SkinBeauty

Classifying Reddit Posts:
Skincare vs Makeup

Presented by Ariana Bina

Problem Statement
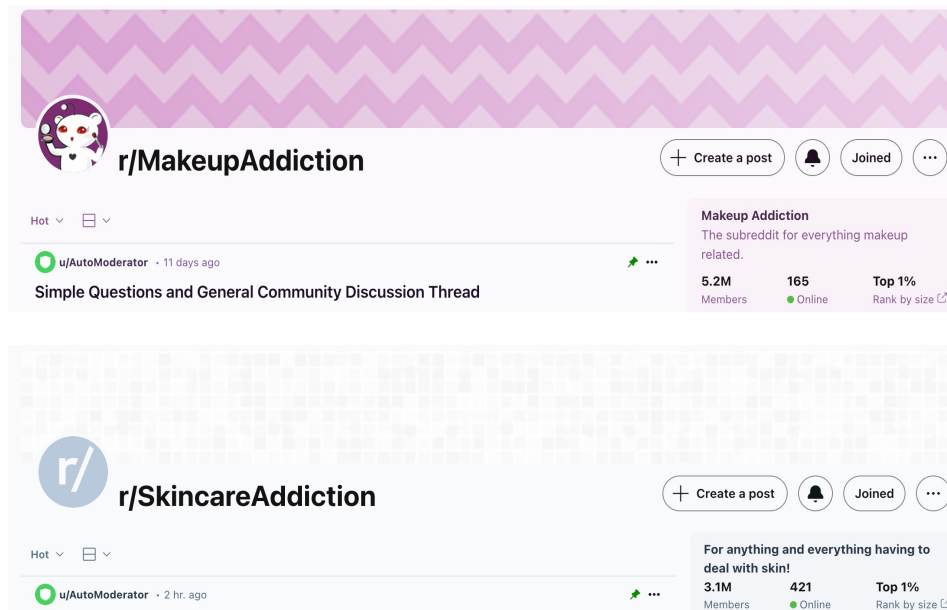
❑ <u>Objective</u>

    o Classify reddit posts based on **titles**

    o **/r/SkincareAddiction**

    o **/r/MakeupAddiction**

❑ <u>Goal</u>

    o Optimize SkinBeauty's marketing efforts

    o Gain insight into the distinct interests of those seeking skincare vs makeup advice.

    o Tailor product offerings to better serve these two distinct customer segments.
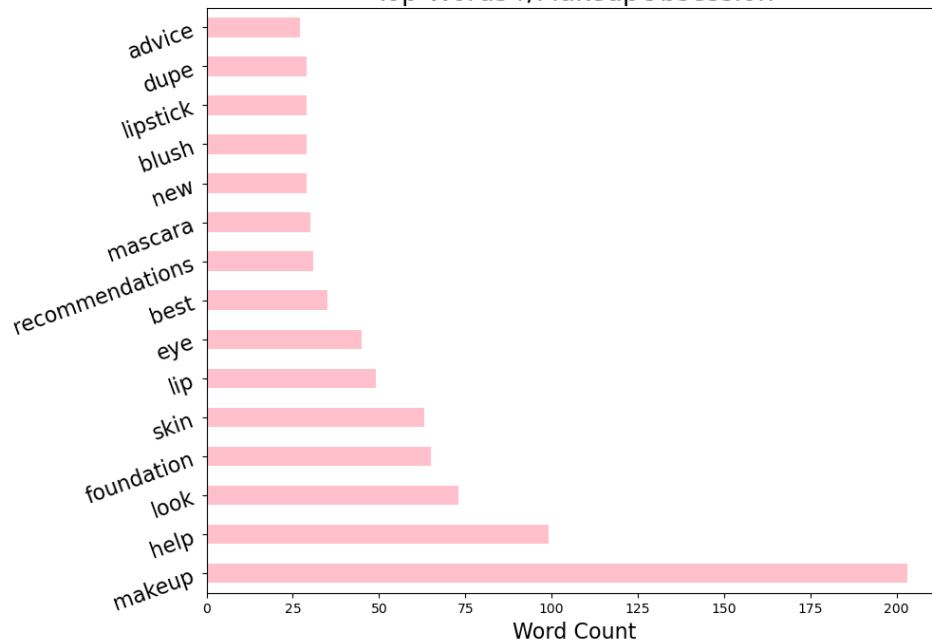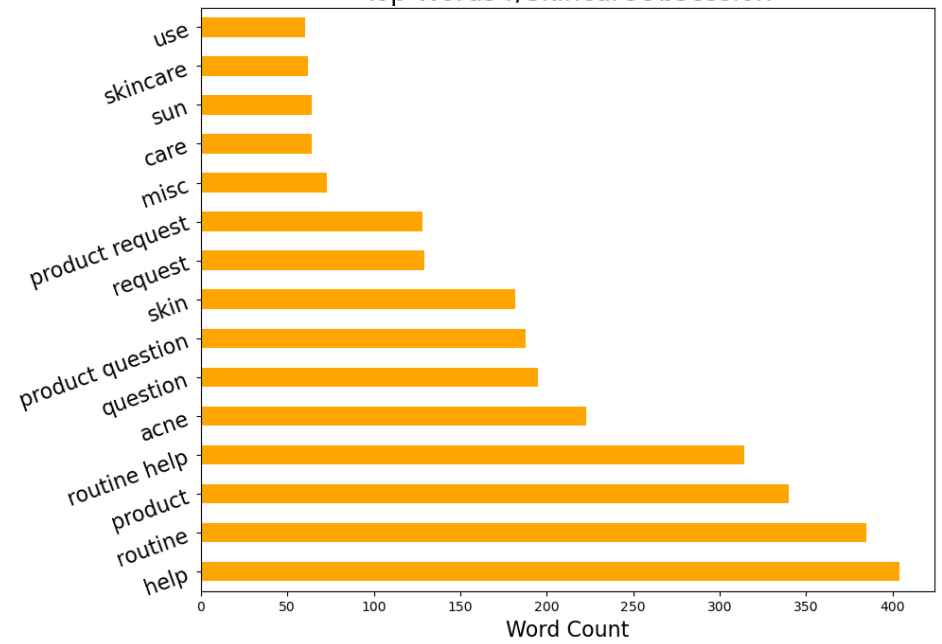
Data Collection

❑Utilized PRAW
❑Total:1,912 posts

15 Most Common Words

Top Words r/MakeupObsession

Top Words r/SkincareObsession

Model 1: Logistic Regression + Count Vectorizer

Best Parameters

- cvec__max_df: 0.9

- cvec__max_features: None

- cvec__min_df: 2

- cvec__ngram_range: (1, 2)

- cvec__stop_words: english

Accuracy	97.39%
Recall	97.41%
Precision	97.35%
F1 Score	0.97

Model 2: Random Forest

Best Parameters

- classifier__max_depth: None

- classifier__min_samples_split: 10

- classifier__n_estimators: 300

- vectorizer__ngram_range: (1, 2)

Accuracy	**97.1%**
Recall	97.73%
Precision	97.64%
F1 Score	0.97

Recommendations

- ❑ Logistic Regression Model

- ❑ Performed the best on the test data, had the highest accuracy and is easily interpretable

- ❑ Less computationally intensive