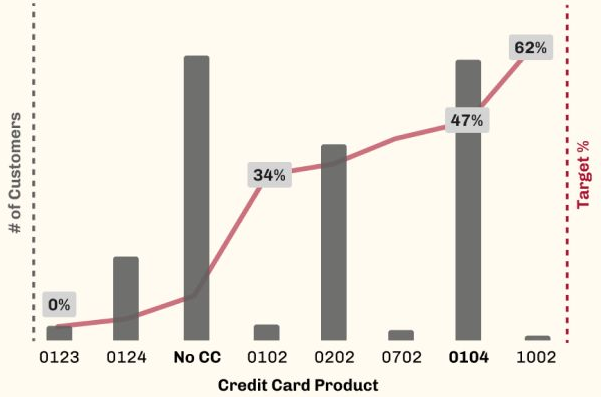
Score	bad	good	Total	Bad accum	Goods accum	% bads	% goods	Lift	% bad accum	% good accum	%pob	KS	%pob accum
1	10	42	52	10	42	19%	81%	3,05	0%	2%	1%	2%	1%
2	40	164	204	50	206	20%	80%	3,04	1%	10%	3%	9%	3%
3	2.040	1.680	3.720	2.090	1.886	55%	45%	1,71	36%	91%	48%	55%	51%
4	505	99	604	2.595	1.985	84%	16%	0,62	45%	96%	8%	51%	59%
5	299	33	332	2.894	2.018	90%	10%	0,38	50%	98%	4%	47%	63%
6	461	24	485	3.355	2.042	95%	5%	0,19	58%	99%	6%	40%	69%
7	797	16	813	4.152	2.058	98%	2%	0,07	72%	100%	10%	27%	80%
8	435	2	437	4.587	2.060	100%	0%	0,02	80%	100%	6%	20%	85%
9	306	1	307	4.893	2.061	100%	0%	0,01	85%	100%	4%	15%	89%
10	845	3	848	5.738	2.064	100%	0%	0,01	100%	100%	11%	0%	100%
	5.738	2.064	7.802				26%						

Features analysis

Credit Card

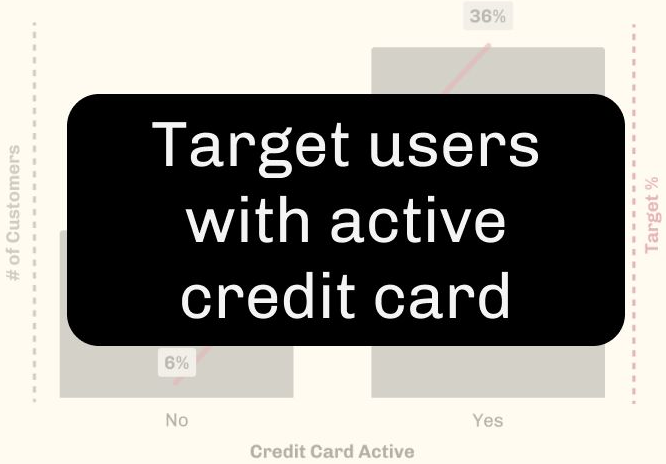


Credit Card Active	# of clients	# of target clients	target %
No	8439	516	6.12%
Yes	17576	6313	35.91%



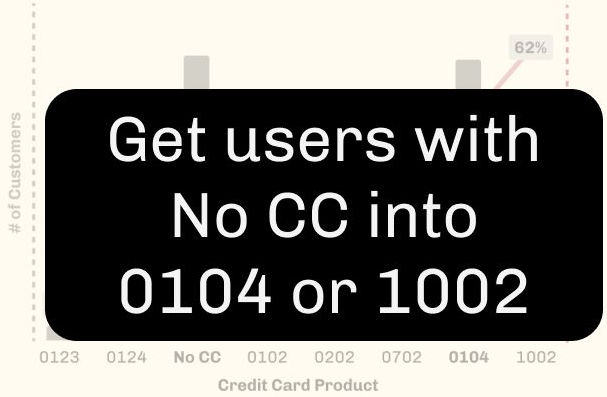
CC Product	# of clients	# of target clients	target %
0123	365	0	0%
0124	2430	15	0.6%
No CC	8430	516	6.12%
0102	404	138	34.15%
0202	5785	2132	36.8%
0702	251	107	42.6%
0104	8283	3885	46.9%
1002	58	36	62%

Credit Card



Target users with active credit card

Credit Card Active	# of clients	# of target clients	target %
No	8439	516	6.12%
Yes	17576	6313	35.91%

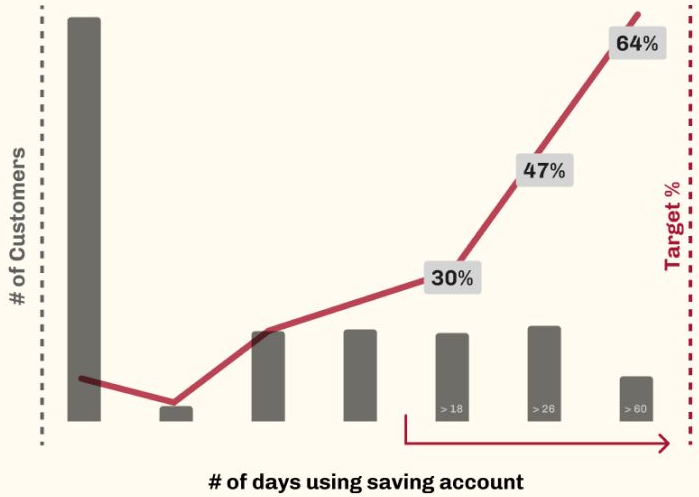


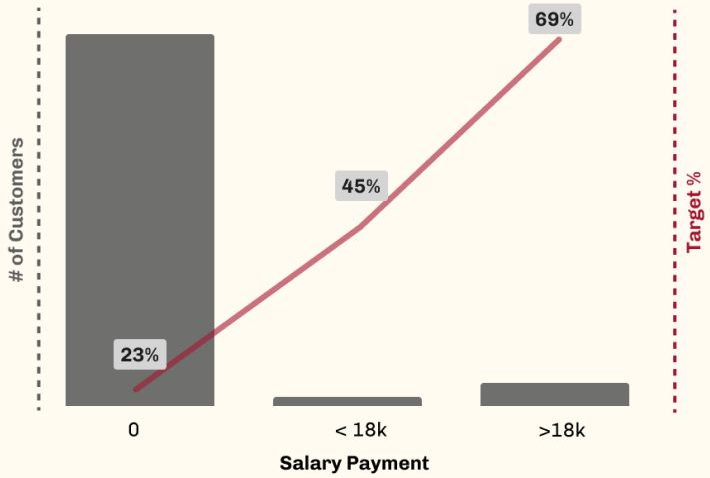
Get users with No CC into 0104 or 1002

CC Product	# of clients	# of target clients	target %
1002	58	36	62%

Target clients with 0104 and 1002 with package ads

Saving Account

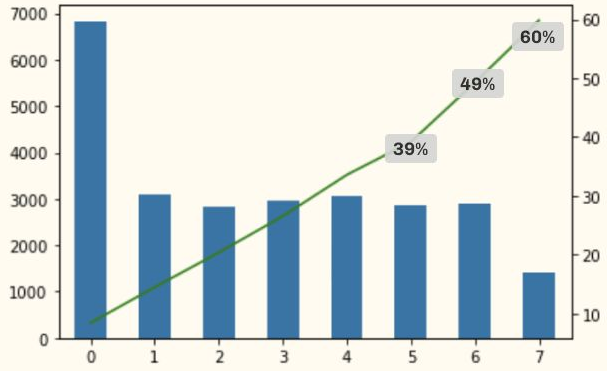




Saving Account

Operations and contact

Operations	# of clients	# of target clients	target %
0	6832	575	8.41%
1 - 2	3092	448	14.5%
3 - 5	2844	579	20.3%
6 - 8	2976	790	26.5%
9 - 14	3047	1023	33.6%
15 - 27	2878	1127	39.1%
28 - 67	2911	1434	49.2%
68 - 401	1426	853	59.8%



Email	# of clients	# of target clients	target %
No	7512	1747	23.25%
Yes	18494	5082	27.47%

Mobile	# of clients	# of target clients	target %
No	3078	647	21.02%
Yes	18494	6182	26.96%

Demographics

Age Group	# of clients	# of target clients	target %
18 - 29	1125	249	22%
30 - 39	5875	1215	21%
40 - 49	7265	1662	23%
50 - 59	5740	1798	31%
60 - 64	2408	776	32%
65 - 69	1991	694	35%
> 70	1602	435	27%

Sex	# of clients	# of target clients	target %
F	11171	2847	25.48%
M	14835	3982	26.84%

Region	# of clients	# of target clients	target %
Buenos Aires	6991	2138	31%
Centro	5768	1116	19%
Norte Grande	4966	1059	21%
Patagonia	2437	821	34%
CABA Centro - Norte	2050	613	30%
Cuyo	1960	490	25%
Amba Resto	1828	592	32%
Sin región	6	0	0%

Conclusions

About the obtained results with the model

LightGBM, for this model, was the best option. Its provides an AUC (Area Under Curve) highly superior compared to ones provided by Random Forest and Logistic Regression.

We conclude that the sales workforce should focus on the clients that made more operations (quantity of all types of transactions); and also that there is no plausible difference between age, living region and sex.

About the practical work experience

We learned how to select, preprocess and transform the given data, how to detect patterns and how to evaluate and understand the results.

Also, we learn about the different algorithms used and how each one perform (AUC and training time).
Finally, we learned how to present data to stakeholders.

UTN.BA
UNIVERSIDAD TECNOLÓGICA NACIONAL
FACULTAD REGIONAL BUENOS AIRES

Thanks

Group 4