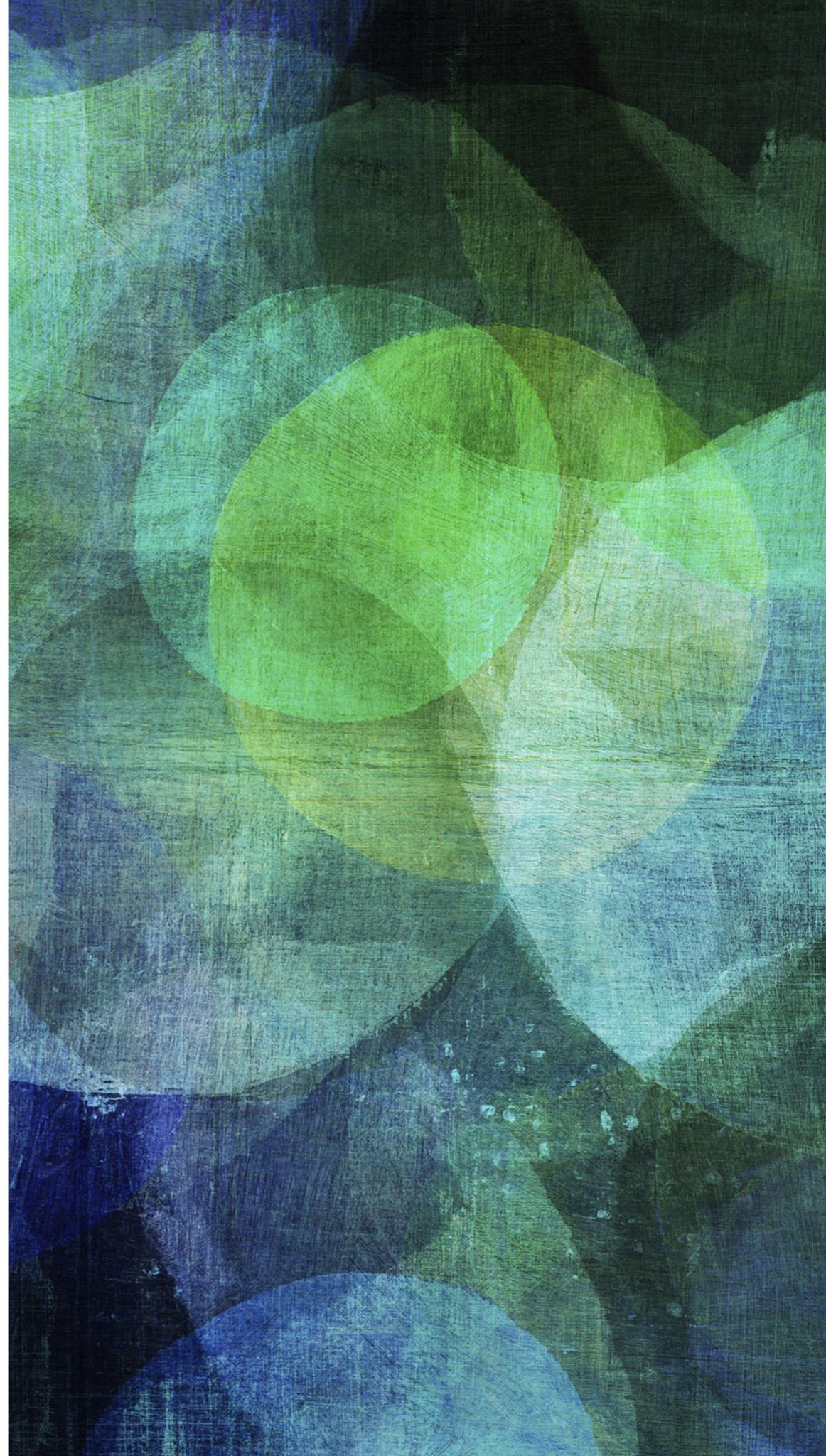


LOANS DEFAULT IN CREDIT ONE

Mary Martinez



THE ISSUE

- Our client Credit One is a company that offers credit scoring service.
- Credit One is in risk to lose the business due to:
 - There has been an increase in the number of customers who have defaulted on loans they have secured from various partners.
- Credit One wants to find a much better way to understand how much credit to allow someone to use or, at the very least, if someone should be approved or not.
- Credit One is requesting our help for the design and implementation of a creative and empirically sound solution.



OUR GOAL

- To build a predictive model for **Credit One** that allow them to know when a loan applicant is likely to default.

INFORMATION PROVIDED - THE DATA

- Credit One provided a database that contains **25 columns** and **30203 entrances or rows**:

Default Payment Next Month								Limit Balance
ID	DPNM	SEX	EDUCATION	MARRIAGE	AGE	LIMIT_BAL		
1	1	default	female	university	1	24	20000	
2	2	default	female	university	2	26	120000	
3	3	not default	female	university	2	34	90000	
4	4	not default	female	university	1	37	50000	
5	5	not default	male	university	1	57	50000	
...
30199	29996	not default	male	high school	1	39	220000	
30200	29997	not default	male	high school	2	43	150000	
30201	29998	default	male	university	2	37	30000	
30202	29999	default	male	high school	1	41	80000	
30203	30000	default	male	university	1	46	50000	

Continue

INFORMATION PROVIDED - THE DATA

Continue

PAY_0	PAY_2	PAY_3	PAY_4	PAY_5	PAY_6	BILL_AMT1	BILL_AMT2	BILL_AMT3	BILL_AMT4	BILL_AMT5	BILL_AMT6
2	2	-1	-1	-2	-2	3913	3102	689	0	0	0
-1	2	0	0	0	2	2682	1725	2682	3272	3455	3261
0	0	0	0	0	0	29239	14027	13559	14331	14948	15549
0	0	0	0	0	0	46990	48233	49291	28314	28959	29547
-1	0	-1	0	0	0	8617	5670	35835	20940	19146	19131
...
0	0	0	0	0	0	188948	192815	208365	88004	31237	15980
-1	-1	-1	-1	0	0	1683	1828	3502	8979	5190	0
4	3	2	-1	0	0	3565	3356	2758	20878	20582	19357
1	-1	0	0	0	-1	-1645	78379	76304	52774	11855	48944
0	0	0	0	0	0	47929	48905	49764	36535	32428	15313

PAY_AMT1	PAY_AMT2	PAY_AMT3	PAY_AMT4	PAY_AMT5	PAY_AMT6
0	689	0	0	0	0
0	1000	1000	1000	0	2000
1518	1500	1000	1000	1000	5000
2000	2019	1200	1100	1069	1000
2000	36681	10000	9000	689	679
...
8500	20000	5003	3047	5000	1000
1837	3526	8998	129	0	0
0	0	22000	4200	2000	3100
85900	3409	1178	1926	52964	1804
2078	1800	1430	1000	1000	1000

GENERAL DATA SCIENCE PROCESS

Step 1: understanding the data

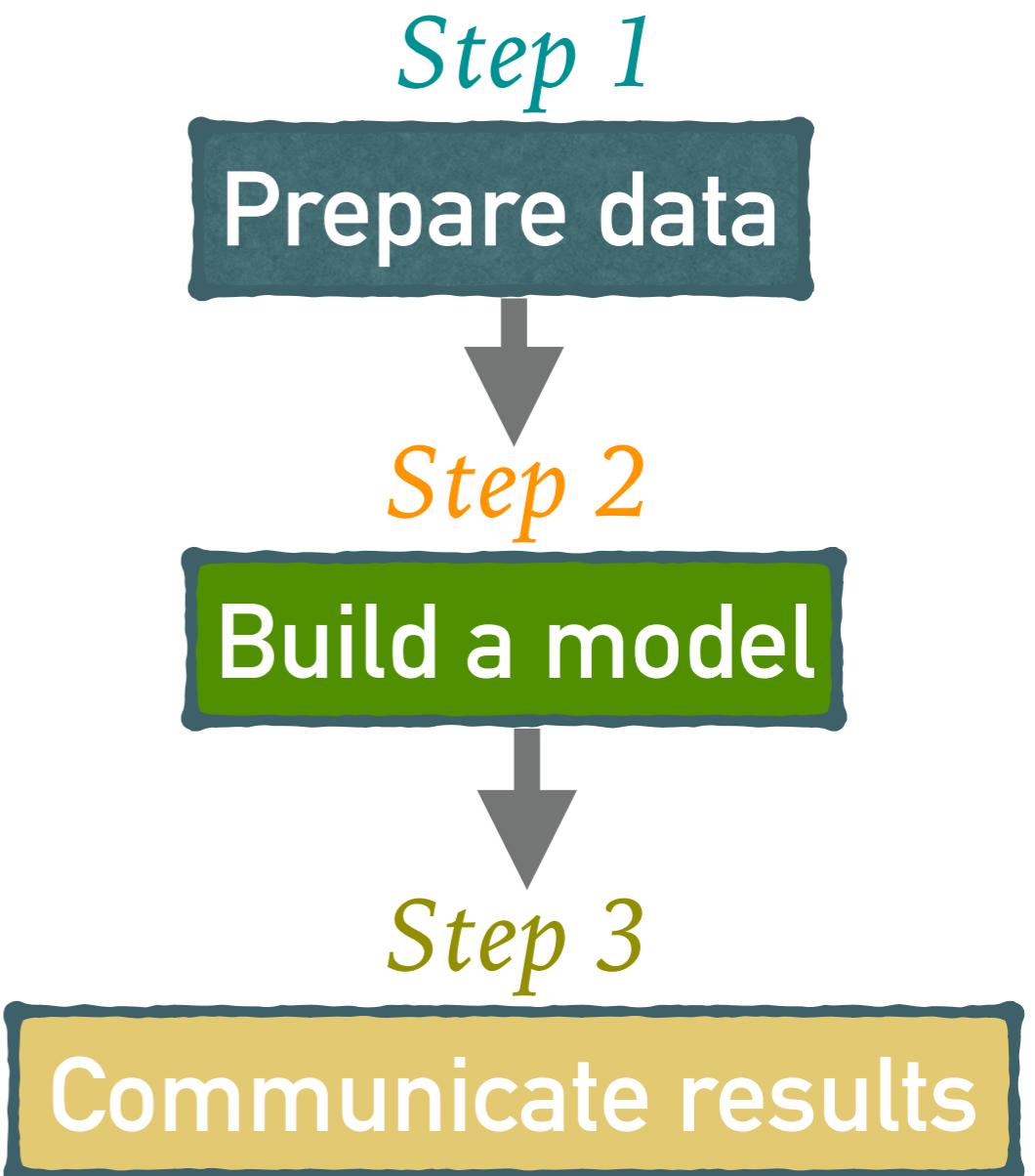
- Do we have all the data required to address the objective?, are there any privacy issues, cleaning the data, etc.

Step 2: model training

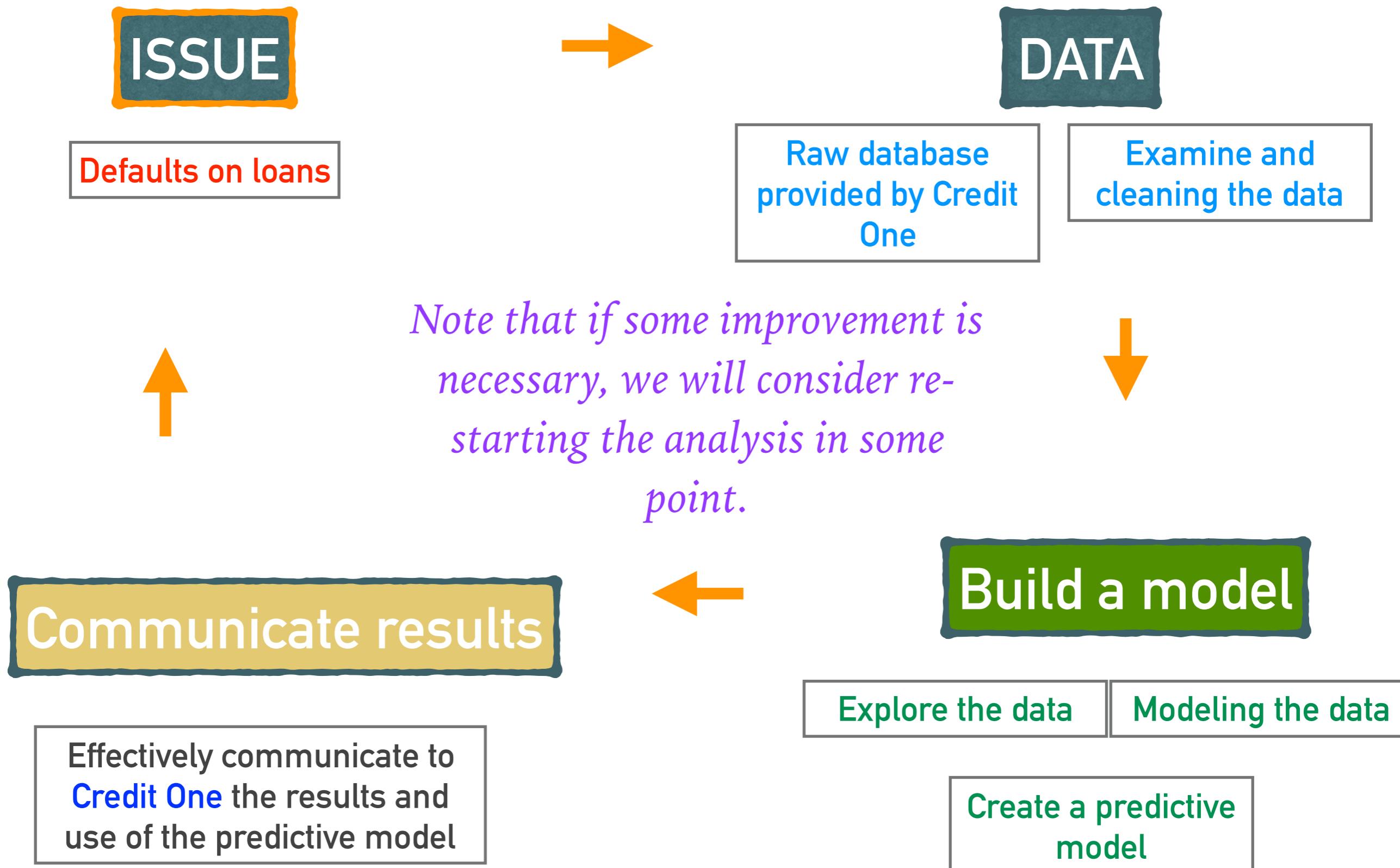
- Choosing a ML model, for example a Random Forest Classifier algorithm.
- Training

Step 3: communicating results

- Present to Credit One the resultant model that predict when to approve a loan



EXTENDED DATA SCIENCE PROCESS



GENERAL COMMENTS

- From an initial insight of the data we note that this doesn't include key information as the date of the transactions, or explain the meaning of the values under the tags PAY_0, PAY_2, etc., for. Example what does mean 0, -2 or 4?
- There are six columns for PAY, BILL_AMT, and PAY_AMT. We noted that for the last two the final number in the name of the column vary from 1 to 6, while for PAY from 2 to 6 and start in 1. This is confusing because it is not clear if there is lost data or if this is a typing error.

PAY_0

PAY_1??

PAY_2

.

PAY_6

BILL_AMT_0??

BILL_AMT_1

BILL_AMT_2

.

BILL_AMT_6

PAY_AMT_0??

PAY_AMT_1

PAY_AMT_2

.

.

.

PAY_AMT_6

