

Predicting Subreddit Based on Machine Learning Models

Presented by Farhad Farahani



Introduction

How to predict a subject with models?

- Online forums and social media platforms like Reddit have become valuable sources of information and feedback for businesses
- Machine learning algorithms have emerged as powerful tools for analyzing and classifying text data
- In this project, we present an approach for collecting the data from **Apple Watch** and **Samsung Galaxy Watch** subreddits and making models based on those data
- With that information we have a better understanding of users which can be used to inform our product development, marketing and customer service strategies

Analysis Description

01

Cleaning and EDA

- Including 400,000 raw data
- Clean Data
- Pre-process Data
- Explore Data

02

Analysis

- Transform Data
- Naïve Bayes
- Random Forest
- Models Evaluations

03

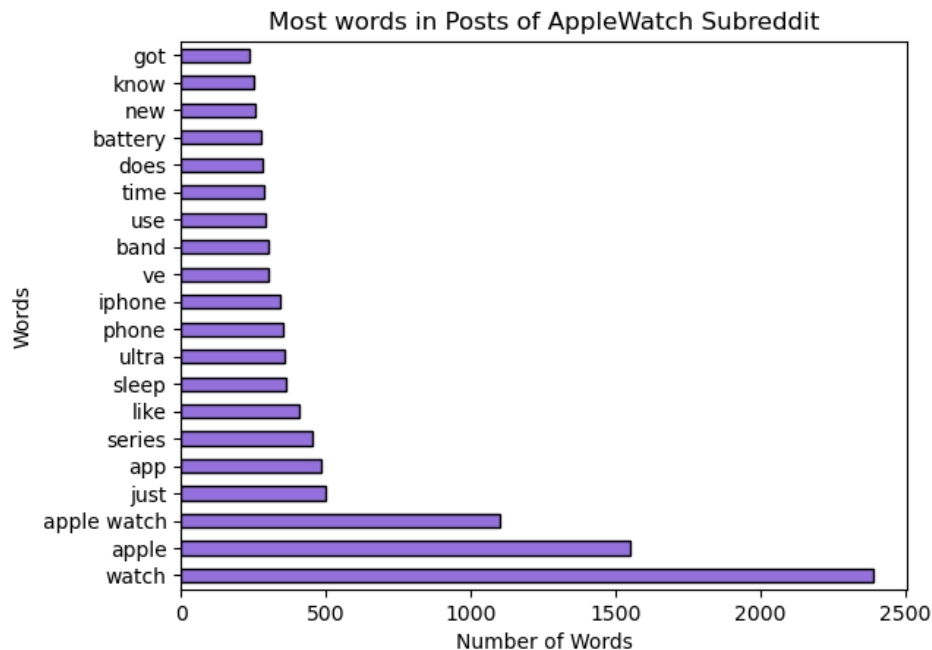
Results

- Most common words
- Confusion Matrix
- Feature Importance

What are the most common words?

In the **AppleWatch subreddit**, people seem to like to talk more about the new series, **Ultra**! And sleep app, bands and battery are other related topics that they're interested in!

In the **GalaxyWatch subreddit**, people are more like to talk about watch faces, battery, updating process and google!



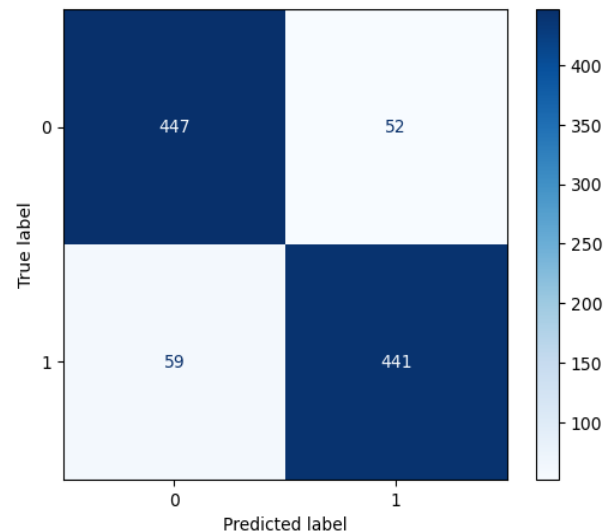
Models

Which one was the best one?

Actually, models (both Random Forest and Naïve Bayes) scores are somehow the **same**, but I chose Random Forest with TfidfVectorizer!

Evaluation Scores	
Accuracy	%89
Precision	%89
Sensitivity	%88
Specificity	%89

Confusion Matrix of Random Forest with TfidfVectorizer

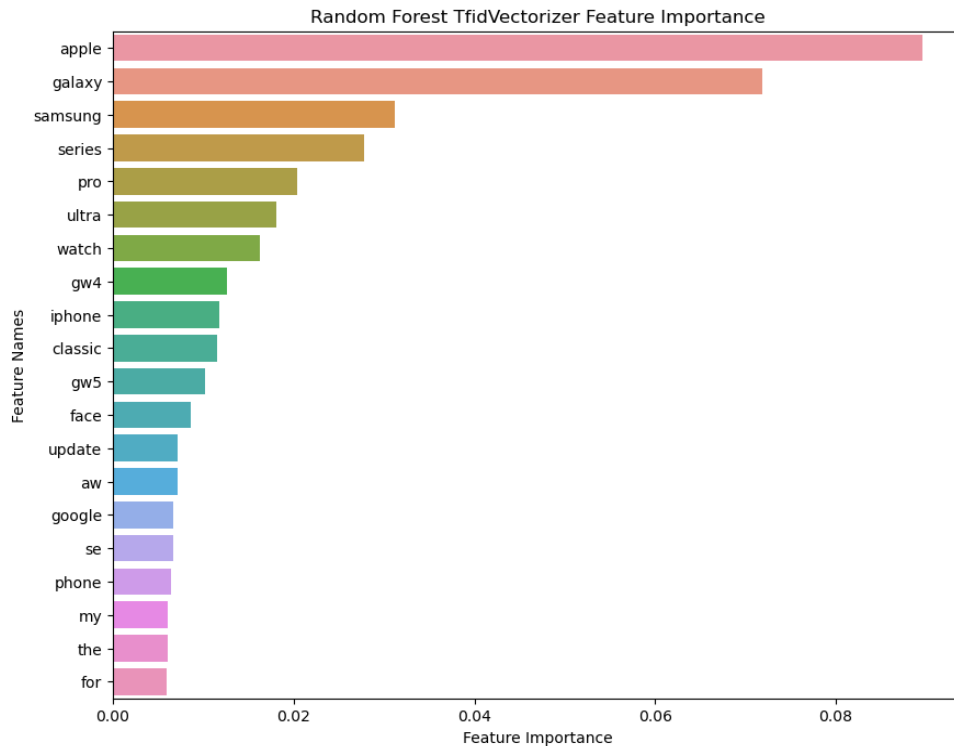


Feature Importance

Words that play a major role in making our models based on our combined dataset:

Apple, galaxy, gw4, gw 5, se, ultra and face!

Watch series and obviously, brands are the most important ones!



Conclusions and Recommendations

- We found that the Random Forest Algorithm with TfidfVectorizer Transformer performed the best, achieving an accuracy of % 89!
- Besides those common words in each subreddit, we also found which words in the combined dataset of those subreddits, played a main role in making our model.
- Use insights gained from our analysis to inform our customer service strategies. By understanding the most common concerns, we can provide better support to our customers.
- Consider the analysis of other platforms related to our products and services (like rate plans, connectivity coverage quality and phone devices). to gain a more comprehensive understanding of customer needs and preferences.



THANKS!

Do you have any questions?

Momenifarahani.farhad@gmail.com



CREDITS: This presentation template was created by **Slidesgo**, and includes icons by **Flaticon** and infographics & images by **Freepik**