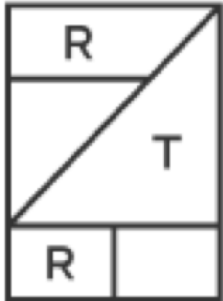


Fit2Wear

Chao Li

Week 5 Data Challenge

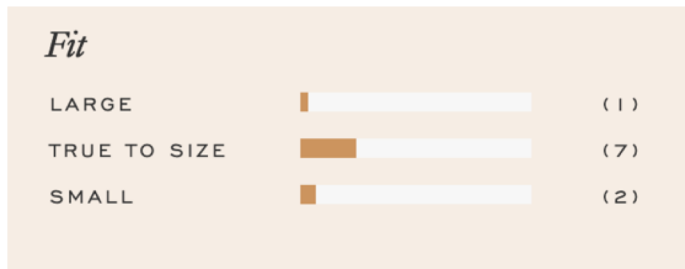
Business value



Problem:

The greatest reason of customer churn is ordering an item that does not fit well.

Vanity sizing: when clothes run larger or smaller in actual measurements than their tag implies.



Goal:

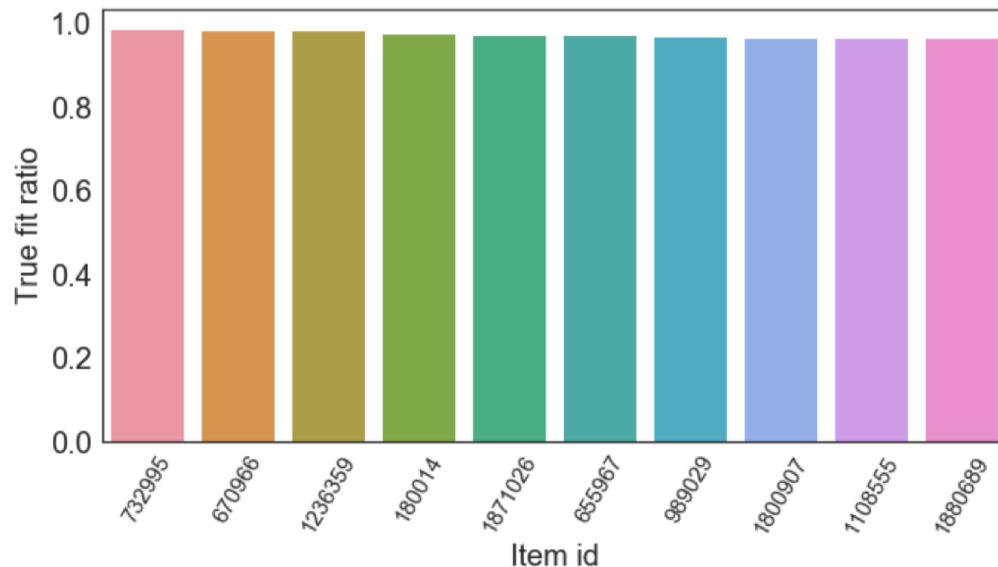
Build a model that empowers a user to discover whether a given item at a given size is likely to fit loosely, well, or tightly.

Which items have the truest fit?

$$\text{True fit percentage} = \frac{\# \text{ Fit}}{\# \text{ Fit} + \# \text{ Large} + \# \text{ Small}}$$

	item_id	Total_number	Fitted_item	Fit_rate
5634	2800454	1	1	1.0
5205	2145536	2	2	1.0
5301	2125454	2	2	1.0
4558	2200369	5	5	1.0
5303	2343038	2	2	1.0
5305	2867464	2	2	1.0
4565	2222459	5	5	1.0
5306	623847	2	2	1.0
5308	1816053	2	2	1.0
4570	2070461	5	5	1.0

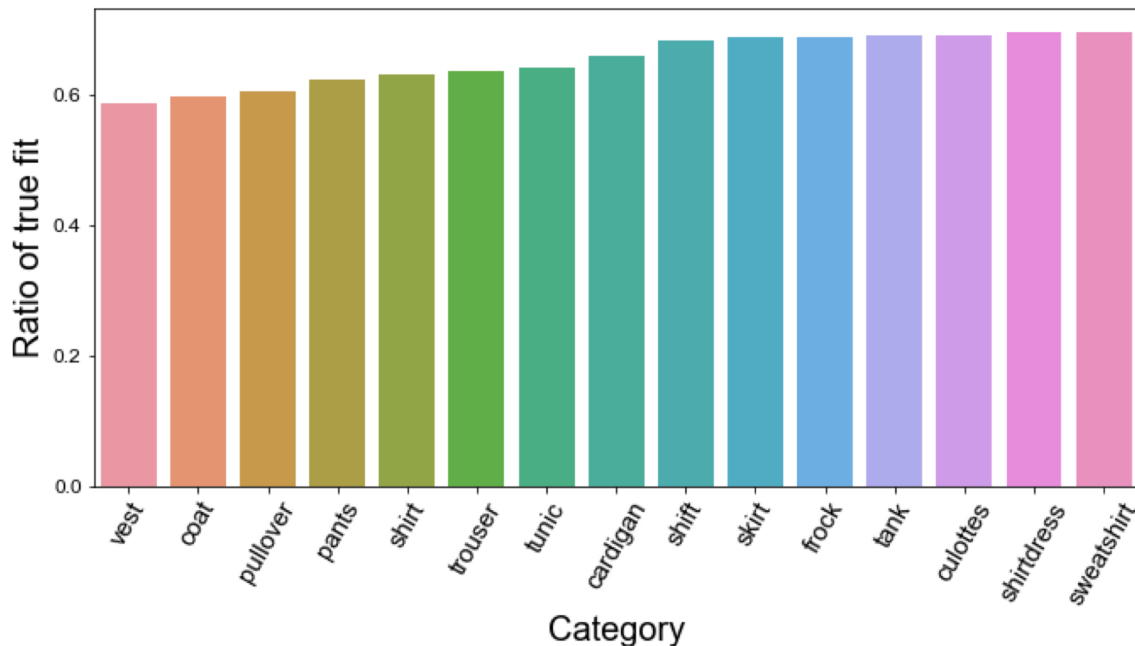
For items with total number > 50:



Which categories have the lowest true fit rate?

For categories with total number > 50:

	category	Total_number	Fitted_item	Fit_rate
58	skort	7	1	0.142857
46	kaftan	17	4	0.235294
51	hoodie	14	5	0.357143
49	tight	15	7	0.466667
61	caftan	4	2	0.500000
41	kimono	30	16	0.533333
18	vest	278	163	0.586331
12	coat	980	584	0.595918
34	pullover	58	35	0.603448
17	pants	422	263	0.623223



Modeling

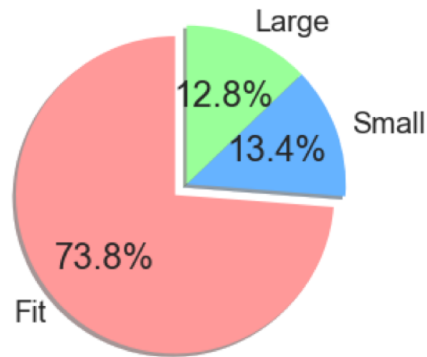
Users

- Age
- Body shape
- Bust size
- Height
- Weight

Items

- Item_id
- Size
- Rating

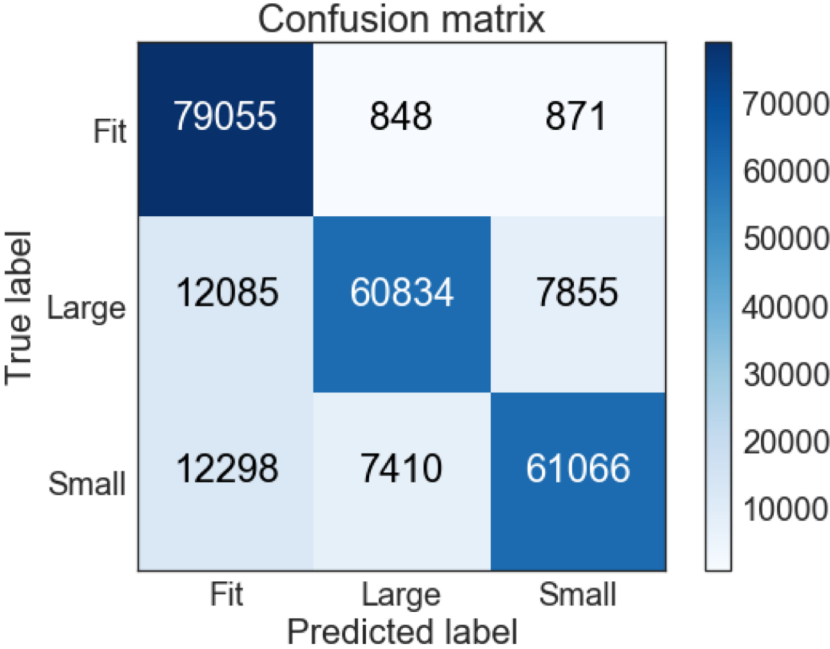
SMOTE
→
LightGBM classifier



Target

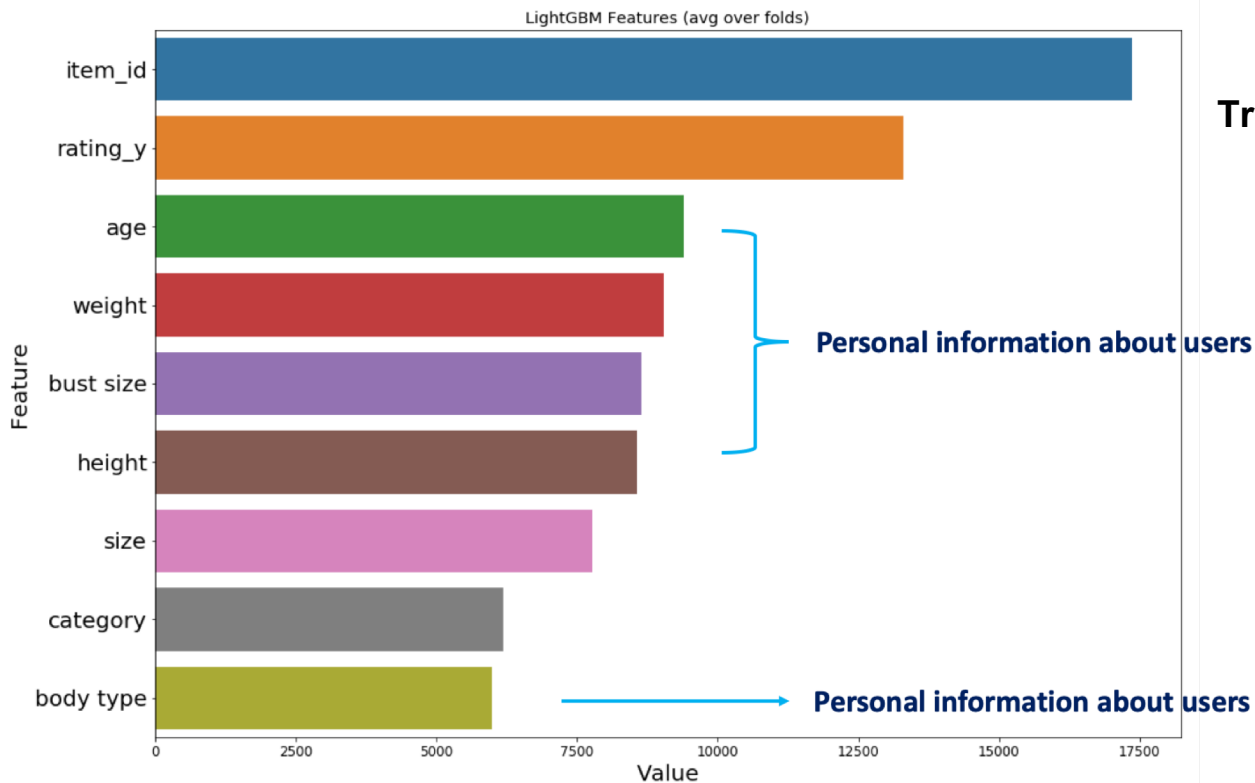
- Fit
- Large
- Small

Model evaluation



Precision	0.80
Recall	0.79

Feature importance



Trade-off:

Reliability of personal information
collected from users



Confidence of the model

Model without features about customers

Precision	0.71
Recall	0.71

