

Data Science in Telecom industry

Presented by Mirjan Miftaraj
11 March 2024

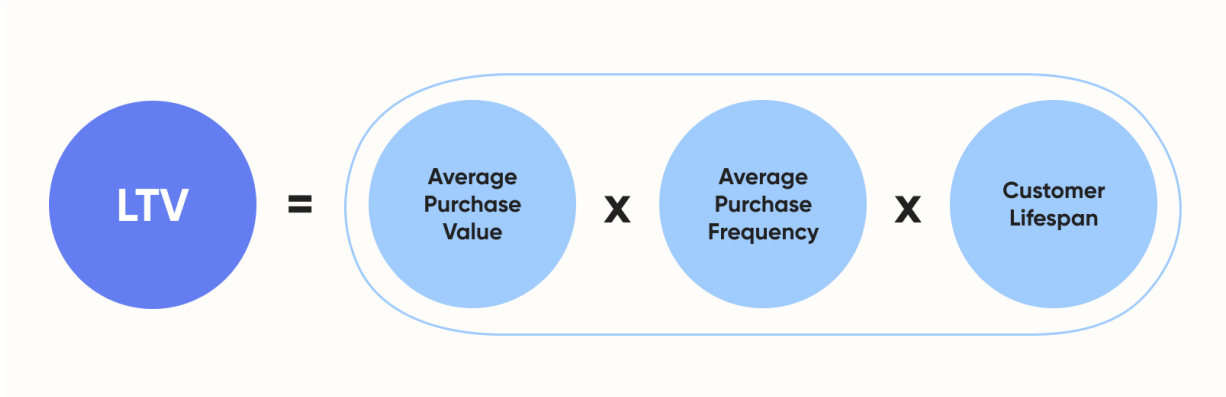


quote

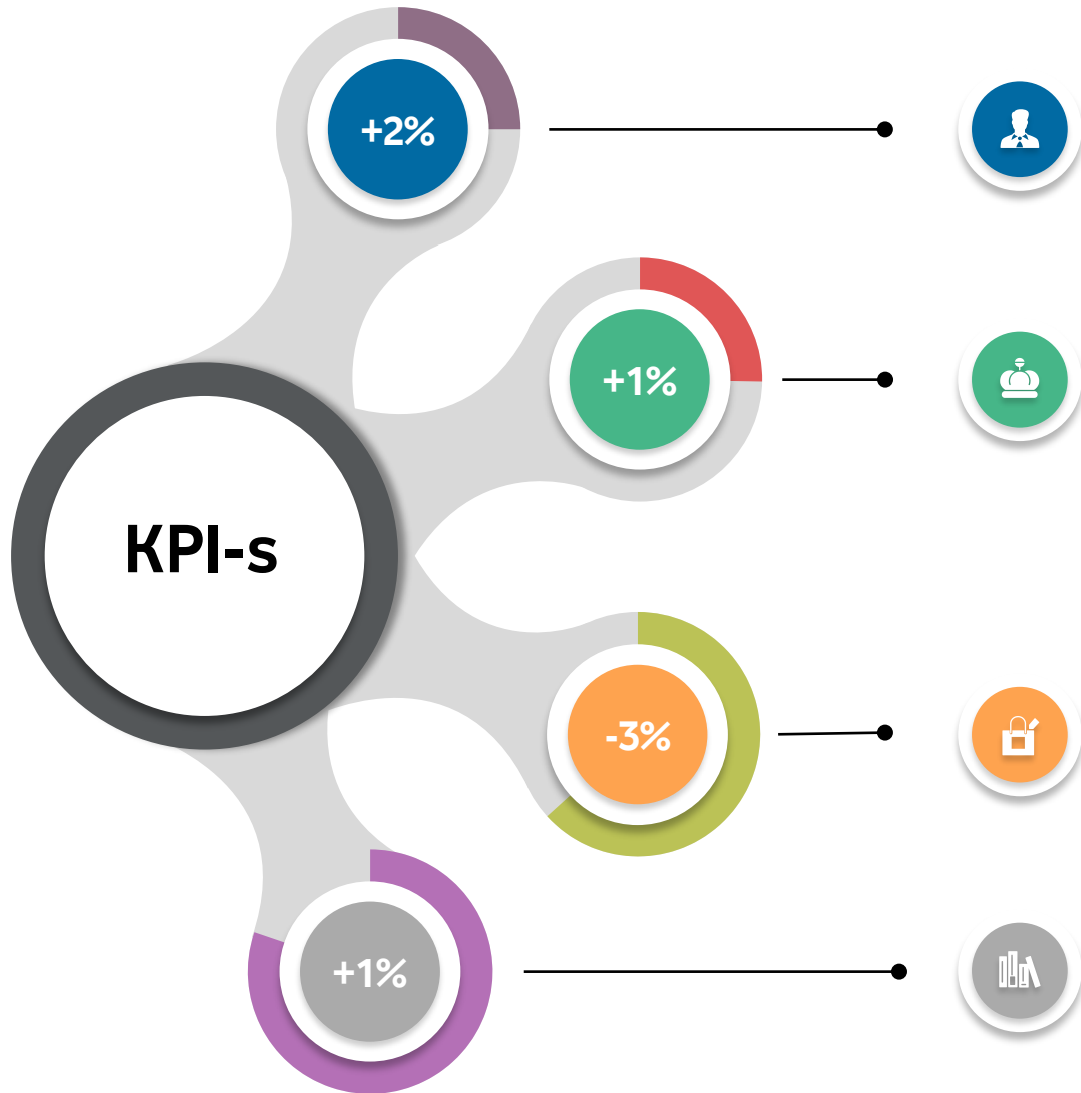
**“It’s easy to lie with statistics.
It’s hard to tell the truth
without statistics.”**



Telecom pillars



Telecom KPI-s



New Connections

Customers connected from the pervious weeks/months

Churn

Customers who are leaving the service or contract

Activity

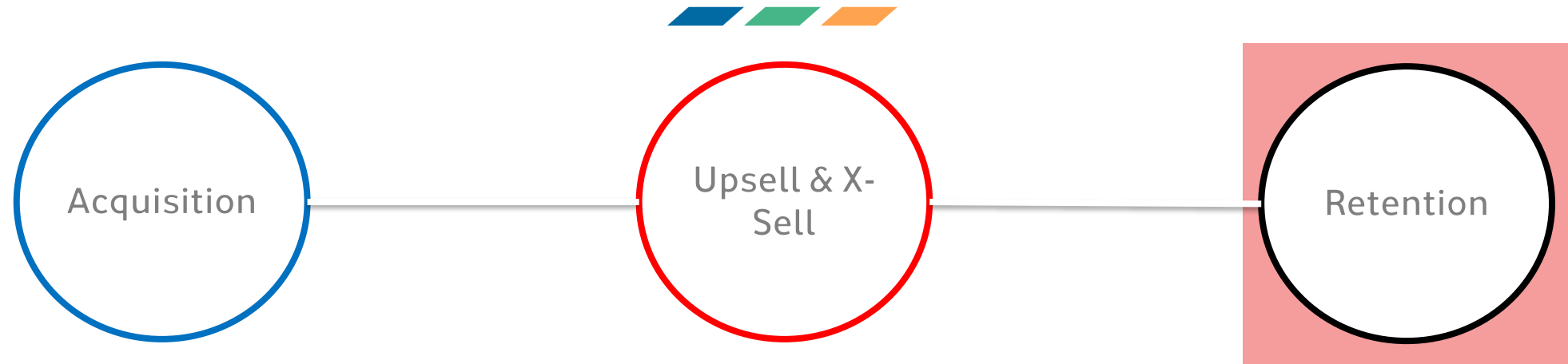
Total activity, including apps, and interconnections

Revenue

Cost and VAT excluding



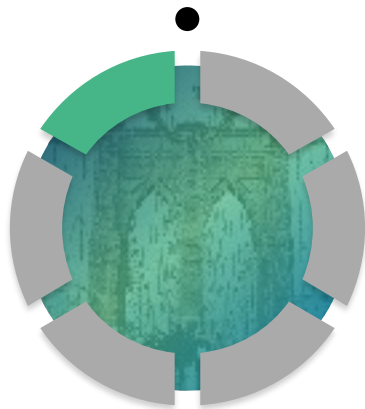
Customer journeys



ATL Products

M4M, MM4M, 1+1, 2+1

L4S, M4S



Rule based



Advanced Analytics



ML Model



sta·tis·tics

**1. The only science wherein
two recognized experts, using
the same set of data may
come to completely opposite
conclusions**



Churn model | Define the problem

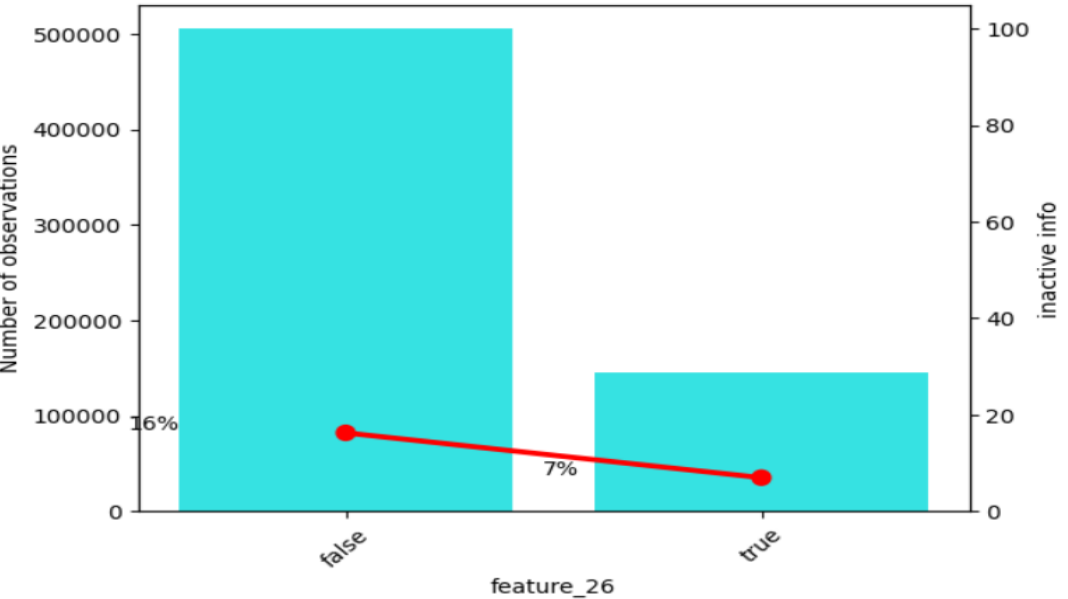
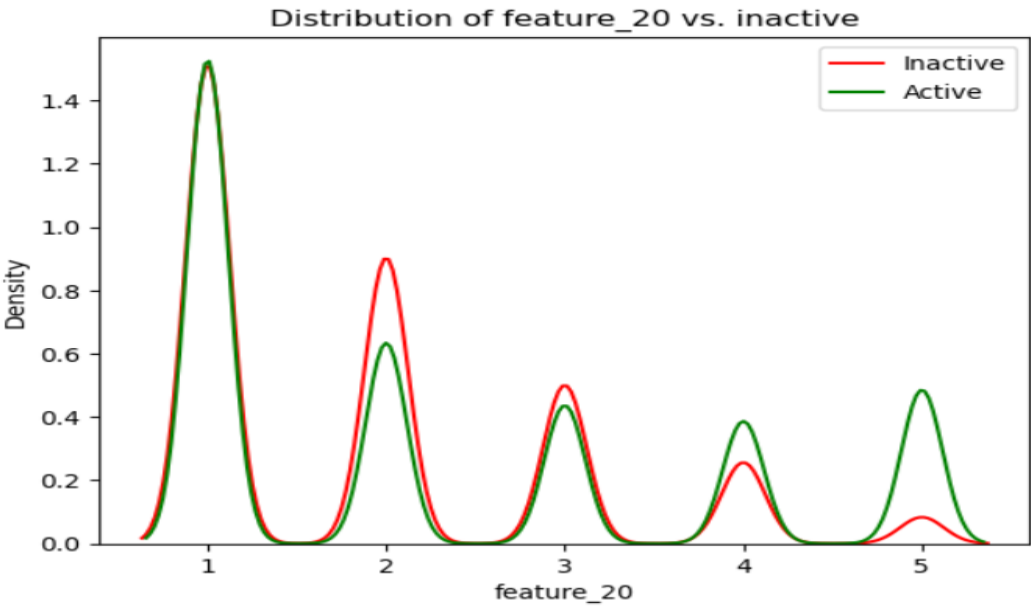
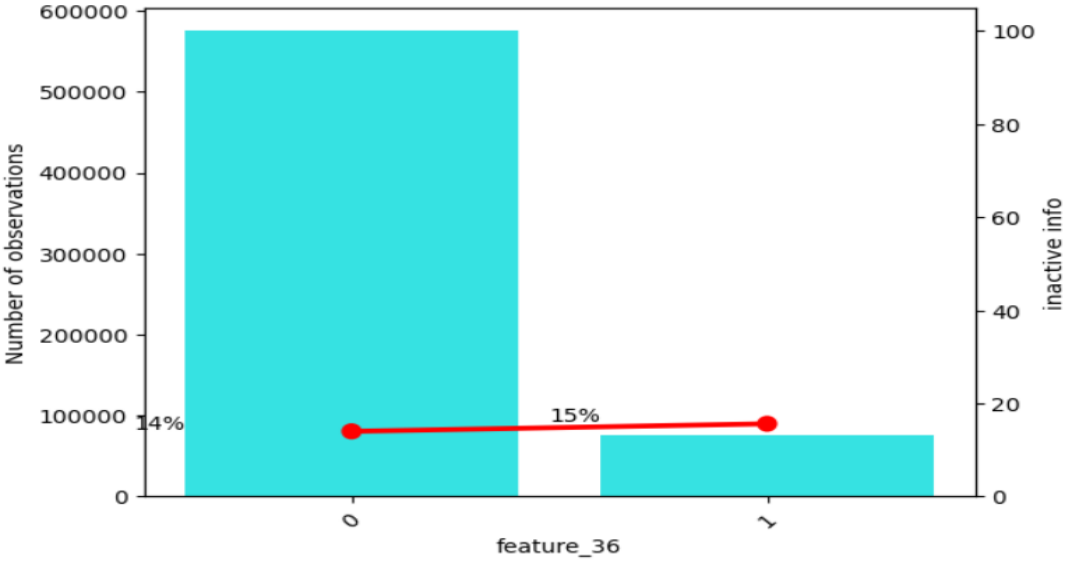
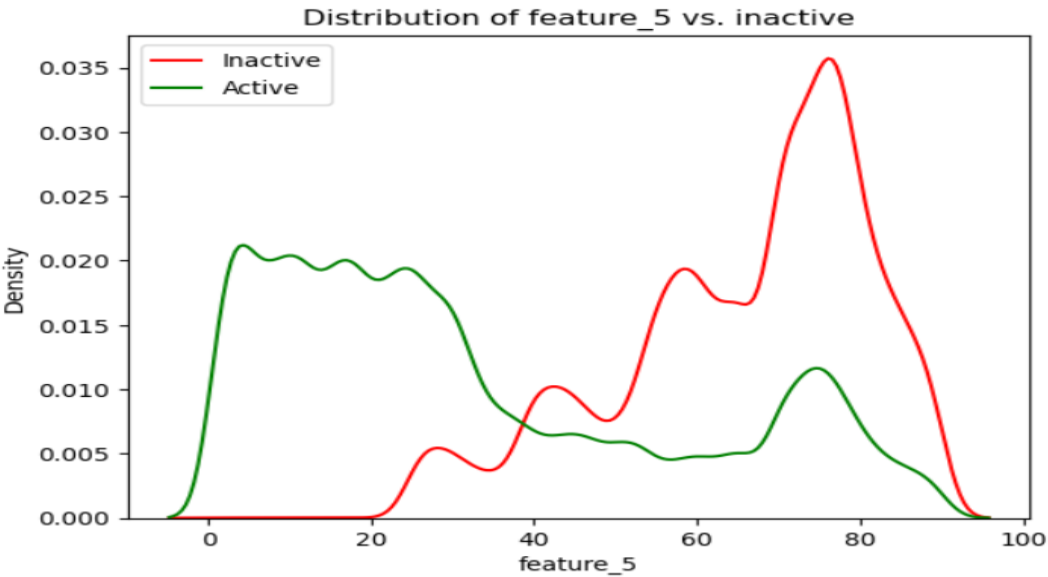


- A churn event is identified when a customer drop the service usage, raises a port-out request or has a valid complain about smth.
- The churn for the last quarter has an increase by 1%.
- The features have different types including continuous and categorical.
- The prediction model is going to be unsupervised learning, binary classification.
- The outcome variable is imbalanced, 85:15.
- **The goal is to find the customers (200K) with highest probability to churn.**

Customer	Age	Roaming	Payments	Tenure	ARPU	Activity Date	Churn
0	66	0	1	60	1424	28	0
1	46	1	3	71	1066	28	0
2	32	1	1	18	839	22	0
3	23	0	5	2	1024	25	1
4	38	1	2	56	1193	24	0
5	60	1	4	9	832	10	1
6	21	1	2	63	647	17	0
7	66	1	3	65	936	11	0
8	34	0	4	7	982	25	1
9	43	0	0	24	1102	11	0
10	48	1	4	33	1301	20	0
11	45	1	2	60	1016	18	0
12	49	0	2	69	1462	6	0
13	31	1	2	46	574	11	0
14	54	1	5	4	1471	20	0
15	18	0	1	15	1136	3	0
16	66	1	1	61	991	12	NaN
17	57	1	2	7	538	4	NaN
18	51	0	1	18	1388	0	NaN
19	47	0	5	32	924	17	NaN



Feature distribution



Machine learning steps



Define the Problem

1st and most important step of the Data Science

Data Preprocessing

Missing values, outliers, combine and join, data mapping, etc.,

Model Building

Different models, hyperparameter tuning, accuracy and error.



Data Gathering

Data can be retrieved from multiple sources: db, files, webpages, text, video and audio, etc.,

Feature Engineering

New features can be created as a combination of original ones. Dimension Reduction, PCA, etc.,

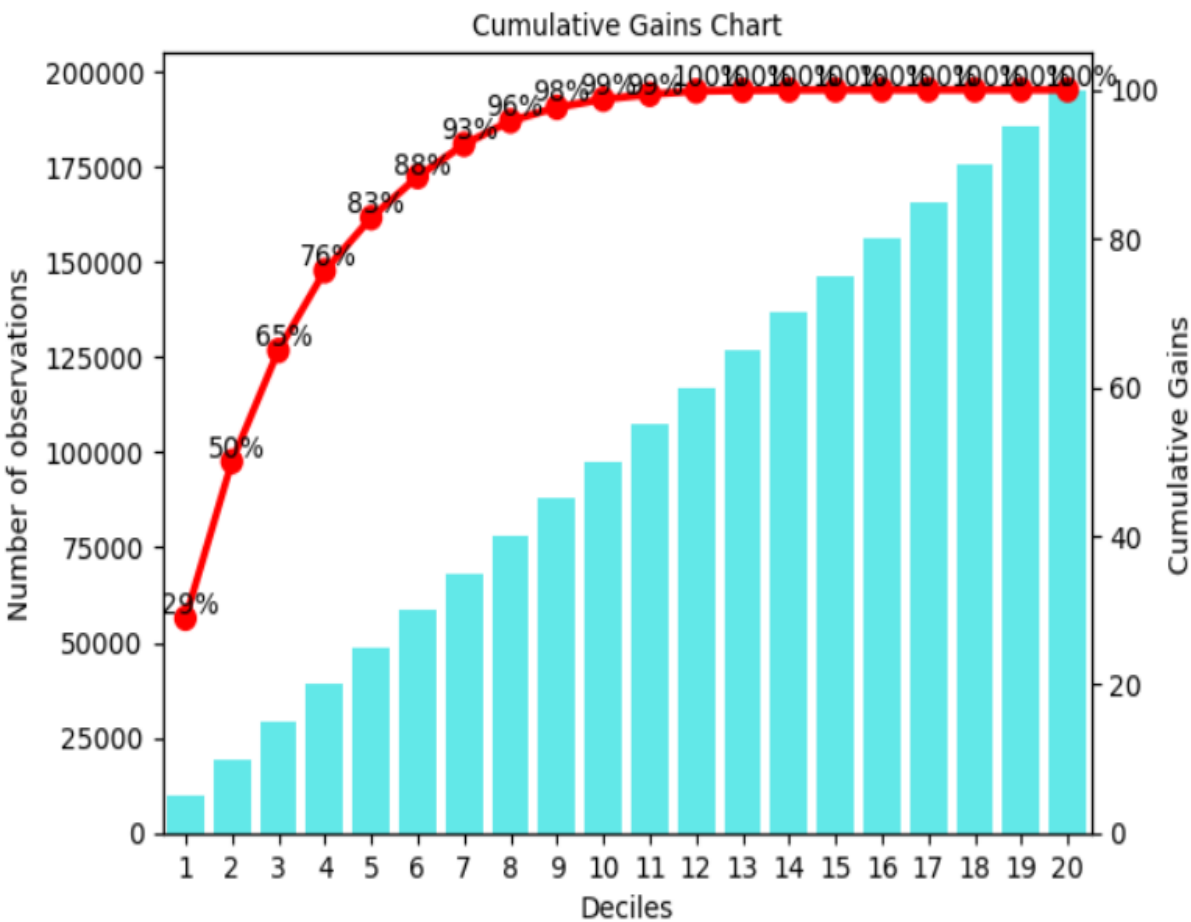
Deploy the Model

Last phase of the journey. It can be anywhere, cloud, app and omni channel

Steps are the same for every Data Science problem, but it can vary by the type of algorithm.



LightGBM | Model results



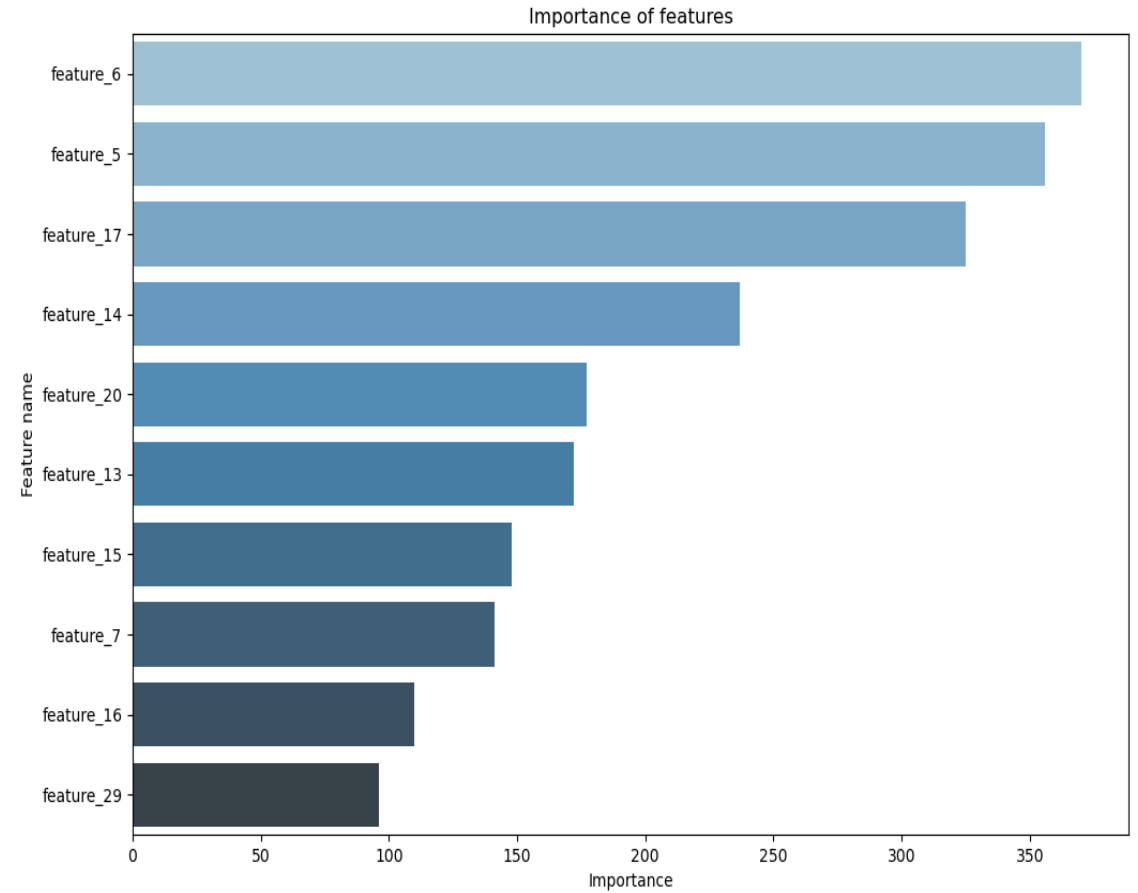
Decile	Count	Actual	Gains	Customers	True Positive Rate
1	9,757	7,835	29%	9,757	80%
2	9,757	5,717	50%	19,514	59%
3	9,757	4,049	65%	29,271	41%
4	9,756	2,911	76%	39,027	30%
5	9,757	1,938	83%	48,784	20%
6	9,757	1,496	88%	58,541	15%
7	9,756	1,179	93%	68,297	12%
8	9,757	855	96%	78,054	9%
9	9,757	496	98%	87,811	5%
10	9,756	296	99%	97,567	3%
11	9,745	177	99%	107,312	2%
12	9,769	121	100%	117,081	1%
13	9,757	22	100%	126,838	0%
14	9,755	15	100%	136,593	0%
15	9,758	5	100%	146,351	0%
16	9,757	1	100%	156,108	0%
17	9,756	0	100%	165,864	0%
18	9,753	0	100%	175,617	0%
19	9,761	0	100%	185,378	0%
20	9,757	0	100%	195,135	0%



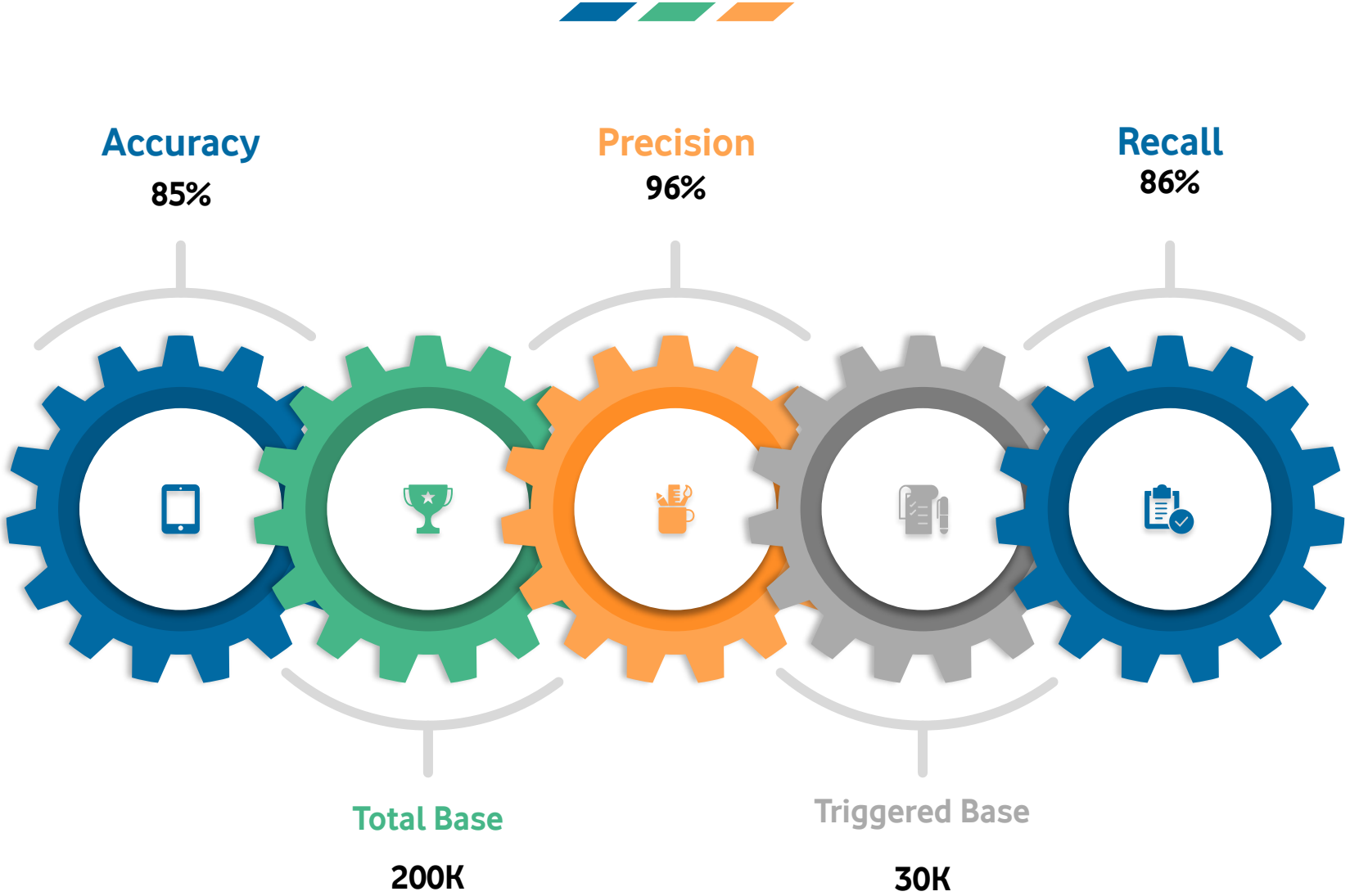
Model results



Decile	Count	Actual	Gains	Customers	True Positive Rate
1	9,757	7,835	29%	9,757	80%
2	9,757	5,717	50%	19,514	59%
3	9,757	4,049	65%	29,271	41%
4	9,756	2,911	76%	39,027	30%
5	9,757	1,938	83%	48,784	20%
6	9,757	1,496	88%	58,541	15%
7	9,756	1,179	93%	68,297	12%
8	9,757	855	96%	78,054	9%
9	9,757	496	98%	87,811	5%
10	9,756	296	99%	97,567	3%
11	9,745	177	99%	107,312	2%
12	9,769	121	100%	117,081	1%
13	9,757	22	100%	126,838	0%
14	9,755	15	100%	136,593	0%
15	9,758	5	100%	146,351	0%
16	9,757	1	100%	156,108	0%
17	9,756	0	100%	165,864	0%
18	9,753	0	100%	175,617	0%
19	9,761	0	100%	185,378	0%
20	9,757	0	100%	195,135	0%



Model results



Questions



Together we can

`str.endswith("The End")`



Mirjan Miftaraj

<https://careers.vodafone.com/>