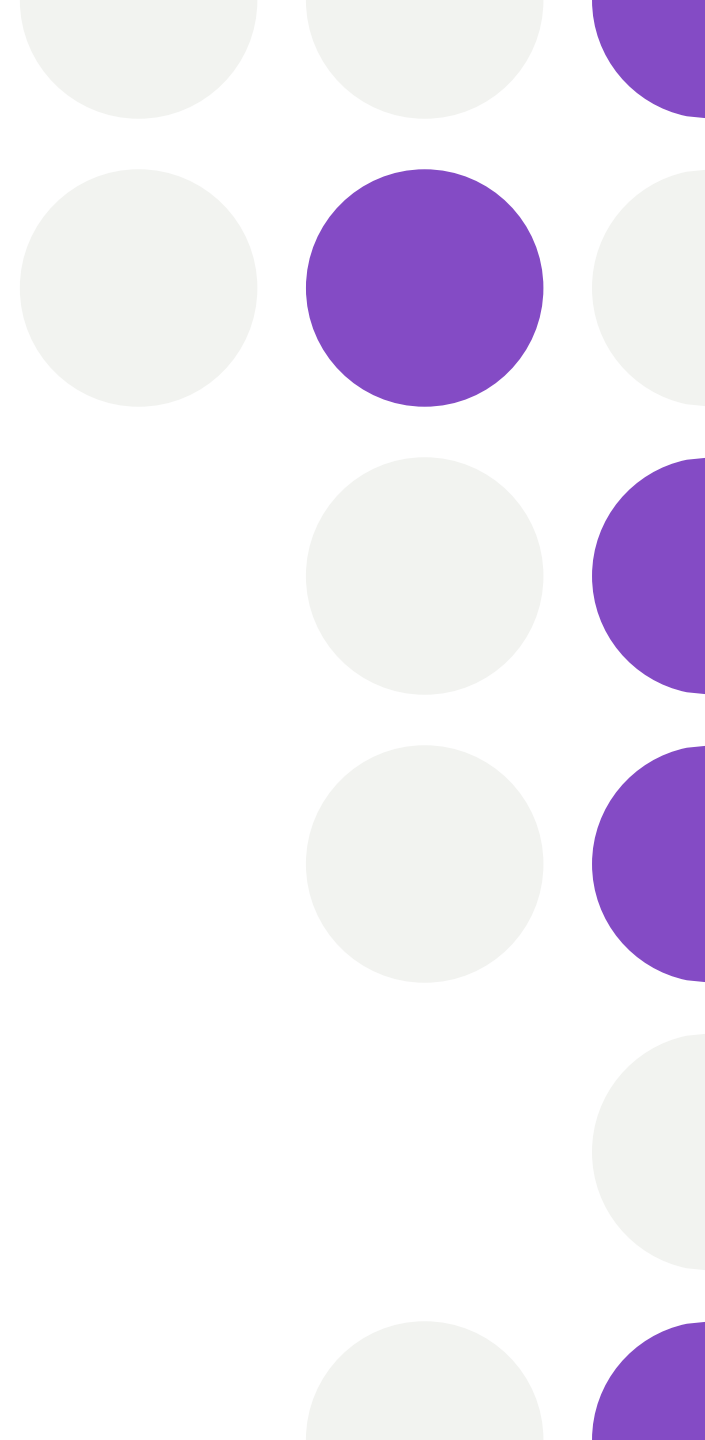
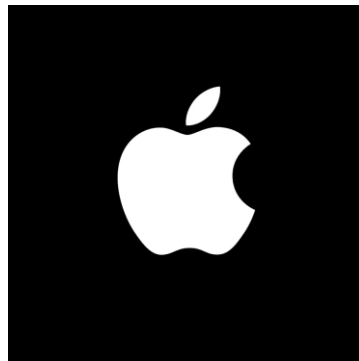


Exploring the Ether: NLP Analysis at SXSW

Faraan Khan

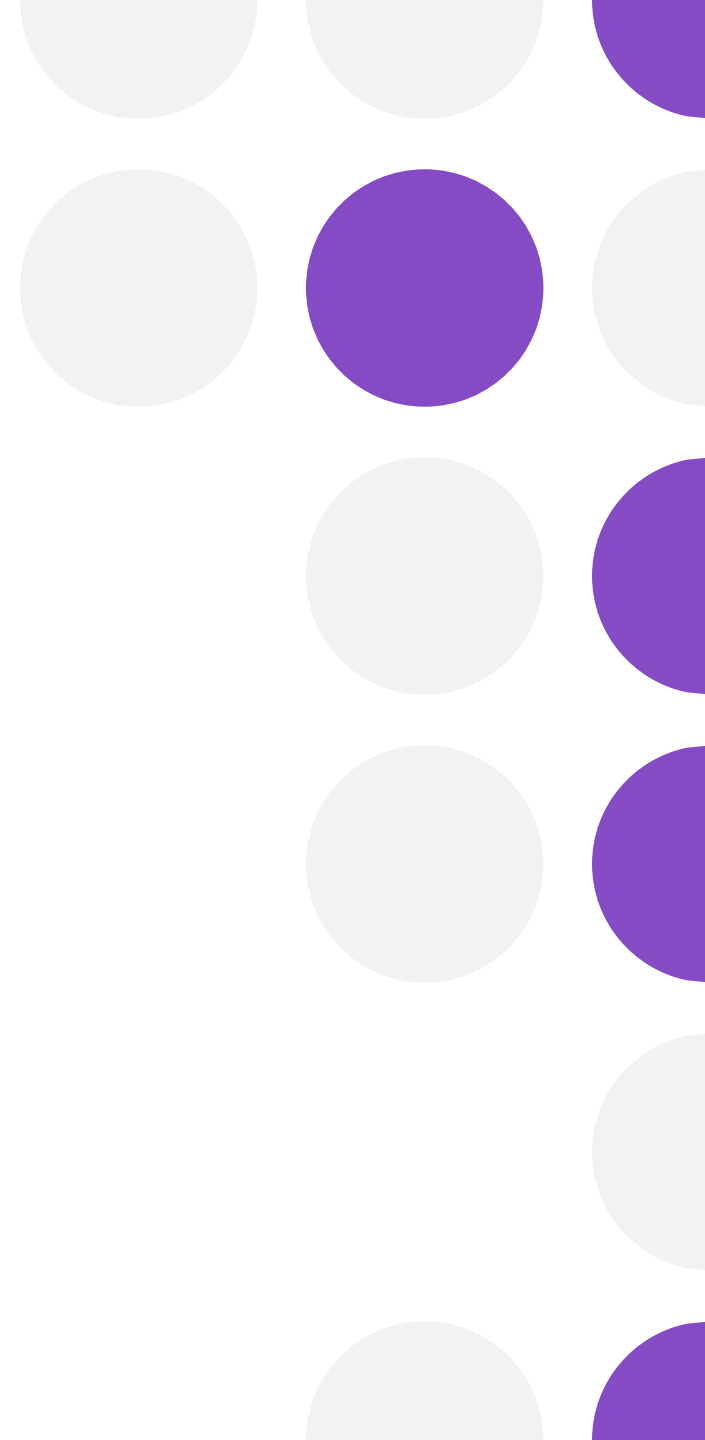
The Value of Brand Equity

Brand equity refers to the value and strength of a brand in the eyes of its customers and in the marketplace. It represents the perception, reputation, and overall influence that a brand has, which goes beyond just the physical products or services it offers.



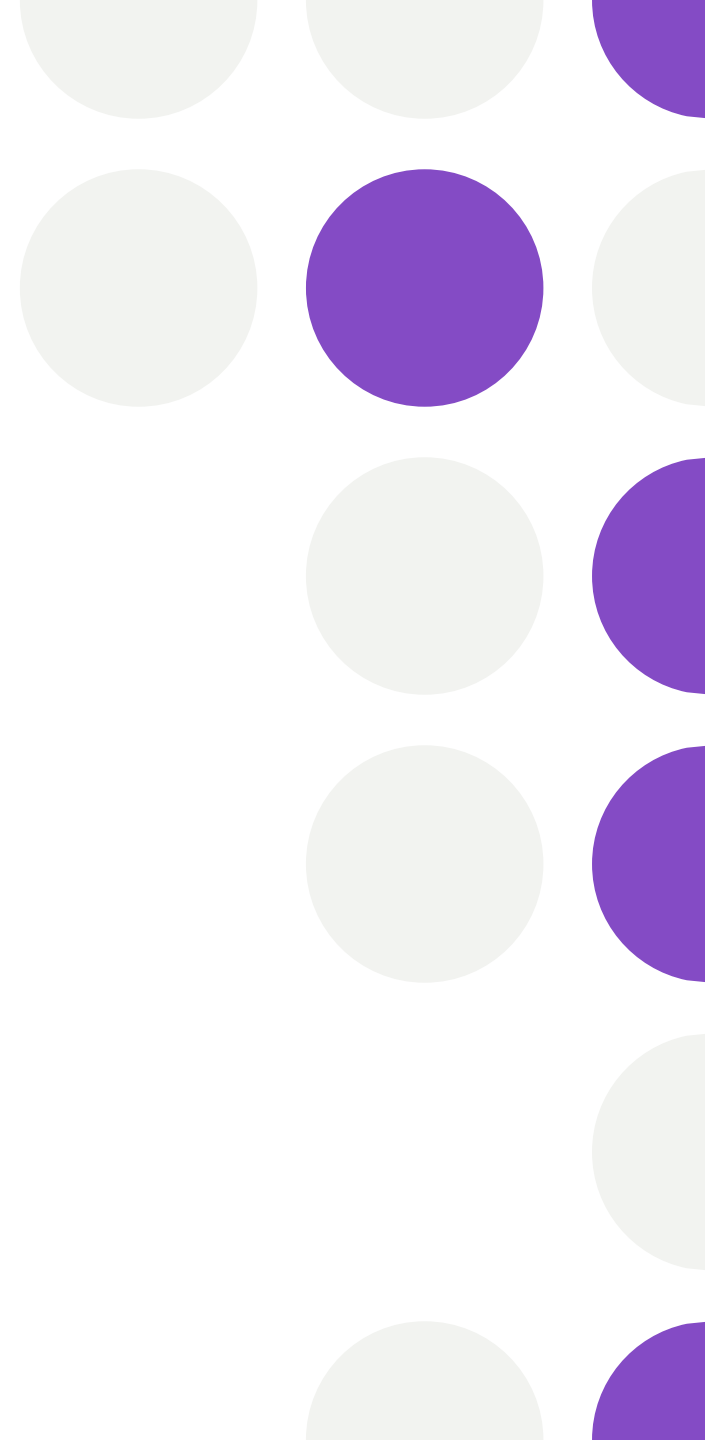
Help Apple at SXSW

- In today's digital world, public perception changes rapidly, and social media sparks it
 - SXSW: annual film, interactive media, music, and tech showcase in Austin, TX
 - Assess tweets about both Google and Apple that occurred at SXSW 2011 and build an NLP model that classifies sentiment
-

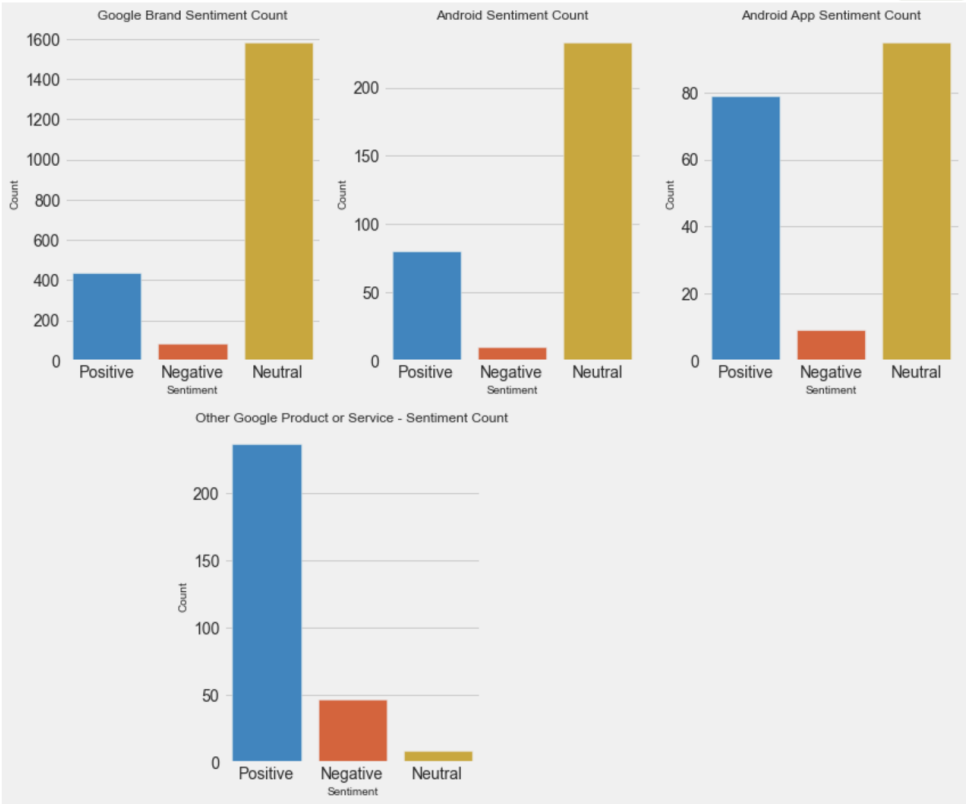
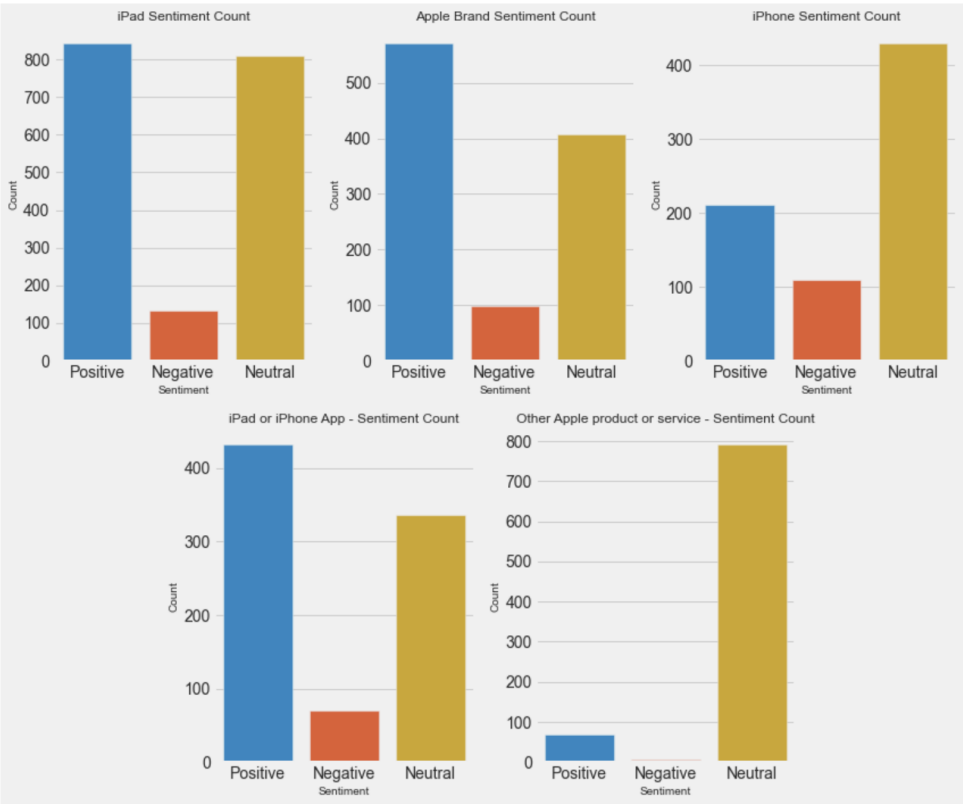


The Data

- After preprocessing and feature engineering:
8294 tweets, 7 columns
 - Neutral: 4759
 - Positive: 2966
 - Negative: 569
 - Combine Neutral and Negative tweets and build a binary classifier – these are the consumers of focus
 - Performance Metric: Precision to minimize false negatives
-

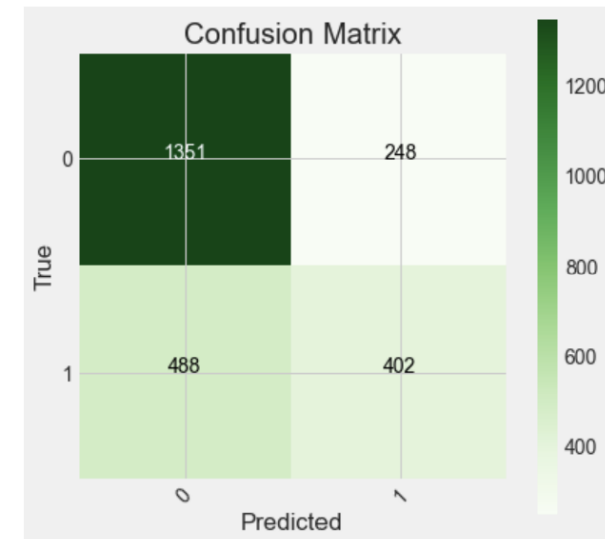


Apple Vs Google



MNB Model

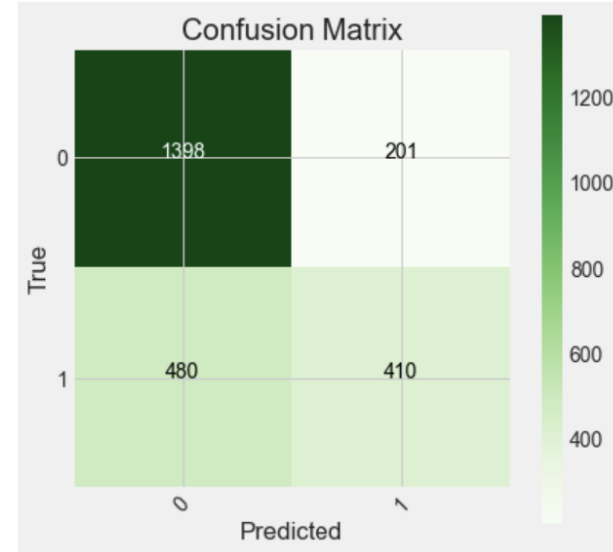
- Ran 3 models; baseline, with **additive smoothing**, and with minority over sampling
- Model with additive smoothing performed best, with a **Precision of .69** and **F1 Score of .69**
- Lowest difference between Training and Testing accuracy, least likely to overfit of the three model types



Training Accuracy: 0.8337639965546942
Cross Validation Accuracy: 0.6801033591731267
Testing Accuracy: 0.7042989152269988
Precision: 0.6930966474598915
Recall: 0.7042989152269988
F1 Score: 0.6915784270490529

Random Forest

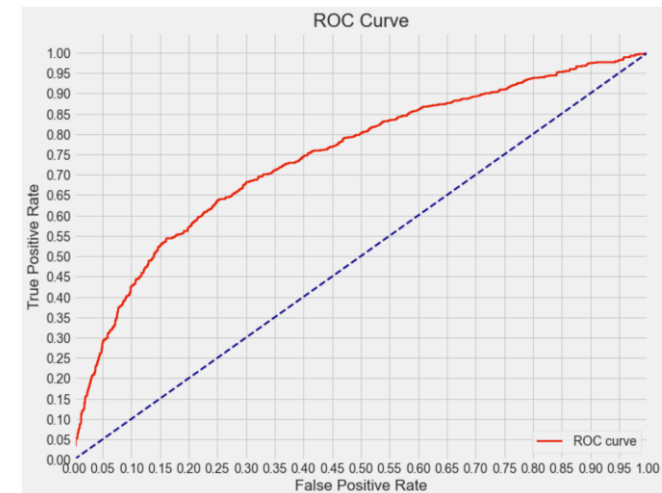
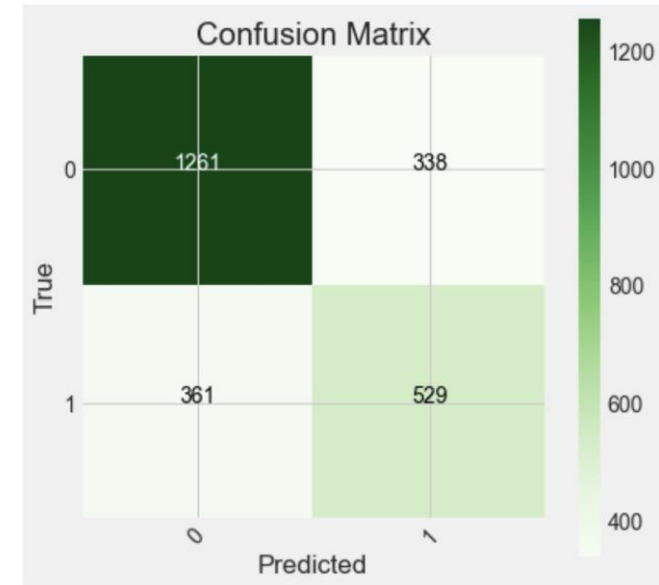
- Ran 2 models – one with the **baseline features**, and one with added features (tweet length and mentions)
- Of the 2, baseline performed best with regard to **Precision (.72)**
- Huge difference between train and test accuracy indicates a likelihood of overfitting



Training Accuracy: 0.965374677002584
Cross Validation Accuracy: 0.6609904769565815
Testing Accuracy: 0.7263961430293291
Precision: 0.7181709843689975
Recall: 0.7263961430293291
F1 Score: 0.7119451403938165

Support Vector Classification

- Ran 2 SVC models, one with linear and one with **rbf kernel**
- Of the 2, rbf performed best with a **precision score of nearly .72**
- Highest **AUC value (.74)**
- Less prone to overfitting than random forest, higher precision performance than MNB



Recommendations/Next Steps

- Deploy the **SVC RBF** model to assess future tweets about Apple – market and engage with Neutral/Negative users
 - Learn from Google – the only area that they dominated positive sentiment was in ‘Other Product or Service’
 - Leverage the brand – consider hosting or sponsoring mission – oriented events
 - Consider developing a multi-class model that can classify “Neutral” users for more targeted marketing strategy
 - There was dissatisfaction with the iPad design, camera, and battery life. Consider surveying users for more feedback
-

Thank you!

Questions?

