

# DSCI6006 Final Churn Model

by Jason Hortsch

# What is churn and why do we care about it?

- Churn is a term for customers who will leave the company
- It is important because it is easier to keep a customer than find a new one
- Potential churners can be incentivized to stay
- A churn model attempts to identify which customers will leave or stay



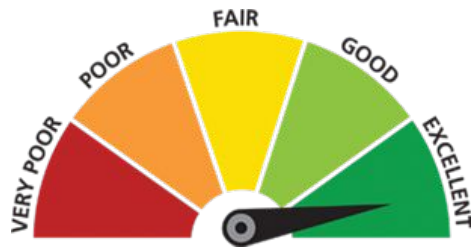
# How is a churn model developed?

- We use a labeled data set
- “Explore” the data and make changes as needed
- We “train” our models on this dataset
- Split data -> train on one portion -> test on remaining portion
- Ideally, by using training/testing, the model will be ready for the real world



# How to model measure success?

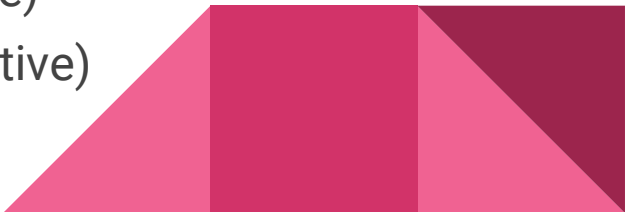
- One metric is simple accuracy
  - What percent of customers in our test set did the model correctly predict
- Can be dangerous - not always the best indicator
- I chose to focus on different metrics that give more information



# Precision / Recall / F1

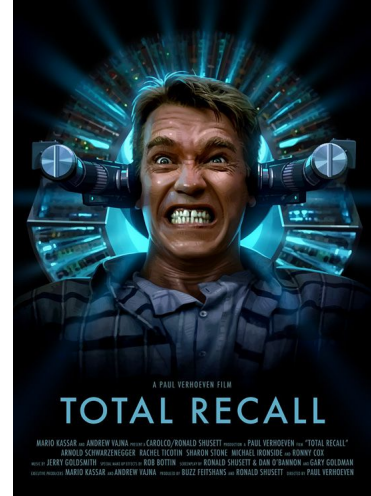
- Consider the classic example of testing for a disease
- There are four outcomes:

	Truly has cancer	Truly has no cancer
Test says cancer	True Positive	False Positive
Test says no cancer	False Negative	True Negative

- **Recall:**  $\text{True Positive} / (\text{True Positive} + \text{False Negative})$
  - **Precision:**  $\text{True Positive} / (\text{True Positive} + \text{False Positive})$
  - **F1:** A combination of the above two
- 

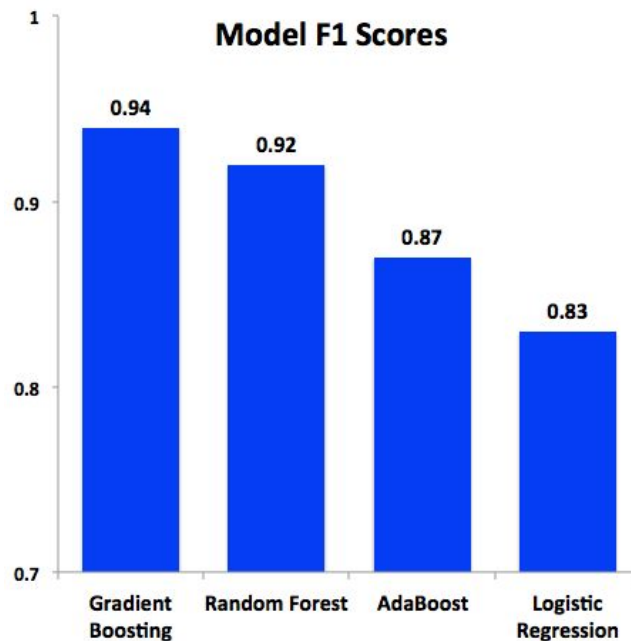
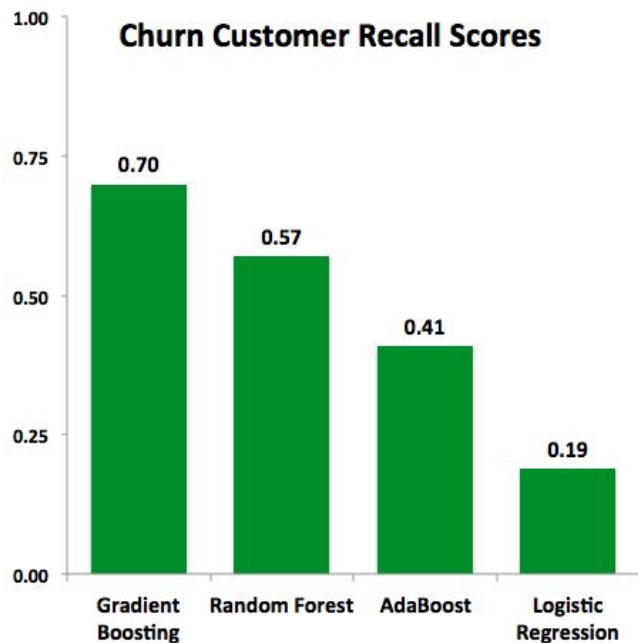
# Explanations and choice of recall

- Consider customers who did churn:
  - **Precision** tells us: Out of customers the model predicted to churn, what percent actually did
  - **Recall** tells us: Out of customers that actually churned, what percent did the model catch
- I chose to focus on the recall score for customers who churned
- Correctly identifying customers who will churn is the key
- This information can be used to increase retention



# Models and performances

Gradient Boosting / Random Forest / AdaBoost / Logistic Regression



# Tuning Random Forest and Gradient Boosting

Gradient Bosting /Random Forest Performances with Tuning:

	Gradient Boosting	Random Forest
F1 Score	0.85	0.80
Churn Customer Recall	0.78	0.70

Gradient Boosting has a slight edge





# Is there a winner?

- Models and performance can be context dependent
- Random Forests are generally considered good for churn
- Must be cautious of overfitting
- The dataset was also not large, so must be careful with 'conclusions'



# Business insights moving forward

- Customers who the model predicts to churn can be targeted
  - Give bonuses / rewards / discounts to encourage them to stay
- Customers who used the service more appeared to be more likely to churn
- More customer service calls for churn customers





# Thank You!

Any questions?