ASIN Buyable Attribute Predicted Analysis in Supply Chain System

Yefei Fan

Background:

This case results from the underperforming NIS DEA. During the daily KPI monitor, we find that the Not In-stock Delivery Estimated Accuracy (NIS DEA) missed the goal (93.5%) very frequently. After diving deep into this problem, we found that in most of the cases, the attribute of the ASINs should have been set to "OB" (obsolete), which means the ASIN would never be purchased from the vendors, while the system still shows the ASIN are procurable and keeps buy-boxes on Amazon.CN. Once customers place orders on these ASIN, the order will never be fulfilled, leading the orders to miss the promises. This is why the NIS DEA is underperforming. In general, the buyable attribute of the ASINs are updated by in-stock managers manually, which is time-consuming and inefficient. With the development of big data, machine learning modelling technics are widely utilized in the IT industry to improve business performance and work efficiency. This project aims to use the machine learning methodologies to solve the misclassification of the ASINs automatically and improve the NIS DEA performance.

Modelling:

- **Data Preparation**

As Media Product Line contributes the most to the NIS DEA due to the misclassification of ASIN buyable attribute and each product group might have different characteristics, this study only focuses on one single product line – Media. 60000 ASINs of Media along with their attributes are selected randomly from the database. In the raw data, "PR" (procurable) ASIN and "OB" (Obsolete) ASIN account for 30000 for each.

- **Variable Selection and Feature Engineering**

Interviews with Media in-stock managers and IPC were conducted to initially scope the relevant features and attributes of the classification of ASIN buyable attribute. In addition, features were also selected based on the supply chain management experience. 38 features were chosen in this initial stage. To narrow down the features used to build the model, descriptive statistical methods were utilized to check the relevance of the features and ASIN buyable attributes.

Last Inbound History refers to the difference between the last date an ASIN was inbound and current date. Missing value of last inbound attributes is set as "2010/1/1" and current observation date is set as "2017/11/26". As Figure 1 shows, the distribution of the OB and PR ASIN regarding inbound history is highly different. Thus, the shorter the ASIN inbound history is, the more likely the ASIN has a PR buyable attribute. In model building stage, if last inbound date is Null, remove the data from training set.
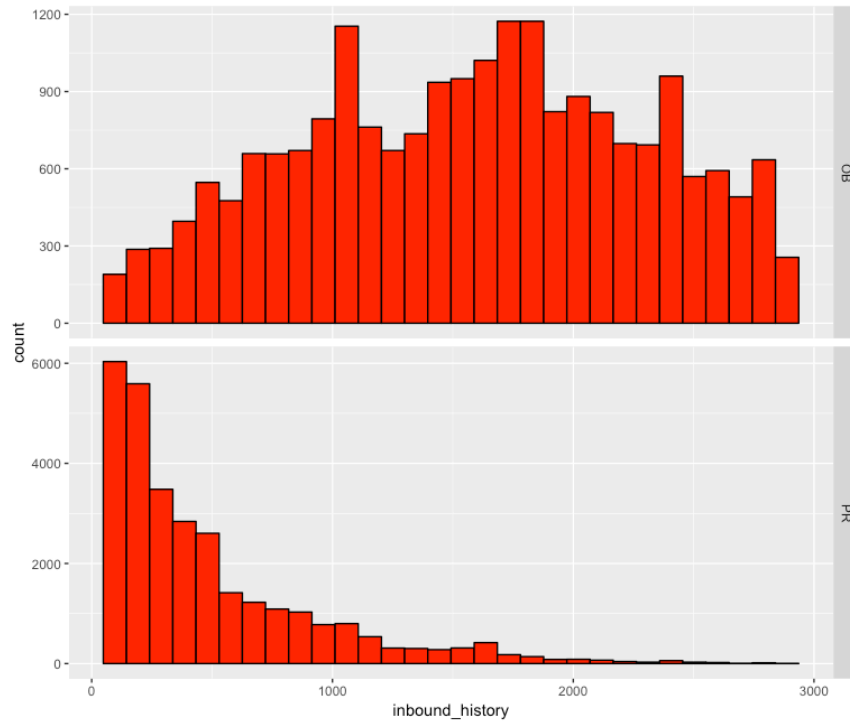
(Figure 1: Inbound history and ASIN buyable attribute)

ASIN History refers to the difference between the ASIN creation date and current date. As Figure 2 shows, ASIN history is highly relevant to the ASIN buyable attribute. One ASIN is very possible to be OB if the ASIN history exceeds a certain range.
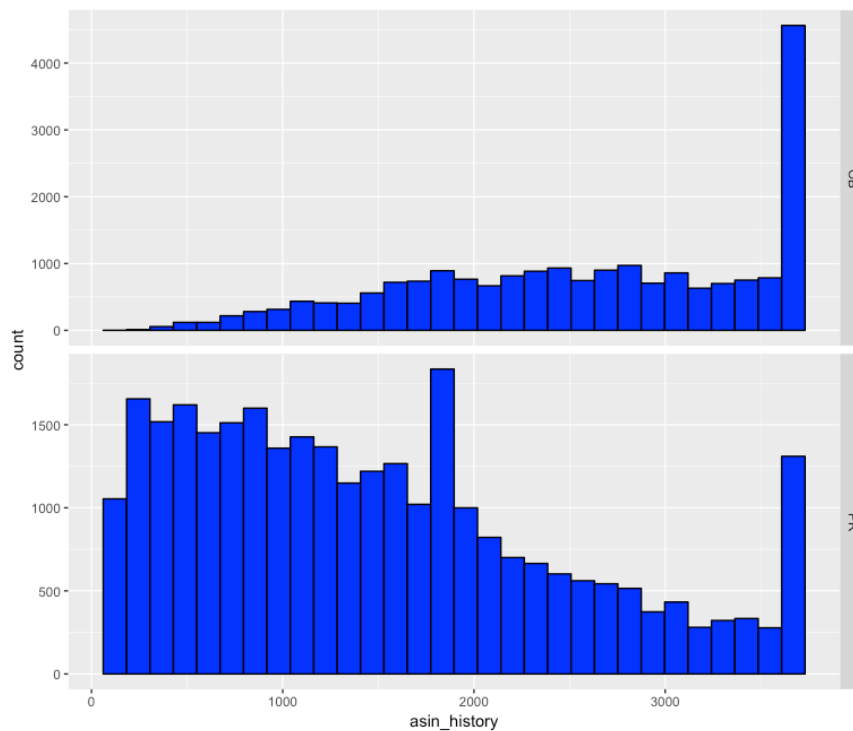
(Figure 2: ASIN history and ASIN buyable attribute)

Total Glance View reflects the popularity on one specific ASIN and indicates ASIN demand. As Figure 3 shows, average total glance view of PR ASIN is 212.02, while average total glance view of OB ASIN is 23.39. Also, p-value of total glance view is very small. Thus, the test rejects null hypothesis and accepts alternative hypothesis: there is significant difference of total glance view between PR ASIN and OB ASIN.

```
> with(nd,t.test(total_gv[label=="PR"],total_gv[label=="OB"]))

        Welch Two Sample t-test

data:  total_gv[label == "PR"] and total_gv[label == "OB"]
t = 7.0626, df = 30193, p-value = 1.669e-12
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 136.2777 240.9740
sample estimates:
mean of x mean of y
212.02004  23.39422
```
(Figure 3: Hypothesis testing of total glance view)

Complete Purchase Orders represents total units of the receive purchase order in the last 90 days and is used to evaluate the vendor performance on one specific ASIN. As Figure 4 shows, average complete purchase orders of PR ASIN is 15.213, however, average complete purchase orders of OB ASIN is 1.127. Moreover, p-value of complete purchase orders is very small. Thus, the test rejects null hypothesis and accepts alternative hypothesis: there is significant difference of complete purchase orders between PR ASIN and OB ASIN.

```
> with(nd,t.test(completepo[label=="PR"],completepo[label=="OB"]))

        Welch Two Sample t-test

data:  completepo[label == "PR"] and completepo[label == "OB"]
t = 7.5815, df = 30316, p-value = 3.514e-14
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 10.44427 17.72755
sample estimates:
mean of x mean of y
15.213042  1.127129
```
(Figure 4: Hypothesis testing of complete purchase orders)

Vender Number means the number of vendors to provide the supply of one specific ASIN. As Figure 5 shows, the mean of vender number of PR ASIN is 0.1865, while the mean of vender number of OB ASIN is 0.01426. Moreover, p-value of vender number is very small. Thus, the test rejects null hypothesis and accepts alternative hypothesis: there is significant difference of vender number between PR ASIN and OB ASIN.

```
> with(nd,t.test(vendor_num[label=="PR"],vendor_num[label=="OB"]))

        Welch Two Sample t-test

data:  vendor_num[label == "PR"] and vendor_num[label == "OB"]
t = 68.088, df = 37279, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.1672741 0.1771900
sample estimates:
 mean of x  mean of y
0.18649528 0.01426323
```

(Figure 5: Hypothesis testing of vender number)

Average Daily Inventory indicates the average daily inventory in the last 90 days and is used to evaluate the inventory depth. As Figure 6 shows, the mean of average daily inventory of PR ASIN is 12.8876, while the mean of average daily inventory of OB ASIN is 0.5042. Additionally, p-value of average daily inventory is very small. Thus, the test rejects null hypothesis and accepts alternative hypothesis: there is significant difference of average daily inventory between PR ASIN and OB ASIN.

```
> with(nd,t.test(inv_avg[label=="PR"],inv_avg[label=="OB"]))

        Welch Two Sample t-test

data:  inv_avg[label == "PR"] and inv_avg[label == "OB"]
t = 17.012, df = 30236, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 10.95669 13.81023
sample estimates:
 mean of x  mean of y
12.8876462  0.5041841
```

(Figure 6: Hypothesis testing of vender number)

After in-depth feature selection and engineering, six features were chosen to build the machine learning model: last inbound history, ASIN history, ASIN demand, complete purchase orders, average daily inventory and vendor number

• Model Building and Result Analysis

Firstly, the raw data with 60,000 ASIN and the attributes are randomly divided into training set and test set. Training set accounts for 75% of total ASINs, test set accounts for the rest 25%. In the study, the model were built with two machine learning. Models: decision tree and random forest.

Decision Tree:
Decision tree model is the model of computation in which an algorithm is considered to be basically a decision tree. In this "decision tree", there are a number of branches or queries as the input and the output is the final decision, during the decision process, every query is dependent on previous queries.

As shown in this decision tree model (Figure 7), average daily inventory, inbound history and complete purchase order are the three significant features that affect the classification of ASIN buyable attribute. In the testing set (Figure 8), the accuracy of the classification is (4146 + 6525)/(4146 + 1130 + 889 + 6525) = 84.09%, which means there is 84.09% chance of predicting the ASIN attribute correctly.
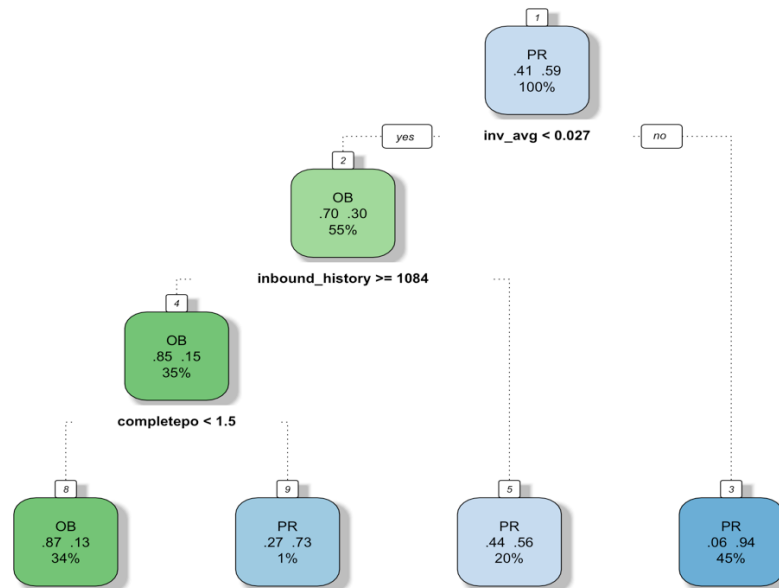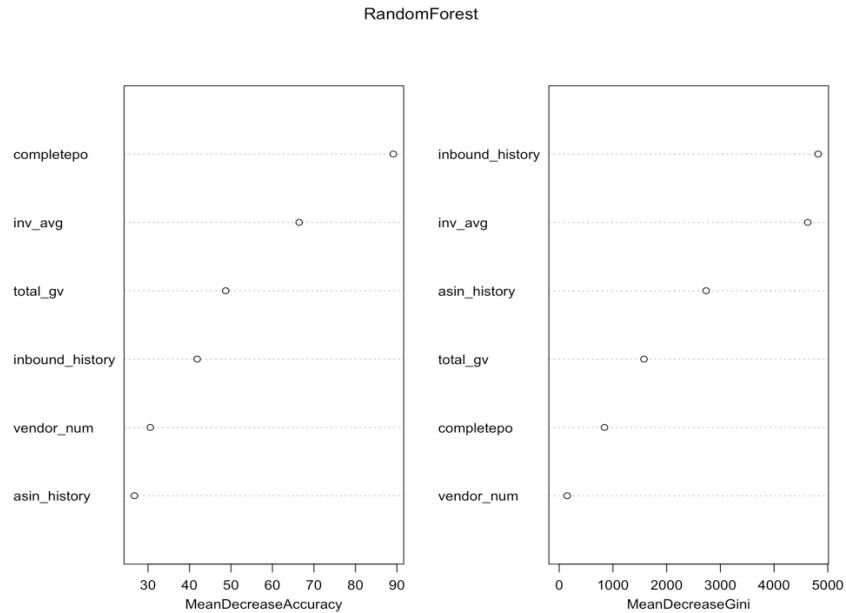

(Figure 7: Decision tree model)

	Reference	
Prediction	OB	PR
OB	4146	889
PR	1130	6525

(Figure 8: Decision tree prediction accuracy)

Random Forest:

Random Forest is an ensemble learning method for classification. It constructs a multitude of decision trees at training time and uses the mode of classifications of the individual trees as the output. This method could correct the overfitting to the training set of decision tree model.

As shown in the random forest model (Figure 9), average daily inventory, inbound history and complete purchase order are the three significant features that affect the classification of ASIN buyable attribute, which is aligned with the results of decision tree. In the testing set, the accuracy of the classification is (4511 + 6624)/(4511 + 790 + 765 + 6624) = 87.75%, which means there is 87.46% chance of predicting the ASIN attribute correctly.

RandomForest

(Figure 9: Random forest model)

	Reference	
Prediction	OB	PR
OB	4511	790
PR	765	6624

(Figure 10: Random forest prediction accuracy)

As Random Forest has higher accuracy than Decision Tree does, we propose to utilize Random Forest to predict ASIN buyable attribute.

Benefits:

Utilizing the machine learning model to predict ASIN buyable attribute regularly would be beneficial in the following aspects:

- NIS DEA improvement

In 2017, the yearly NIS DEA is 91.6%, which missed the goal by 190 bps. This machine learning ASIN attribute prediction model would correct the ASIN buyable attribute which is misclassified and prevent the not-buyable ASIN from being ordered by customers in a timely manner. The estimated NIS DEA would be improved by 100 bps per week.

- Stuck Order Reduction

As of Dec 2017, there are 84,704 units pending in the system, in which 97.97% of backing units are pending shipment and missing Promise due to normal inventory availability issues. By implementing the random forest machine learning method, we would be able to detect the ASIN with misclassified buyable attribute in a timely manner. With the correction of buyable attribute, the about 40,000 existing stuck units would be removed from the backlog, greatly reducing the risks of customer contacts and complaints.

- Work Efficiency Improvement

Amazon.CN has millions of ASIN. It is time-consuming for in-stock managers to check the attribute of each ASIN and correct the misclassified attribute manually. The random forest machine learning method would initially narrow down the ASIN list of which buyable attribute has high probability of misclassification. In-stock managers just need to manually check the attribute of the short list. This machine learning methodology would save in-stock managers 50% labor hour, greatly improving work efficiency,