

Price Change Prediction of Electronics in online shopping

PROBLEM

Background:

theBargain is an eCommerce Predictive Analytics company with an objective to insure that a Shopper buys products when the Price is Right. Their tool helps the shoppers to quickly focus on the best product choices available in the market. Our "Price Guarantee Engine" takes into consideration a comprehensive set of reliable and up-to-date data.

Goal:

The goal is to accurately capture the 10% of most likely price increases (PriceUp=1). The performance measure used will be top 10% lift. This is computed by sorting the records according to PriceUp (probabilities), and counting how many among the 10% records with the highest probabilities were actually price increases. The higher the value, better the score.

Research Questions

- What is the impact of skewed target variable on the model, how to handle the skew ness?
- Is the probability of price change higher on some sites?
- Is the probability of price change higher for certain brands and category?
- Is it possible to capture the 10% of most likely price increases (PriceUp=1) using top 10% lift?

DATA

Data is made publicly available by theBargain.in website and has been collected from <https://inclass.kaggle.com/c/price-change-prediction-of-electronics-in-online-shopping>

Name	Description
RowID	index number
name	Product name
brand	Brand of the Product
color	Color of the Product
freeShipping	1 = YES, 2 = NO, 0 = Data unavailable
inStock	1 = YES, 2 = NO, 0 = Data Unavailable
avRating	Average rating of the product
reviewCount	No. of users who rated the product
listPrice	Price of the product on "date"
shippingPeriod	Shipping period of the product
date	Timestamp of the product and price information (mm/dd/yyyy hh:mm)
siteName	Name of the website from which the product is sold
category	Category of the product
TimeNextPrice	number of days until the next available price information
group	always a value of 1 (required for administrative reasons).
PriceUp(Target)	whether the product price went up ("1") or not ("0") on the next time the price information is recorded ("TimeNextPrice")

	Proportion	%
0		95.36
1		4.58
11		0.05

PriceUp

0 1 11

ANALYSIS

How to compute Missing Values?

Following computation methods that make sense with the variable descriptions are used:

- Color = Mode (Black)
- avRating= 2.5 (Median)
- ReviewCount = Median
- Shipping period = Mode(2-3 days)

How to handle Skewness in Target Variable?

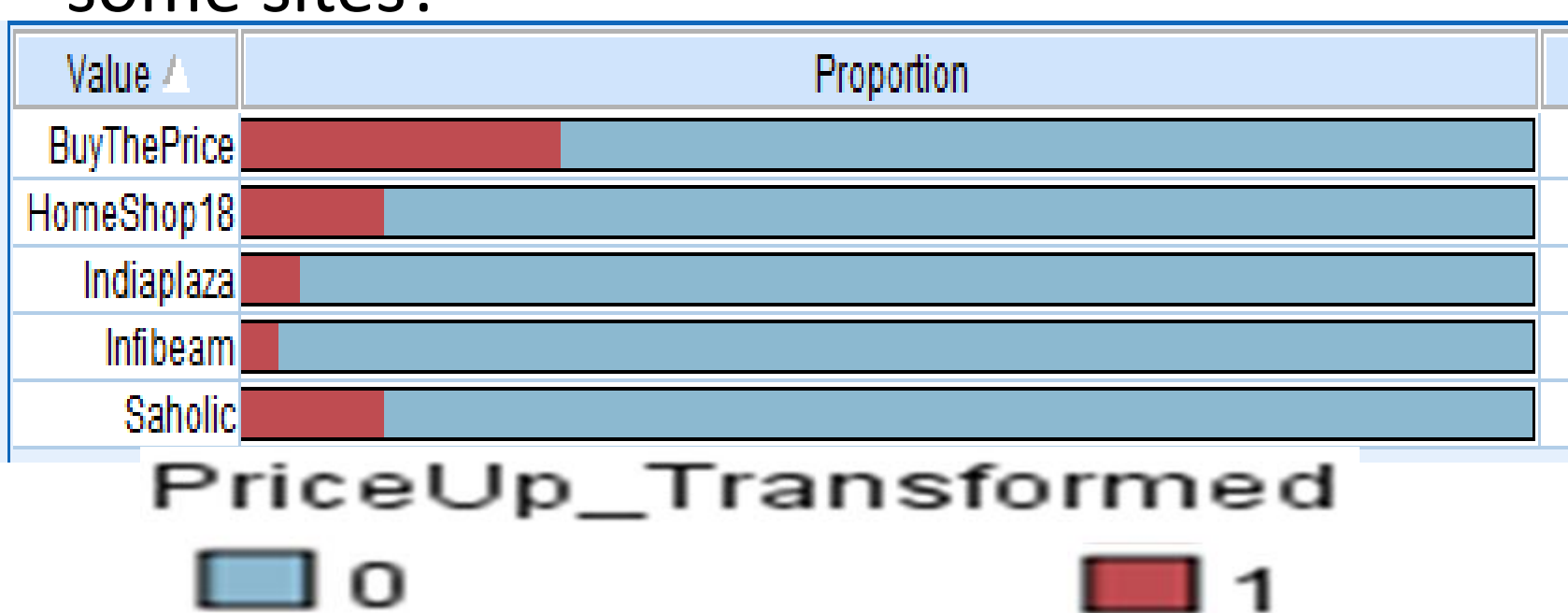
In order to avoid misclassification, we need to Reweight priceup variable with misclassification costs.

Misclassification Costs

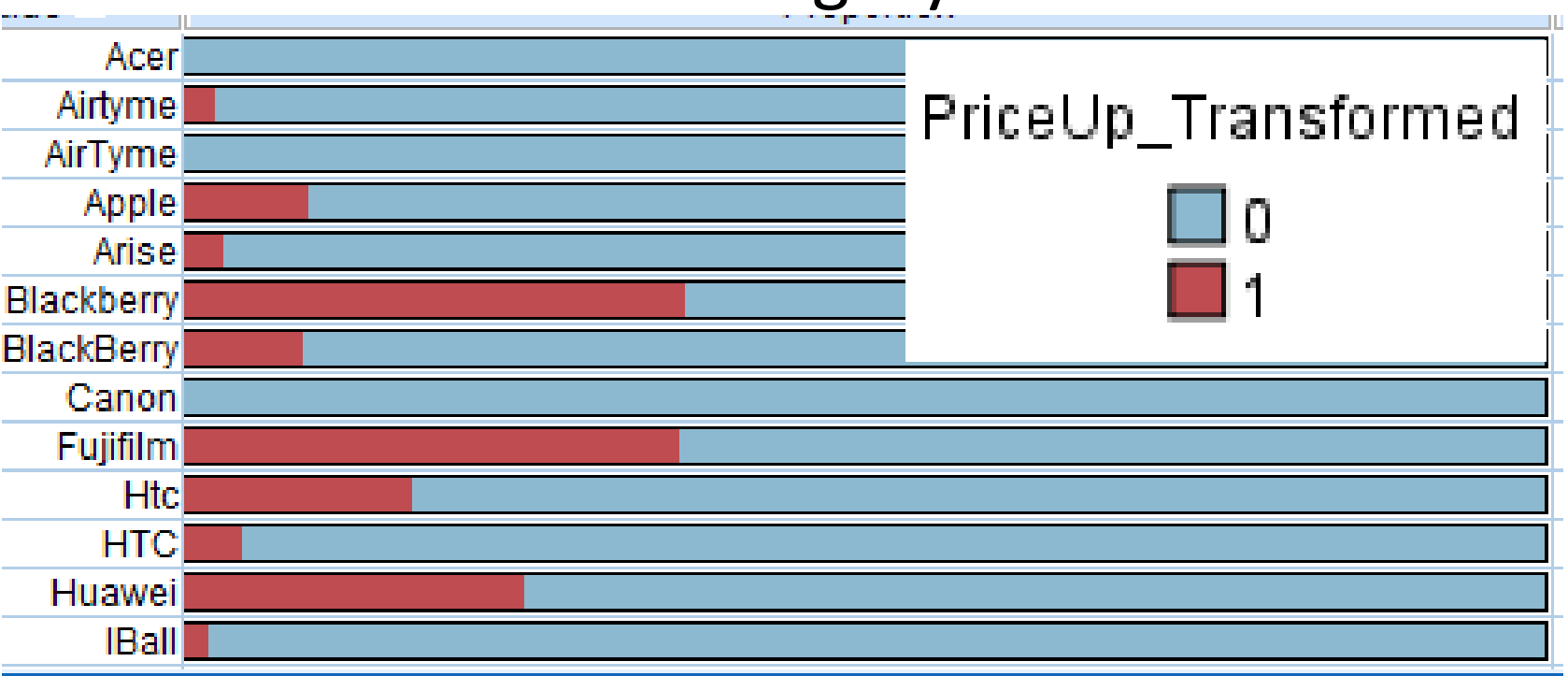
☒ Use misclassification costs

	0	1
Actual 0	0.0	1.0
Actual 1	19.0	0.0

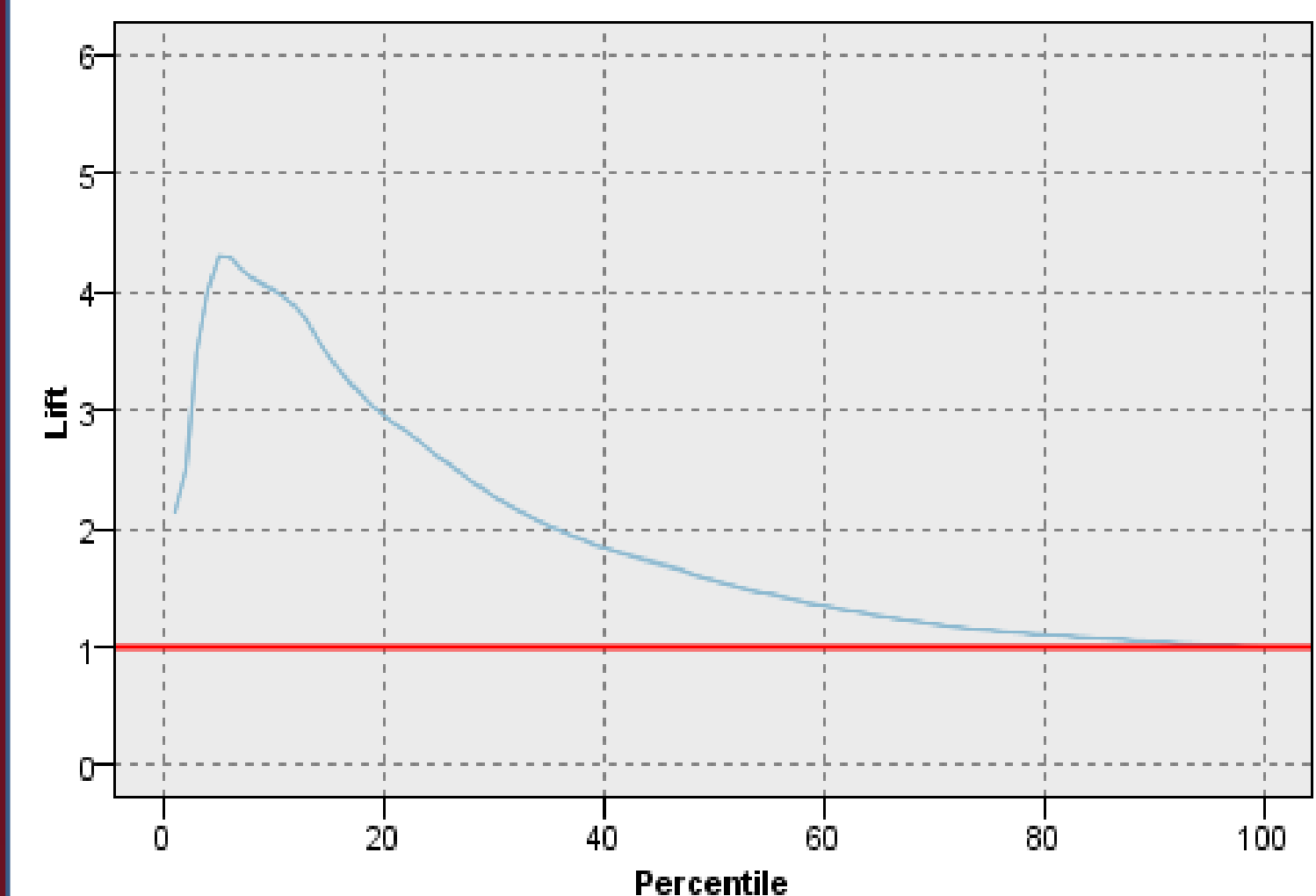
- Is the probability of price change higher on some sites?



- Is the probability of price change higher for certain brands and category?



- Through, final model, CHAID, we should get ~40% of success.(Lift Chart) using top 10% of the items .Without the model, we would get 10% of success only



```
siteName in ["BuyThePrice"] [Mode: 0]
TimeNextPrice_transformed <= -0.118 [Mode: 0] => 0
TimeNextPrice_transformed > -0.118 [Mode: 0] => 1
siteName in ["HomeShop18"] [Mode: 0] => 1
siteName in ["Indiaplaza"] [Mode: 0]
Shipping_Days = 3 [Mode: 0]
Brand_transformed in ["apple" "lg" "micromax" "nokia"] [Mode: 0]
Brand_transformed in ["blackberry" "htc" "iball" "intel" "karbonn"] [Mode: 0] => 1
Brand_transformed in ["lenovo" "motorola" "samsung" "sony" "sony ericsson" "spice"]
Shipping_Days = 4 or Shipping_Days = 6 [Mode: 0] => 1
siteName in ["Infibeam" "Saholic"] [Mode: 0]
Shipping_Days = 2 [Mode: 0] => 0
Shipping_Days = 3 [Mode: 0]
TimeNextPrice_transformed <= -0.517 [Mode: 0] => 0
TimeNextPrice_transformed > -0.517 and TimeNextPrice_transformed <= 0.049 [Mod
TimeNextPrice_transformed > 0.049 and TimeNextPrice_transformed <= 0.215 [Mod
TimeNextPrice_transformed > 0.215 [Mode: 0] => 1
Shipping_Days = 4 or Shipping_Days = 5 [Mode: 0] => 1
```

Conclusion

- We determined that online price of the item is related to site name, brand name and category of the product. With our model, 21 out of top 10% items(39 items) are most likely to undergo price change.