

Which customers are unhappy?

Nadya Agrawal
March 2023



I understand your frustration.

1. Overview & Business Problem
 2. The Data
 3. The Method
 4. The Model
 5. Findings
6. Recommendations
 7. Next Steps

Overview & Business Problem

Our e-commerce company loses money when we lose customers.

How can we identify which customers will leave us before they do?

Through machine learning, we can determine the best predictors for churn.

Intuition: transactions and website usage will be the most important.

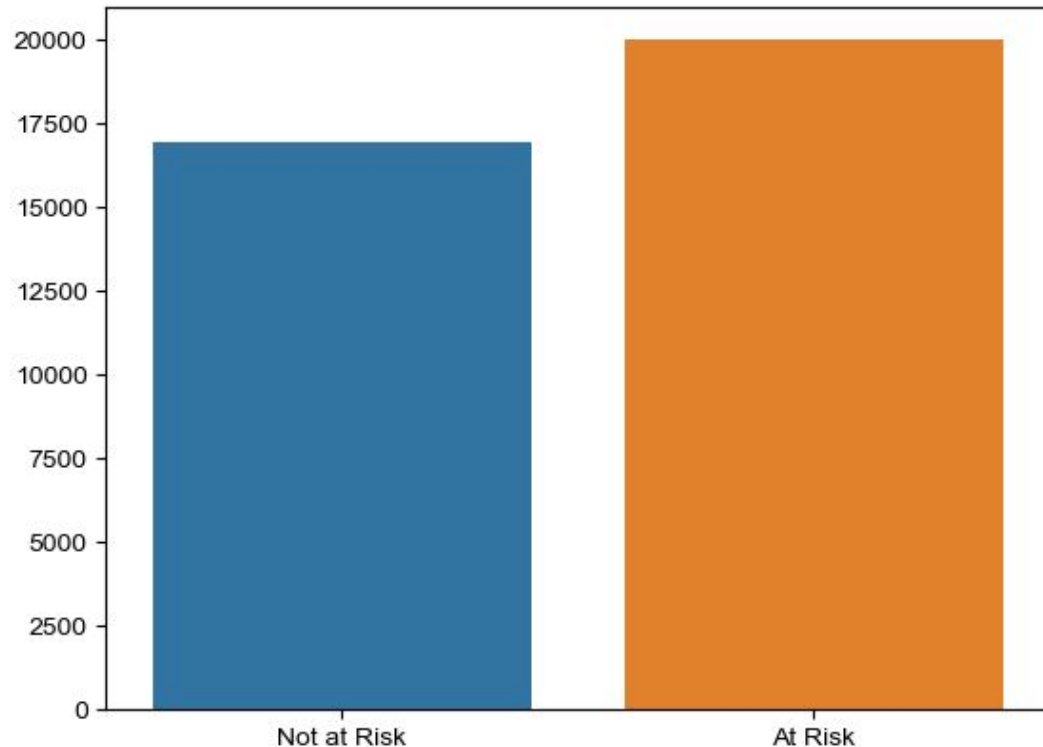
The Data

The data comes from the **“Customer Churn”** dataset, posted by Pawan Trivedi on Kaggle.

It contains data on membership, website use, transactions, and more.

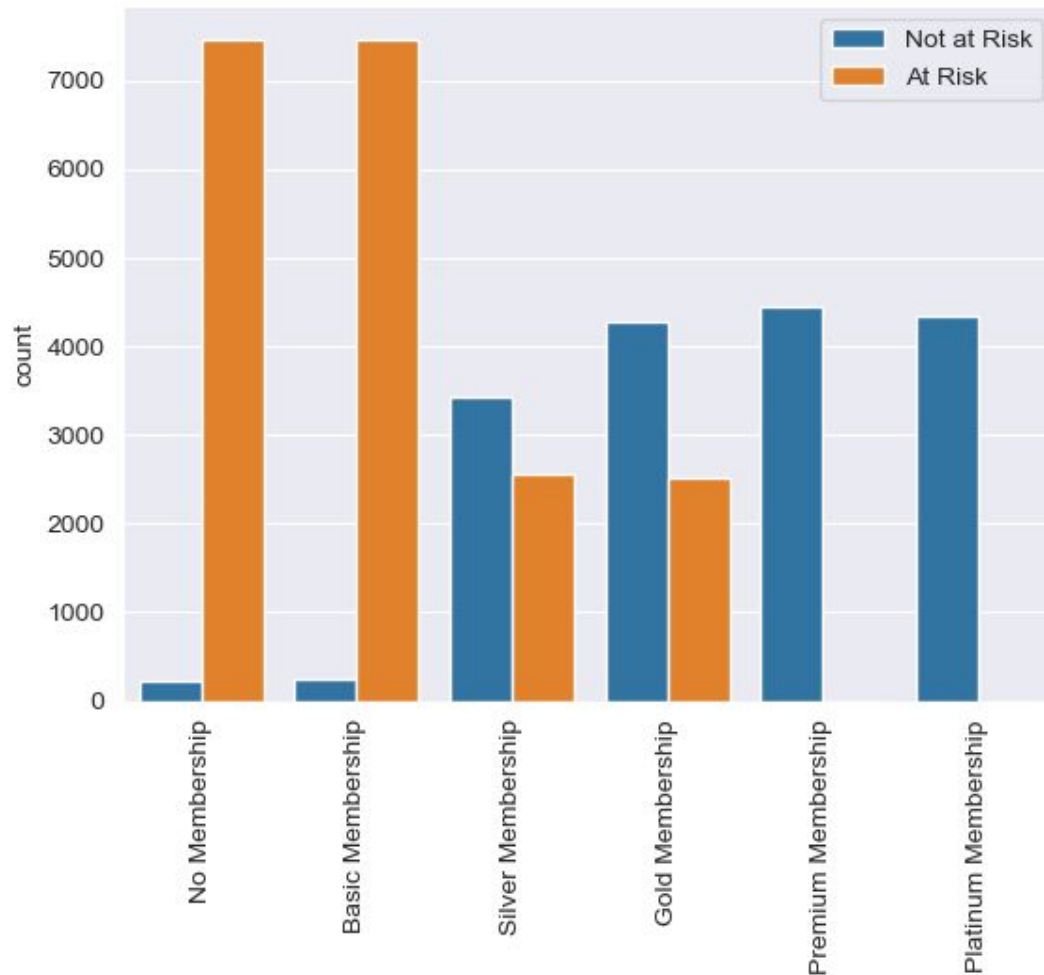
Data Split

Pretty even split between those at risk of churn and not



Membership

- Six membership types
- Most are in the “**basic**” or “**no**” membership categories, these are **most are at risk** of churn
- The data shows everyone with a “**platinum**” or “**premium**” memberships are **not at risk** of churn



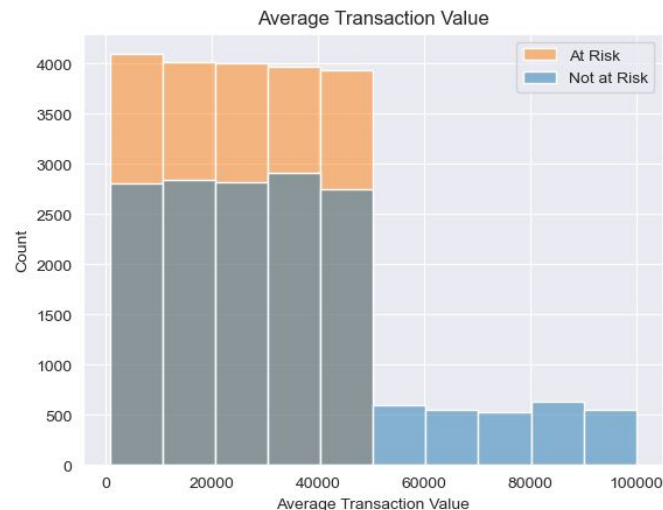
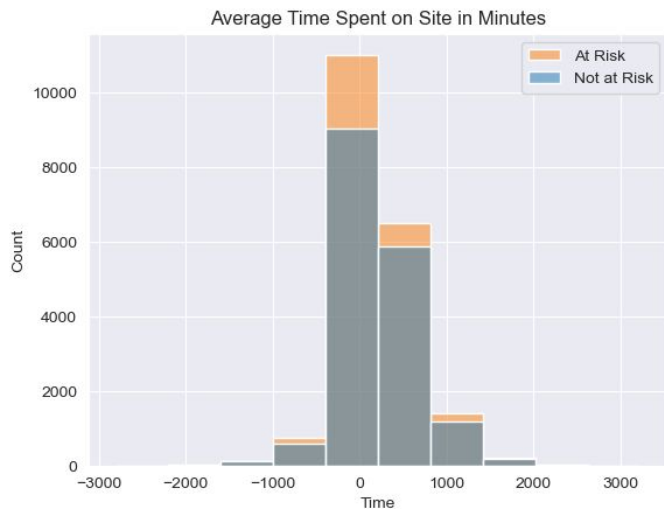
Wallet Points

The target seems almost sectioned when it comes to points



Time & Money

Thought time and money spent on the site would be best indicators, but doesn't seem to be the case.



The Model

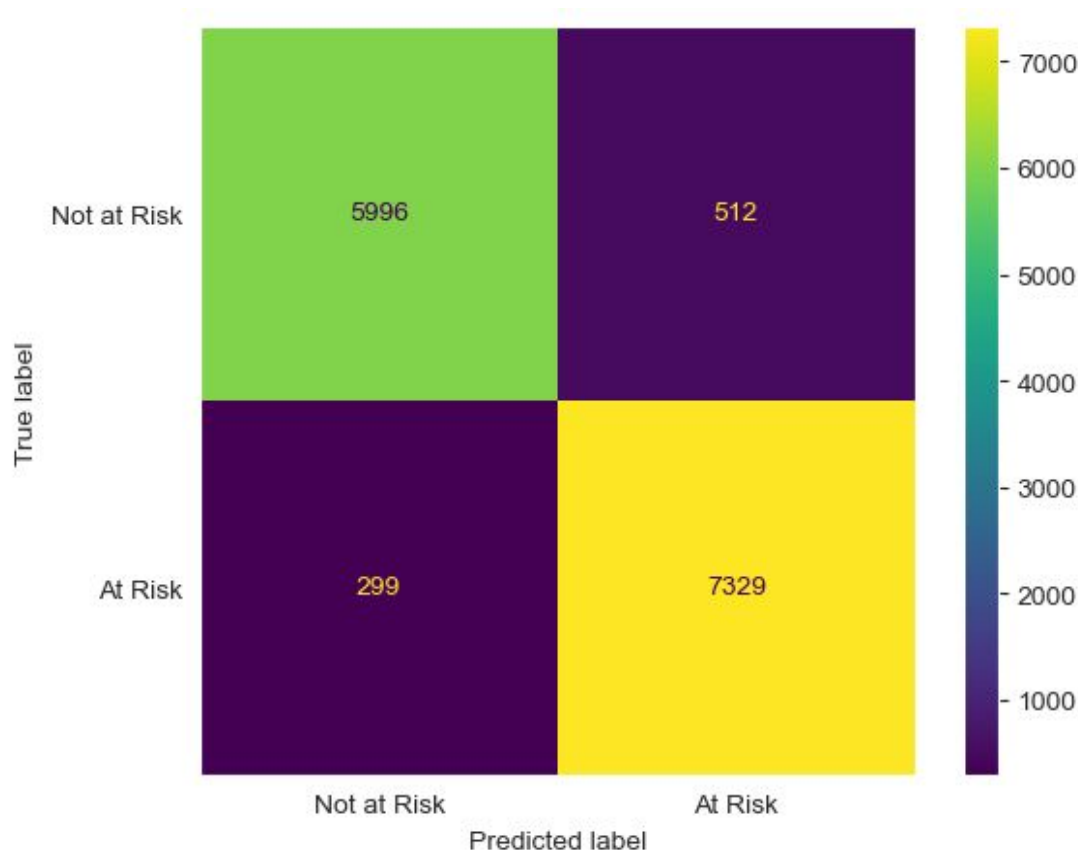
- Ran a range of models, found XGBoost Classifier to be the best with 96% recall score and successfully predicting 94% of test data
- Removed negative values since unreadable
- Wanted to minimize false negatives

Confusion Matrix

XGBoost Classifier **minimized false negatives** the best compared to other models.

Received a 96% recall score.

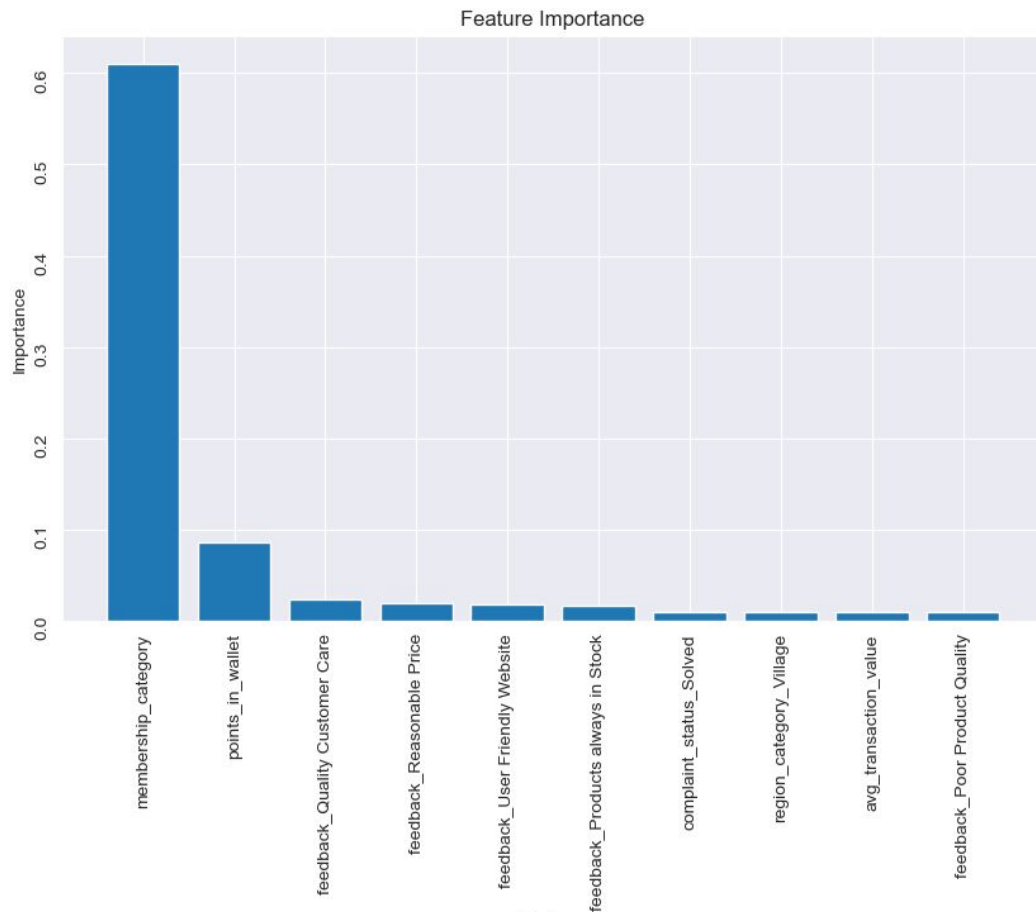
Successfully predicted 94% of test data.



Features

Membership category is by far the most important feature in our model.

Followed by **wallet points** and if they've complained about **customer care**.



Wallet Points (again)

Wallet points is the second most important feature. Earlier we saw sectioning in the data.

Now, looking at the bins at risk customer fall into, most hold between **500-750 points**.



Recommendations

- Use **XGBoost Classifier** to predict churn risk
- Focus marketing on customers that are in **lower membership categories** or do not hold a membership
- Re-target customers who have between **500-750 points in wallet**
- Re-target customers who have given feedback on **customer care quality**

Next Steps

1. **Solve for overfit with additional recent, high-quality data.**
2. **Look at previous marketing to at-risk customers and determine whether re-targeting helps or hurts.**

THANK YOU

Nadya Agrawal



/npagrawal

nadyaagrawal@gmail.com