Making better Ratings

Creating a “True” Reviews Rating with NLP

By Ivan Zakharchuk

Why Reviews Are Important?

For businesses

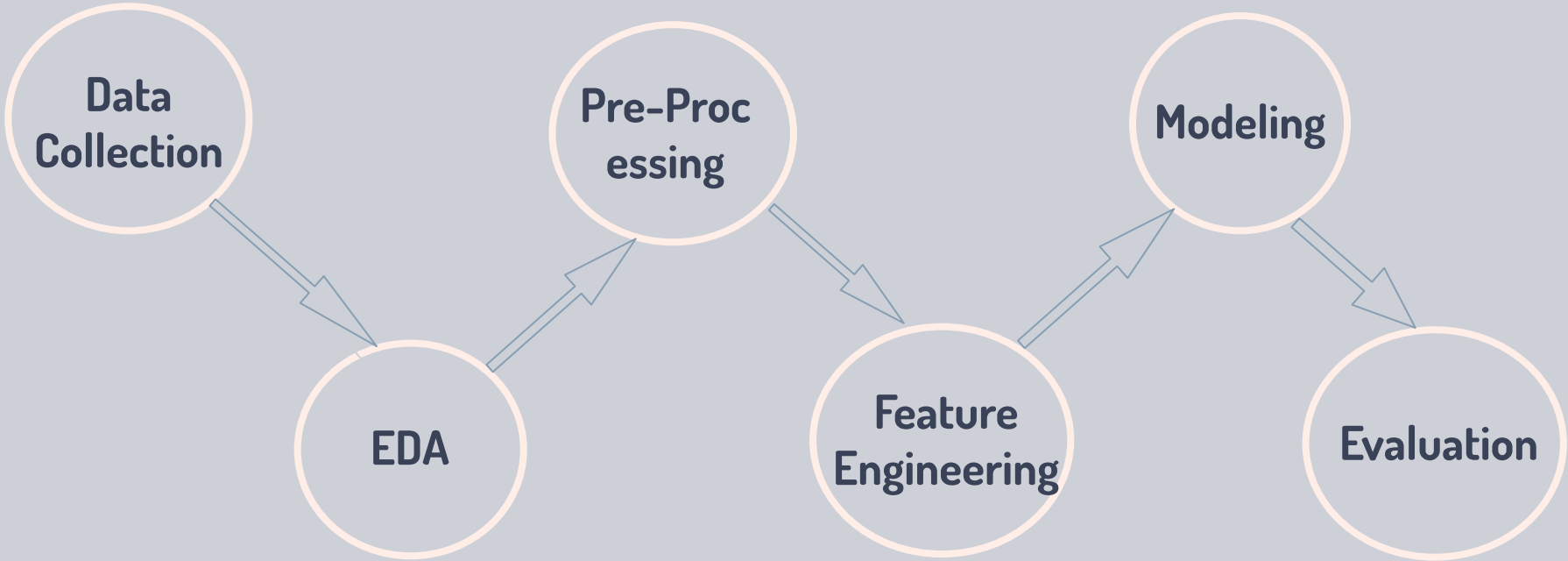
- helps customers decide on business services
- proves the credibility of businesses
- helps identify areas for businesses to improve in

For customers

- shows that business is real
- proves the extent in which a business is invested in maintaining its reputation
- validates purchasing decision

As customers, we use ratings and reviews to get a better idea about businesses or services, but oftentimes the ratings that people leave differ from the depth of the reviews they post. That's why it is important to have a classification system that can correctly rate and review.

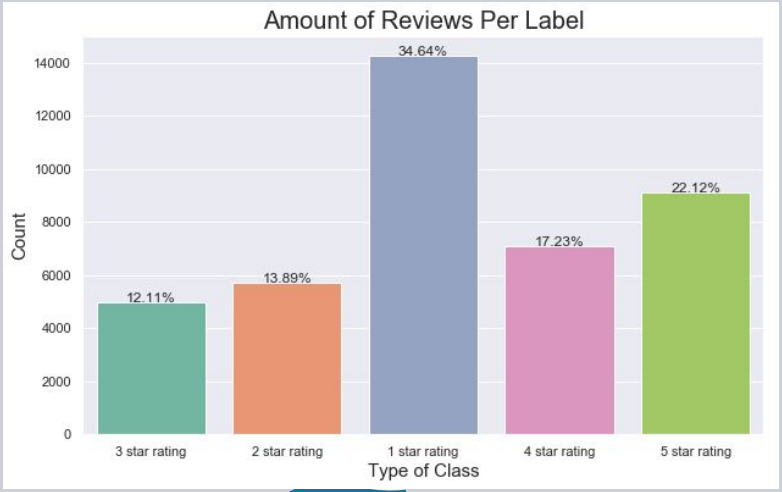
Roadmap


Data Collection
EDA
Pre-Proc essing
Feature Engineering
Modeling
Evaluation

Data

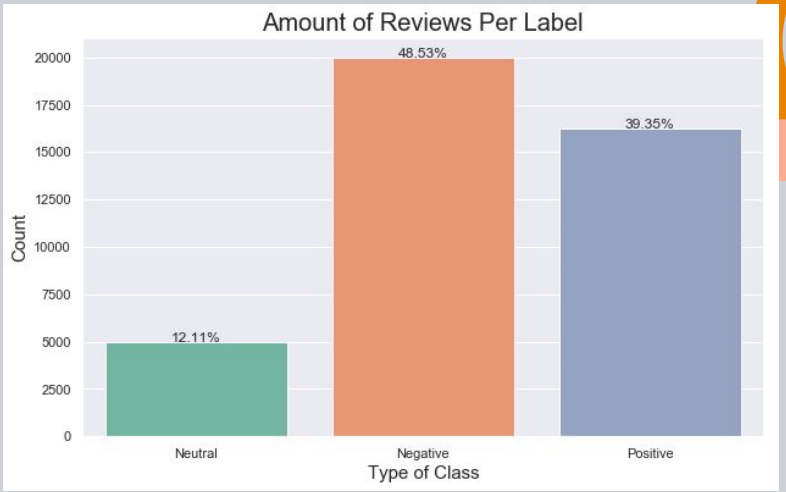
Number of reviews:

More than 41 000 reviews scraped from Yelp source

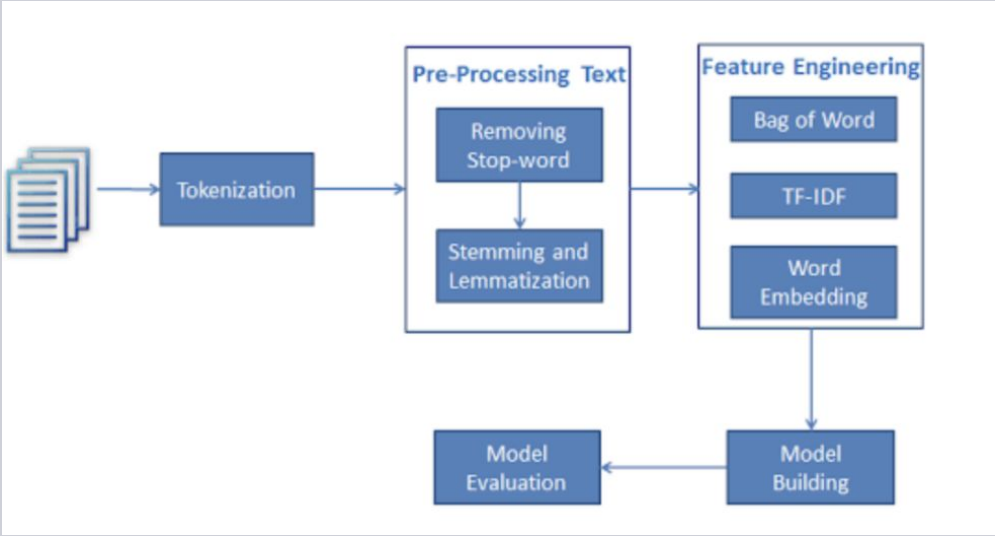


Number of businesses:

More than 1 600 businesses



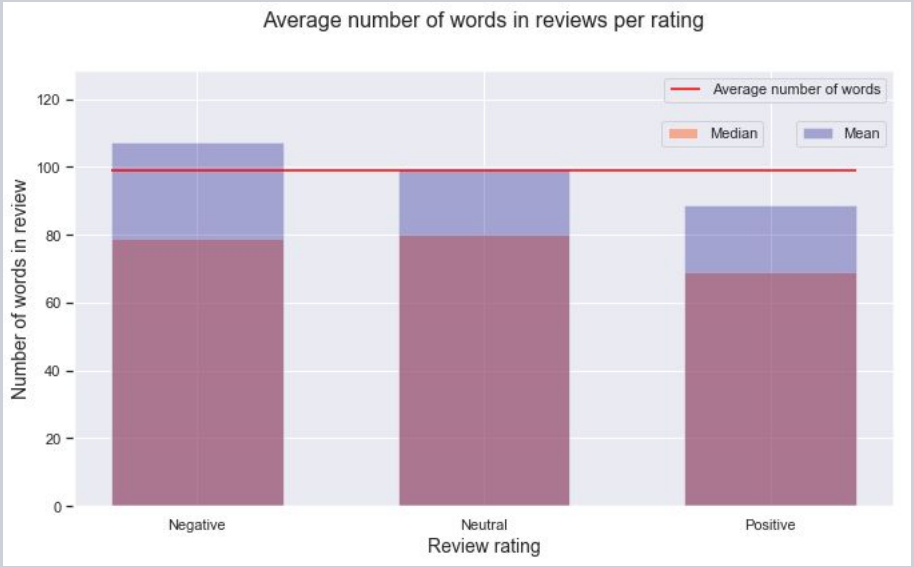
Methodology

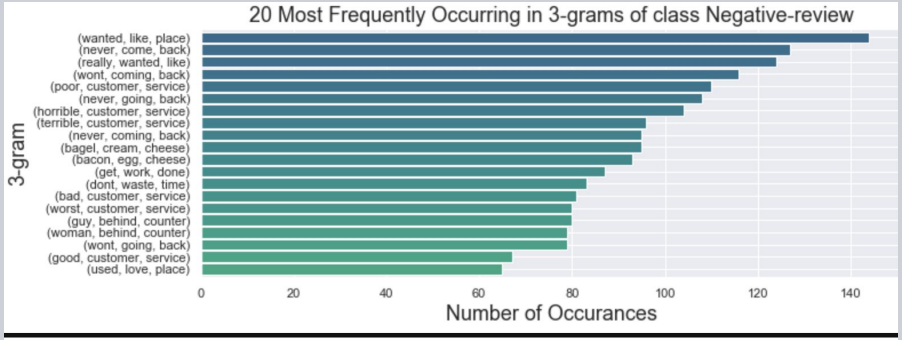


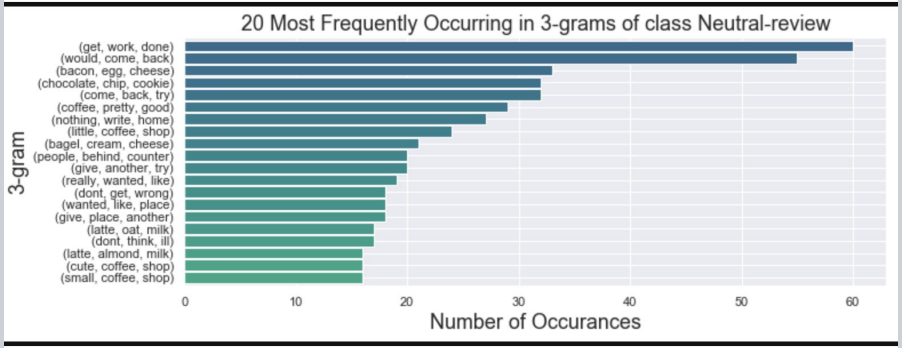
NLP allows computers to interact with text data in a structured and sensible way. With NLP, computers are taught to understand human language, its meaning and sentiments. In order to translate complex natural human language into systematic constructed features, we need to follow some major steps.

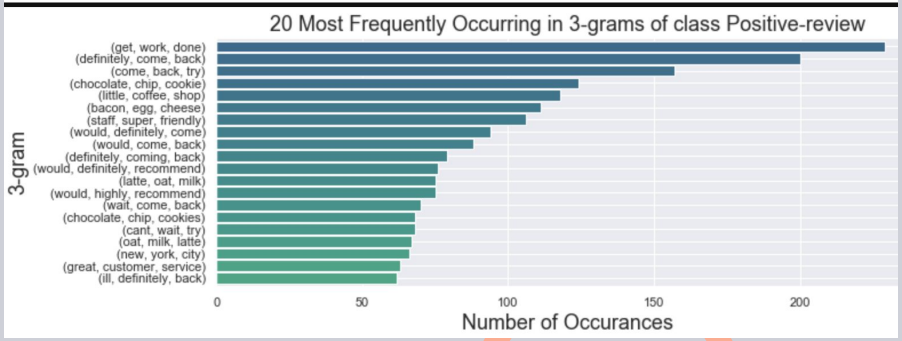
Analysis

Analysing occurrence 3-gram words in each class and length of reviews per class

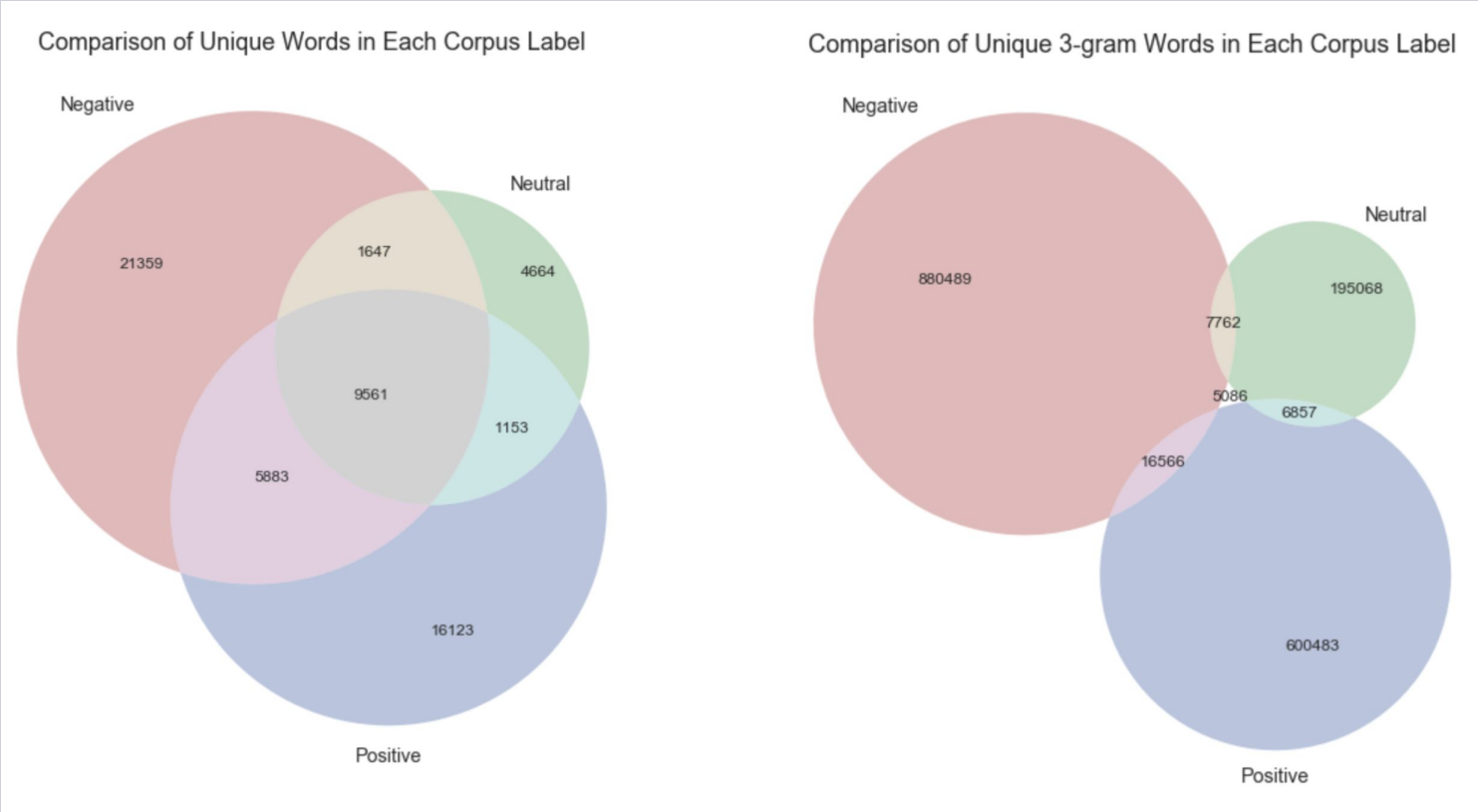








Analysis



Venn Diagrams for Unigrams and 3-gram

Results and Evaluation

Random Forest

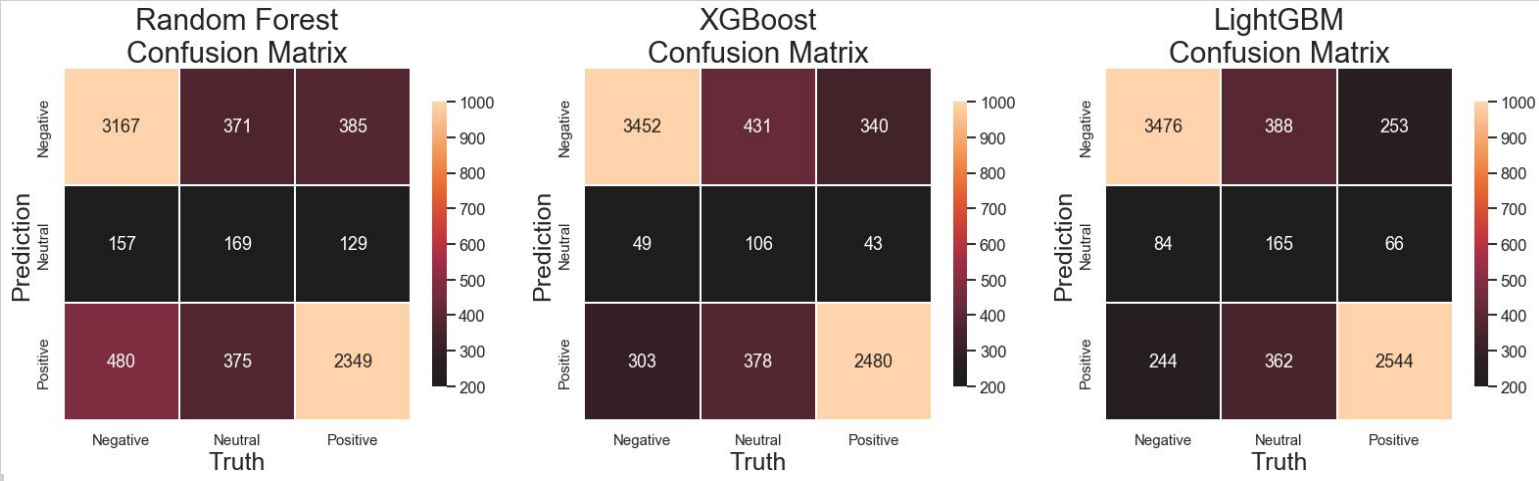
F1 - score : 0.73
Accuracy : 0.74

XGBoost

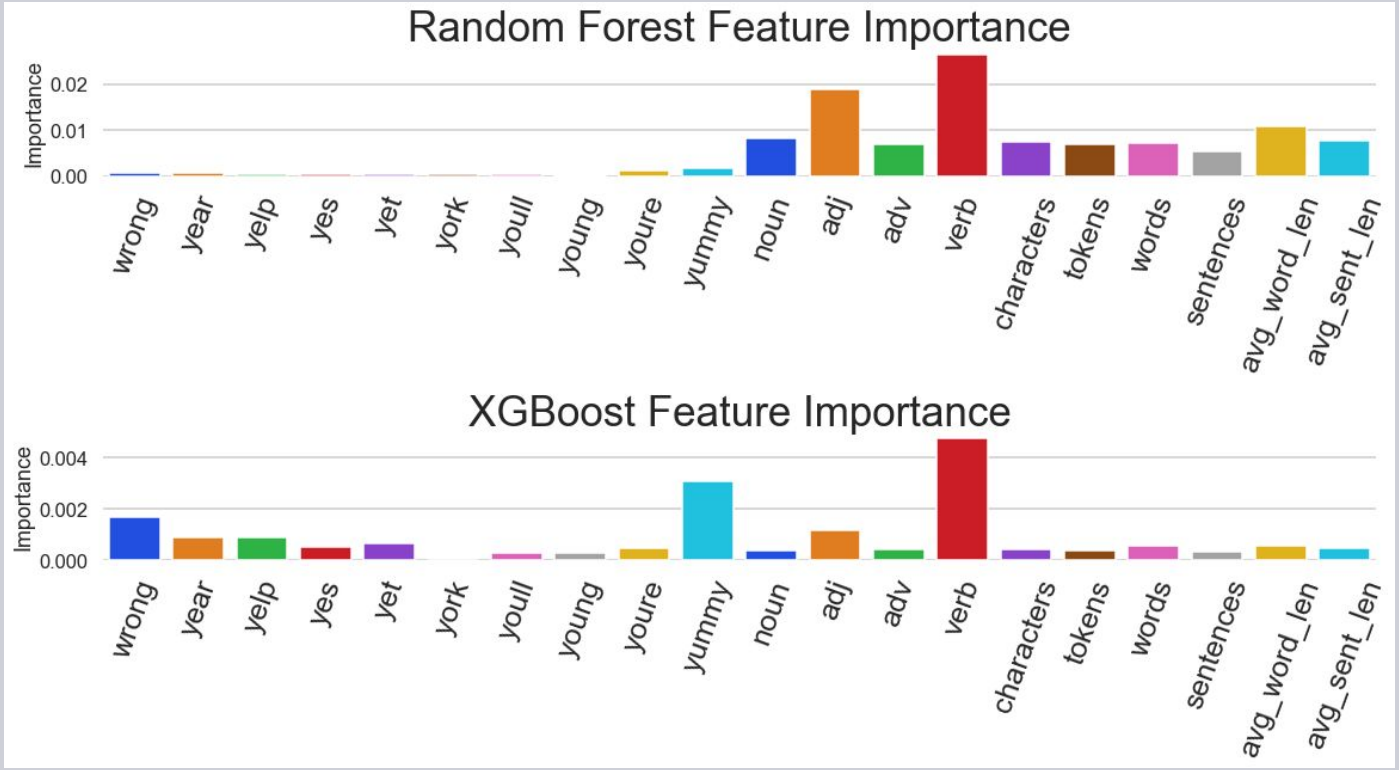
F1 - score : 0.76
Accuracy : 0.79

LightGBM

F1 - score : 0.80
Accuracy : 0.82



Future Importance



Conclusion

Based on result final model is LightGBM
with following results:
F1 - score : 0.8 and
Accuracy : 0.82

Next Steps

- Collect more data labeled as 'Neutral' to balance classes
- Apply additional preprocessing steps to remove specific unigram and bigram words
- Apply sequential based models for text classification (LSTM and RNN).

Contacts:

- Email : ivan.zakharchuk@gmail.com
- Linkedin : https://www.linkedin.com/in/ivan-zakharchuk
- Github : https://github.com/vanitoz