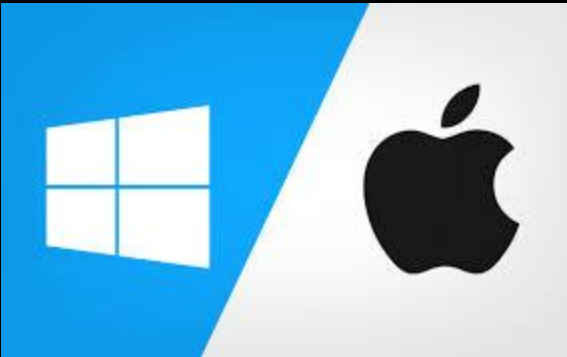# Webscraping & Classification Model - PC vs Mac Gaming

Pablo Rivera

February 2020

Agenda

- Problem Statement

- Key Objectives

- State of the Market

- Classification Modeling & Evaluation

- Recommendations & Next Steps

Problem Statement & Key Objectives

Context:

For too long, PC's dominance in the Gaming, Enthusiast, and eSports space has gone unchecked.

Tim Cook, CEO of Apple is seeking to explore new market opportunities in order to reverse stagnant Apple unit sales. Therefore Apple has recently hired a new executive to spearhead this initiative, named ***Cim Took***.

Mr Took has scheduled time with you to discuss initial steps and analyses he will need. In order to prepare for this meeting you turn to your favorite source of consumer sentiment: ***Reddit***

Subreddits evaluated:

- **PC Gaming Master Race**

- **MacGaming**

Key Objectives:

- ***PULSE OF THE CUSTOMER: Begin prototyping a Classification System for ongoing analysis of platform-agnostic gaming consumer comments. Eventually this will expand to include sentiment analysis across various platforms (not just reddit) and customer targeting for ad campaigns***

- ***Evaluate model – how accurately can it identify PC Gaming apart from Apple Gaming***

Executive Summary

- Key objectives: Prototype a Classification System for ongoing analysis of gaming consumer comments to track performance of Apple Gaming initiative. Eventually model will expand to include sentiment analysis across various platforms (not just reddit) and customer targeting for ad campaigns

- PC Gaming Hardware Market is a large opportunity estimated to be over $23 Billion in 2017, shows strong growth and is dominated by PCs. Apple does not currently have strong performing or price-effective solutions for this space

- PCs also dominate the Reddit gaming community as well as high visibility (60 million+ viewer) eSports events

- Over 3.5 Million records were scraped from Reddit however after duplicates and balancing classes 2100 per class were kept

- 7 distinct machine learning models were applied and SVC exhibited the highest accuracy and strongest ability to distinguish between PC & Apple Gaming with an R^2 score of 0.8980 and ROC Curve with AUC 0.947 (showing strong performance with Sensitivity and Specificity)

- Model Coefficients & Insights suggest gamers prefer custom builds with RGB, high performance and SSDs

- Recommendations:
    - Customer Survey to corroborate insights on customer gaming needs (e.g. FPS, CUDA cores, compatibility issues) to influence roadmap
    - Continue building out robustness of classification model: additional feature reduction/refinement, additional feature engineering, lemmatization, experiment with different platforms such as Amazon reviews, Newegg reviews, Twitter
    - Begin cultivating esports partnerships and cyberathlete sponsorships

Background & State of the Market

State of the Market: Apple 19Q3 Unit Shipments & Unit Share Declined Compared To 18Q3

Worldwide Top 5 PCD Companies, 2019Q3
Unit Market Share (Traditional PCs)

Source: IDC 2019
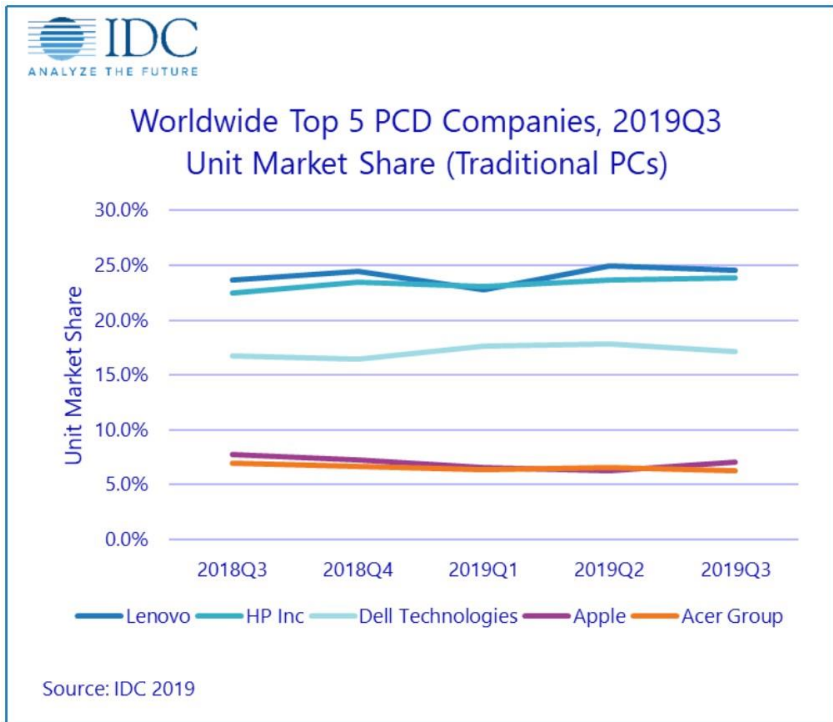
Top Companies, Worldwide Traditional PC Shipments, Market Share and Year-Over-Year Growth, Third Quarter 2019
(Preliminary results, shipments are in millions of units)

Company	3Q19 Shipments	3Q19 Market Share	3Q18 Shipments	3Q18 Market Share	3Q19/3Q18 Growth
1. Lenovo	17.3	24.6%	16.2	23.6%	7.1%
2. HP Inc.	16.8	23.8%	15.4	22.5%	9.3%
3. Dell Technologies	12.1	17.1%	11.5	16.8%	5.3%
4. Apple	5.0	7.1%	5.3	7.7%	-6.1%
5. Acer Group	4.4	6.3%	4.8	7.0%	-7.2%
Others	14.8	21.1%	15.3	22.4%	-3.2%
Total	70.4	100.0%	68.4	100.0%	3.0%

Source: IDC Quarterly Personal Computing Device Tracker, October 10, 2019

Includes Desktops, Laptops, Workstations

Source: IDC, https://www.idc.com/getdoc.jsp?containerId=prUS45584619;

PCs Dominate High-Growth eSports Events And Corresponding Products

- **2.5 BILLION** Video Gamers all over the world

- 62% of customers prefer playing games on **PC / Mac**

- Total PC Gaming Hardware Industry Revenue in 2017 was **$23 Billion**

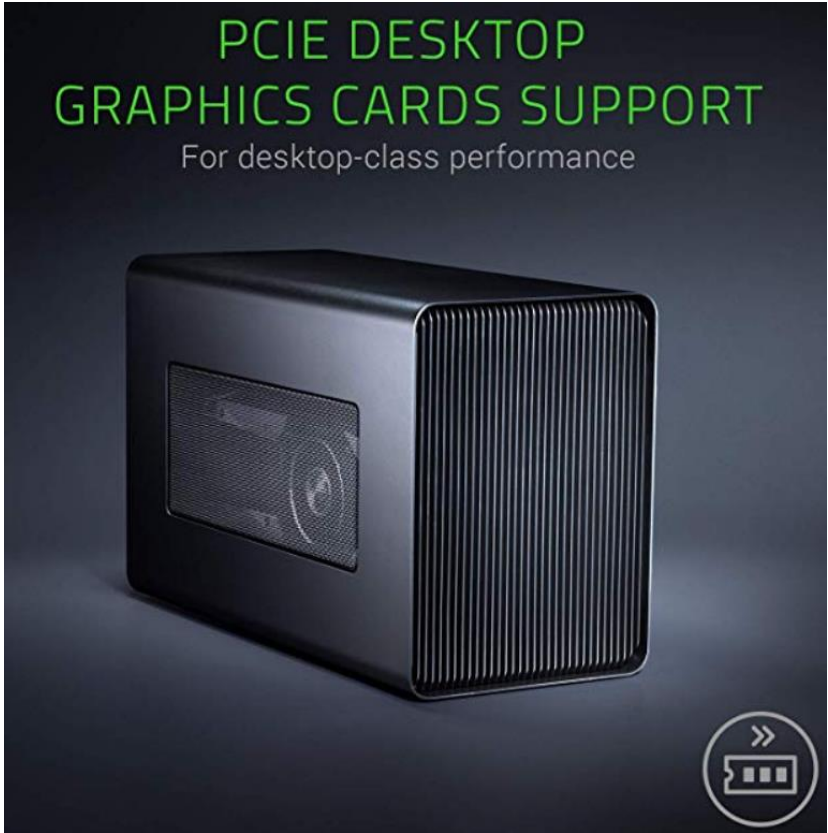- Steam game sales revenue reached **$4.3B** in 2017, growing 19% YoY

100 MILLION unique viewers watched the 2019 League of Legends tournament which had a $2.5M grand prize (that's more than the Super Bowl)

State of the Market: Apple Gaming

$1500
Toaster $299
Just for the enclosure.
$1200 for a topline GPU

Your bag total is $57,850.98.

Get free shipping and free returns on all orders.

Check Out

Mac Pro	1 ⌄	$54,447.98
Show product details ⌄		Remove

 Add AppleCare+ for Mac Pro for $299.00   Add
Get up to three years of technical support and accidental damage coverage.
Learn more ›

 Add a gift message   Add

 Order today, delivers: Pickup:
Feb 17 - Feb 24 - Free   Unavailable for pickup at Apple
Delivery options for: 02476†† ⌄   Burlington
                                   Show more stores

Subtotal	$54,447.98
Shipping	FREE
Estimated tax for: 02476†† ⌄	$3,403.00
Have a promo code? Enter it now ⌄	
Total	**$57,850.98**

Note: $3,400 in state sales tax

Source: Apple.com; Newegg.com; Amazon.com

PCs Also Dominate Reddit Gaming Community

Subreddit Statistics

■ PC Gaming
■ Apple Gaming

3.2M

0.3M

Subscribers
Higher is better

6,347.0

4,104.0

65.0

22.0

636.0

9.0

Rank
Lower is better

Daily Comments

Daily Posts

Subscribers

30k
25k
20k
15k
10k
5k

2013 2014 2015 2016 2017 2018 2019 2020

3M
2.5M
2M
1.5M
1M
0.5M
0

■ PC Gaming
■ Apple Gaming

compar

2013 2014 2015 2016 2017 2018 2019 2020
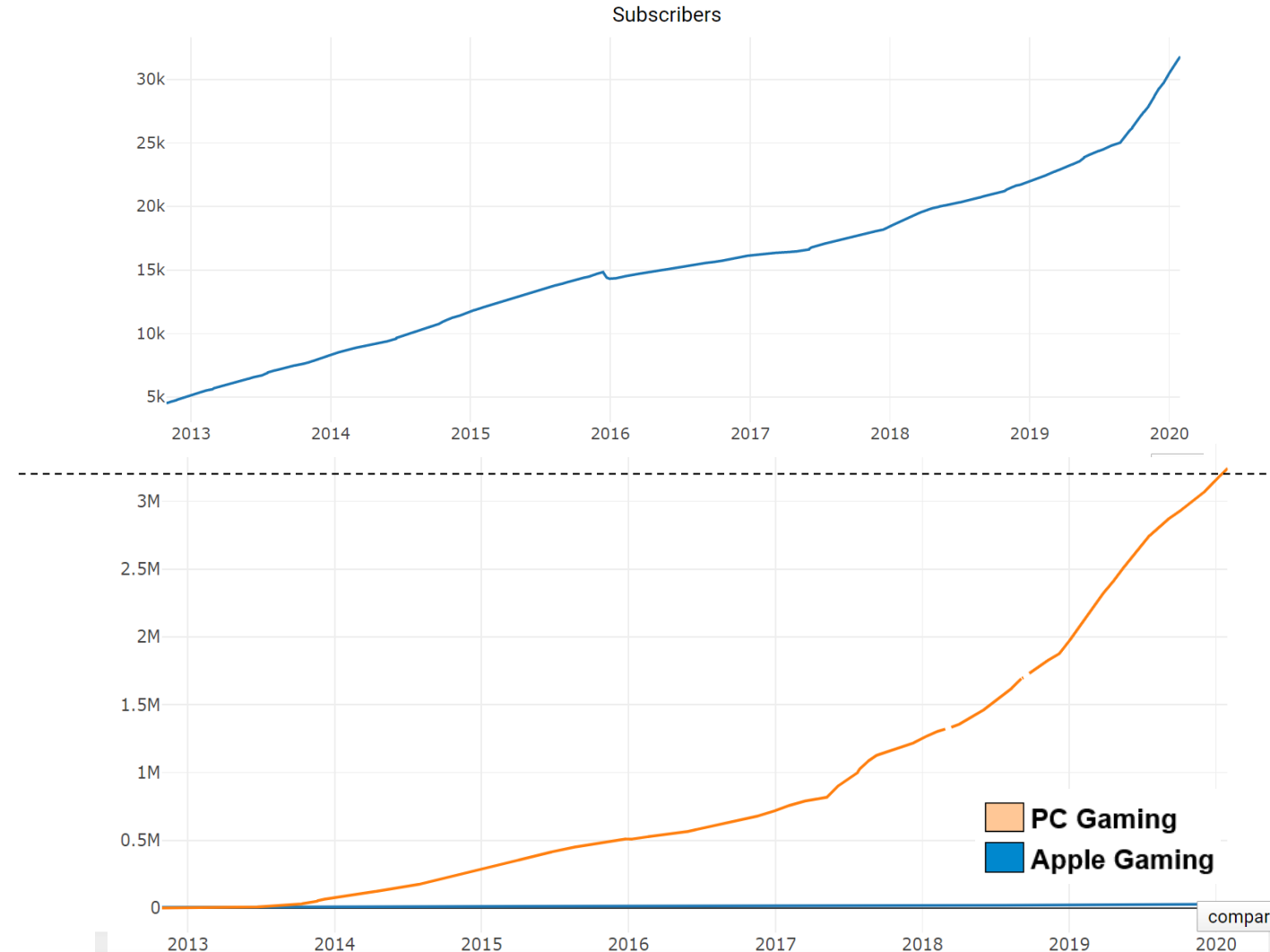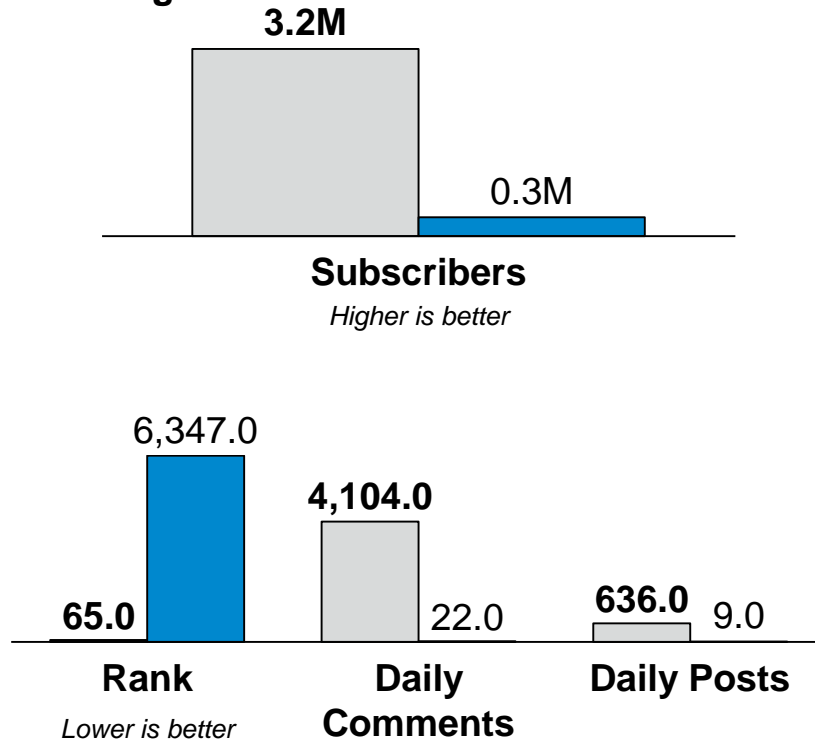
Modeling

Model Overview

Reddit Data Collection

- PushShift API
- 3 million+ records (5GB) were collected
- Over 1800 days

- Apple only 2000 uniques
- In order to maintain balanced classes, PC reduced to 2000

Cleaning, Processing, Vectorization

- Deduplication
- English language standardization

- TfidfVectorizer
- English stopwords included
- Removed irrelevant words/features

Modeling & Hypertuning

- Support Vector Machines 89.804% test accuracy (best model)
- Logistic Regression
- XGBoost Classification

- Gaussian Naïve Bayes
- AdaBoost
- CatBoost
- Voting Classifier Ensembling

Model Evaluation: High Accuracy & Strong Ability To Distinguish Between Classes (Mac vs PC)

Model Accuracy Scores (R^2)

TRAIN
TEST

	TRAIN	TEST
SVC	**0.990**	**0.898**
Logistic Regression	0.954	0.892
Vote Classifier Ensembling	0.957	0.882
CatBoost	0.921	0.869
Gaussian Naïve Bayes	0.985	0.866
AdaBoost	0.873	0.863
XGBoost Classifier	0.852	0.841
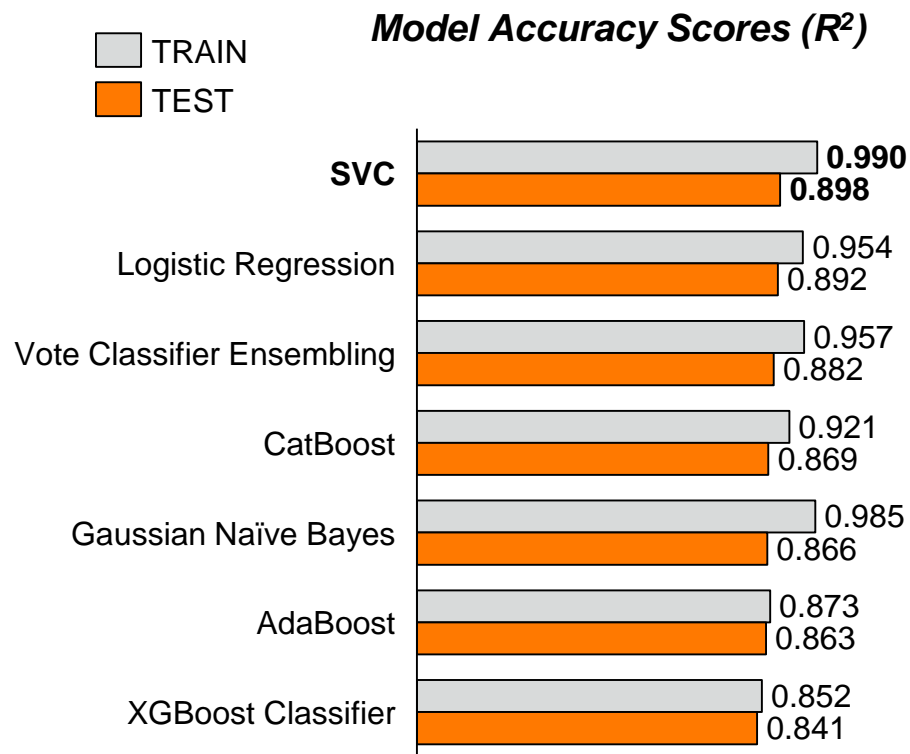
Baseline Score is 0.5
This is driven by the model's balanced classes

Distribution of P(Outcome = 1)

Outcome = 0 — MAC
Outcome = 1 — PC

Frequency

Predicted Probability that Outcome = 1
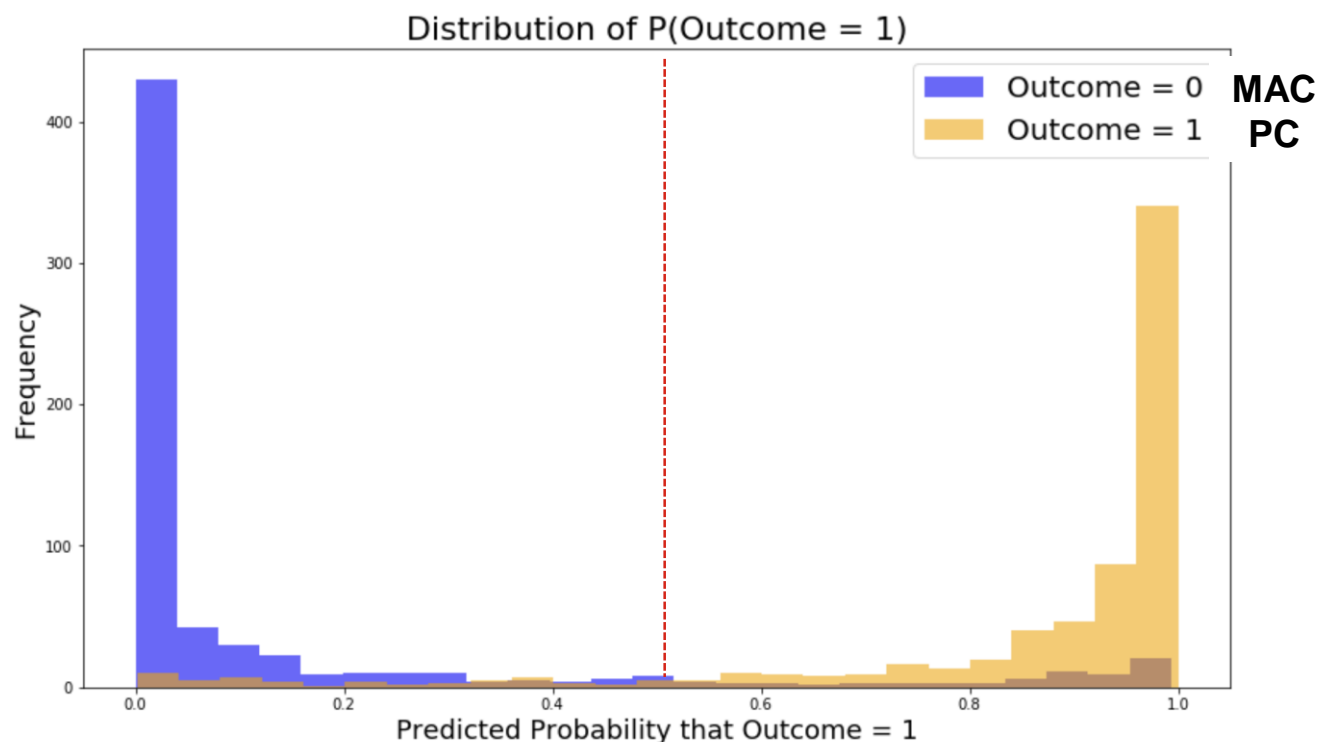
Model Evaluation: Effective Distinction Between Classes

ROC Curve with AUC = 0.947
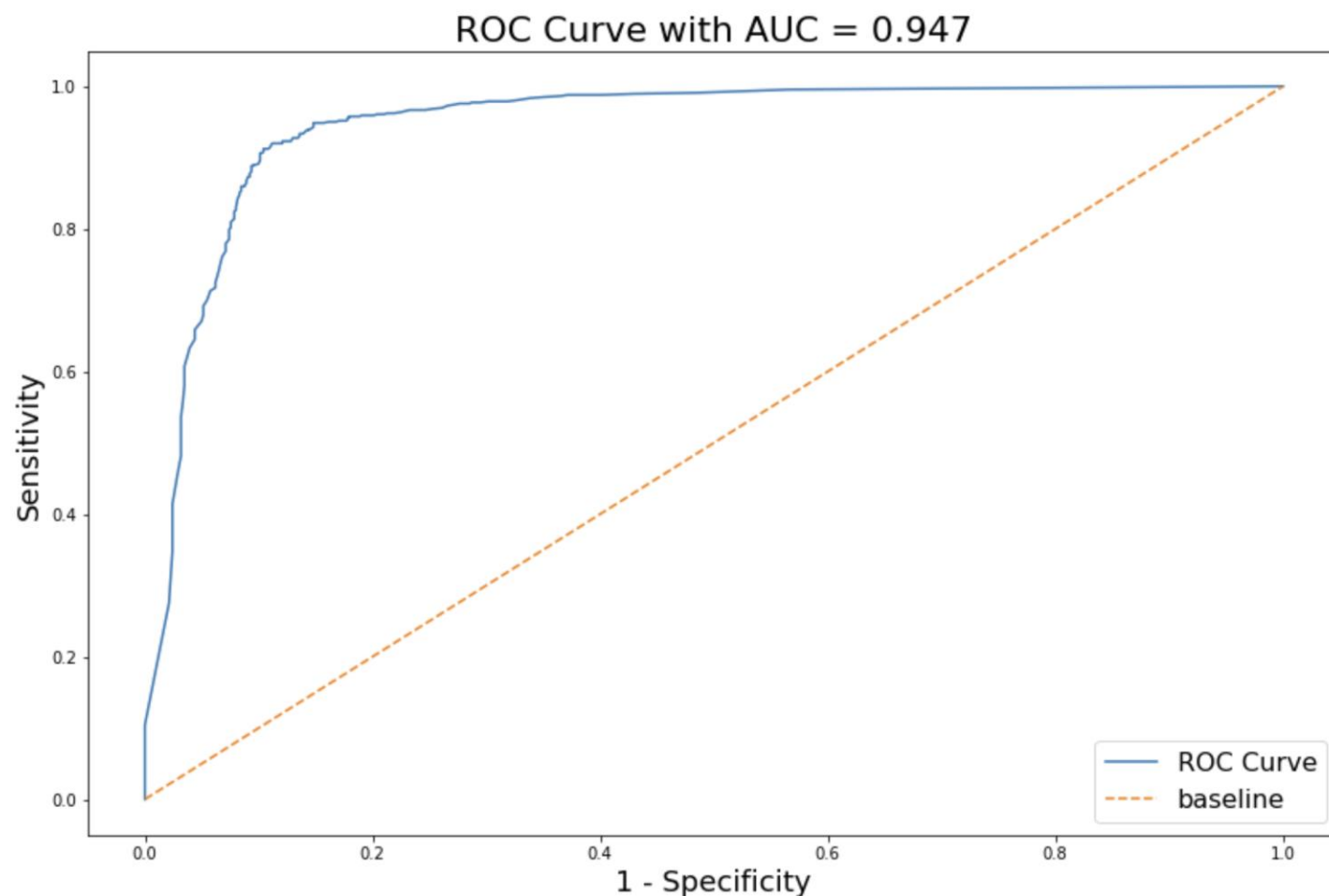
0.932
- **Sensitivity** – how well the model classifies PC Gaming

0.864
- **Specificity** – how well the model classifies Mac Gaming

Insights: Leading Coefficients Suggest Gamers Prefer Custom Builds With RGB Lights, High Performance And SSDs

Leading words that had the highest coefficients (absolute values):

- **Dead** – researching own build after dead Apple product

- **Questions** – on custom builds

- **GPU CPU** – seeking methods of improving performance

- **Finally complete** – indicates custom PC build

- **Z370** – intel chipset

- **Rgb strips** – apple has no RGB

- **Xbox** – games seeking better performance/gameplay experience than console

- **New ssd**

- **Corsair**

- **Halo**

- **Underperforming** – Apple customers

- **Arrived today** – components arriving for custom PC builds

- **Final build**

- **Egpu** – Apple install base researching gameplay improvement options

- **Video card**

- **Low gaming settings -** Apple

- **Feels good** – customer satisfied with PC custom builds

Note that the best model was SVC (black box model)
These coefficients are derived from Logistic Regression model

14

Recommendations & Next Steps

- Survey to bring out insights on customer gaming needs (e.g. FPS, Cuda Cores, compatibility issues) to influence roadmap

- Customer segments to include:

    - Existing PC Gaming enthusiasts

    - Existing Apple fanboys

    - Esports athletes

    - Regular consumers (gauge gaming appeal)

- Begin working on eSports partnerships and cyberathlete sponsorships

- Continue building out robustness of classification model: additional feature reduction/refinement, additional feature engineering, lemmatization, experiment with different platforms such as Amazon reviews, Newegg reviews, Twitter

- Apple to further evaluate addressing custom-builds either through easier installation of operating system or partnerships with component manufacturers (e.g. hardware + MacOS bundling), and RGBs...

- Build financial multi-year projections to estimate Apple entry and cannibalization of PC Gaming market

Appendix