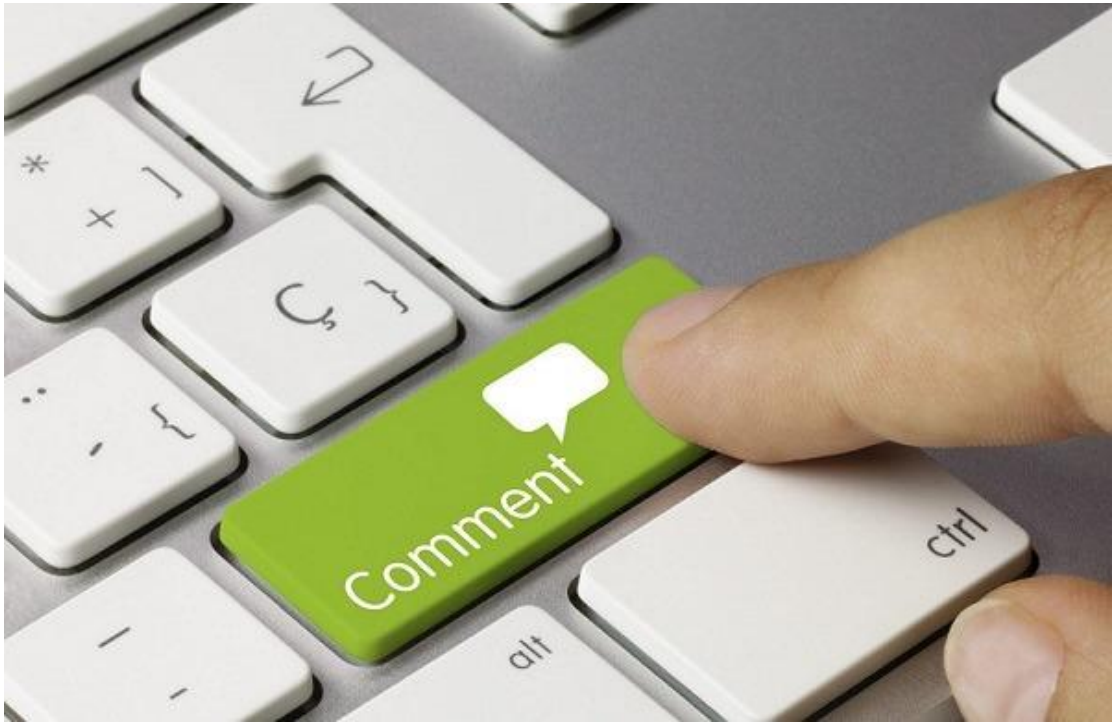
Toxic Comments Classification

24th May 2022

Karam Shbeb
Kamil Sabbagh

Introduction to the topic



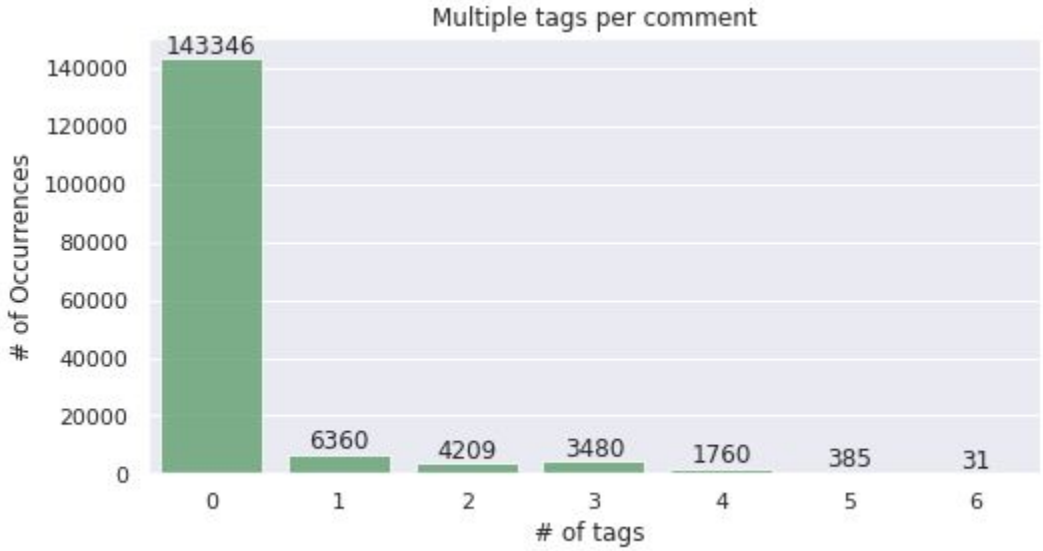
Data

Kaggle collected comments from wikipedia and created the data set for it’s challenge

It contains more than 300 000 comments in both training and testing datasets

comment_text	toxic	severe_toxic	obscene	threat	insult	identity_hate
Explanation\nWhy the edits made under m...	0	0	0	0	0	0
D'aww! He matches this background colou...	0	0	0	0	0	0
Hey man, I'm really not trying to edit ...	0	0	0	0	0	0
"\nMore\nI can't make any real suggesti...	0	0	0	0	0	0
You, sir, are my hero. Any chance you r...	0	0	0	0	0	0

Understanding the data



Understanding the data

Review of existing approaches

There was a challenge by Kaggle 4 years ago

- Contestants used methods such as : Logistic regression, transformers, SVM , LSTM and GRU.
- Best accuracy was 98,8%
- TF-IDF tokenizer
- Adding username as a parameter leads to overfitting

Our Solution

01 Preprocessing

02 Vectorization

03 Modeling

04 Evaluation

Preprocessing

- All letters to lowercase
- Remove punctuation, newlines, spaces and usernames
- Apostrophe replacement (e.g: can't --> can not)
- Remove stop words

Vectorization

- TF-IDF
- Word2vec

Modeling

Binary Relevance	Classifier Chains	Label Powerset
LR and SVM	LR	We didn’t do it
The simplest one	Feed the next one	Too much computation

Evaluation

Recall

How many true positives

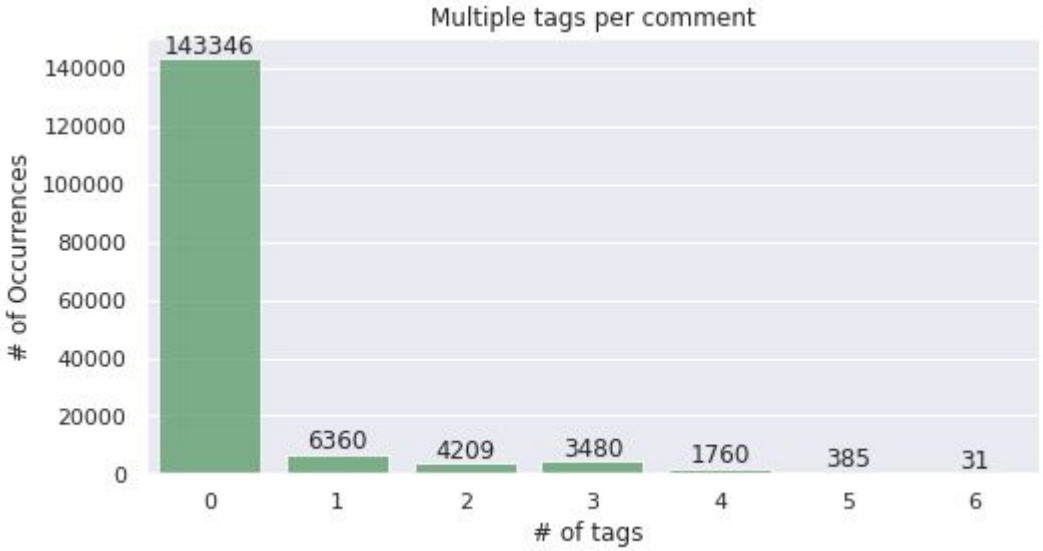
F1-score

Combines recall and precision

Hamming loss

Imbalanced data

Imbalanced data

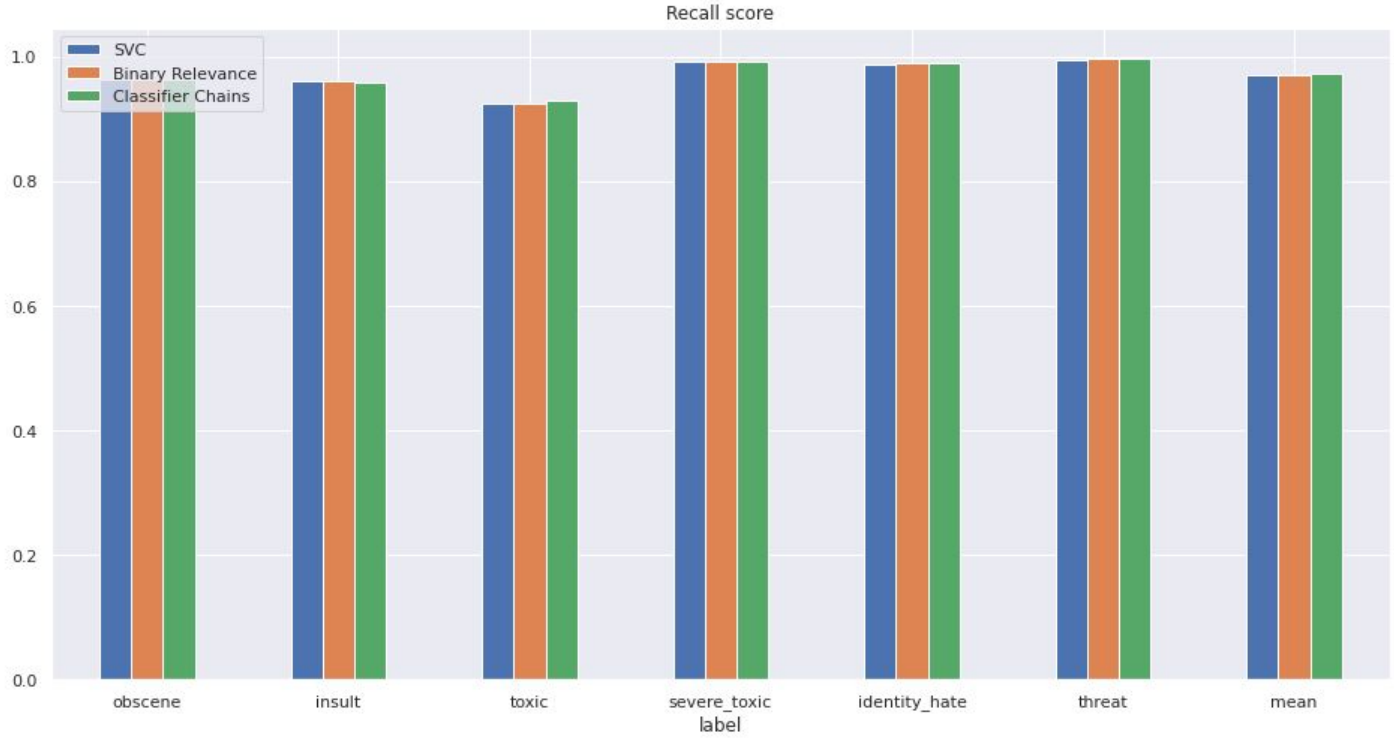


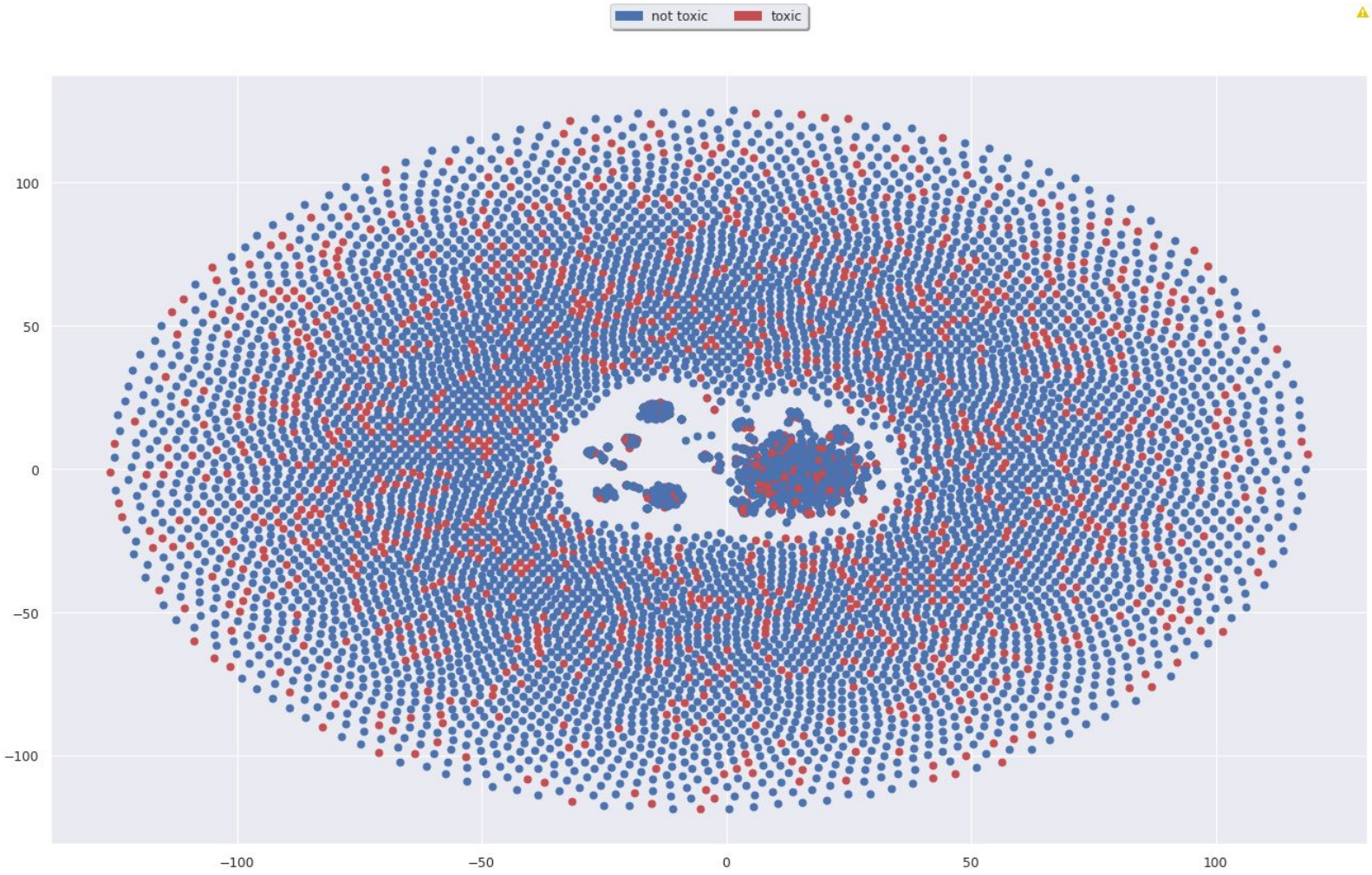
Logistic regression using Classifier chains

hamming loss is 0.028421228130503195

	label	F1	Recall
0	obscene	0.963574	0.963534
1	insult	0.958019	0.958830
2	toxic	0.931658	0.929069
3	severe_toxic	0.992373	0.992029
4	identity_hate	0.988603	0.989950
5	threat	0.995748	0.996061
6	mean	0.971663	0.971579

Recall


Recall score
SVC
Binary Relevance
Classifier Chains
1.0
0.8
0.6
0.4
0.2
0.0
obscene
insult
toxic
severe_toxic
identity_hate
threat
mean
label


not toxic
toxic
100
50
0
−50
−100
−100
−50
0
50
100

Sorry for this toxic slide

Conclusion

Thanks

Do you have any questions?

k.sabbagh@innopolis.university

k.shbeb@innopolis.university