

NLP: Sentiment Analysis

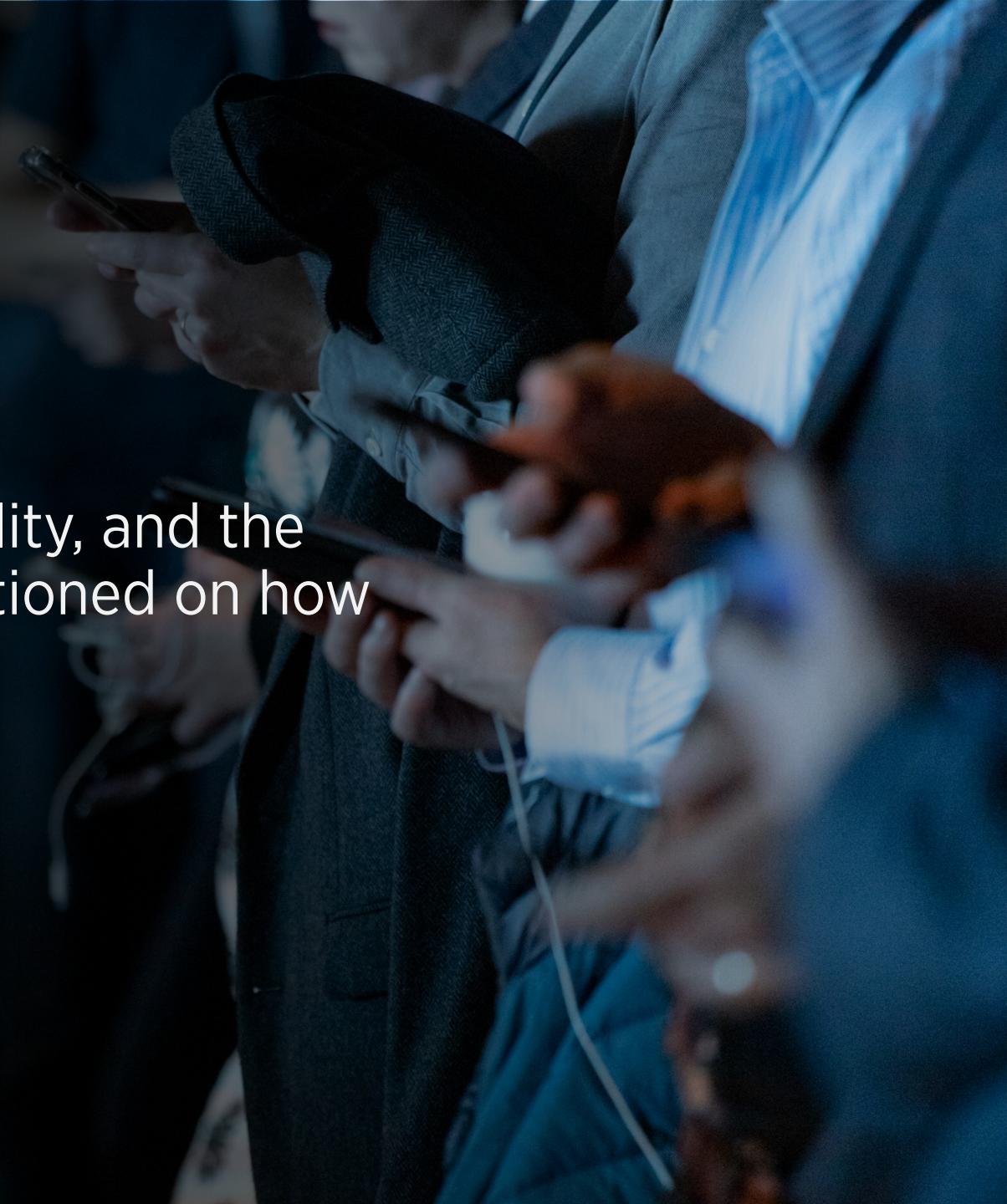
Analyzing Product Sentiments from Tweets

Sung Bae

The Importance of Communication

“Our beliefs and perceptions of reality, and the choices we make, are largely conditioned on how others see and evaluate the world”

- Bing Liu , University of Illinois, Chicago



Twitter

145 million daily users

22% of Americans are on Twitter

500 million tweets sent each day

65.8% of US companies use Twitter for marketing

80% of Twitter users have mentioned a brand in a tweet

77% of Twitter users feel more positive when their tweet
has been replied to.

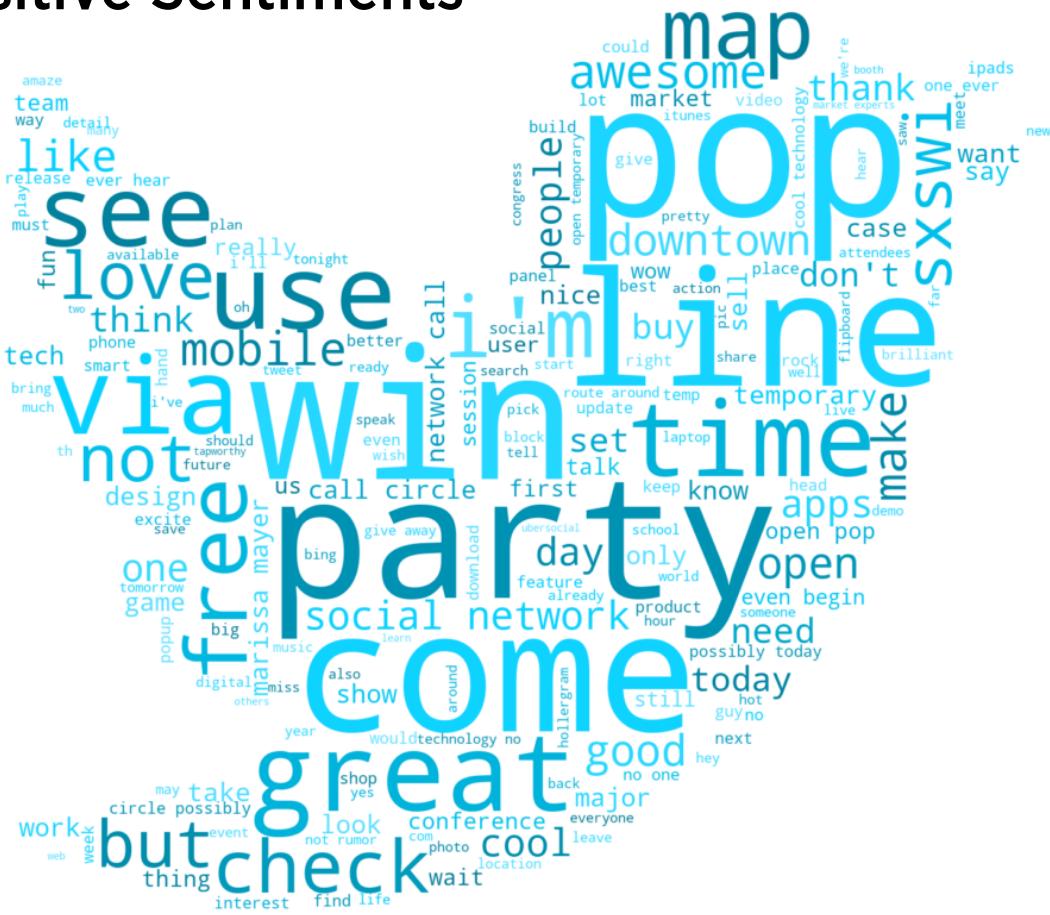
Goals

Create a model that can

- [1] Flag the company when there is a negative tweet about their specific product
- [2] Correctly classify tweets into negative, neutral, and positive sentiments
- [Bonus] Correctly identify which product

EDA: Positive and Negative Sentiments

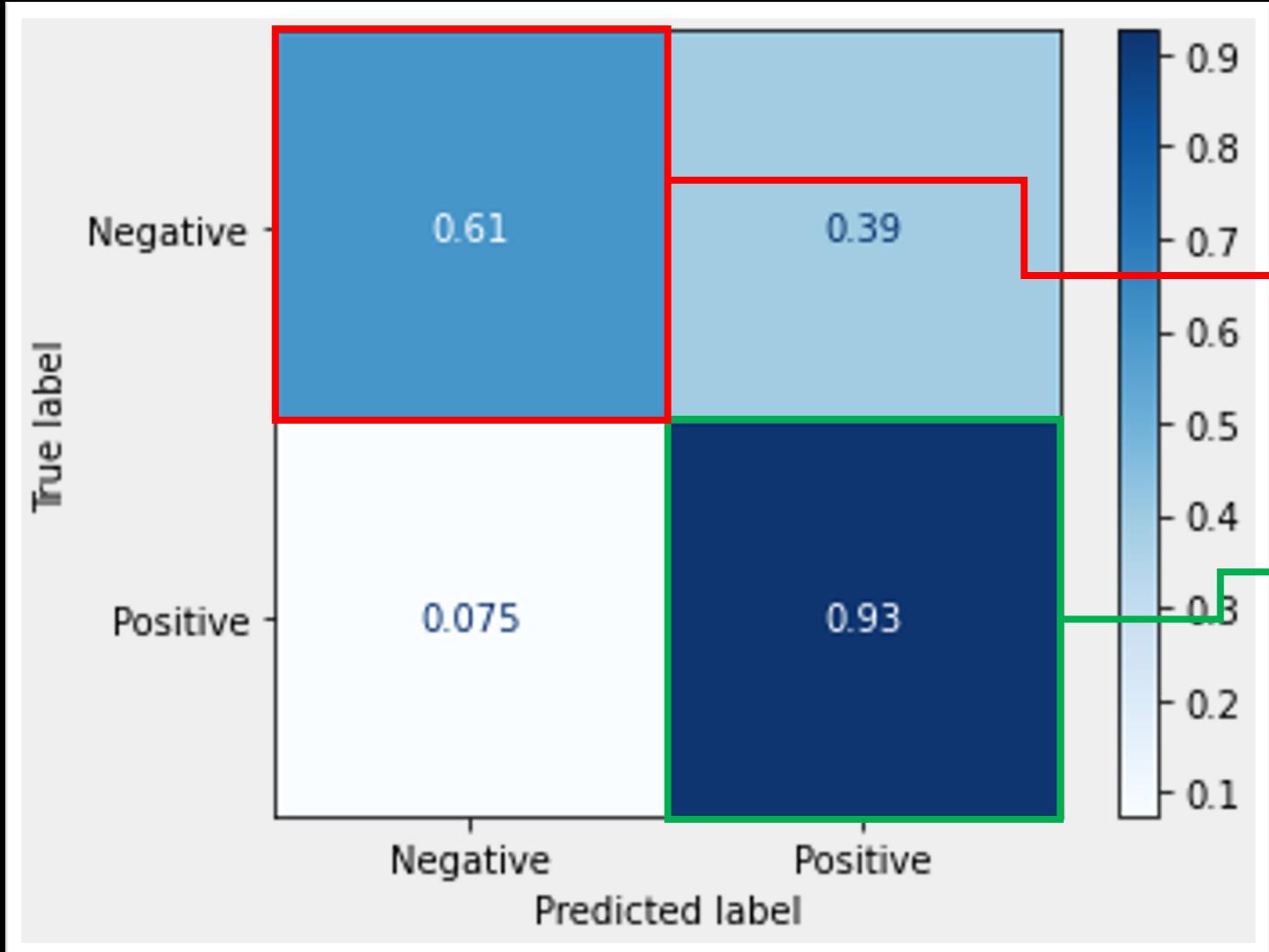
Positive Sentiments



Negative Sentiments



Model: Best Accuracy (SVM_nltk_tfidf)



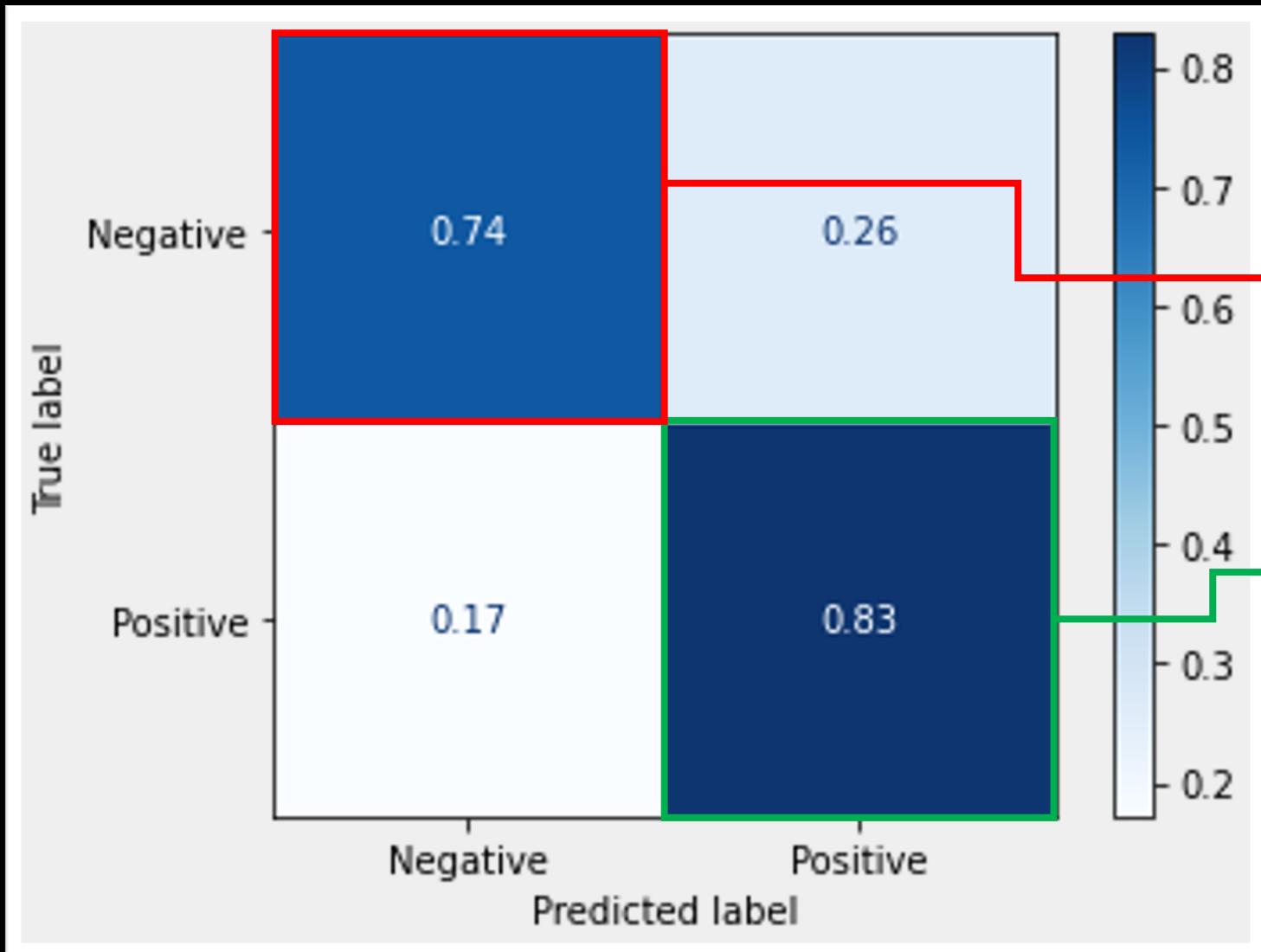
Overall Accuracy = 88 %

Negative Recall = 61 %
61% of negative sentiment tweets were correctly identified

Positive Recall = 93 %
93% of positive sentiment tweets were correctly identified

Reminder:
Negative recall is our priority!

Model: Best Negative Recall (LR_spacy_we)



Overall Accuracy = 81 %

Negative Recall = 74 %

74% of negative sentiment tweets were correctly identified

Positive Recall = 83 %

83% of positive sentiment tweets were correctly identified

Note:

Accuracy isn't everything!

Results: Binary Models Comparison

label	negative_recall	positive_recall	test_accuracy	average_time
LR_spacy_we	0.737226	0.828000	0.813980	0.018581
NB_nltk_tfidf_smote	0.708029	0.830667	0.811725	0.000474
LR_nltk_tfidf_smote	0.642336	0.893333	0.854566	0.000508
SVM_nltk_tfidf_smote	0.605839	0.925333	0.875986	0.000522
NB_spacy_tfidf_smote	0.540146	0.597333	0.588501	0.011910

Best Negative Recall:

[LR_spacy_we]

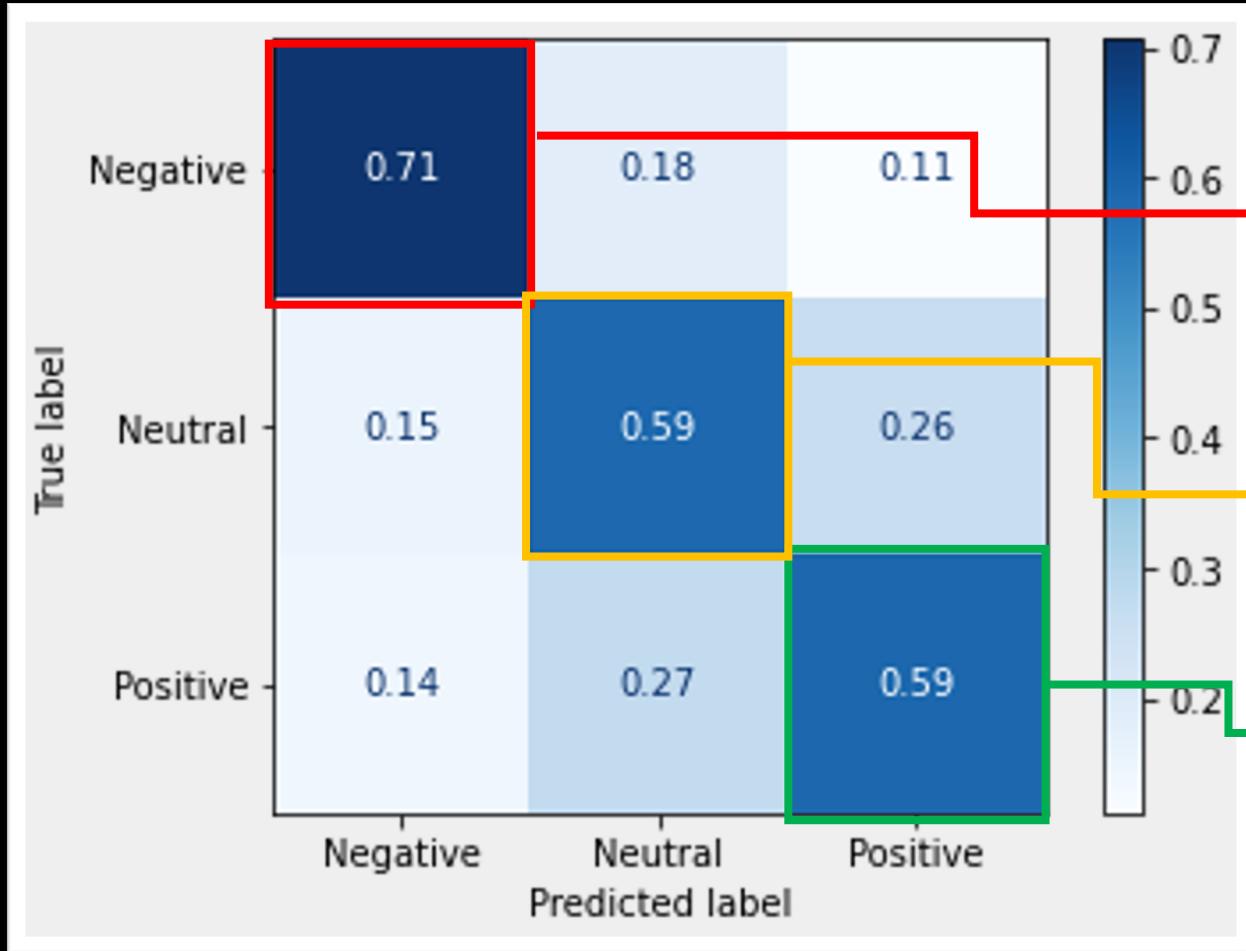
Logistic Regression - Word embedding

Best Computing Time:

[NB_nltk_tfidf_smote]

Naïve-Bayes' – TF-IDF with SMOTE

Model: Multiclass Model



Overall Accuracy = 60 %

71% of negative sentiment tweets were correctly identified

59% of neutral sentiment tweets were correctly identified

59% of positive sentiment tweets were correctly identified

Conclusion:
The result becomes too messy.

Conclusion

Binary Models

- Best Negative Recall:
 - 81% accuracy with 74% negative sentiment recall
- Best Accuracy:
 - 88% accuracy with 61% negative sentiment recall
- Fastest:
 - 81% accuracy with 71% negative sentiment recall

Multiclass Sentiment

- 60% accuracy with 71% negative sentiment recall

Recommendations 1 – Binary Sentiments

- Multiclass not only complicates the model itself, but also it dilutes negative and positive sentiments together, therefore, we recommend using binary sentiment classification model

Recommendations 2

Best Negative Identifier

Model LR_spacy_we

logistic regression using word-embedding mode

Pros:

- Best Negative Recall (74%)
- Good Overall Accuracy (81%)

Cons:

- Takes longer time (40x longer)

Recommendations 3

Fastest Identifier

Model NB_nltk_tfidf_smote

Naïve Bayes' using TF-IDF model with SMOTE

Pros:

- Best Computing Time
- Good Negative Recall (70.8%)

Cons:

- Uses SMOTE - Synthetically generated data
- *Further studies need to be done for consistency

Future Direction

[1] Further data acquisition:

- Due to imbalance in data, our model suffered. Acquiring similar datasets with more balanced classes would increase our model's performance and reliability.

[2] Real time Twitter analysis

- Have this model refined and put in a production so that it can monitor tweets real time and analyze and negative flags about any particular product.

[3] Research various costs between models

- Different models have different strengths and weaknesses, and we believe researching different costs including computing and labor costs using different models would allow us to choose a better model

[4] Try different NLP techniques

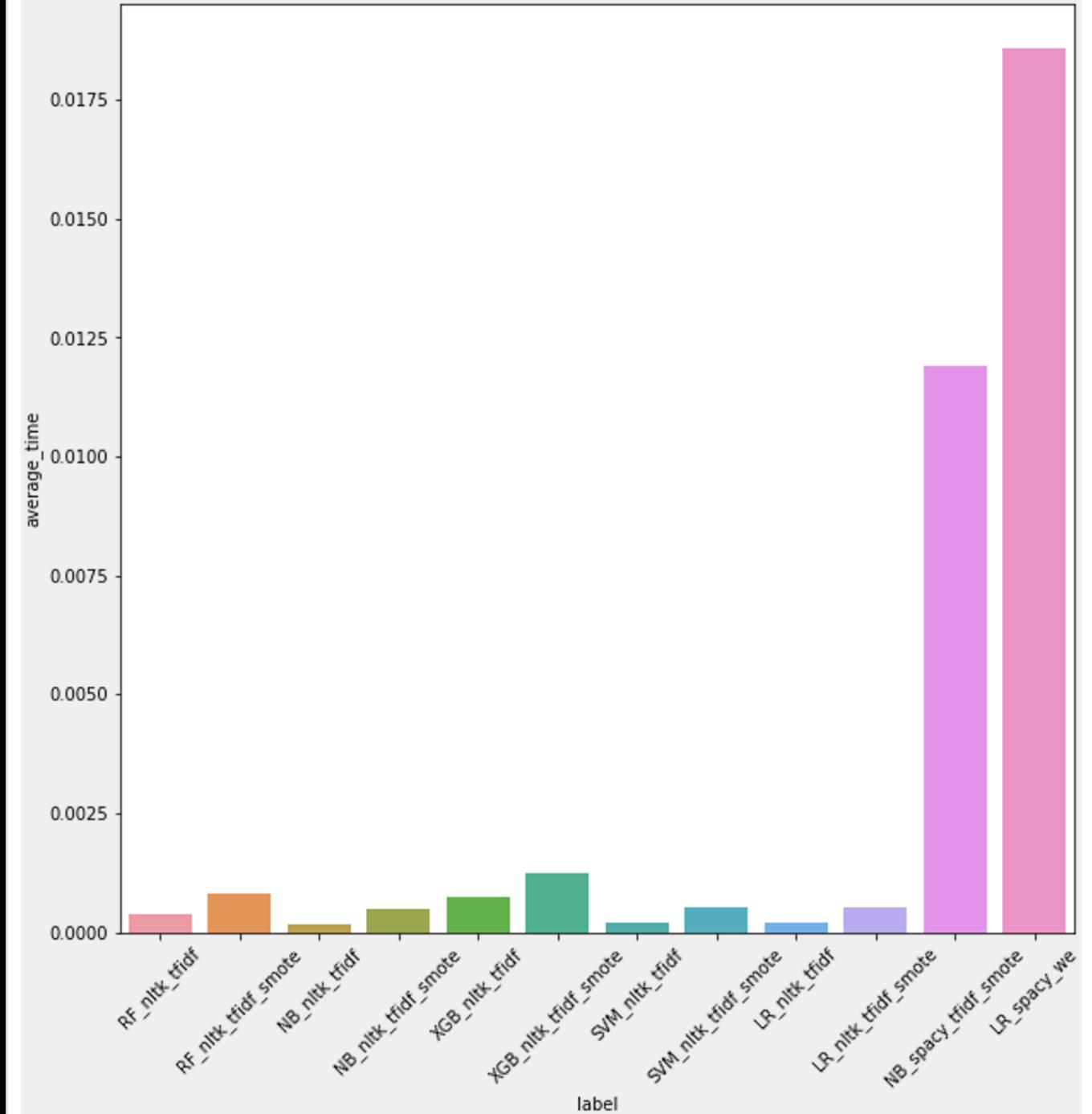
- There are so many different NLP techniques and they are growing due to works of so many different research entities. It would be very valuable to try different approach including skip-gram and even deep learning NLP for this project.

Thank you for listening

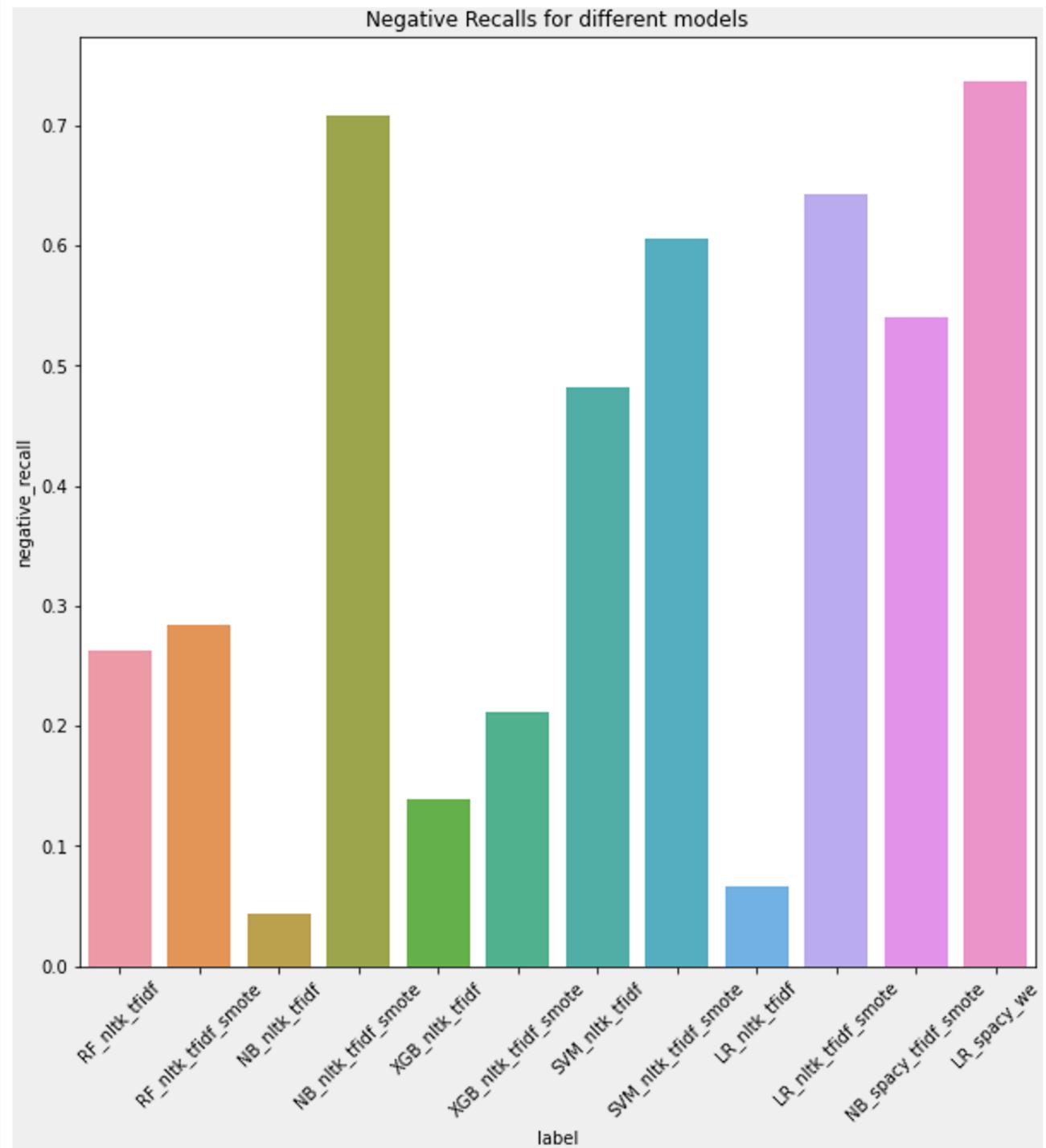
Appendix

Average Computing Time

Average Computing time for different models



Negative Sentiment Recalls



Model Results: Product Predictor

Overall Accuracy = 91 %

Comment:
Products under same brand are confused by the model

Example.
Android and Android App

