

Online Shopping Intent



- ▶ Tony Ghabour
February 12, 2020

Objective



To predict with high **precision** AND **recall** whether online shopping sessions will result in purchase.

- ▶ **Precision:**

 Too many false positives → inflated revenue estimates

- ▶ **Recall:**

 Too many false negatives → insufficient inventory

Data



Scope

- ▶ 12K + records
- ▶ 17 features
- ▶ Binary Classification



Software

- ▶ Browser
- ▶ Operating System



Visitor Type

- ▶ New vs. Returning
- ▶ 20 distinct traffic types



Page Counts

- ▶ Informational
- ▶ Administrative
- ▶ Product-Related



Time-Related

- ▶ Month of session
- ▶ Browsing time by page type



Other

- ▶ Region
- ▶ Bounce/Exit Rates



Google Analytics



Limitations



Unbalanced Data

- ▶ Prevalence of positive class only about 15%.
- ▶ Oversampling employed



Censored Features

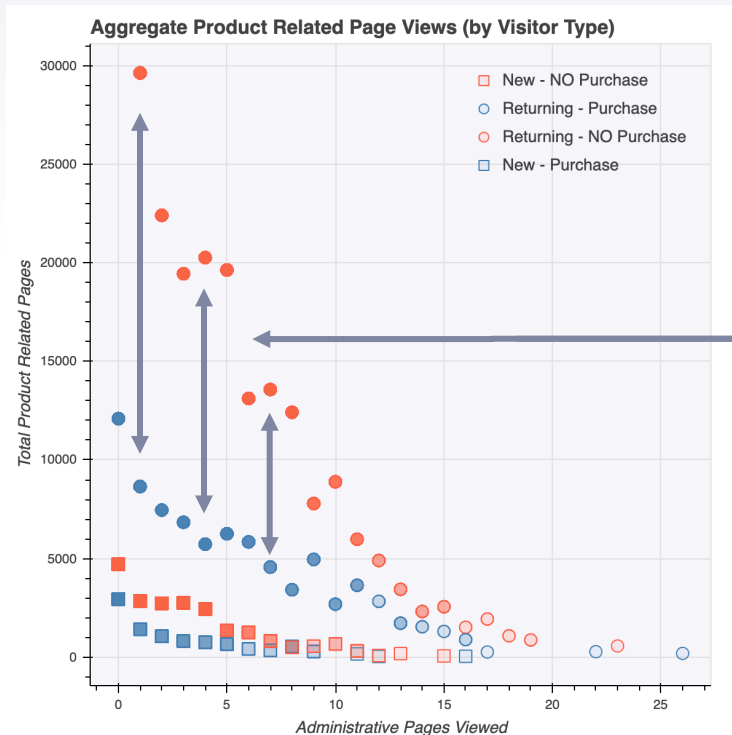
- ▶ Encoded features (e.g. "Browser" = 1)
- ▶ No "real-world" meaning to certain inputs
- ▶ Difficult to interpret



Missing Features

- ▶ No information regarding website
- ▶ Guest vs. account-holder
- ▶ No price or product information
- ▶ Vague domain

Analysis



Among visitors of the same type (new vs. returning), **fewer** product-related page views suggests more likely to purchase! (Maybe?)



More pronounced among returning visitors but holds for new visitors as well.



Clear divide between those who shop with intent vs. those who are searching for something.

Modeling



| | | | |
|-----------|------------------------------------|--------------------------|-----------------------------------|
| | Dummy (Strategy = "Stratified") | Random Forest (Tuned) | Ensemble (Stacking Classifier) |
| Accuracy | 0.728 | 0.904 | 0.902 |
| Precision | 0.136 | 0.726 | 0.709 |
| Recall | 0.145 | 0.603 | 0.607 |
| F1-Score | 0.140 | 0.659 | 0.654 |

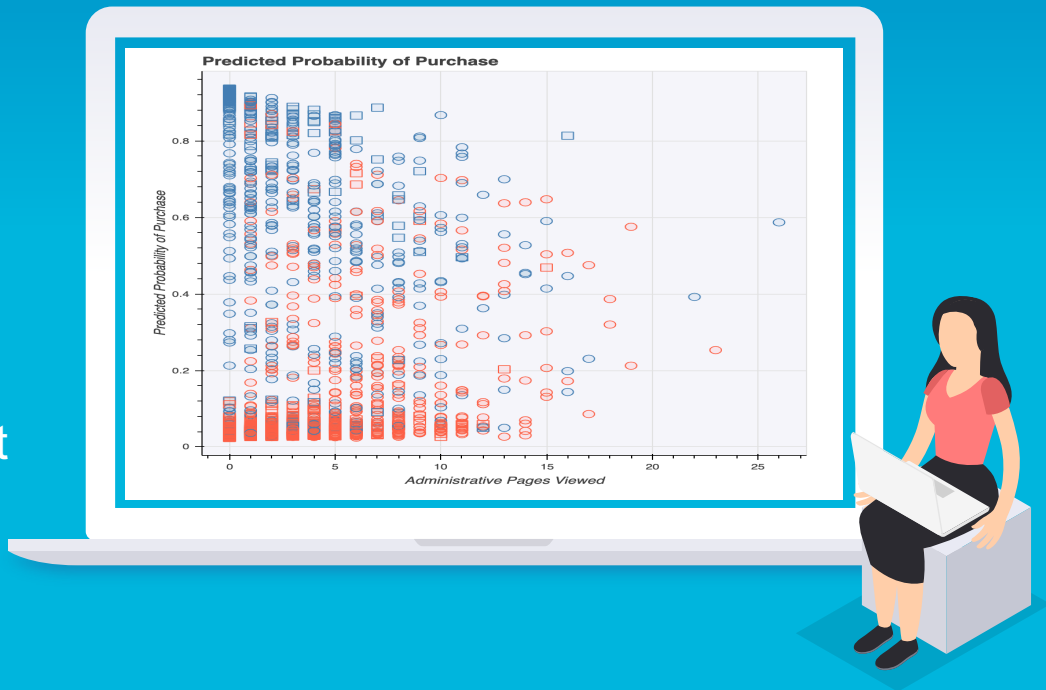
- ▶ All models reflect use of:
 - 7 engineered features
 - 30% holdout test set
 - Standard scaling
 - Random oversampling

Data Visualization

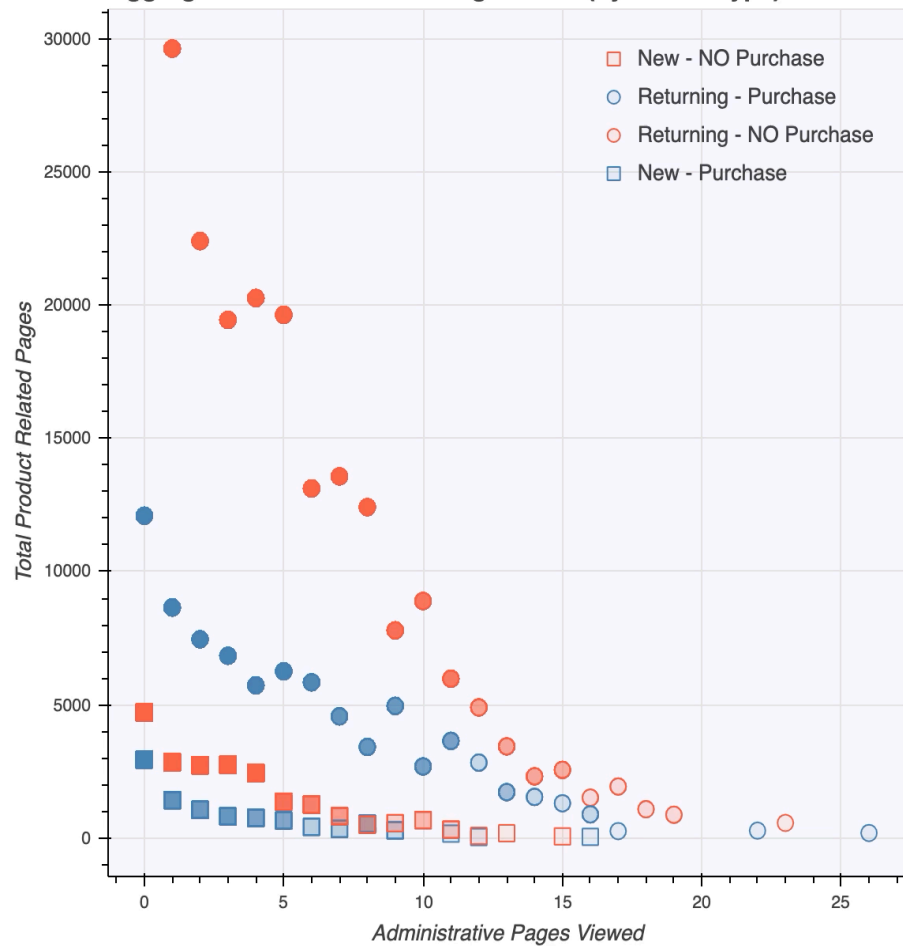
Can visitor type and browsing behavior predict probability of purchase?



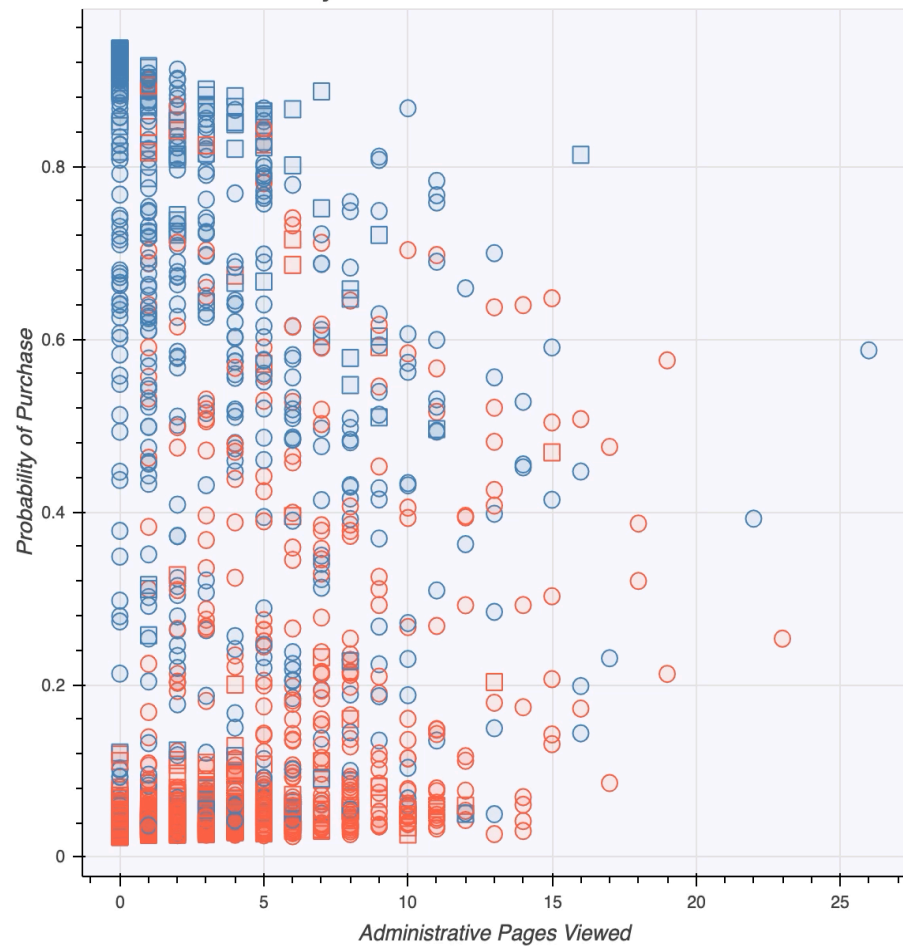
- ▶ Linked Exhibits
- ▶ Sample Select
- ▶ Shared Axis
- ▶ Tooltips
- ▶ Pan & Zoom
- ▶ ...and more



Aggregate Product Related Page Views (by Visitor Type)



Predicted Probability of Purchase





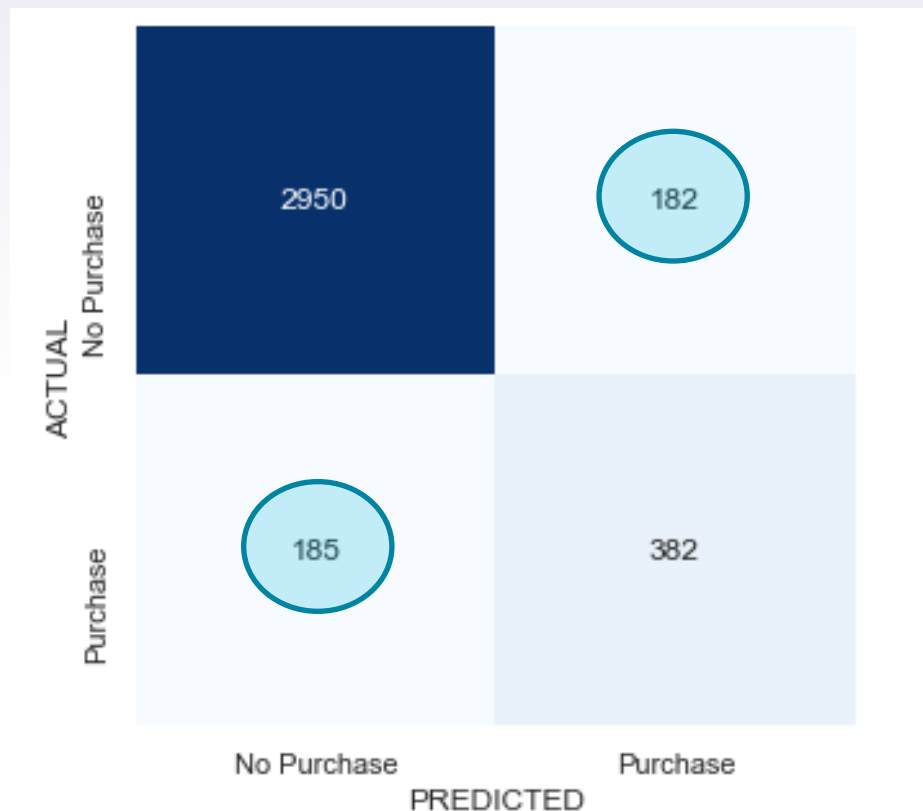
3 False Negatives



3 (net) False
Negatives

Better to be lucky than good.

- ▶ Reasonably accurate on its own
- ▶ Net aggregate results even better
 - 185 False Negatives
 - 182 False Positives
 - 3 Net False Negatives**
- ▶ **Meets original objective of balancing projected revenue and inventory**



THANKS!

Any questions?

You can reach me at:

- ▶ tony.ghabour@gmail.com
- ▶ www.linkedin.com/in/anthony-ghabour

