Machine Learning Supervised Project

Baigalmaa BAATAR
30 juin 2020
Jedha Bootcamp

Conversion rate on site

- Which feature is more important for conversion rate?
- How we can improve conversion rate?
- Could we predict conversion rate?

Which features we have?

Country: China, UK, US, Germany

Age

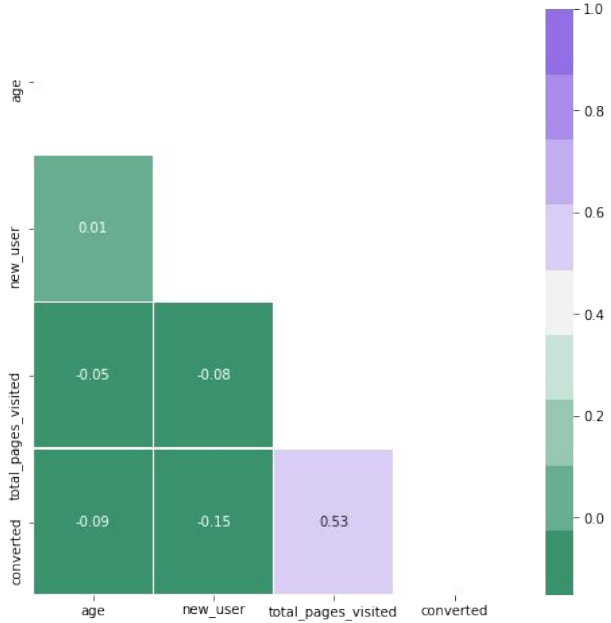
New user or old user

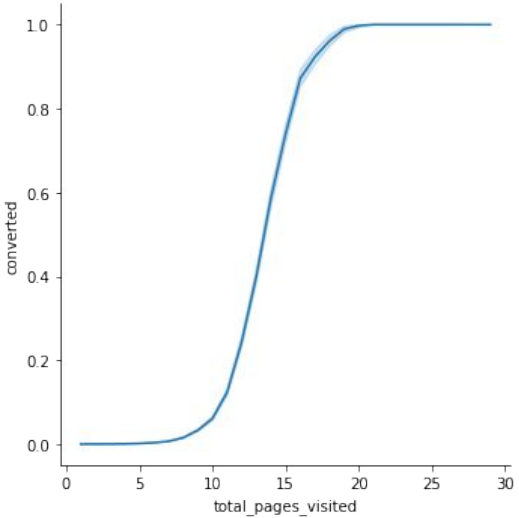
Source: Direct or Seo, Ads

Total pages visited

Converted or not (bought the item or not)

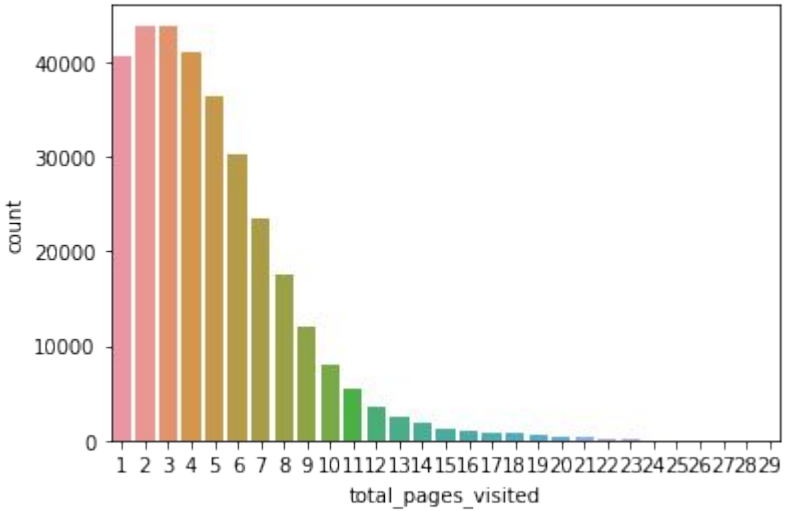
It seems Total pages visited has more conversion rate





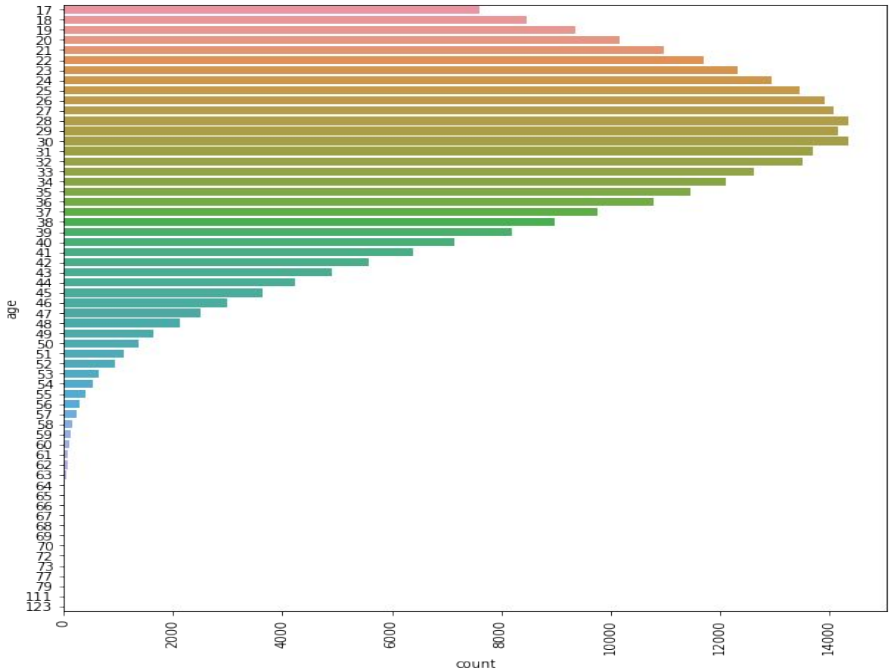
We will examine if it is true or not

Let’s see more on total pages visited



Most users visit 3-4 pages!

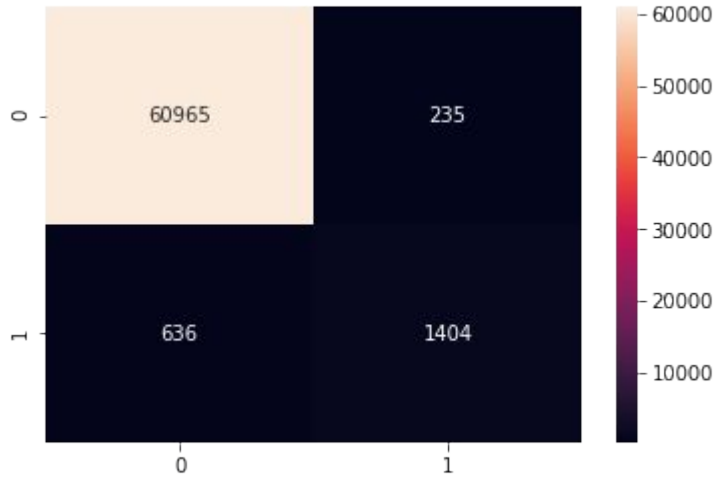
Let’s see if we can predict using Logistic Regression



Most users are young people

Logistic regression

We calculated more on zero (not converted) than converted. Hmm

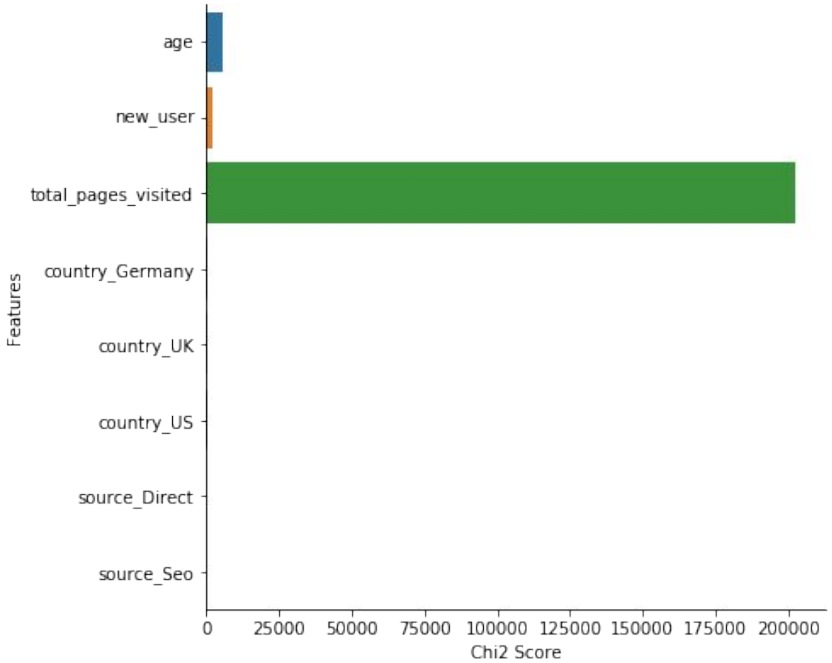


Logistic regression

What is most important feature?

As we predicted!!!

Total pages visited!

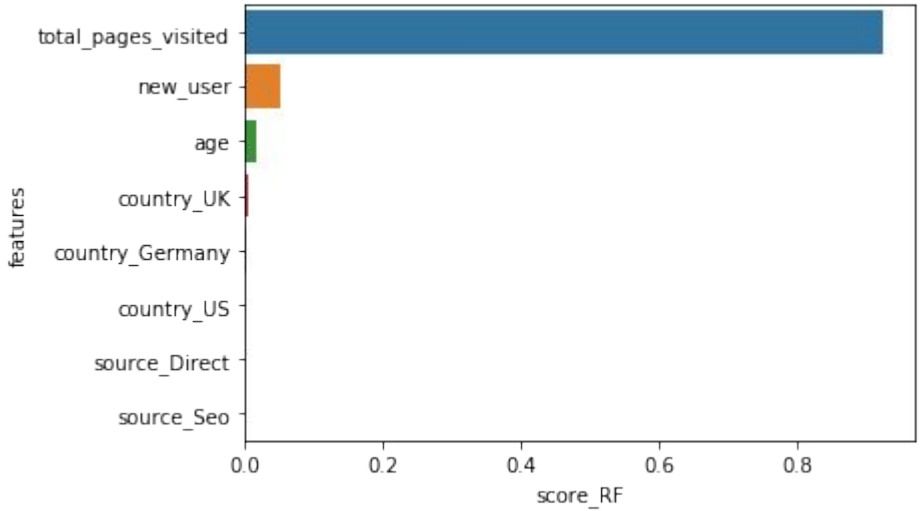


Logistic regression

What is most important feature?

As we predicted!!!

Total pages visited!



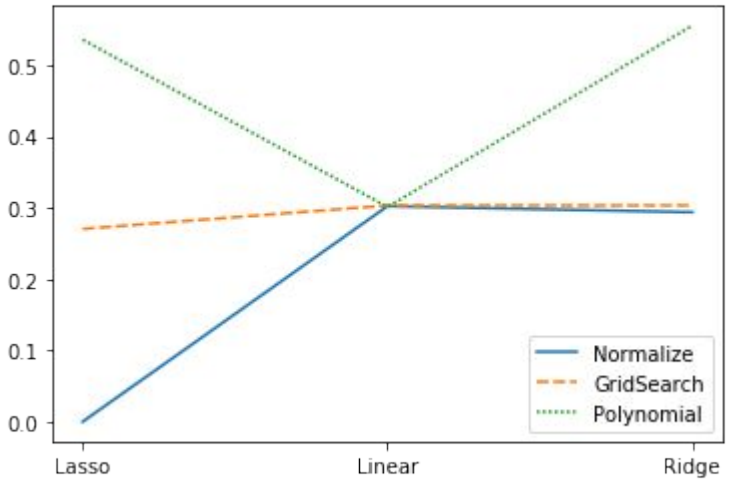
Polynomial feature

Models:

1. Linear regression
2. Ridge
3. Lasso

Features:

1. Nothing
2. Just GridSearch
3. With Polynomial



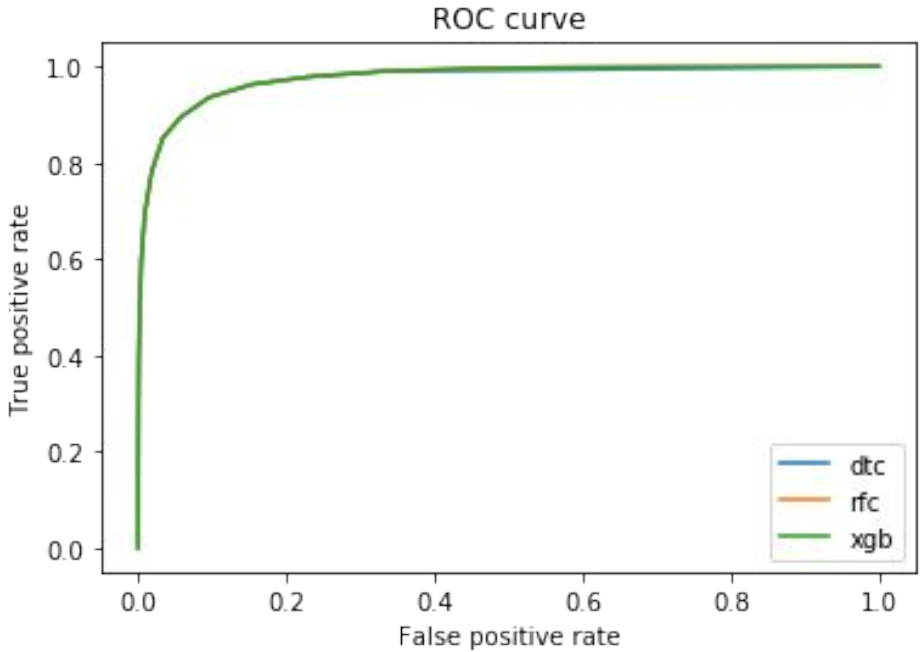
Decision Tree

As expected we predicted more non converted.

	precision	recall	f1-score	support
0	0.99	1.00	0.99	61200
1	0.86	0.63	0.73	2040
accuracy			0.98	63240
macro avg	0.92	0.81	0.86	63240
weighted avg	0.98	0.98	0.98	63240

ROC

XGB, Random Forest, Logistic regression plot



Conclusion

- Total pages visited seems most important feature
- New users
- Also age (mostly young people)