Bank term deposit modelling

ALEJANDRO MARTINEZ
JUL 20TH 2025

Data Science Solution Outline

Developed a solution that can help optimize the targeting of customers that are more propense (more likely) to get a term deposit when prompted by marketing phone calls.

This model can **add value by reducing costs** of man staff making calls to customers

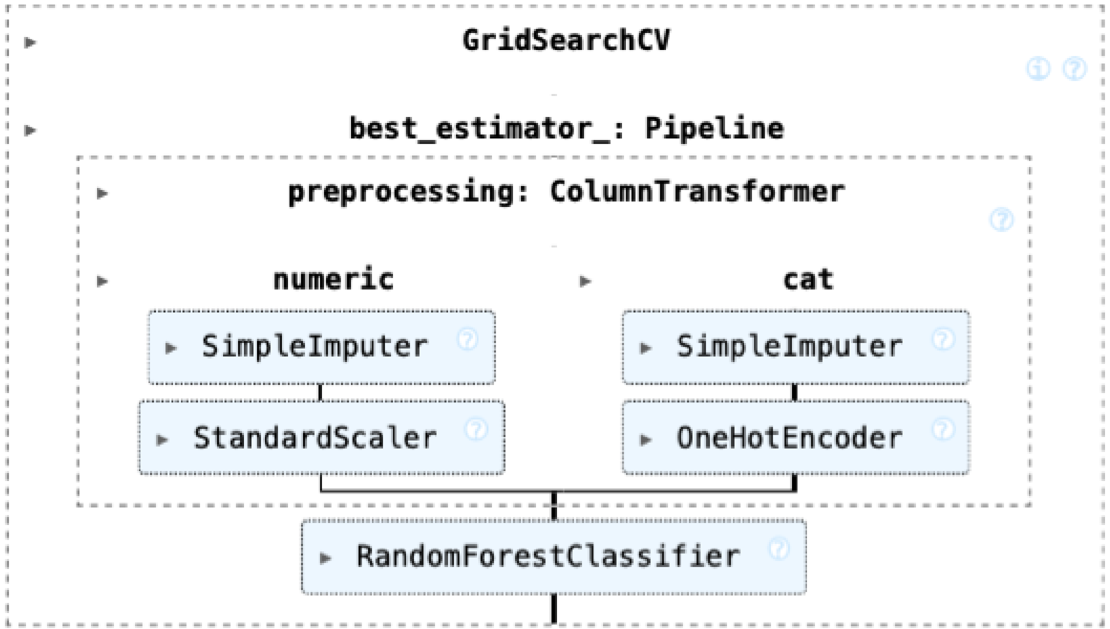
The model can predict if the customer is likely to get a term deposit. So the marketers can focus is only calling this high likely customers instead of wasting resource and time on customers that are not likely to get the product.

How was it developed?

- Taking data from direct marketing campaigns we have trained a machine learning model using data history and descriptors for the customers and call logs.
- Using this specific set of datapoints:
- 'age', 'balance', 'day', 'duration', 'campaign', 'pdays', 'previous', 'job', 'marital', 'education', 'default', 'housing', 'loan', 'contact', 'month', 'poutcome'
- The dataset description can be found here:

  https://archive.ics.uci.edu/dataset/222/bank+marketing

- The modelling technique used was a Random Forest, with K-fold validation. Using hold-out test set for evaluation of the accuracy of the model.

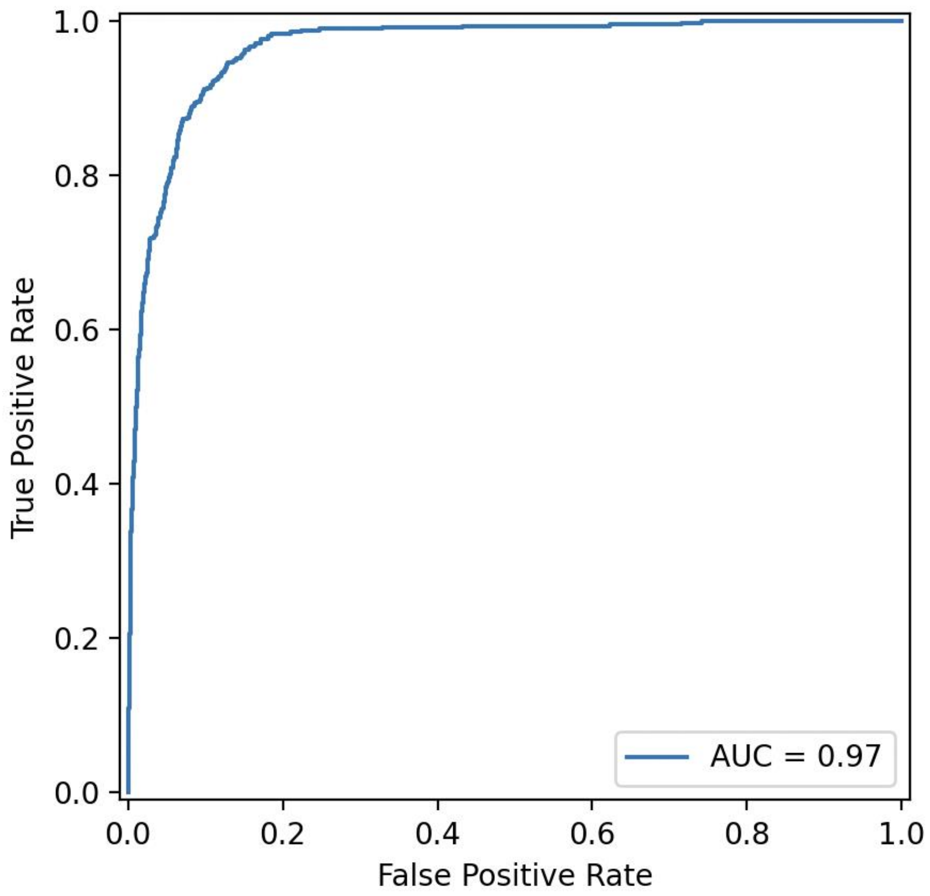


	age	job	marital	education	default	balance	housing	loan	contact	day	month	duration	campaign	pdays	previous	poutcome	y
0	58	management	married	tertiary	no	2143	yes	no	unknown	5	may	261	1	-1	0	unknown	no
1	44	technician	single	secondary	no	29	yes	no	unknown	5	may	151	1	-1	0	unknown	no
2	33	entrepreneur	married	secondary	no	2	yes	yes	unknown	5	may	76	1	-1	0	unknown	no
3	47	blue-collar	married	unknown	no	1506	yes	no	unknown	5	may	92	1	-1	0	unknown	no
4	33	unknown	single	unknown	no	1	no	no	unknown	5	may	198	1	-1	0	unknown	no
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
45206	51	technician	married	tertiary	no	825	no	no	cellular	17	nov	977	3	-1	0	unknown	yes
45207	71	retired	divorced	primary	no	1729	no	no	cellular	17	nov	456	2	-1	0	unknown	yes
45208	72	retired	married	secondary	no	5715	no	no	cellular	17	nov	1127	5	184	3	success	yes
45209	57	blue-collar	married	secondary	no	668	no	no	telephone	17	nov	508	4	-1	0	unknown	no
45210	37	entrepreneur	married	secondary	no	2971	no	no	cellular	17	nov	361	2	188	11	other	no

How good is the model?

- The models achieves an overall AUROC curve of .97 on a hold-out (unseen) dataset. This means the model predictive power is very good and it can be used with confidence to localize the most propense customers.

- Note the development time was restricted to only few hours of training. Possible improvements as more hyperparameter tuning and more complex modelling techniques could raise the performance of the model even further if ore time allowed for development.



How can we use this model for optimizing out marketing?

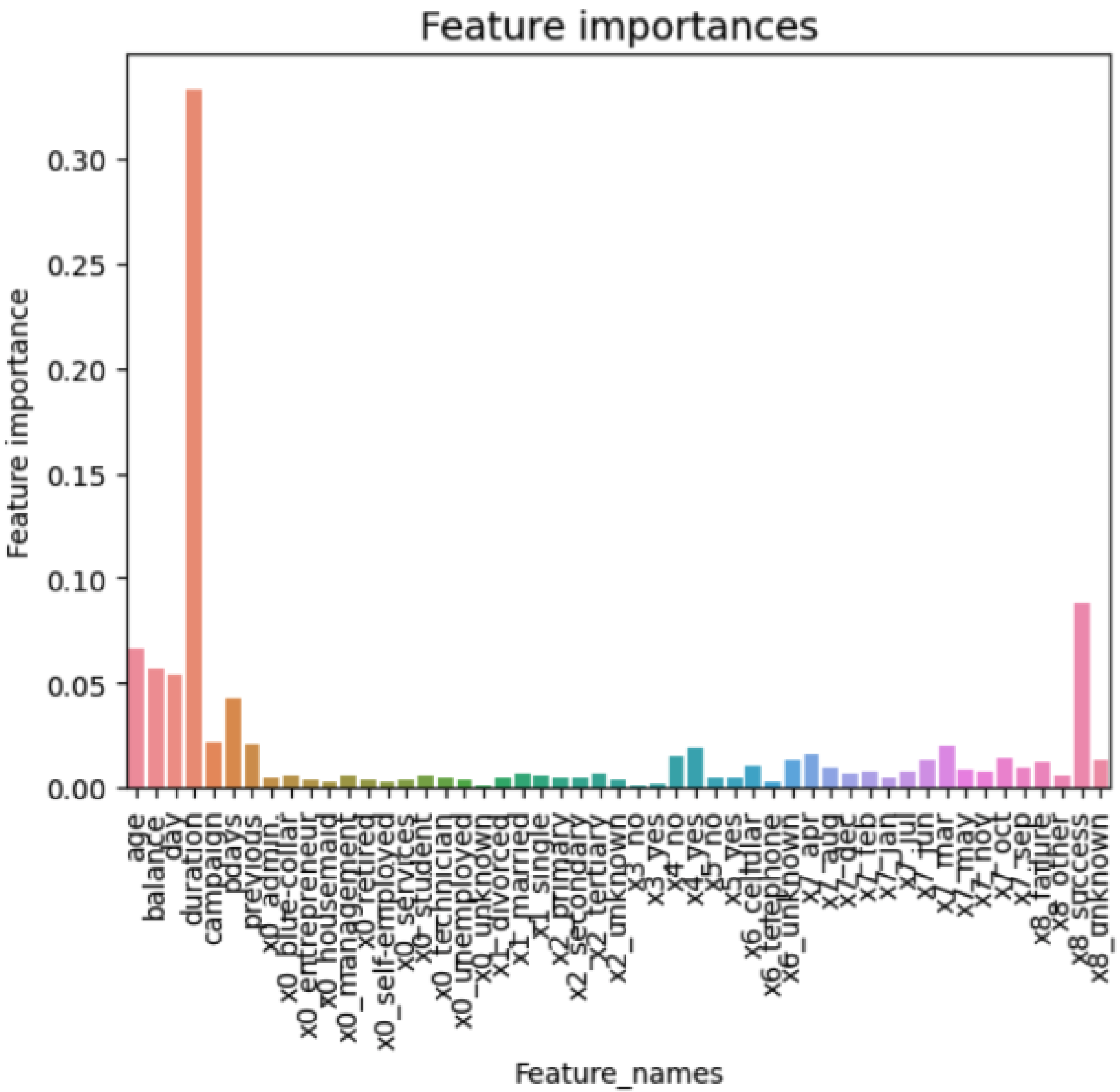
The model provides us with the probability that the customer will take on the product. By only calling the most propense customers we can guarantee that the number of call will be reduce to only target the customers that will actually take on the term deposit

The marketers can use for instance. Only the top 10% most propense costumers and focus only on them. This will generate a decrease in cost hence optimizing the marketing budget.

The prediction for the model are ready to used on new data using the scripts generated as part of this workl.

Some Insights on the model's feature importance

- After fitting the model we can extract the most important features and gain some insights on what is the most predictive features and sense check our model as well. This was done as part of the model post analysis. Localizing that the 3 most important features were:
- 1. The duration of the last call to that customer
- 2. If the outcome of the previous marketing campaign was a "success"
- 3. The age of the customer



- Thank you