

Identifying sentiments in tweets

Overview

- Project objective
- The data
- Method and approach
- The final model
- Results and implications

Business Problem

Objective: **Categorize** tweets by **sentiment** expressed

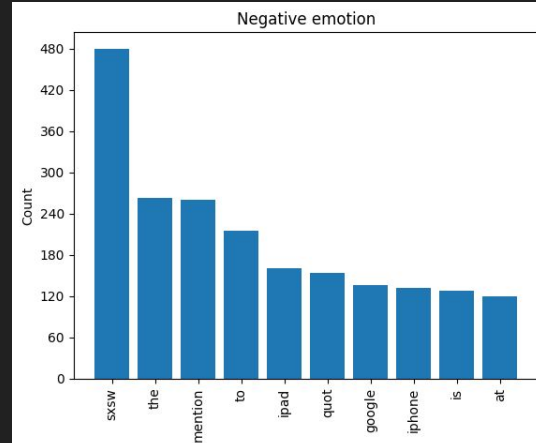
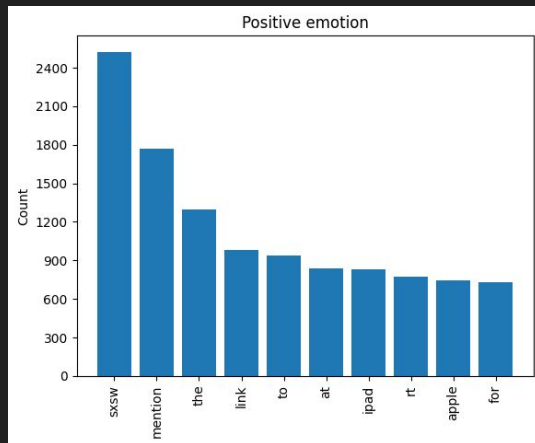
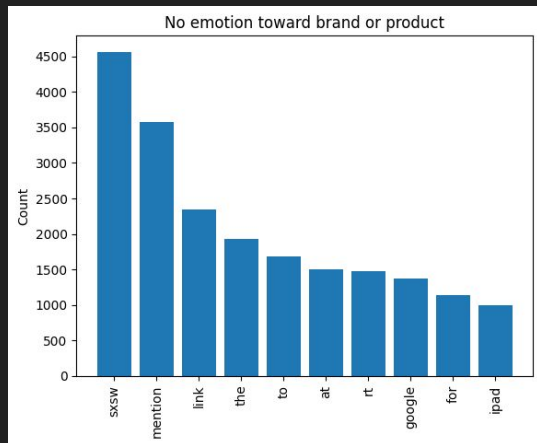
Stakeholder: Twitter's advertising team and prospective clients

- Help the advertising team identify positive and negative tweets
- Help inform clients what **twitter users like** and **dislike** about their products
- Clients can **refine** their **advertisements** accordingly

Data

- Almost **10,000** real tweets about various products
- **Contents** of the tweet and **sentiment** expressed (positive, neutral, negative)

Distribution of most frequently used words in each category:



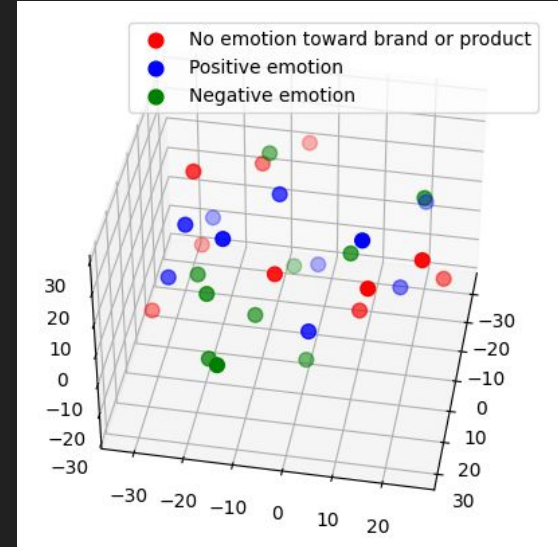
Modeling method

Support Vector Machine

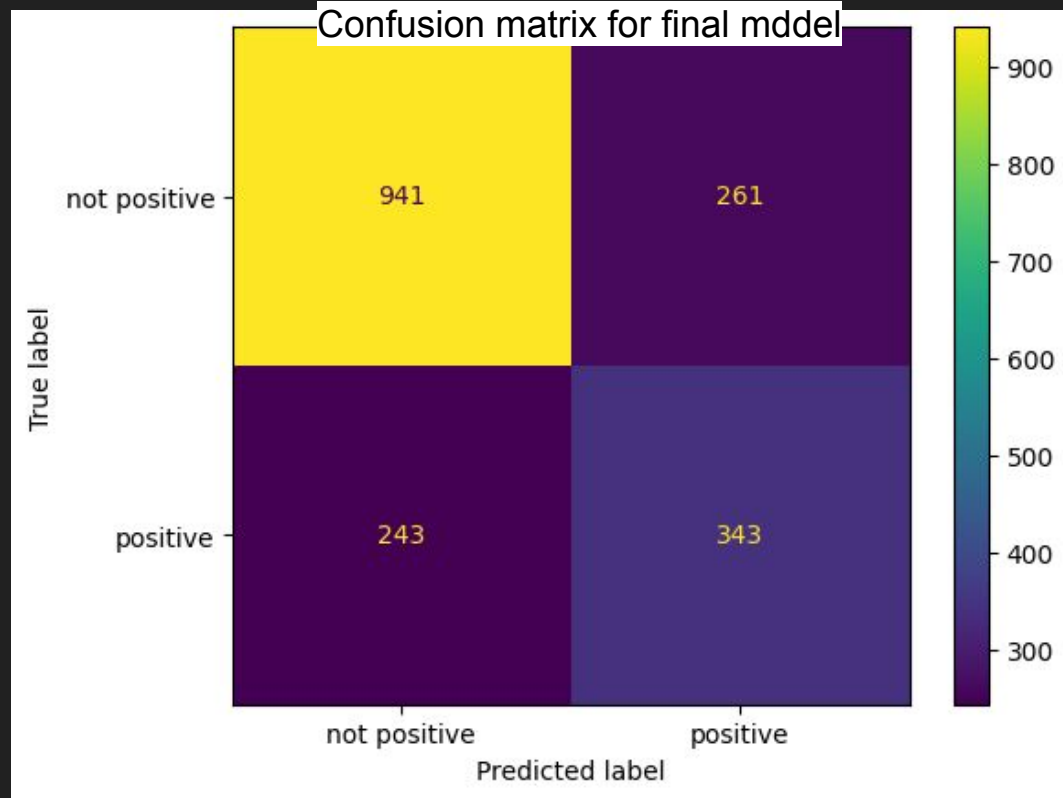
- Maps data into multidimensional space
- Find the **best line of separation** between categories

Binary classification

- Sorts tweets into **positive** and **not positive**
- Prioritizes identifying positive tweets correctly



❖ ~ 72% accuracy, ~60% of positive tweets identified



Recommendations

- ❖ Model to use: binary classification with support vector machine
- ❖ Analyse tweets identified as **positive** to find points to **include** in promotion
- ❖ Analyse tweets identified as **non-positive** to find points to **exclude** or **improve** upon
- ❖ **Further filtering** is recommended

Next Steps

- ❑ Obtain more data points about negative tweets
- ❑ Add more features to train the model
- ❑ Expand into multiclass classifier

Thank you for listening!

Beatrix Wong

Email:

beatrix.wmh@hotmail.com