

Predicting Shipping Order Delivery Status

Morgan A. Keith

A dark blue diagonal gradient bar that starts from the bottom left and extends towards the top right, covering the lower half of the slide.

Business Understanding

The project will help predict a packages deliverable status as either Late Delivery, Advance Shipping, and Ship on time. I will show the overall delivery status and a few features values delivery status relation.

Data

The data being used is information from a company's supply chain about shipping orders like payment type, customer information, order information, delivery dates and status. The timeframe of the model is from 2015-01 - 2018-02.

- Some features were removed because they cause data leakage, redundant information, or added too many features when converting categorical data.

Method

This project will observe the delivery status of orders compared to some features, and create a model to predict if the order is late, on time, or in advance. I compared delivery status to each month, shipping mode, market, and customer segment. Each feature showed about half the orders are late delivery. For predicting models, I used logistic regression as a baseline, then random forest and XGBoost, and random search to optimize parameters.

Exploratory Data Analysis

- Each month there were about 5000 orders, with between 2500-3000 late orders (57.3%).
- Each year delivery status stayed the same at late delivery (57%), advance delivery (~24%), shipping on time (~18%)
- For Shipping Mode, Standard Class has the most orders (60%), and with 40% late deliveries. Other shipping modes late delivery percentage are First Class (100%) , Second Class (80%), and Same day (52%).
- Orders being delivered to Latin America and Europe the most, though all Markets show the same late delivery status percentages with late delivery (~57%), advance shipping (~24%), and shipping on time (~18%).
- For Customer Segment, Consumers are ordering the most though delivery status among segments are consistent with late delivery (~57%), advance shipping (~24%), and shipping on time (~18%)

Random Search XGBoost

XGBoost

	precision	recall	f1-score
Advance shipping	0.58	0.63	0.60
Late delivery	0.75	0.85	0.80
Shipping on time	0.66	0.30	0.41
accuracy			0.70



Conclusion

In conclusion it seems our models are able to predict late delivery well, advance shipping was predicted moderately well, and shipping on time had poor predictability.

- Shipping on time also shows poor recall throughout the models, though with reasonable precision, it's possible shipping on time shares a lot of predictor traits with late and advance shipping (leading to false negatives).
- Advance Shipping tended to have a moderate difference between precision and recall. Its precision is lower than the recall, the model may have trouble separating between shipping on time and advanced. Especially since advance shipping is relative to shipping mode, it can only happen for standard shipping in this data.
- Late Delivery was able to be predicted well, perhaps better than advance and on time shipping because it has more clear predictors to help the model accuracy.

Next Steps:

- Overall, there is over half late deliveries, some features delivery status were looked at, but perhaps comparing a customer's information type to their location, and see if that has an affect on delivery status. This could help identify why there are late deliveries.

The End

Morgan A. Keith
e-mail: Morgan A. Keith