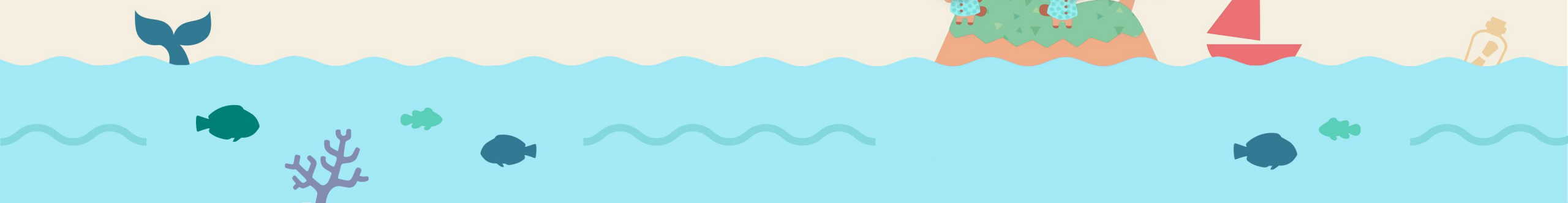


# WELCOME TO THE JUNGLE: MAKING SENSE OF ANIMAL CROSSING REVIEWS

Data Science Live

Scott Yang, Jonerik Blank, and Peter Zhang



# Background

## Worldwide Media Revenue 2020

Movies  
\$45 B

Music  
\$51 B

Video Games  
\$160 B

Video Games are a  
Huge Market



Over 32 Million  
Copies Sold in  
2020

Animal Crossing: New  
Horizons Broke Records

A recent study  
found that eWOM  
impacts video  
game sales  
more than  
twice as much  
as the next best  
quality cue

Quality Cues are  
Important

# Abstract

**Goal: Predict player sentiment before release to maximize probability of positive reception.**



Clearly managing critical reception is vital to a game's market performance. Leveraging a dataset comprised of both critical and user reviews, we examine how various NLP methods can be used to produce accurate predictive models for both positive and negative reviews using specific words. Developers could, in the future, use these models during playtesting to gauge the likely critical reception of their title prior to release.



# Exploratory Data Analysis

## Dataset:

2,999 User Reviews

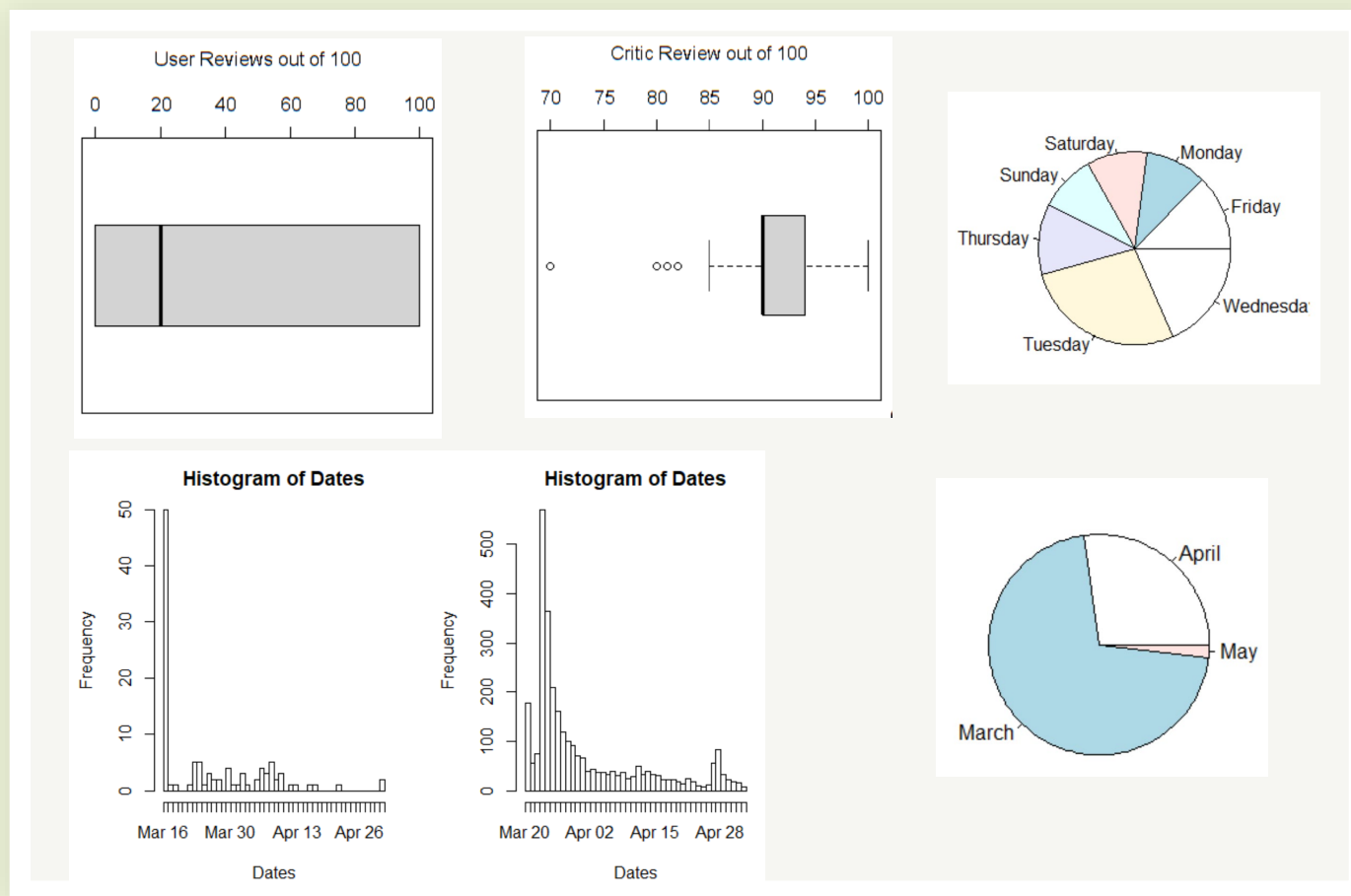
107 Critic Reviews

Scored From 1-100

Includes Date and Text of Reviews

>85 Score = Good (critic)

>80 Score = Good (user)





# Methods Used

## LASSO

User AUC: 0.91  
Critic AUC: N/A

## Keras Neural Network

User Acc: 0.87-0.88  
Critic Acc: 0.84-0.9

## Simple Trees

User AUC: 0.80  
Critic AUC: 1

## Random Forest

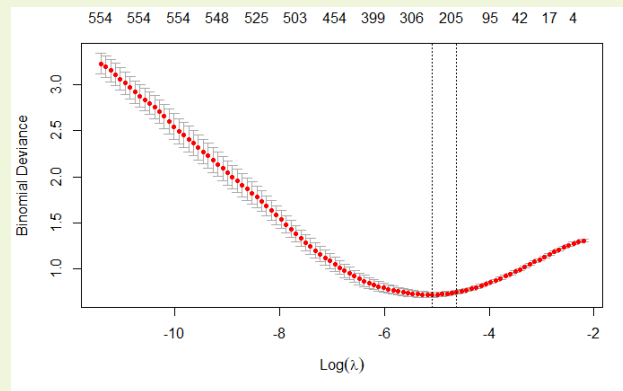
User AUC: 0.92  
Critic AUC: 0.55



# LASSO Results

- 16043 predictor terms in initial user bag of words → Shrunk to 823\* predictors
- 1304 terms in initial critic bag of words → Shrunk to 433\* predictors
- Small critic dataset was not viable – too few observations. Users reduced to 186 non-zero predictors, AUC 0.9146; Misclassification error of 0.218
- 7:3 for training to validation split
- **Positive Words:** Praise, unique characteristics of game
- **Negative Words:** Perceived values of company, repetitiveness

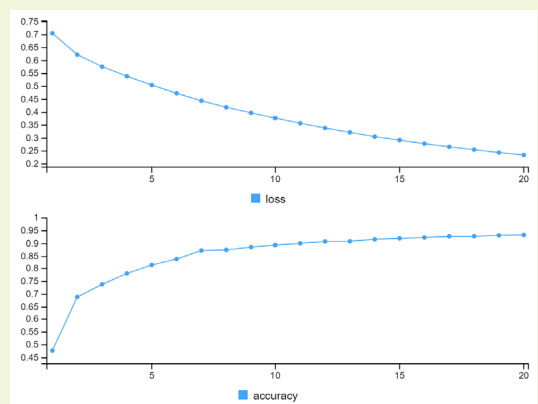
\*must appear in at least 1% of documents



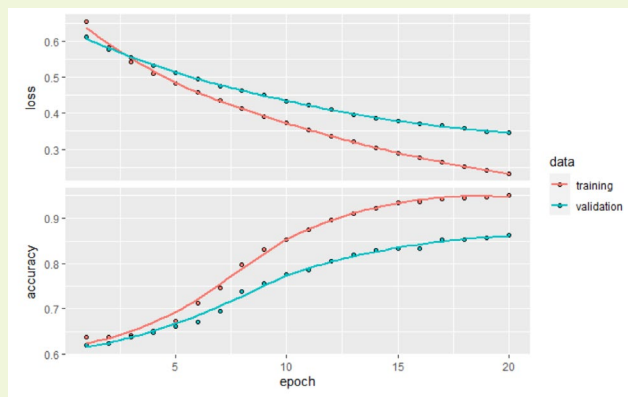


# Keras Results

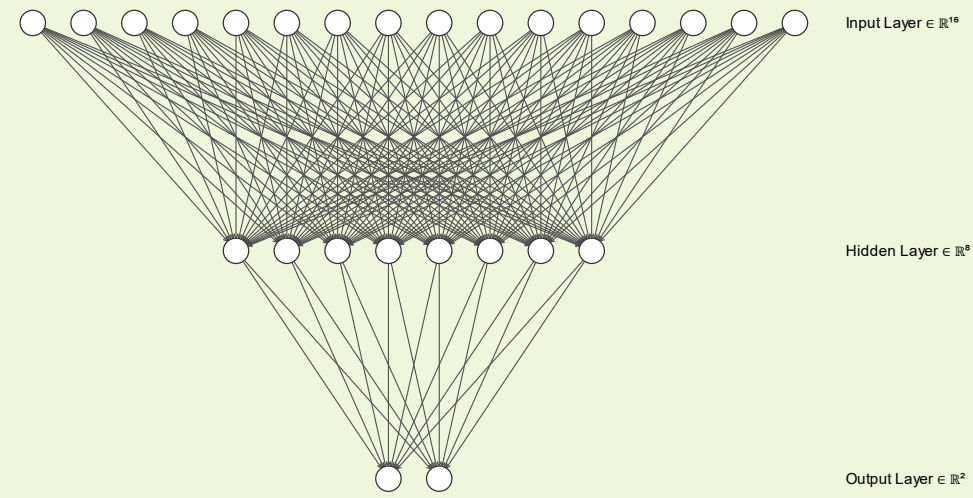
- For Neural Network, model architecture was 2 layers deep. 1st layer at 16 neurons, 2nd layer had 8 neurons. Inner layers used ReLu function, final output used softmax
- Using CUDA (GPU compute) for Neural Network (via KERAS) sped up processing
- 20 epochs produced best accuracy (diminishing returns around 10), we end with 0.87-0.886 accuracy for users; adding layers did not improve accuracy noticeably; bigrams-quadrograms did not improve accuracy
- 6 epochs produced best accuracy for critics (unlike LASSO, still useable even w/ small dataset); 0.84-0.90 accuracy



Validation Data Only



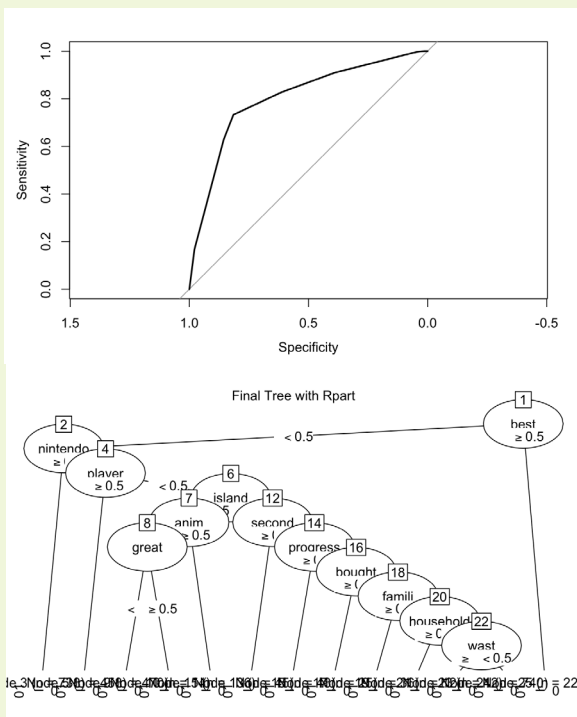
Test Data (w/ internal validation split)



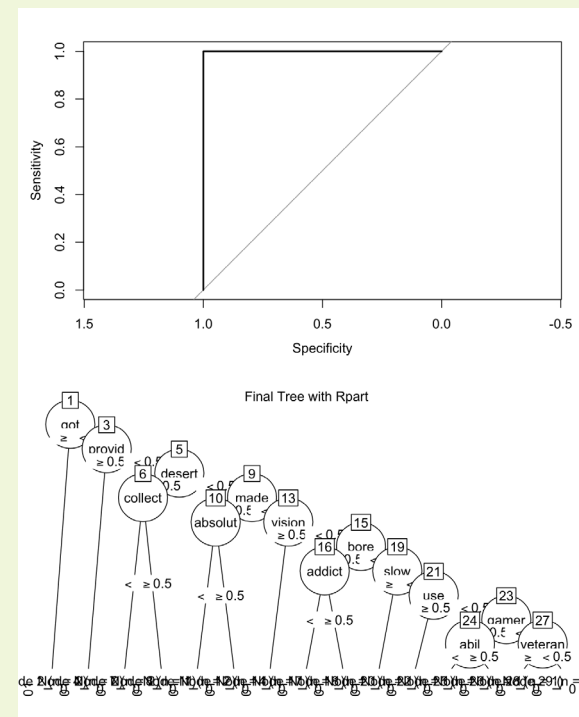


# Tree Results

Trees – ran simple tree with all predictors on both user and critic reviews



Roc curve and tree for user reviews, auc = .8035



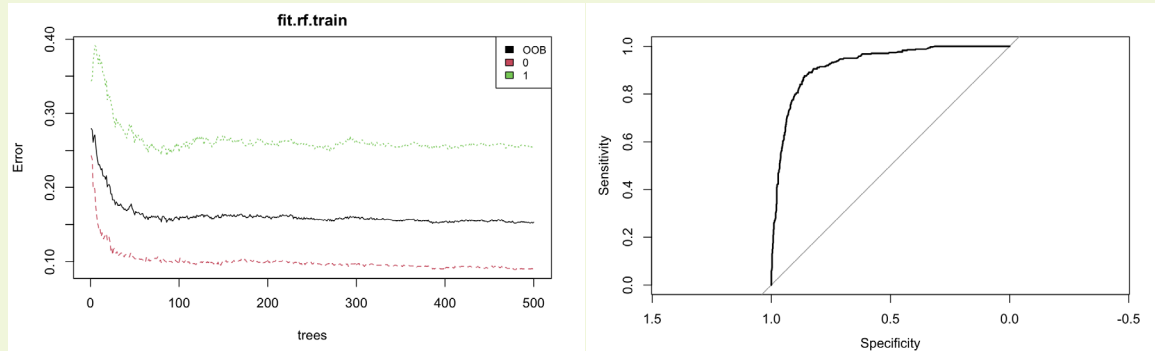
Roc curve and tree for critic reviews, auc = 1



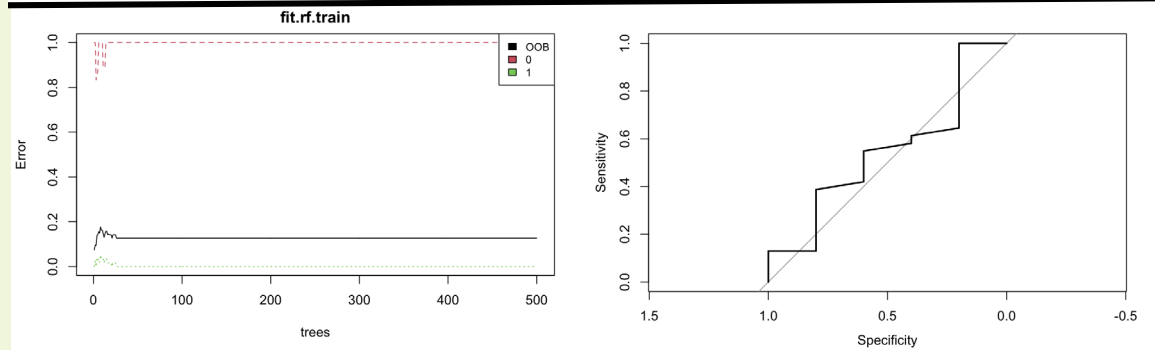


# Random Forest Results

- Ran Random Forest w/ 67%/33% training/test split for user and critic reviews
- Very good results for user reviews (AUC = 0.92)
- Critic reviews: Poor results, likely due to the small size of the critic review dataset (107 total datapoints, 71 train, 36 test)





**User reviews: ROC = 0.9245,  
Misclassification error = 0.135**



**Critic reviews: ROC = 0.5452,  
Misclassification error = 0.1389**



# Final Model Comparisons

Model	Accuracy(Users)	Accuracy(Critics)	Interpretability	Recommended Use
LASSO				Understand user sentiment
Neural Net				Potential Accuracy – Be cool
Trees				Mix of Accuracy and interpretability
Trees – Random Forest				Correctly classify good vs bad reviews



# Takeaways and Next Steps



## Key Takeaways

- (-) sentiment focused on perceived ethics of company rather than game quality
- (+) sentiment praised unique game elements – make sure players are noticing during play testing!
- Field expertise needed to interpret word cloud!
- Neural Network and Pre-trained models allow used of smaller datasets (e.g. critics); however, there is loss of interpretability
- More complexity does not necessarily lead to better results
- *Don't accept usurious loans from Tom Nook*

## Next Steps

- Use KERAS built in tokenizer rather than TM
- Use BERT or GPT pre-trained models rather than our own architecture (classifier models)
- Ensemble to average different methods
- Use different game genres
- Re-scrape Metacritic for new user reviews and see if model predicts equally well on unknown Animal Crossing reviews
- Compare processing speed with CUDA (GPU) enabled vs CPU only

# Q&A







Thanks for  
Listening!