**Buy Accuracy for Suggested Offer Algorithm Release 1, Version 4.0**

All information below is based on the following process:

1. All tables related to the suggested offer algorithm were compiled based on item activity during the 13 months from 3/1/2018 to 3/30/2019.
2. Catalog title performance was measured for all items priced at all locations from 4/1/2019 to 4/30/19, during the time period from 4/1/19 to present. The same metrics were calculated as used in the suggested offer algorithm (total accumulated days on shelf with trash penalty divided by the number of items priced or scanned during the time period), and buy percentages were assigned using the same method.
3. Predictions made by the table calculated in step 1 were compared to the catalog title performance measured in step 2 to calculate accuracy. Two types of accuracy evaluation are possible: continuous (comparing predicted versus actual average accumulated days on shelf) and categorical (comparing predicted buy grades to actual buy grades).
4. Multiple categorical accuracy evaluations were performed: for instance, chain-level predictions to chain-level performance, chain-level predictions to location-level performance, and location-level predictions to location-level performance.
5. (In progress) Multiple continuous accuracy evaluations are also in progress, using actual error and root mean squared error (RMSE) as evaluation metrics.

**Limitations**

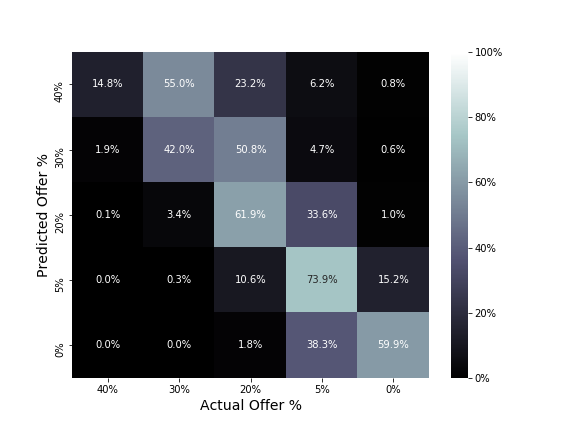
Several limitations to our ability to assess the accuracy of our buy algorithms exist, though they seem unlikely to be insurmountable.

1. Accuracy measurements are currently based on a single roll-up of the item lifecycle and buy algorithm tables. In production, these tables are updated and re-rolled each week with current information. Using a single roll-up to measure accuracy is therefore of limited use. Applying it to only one week results in a large number of titles with insufficient number of items priced to measure algorithms’ predictive accuracy. Applying it to one month or more means that some of the predictions being evaluated are out-of-date. A means to fully replicate the process of generating the algorithm at weekly intervals, but for historical data, is necessary to for a thorough assessment of predictive accuracy.
2. There are two known flaws in reconstructing the base algorithm data as it would have appeared historically:
   1. Reconstruction of which shelf scans would have been current and what would have been historical is not currently implemented. It does not seem that any algorithm yet implemented has discriminated between current and historical scans in any of their calculations, but this assertion could be incorrect.
   2. While most item status flags in the SipsProductInventory table are based on transfer events, and therefore have a traceable history, the flag “M” does not have a traceable history. Currently, it appears that no table exists that tracks when items are flagged as “M” (missing) or which system it was flagged by.

**Chain-level Predictions**

Chain-level suggested offers measured against chain-wide performance yielded accurate buy grade predictions 61.8%of the time for catalog titles of “general” (not paperback or CD) binding, represented by 5 or more copies priced between 4/1/19 and 4/30/19.

Suggested offer accuracy by buy grade can be seen in the following charts:



The chart to the left displays accuracy by offer percentage. A perfectly accurate algorithm would show white squares of 100% across the top-left to bottom-right diagonal, with all other squares being black with values of 0%.

When 40% was the offer predicted, the algorithm was accurate 14.8% of the time. 55.0% of the time, an offer percentage of 30% would have been accurate. 23.2% of the time, an offer percentage of 20% would have been accurate.

Predictions for buy offers of 5% were correct 73.9% of the time, which was the highest accuracy of any buy grade.

A more useful, but slightly more difficult to interpret, way of displaying the same chart is for each square to display the percentage of total items evaluated that it represents:

A picture containing monitor, text

Description automatically generatedIn the chart to the right, a perfectly accurate algorithm would be represented by white and grey squares along the top-left to bottom-right diagonal, with all other squares being black with 0% values.

it is apparent that 27.2% of all items evaluated were accurately assessed offers of 20%, closely followed by accurately assessed offers of 5%.

14.8% of all inaccurate offers were predicted as 20% while their actual performance justified an offer of 5%.

7.3% of all inaccurate offers were predicted as 30% where an accurate offer would have been 20%.

40% offer made up a very low percentage of chain-level predictions.

**Location-level Predictions**

Location-level suggested offers measured against location-specific performance yielded accurate buy grade predictions 42.9%of the time for catalog titles of “general” (not paperback or CD) binding, represented by 5 or more copies priced between 4/1/19 and 4/30/19. This lower level of accuracy compared to chain is due to the fact that each location offer is (necessarily) made based on fewer items than they are at chain level and therefore tend to have higher levels of variance in item sales patterns. This does not necessarily mean that location-level predictions are less accurate for each location than chain-level predictions, as will be shown below.

A screenshot of a cell phone

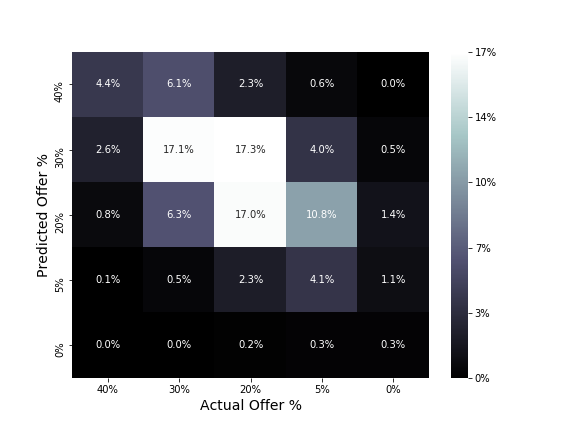
Description automatically generatedEvaluating accuracy using the same charts as used above:

At a location level, accuracy is much more even across all grades, though the overall level of accuracy is lower.

75-95% of item grade predictions are within one grade of actual performance.

Location-level predictions were most accurate for offers of 5%, being correct 51.1% of the time

Location-level predictions were least accurate for offers of 40%, being correct 32.8% of the time.



At a location-level, offers of 40% and 30% are predicted more frequently, and offers of 0% are predicted far less frequently.

The primary source of error in location-based offers is 17.3% of offers inaccurately classified as 30% which should have been classified as 20%

The secondary source of error is 10.8% of offers inaccurately classified as 20% which should have been classified as 5%

These two together account for 28.1% of the inaccuracy in location-based offers.

**Location-Level Predictions w/ Chain-Level Default**

Location-level suggested offers, substituting chain-level offers where location offers are absent, measured against location-specific performance yielded accurate buy grade predictions 42.9%of the time for catalog titles of “general” (not paperback or CD) binding, represented by 5 or more copies priced between 4/1/19 and 4/30/19. This lower level of accuracy compared to chain is due to the fact that each location offer is (necessarily) made based on fewer items than they are at chain level and therefore tend to have higher levels of variance in item sales patterns. This does not necessarily mean that location-level predictions are less accurate for each location than chain-level predictions, as will be shown below.