

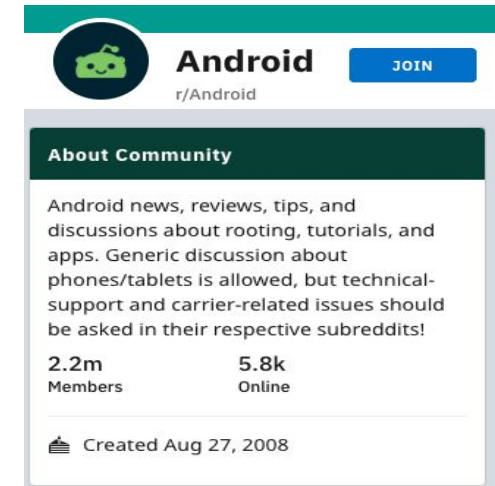
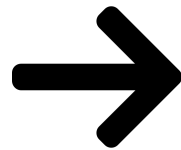
# Selecting the Perfect Operating System: iOS or Android?

Mariam Javed

# Problem Statement

Using the Reddit Web API, web scrape posts from the r/iOS and r/Android subreddits to determine unique sentiments and develop a classification model. Apply this model to the scraped r/gadgets posts to better understand which OS is most trending.

Once we determines the preferred mobile app platform/operating system, we're hoping this information will feed into other key considerations and help StudioM determine how to promote our app and what other key functionalities will be involved.

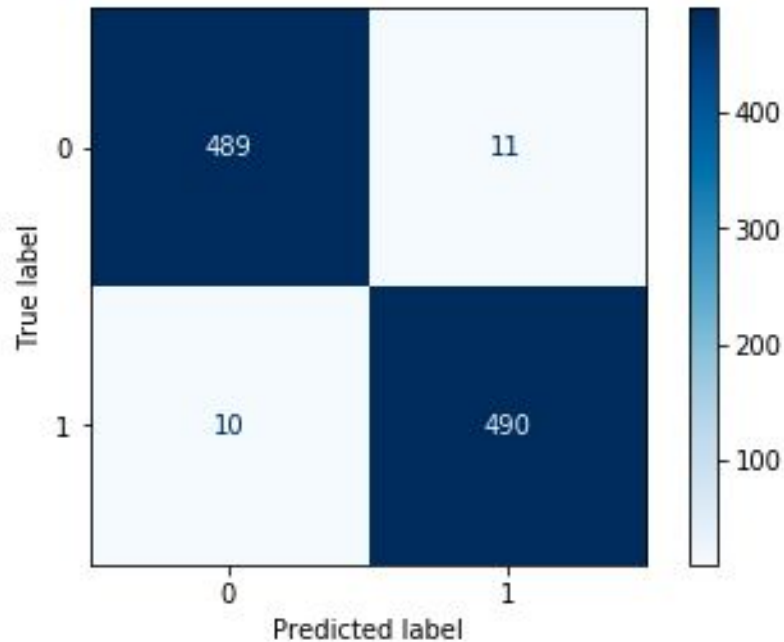


# Training Model: iOS and Android Common Words

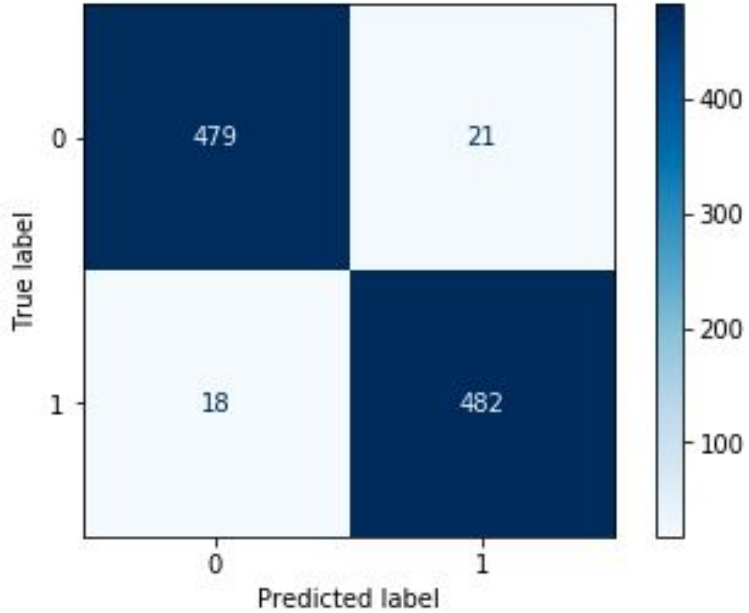


# Training Model: Metrics

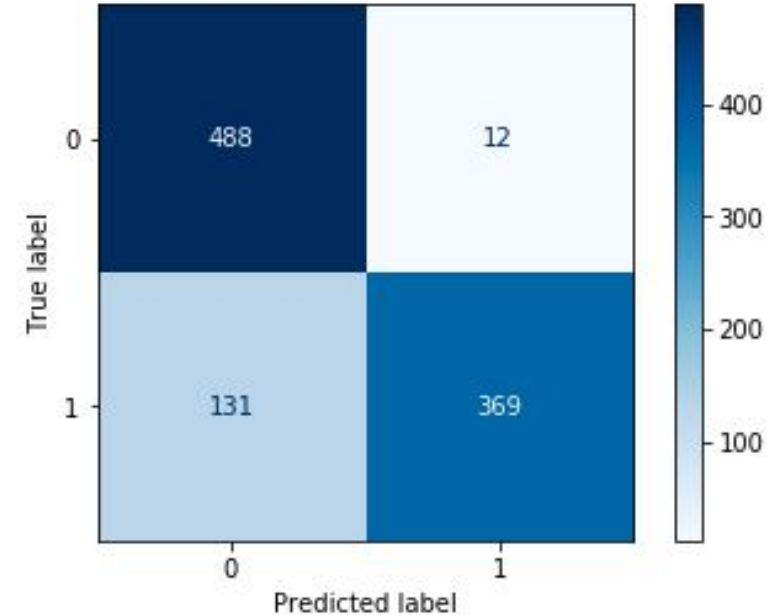
## 1 Logistic Regression



## 2 Naive Bayes

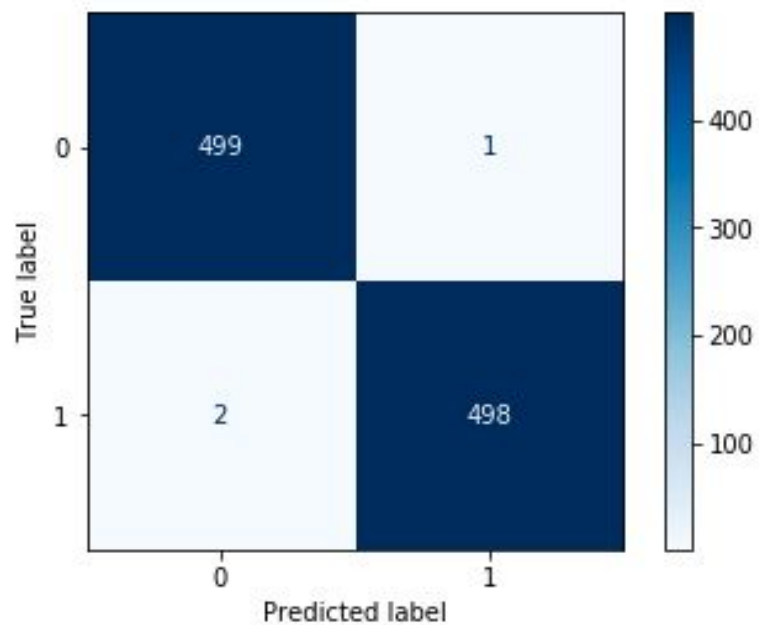


## 3 Random Forest

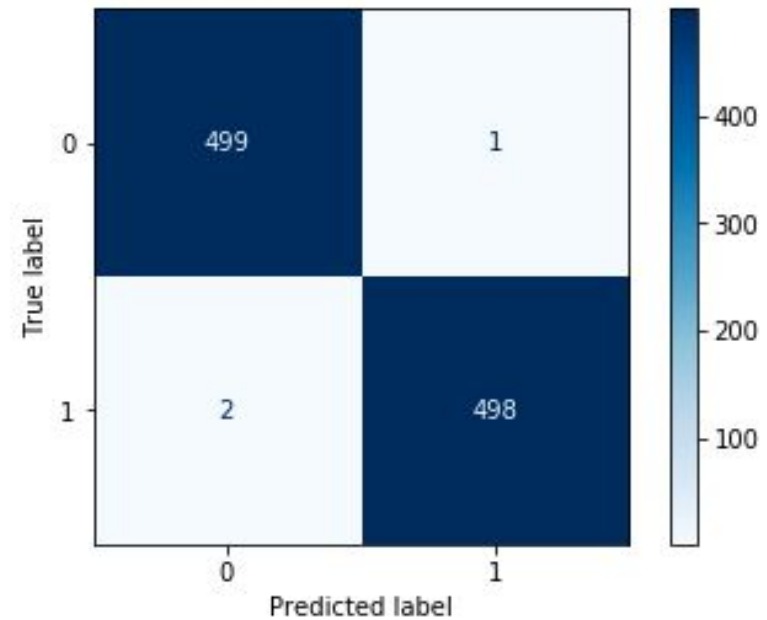


# Training Model: Metrics (cont'd)

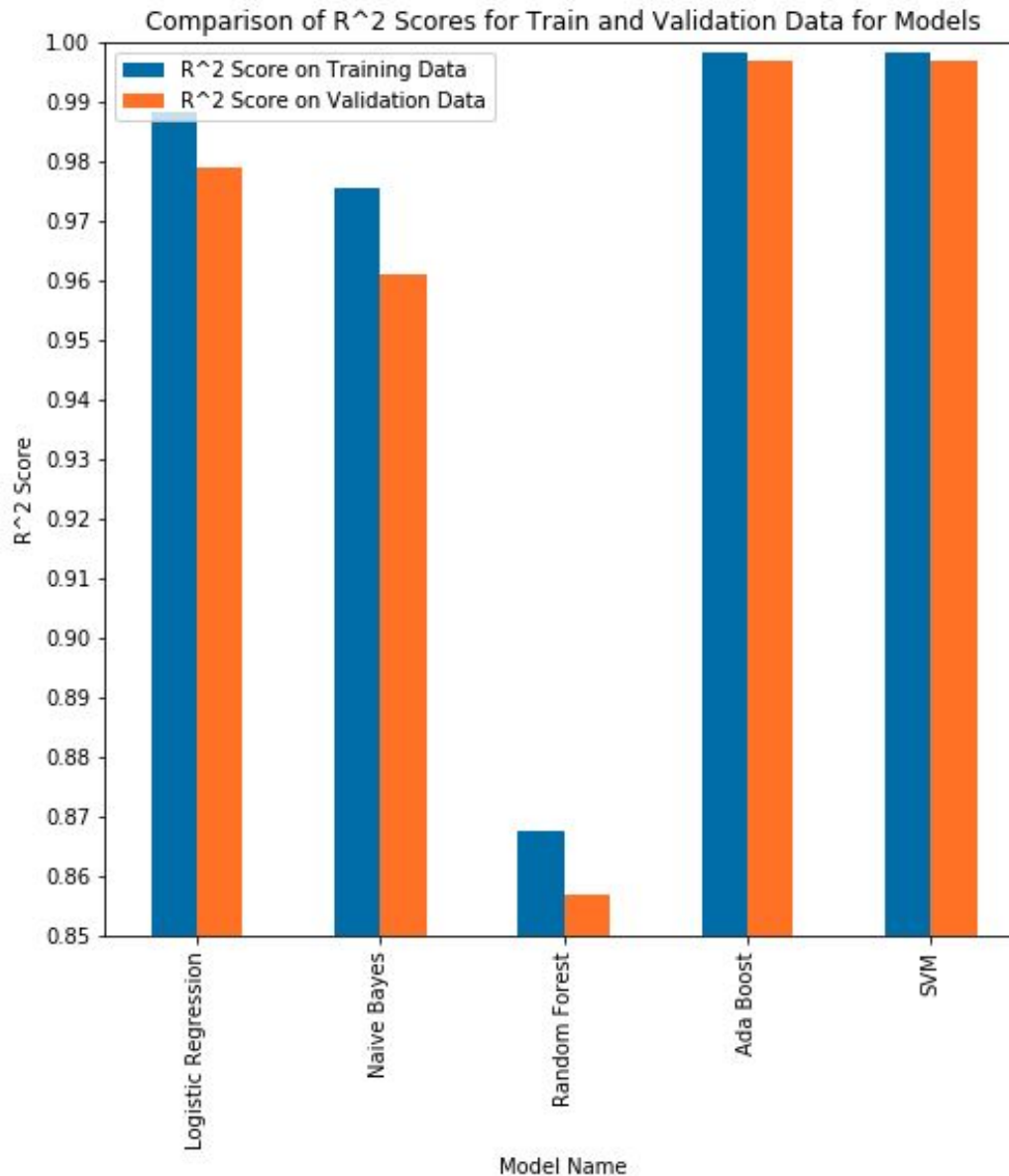
## 4 AdaBoost



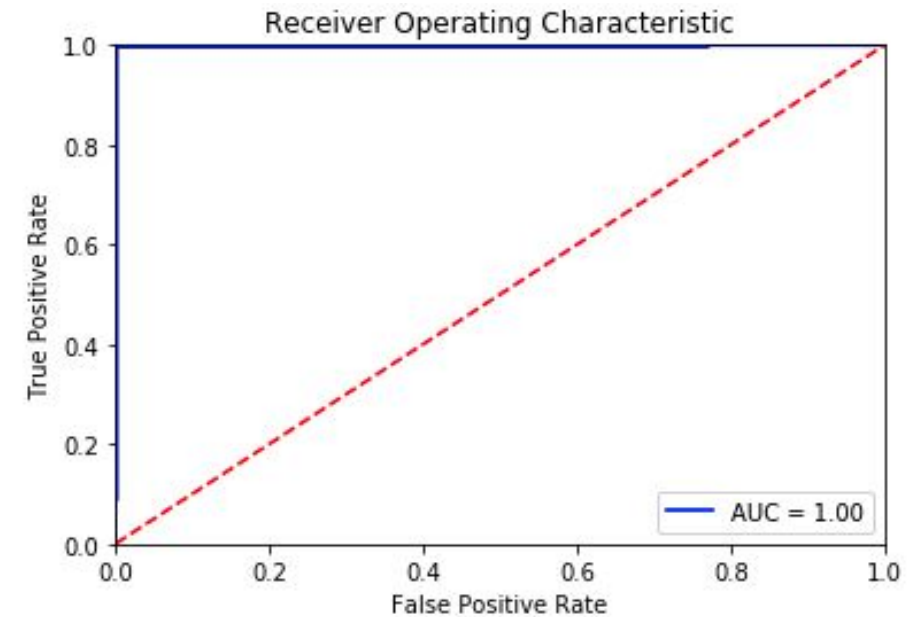
## 5 SVM



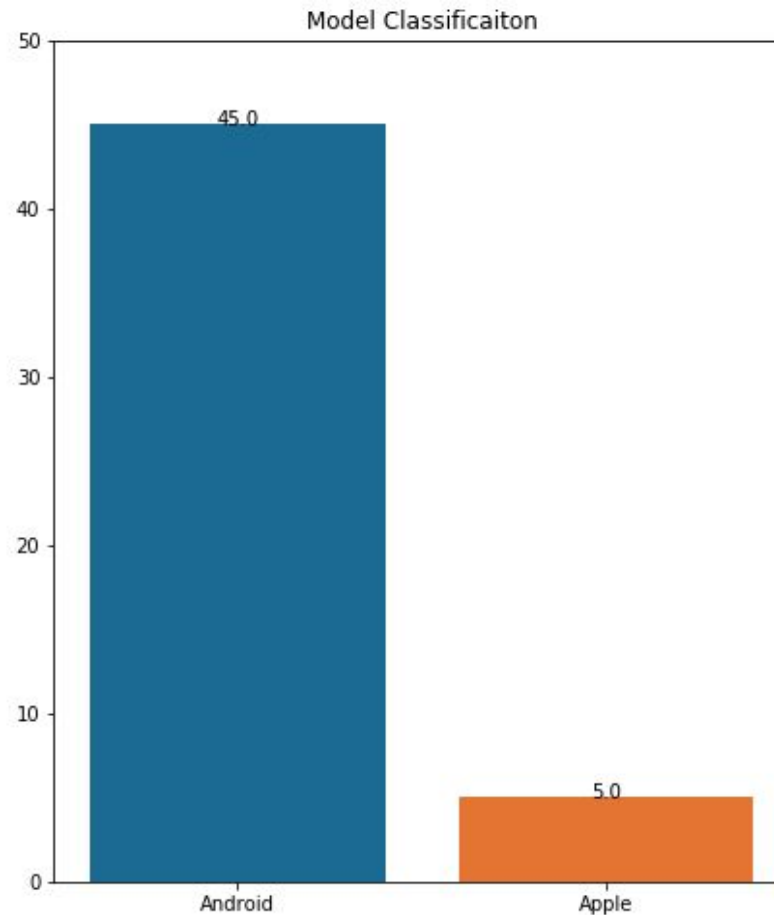
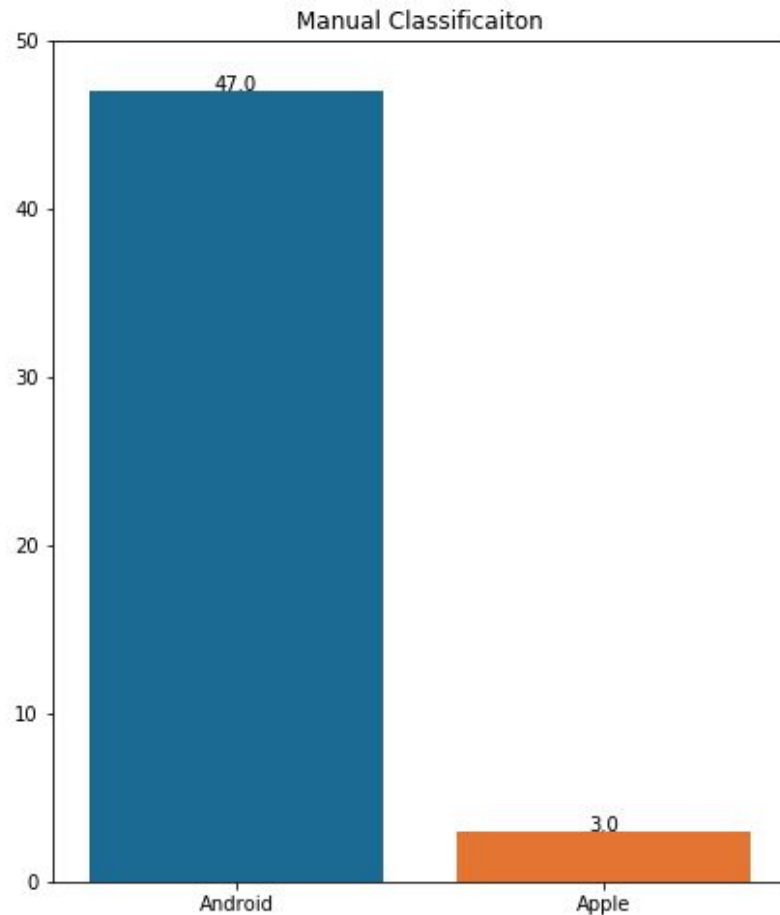
# Training Model: Selecting Model



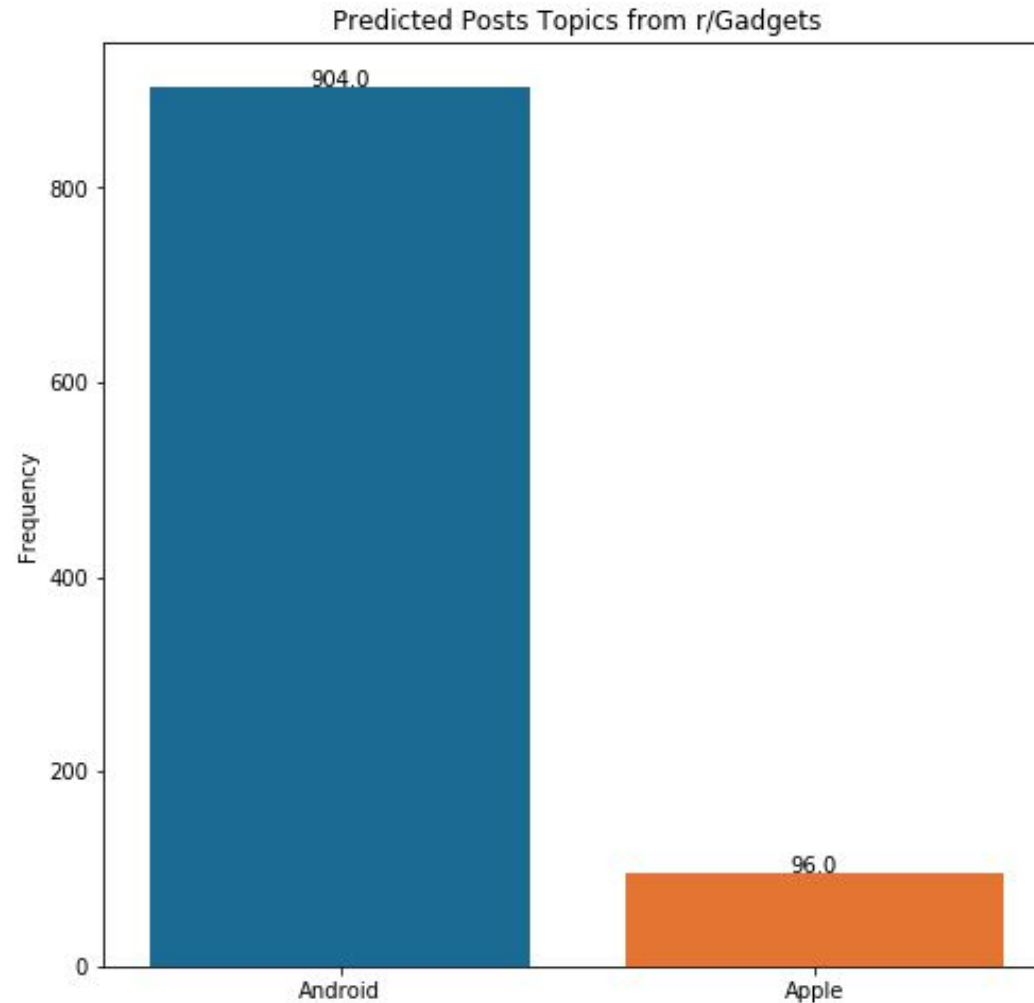
ROC for SVM Model:



# Testing Model: r/gadgets Subreddit Test50



# Testing Model: r/gadgets Predictions





# Conclusion and Key Findings

- Based on the classification model, the 'Android' OS is significantly discussed more and trending on the r/gadgets subreddit. Since StudioM only has the capacity to develop an app for one OS at this stage, we are going to use this experiment to support our business decision and move forward with creating an app for Android.
- The SVM model and AdaBoost models had almost identical Train and Validation scores of 0.99 for both models, however, SVM had a slightly better cross-validation score and our ROC curve looked near perfect. For this reason, we chose the SVM model as our final classification model.
- Since StudioM will be creating an app for Android, it's important the app does the following things as per suggestions made by [AppSamurai](#):
  - Follows Android design principles
  - Sets up Google Beta testing
  - Social Media Advertising
  - Try and get featured on the r/gadgets or r/Android subreddit pages as these have high volume Android customers.
- Since our classification model was near perfect, it would be beneficial to pull out some of key sentiments like Apple/iOS/Android and then run our model without these keywords and see how well it performs in that situation.

# Questions?