

Depression detection in social networks

Martí Caixal i Joaniquet





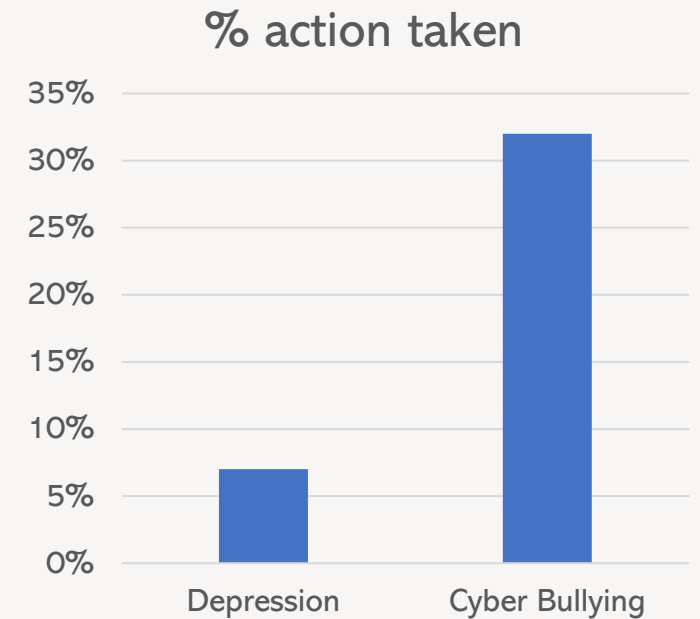
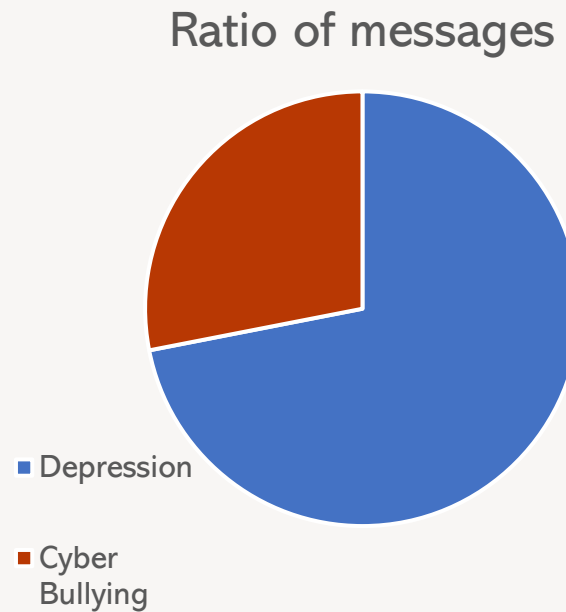
Agenda





What problem are we facing?

- More depression messages than other types
- Yet, they receive less attention





Objectives

NLP PROBLEM

Find out which method gives the best results and how they differ

Shallow Learning:

- Naïve Bayes
- Decision Tree
- Random Forest
- SVM
- KNN
- Hyper Parameter Search

Deep Learning

- RNN
 - RNN LSTM
 - RNN GRU
- Transformers (BERT)



Planification

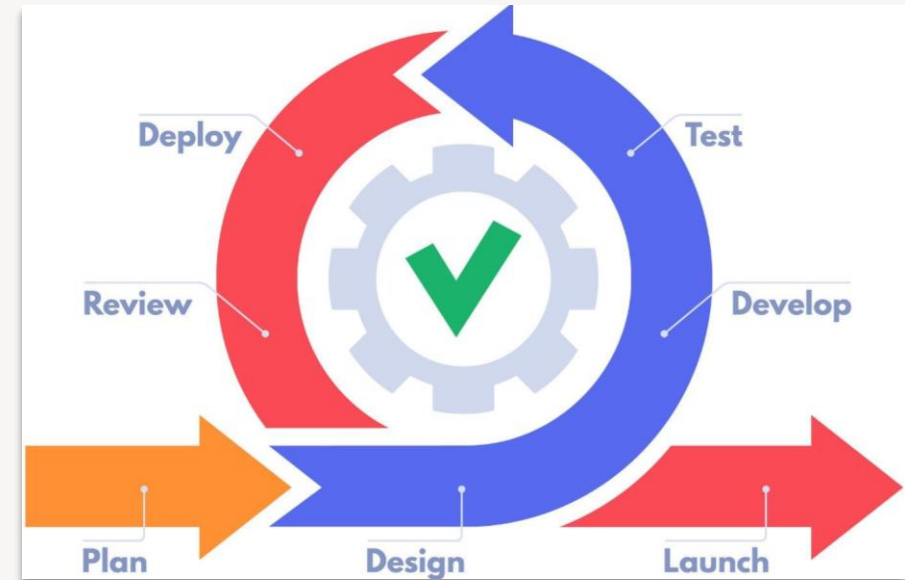
Short iterations

Good control of flow

Friendly to changes

Independent subobjectives

Easy to detect errors



Agile Methodology



Data used

Mental Health Twitter (Twitter 3)

- Only the message and label

Depression Twitter (Twitter Scale)

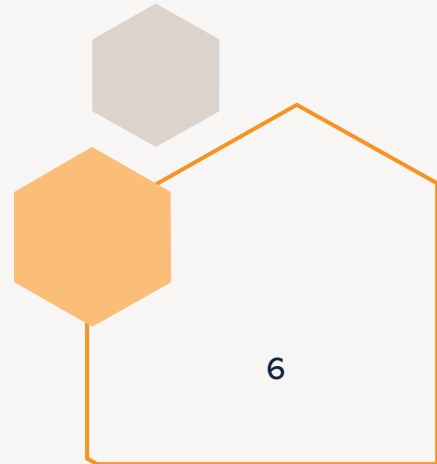
- Labeled in a scale from 1 to 5

Depression Reddit (Reddit)

- Already cleaned

- 10000 messages
- 2 classes
- Unbalanced (80/20)

*“ @cosmicgirlie Thinking of you.
Everything crossed Turn baby turn! “*





Data used

Mental Health Twitter (Twitter 3)

- Only the message and label

Depression Twitter (Twitter Scale)

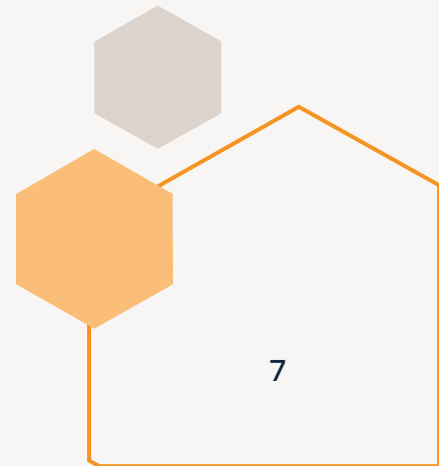
- Labeled in a scale from 1 to 5

Depression Reddit (Reddit)

- Already cleaned

- 45000 messages
- 4 classes (Scale from 0 to 3)
- Unbalanced (40/20/30/10)

*“ humm dodgers scored a hr stupid
dodgers i hate them”*





Data used

Mental Health Twitter (Twitter 3)

- Only the message and label

Depression Twitter (Twitter Scale)

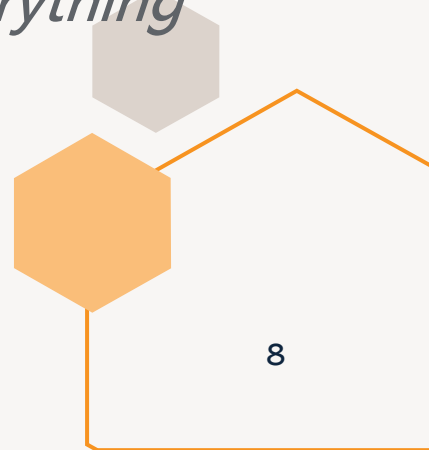
- Labeled in a scale from 1 to 5

Depression Reddit (Reddit)

- Already cleaned

- 40000 messages
- 2 classes
- Unbalanced (60/40)
- Already cleaned

“ i used to be highly functional before but it now i can barely function at all i take everything just...”





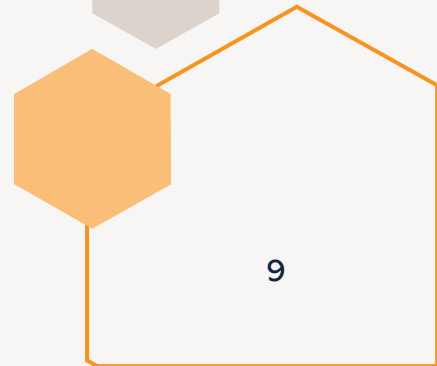
Data used

Unbalanced,
target class being minority:

- × Undersampling
- × Oversampling



- Recall instead of accuracy
- Macro average





Initial preprocessing

Delete usernames

Delete Stop Words

Delete numbers

Lemmanization

Delete punctuation



Specific approaches

Shallow Learning

- Bag of Words
- TF-IDF

Deep Learning

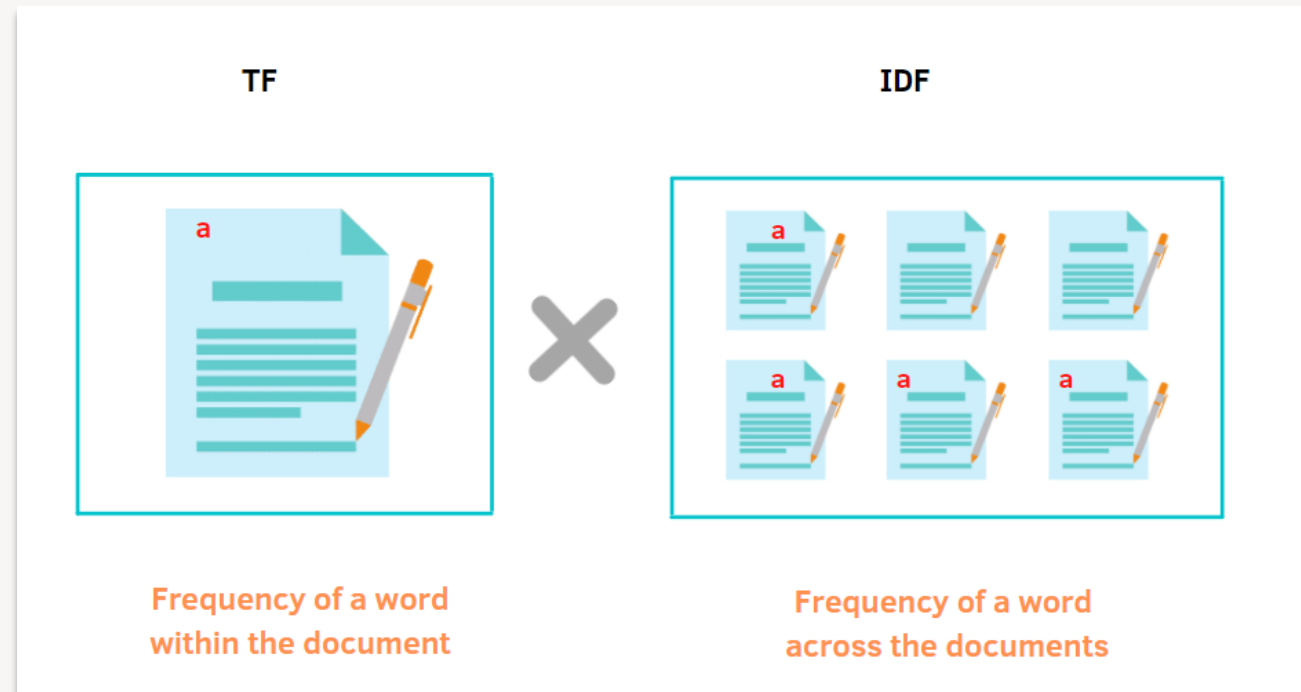
- Word Embedding (GloVe)



Bag of Words

Document	the	cat	sat	in	hat	with
<i>the cat sat</i>	1	1	1	0	0	0
<i>the cat sat in the hat</i>	2	1	1	1	1	0
<i>the cat with the hat</i>	2	1	0	0	1	1

TF-IDF



Shallow learning results

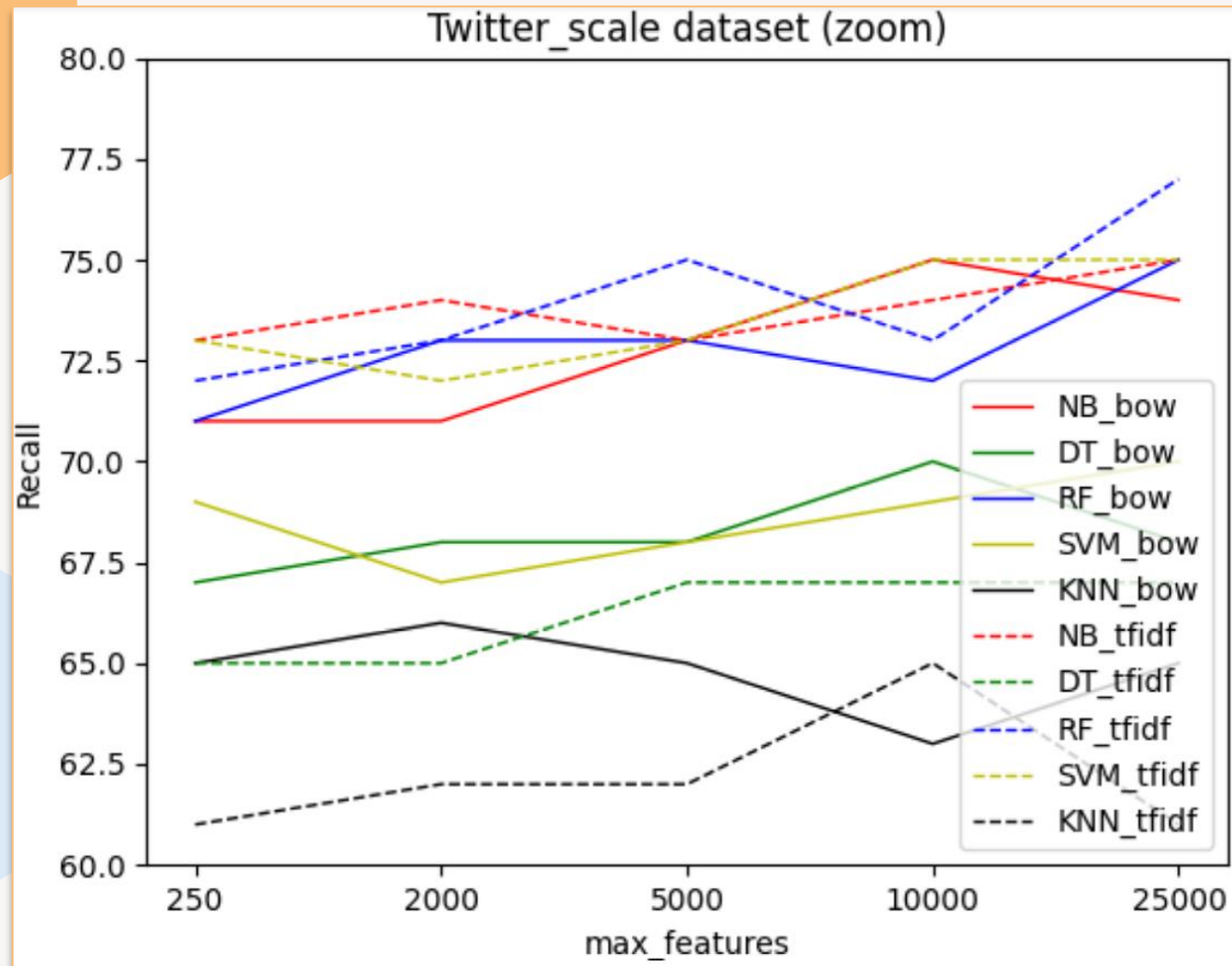
- Scikit Learn library
- Default parameters

- Naïve Bayes
- Decision Tree
- Random Forest
- SVM
- KNN

- Hyperparameter Search



TF-IDF vs BoW & feature size



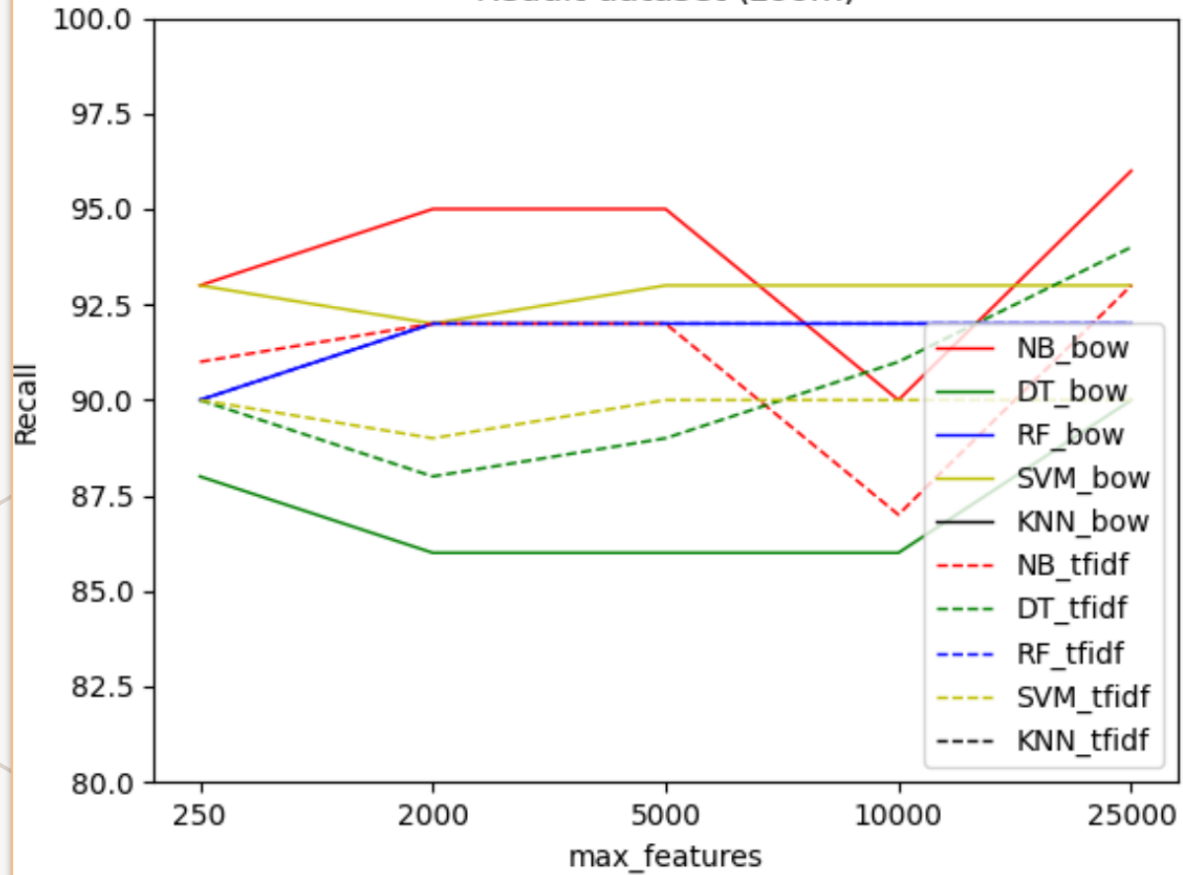
TF-IDF - - - -

BoW ————

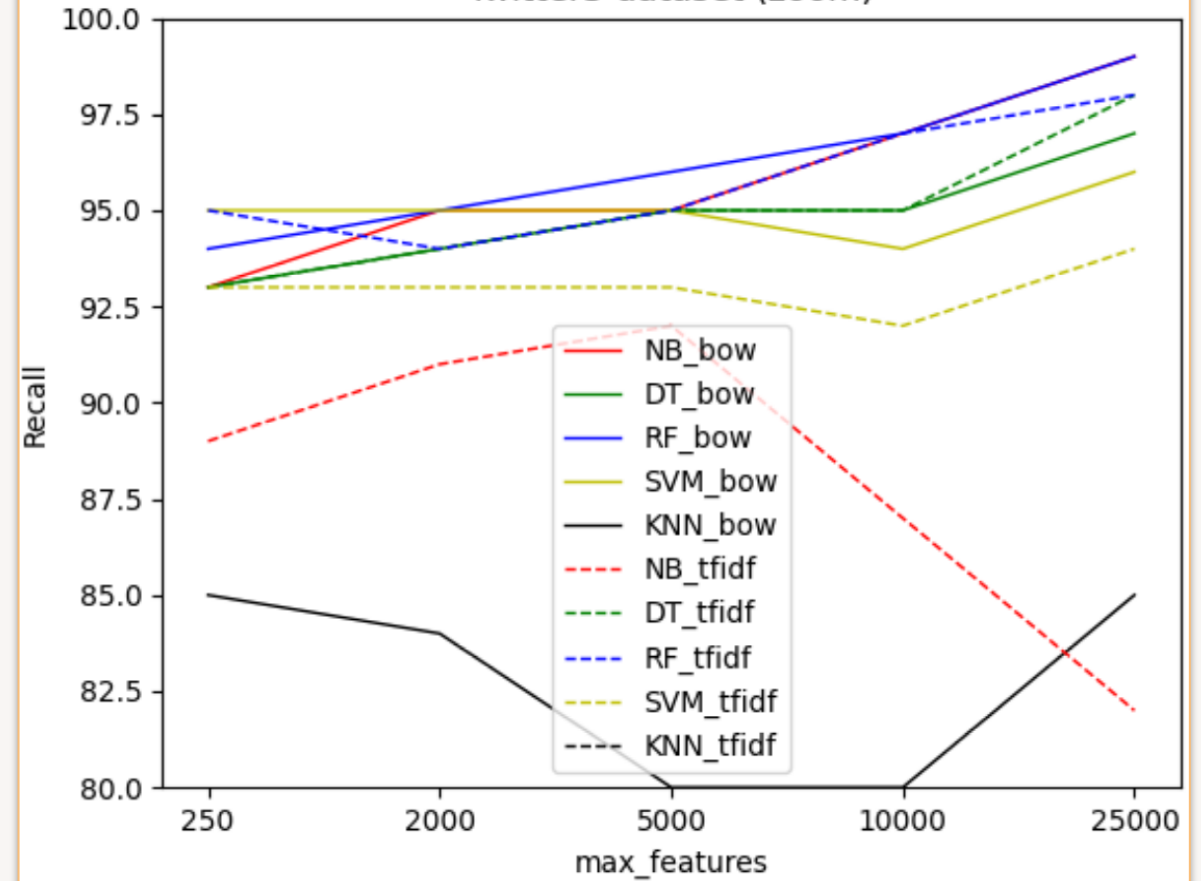
- ✓ TF-IDF slightly better
- ✓ Increase number of features slightly improves results

Results

Reddit dataset (zoom)

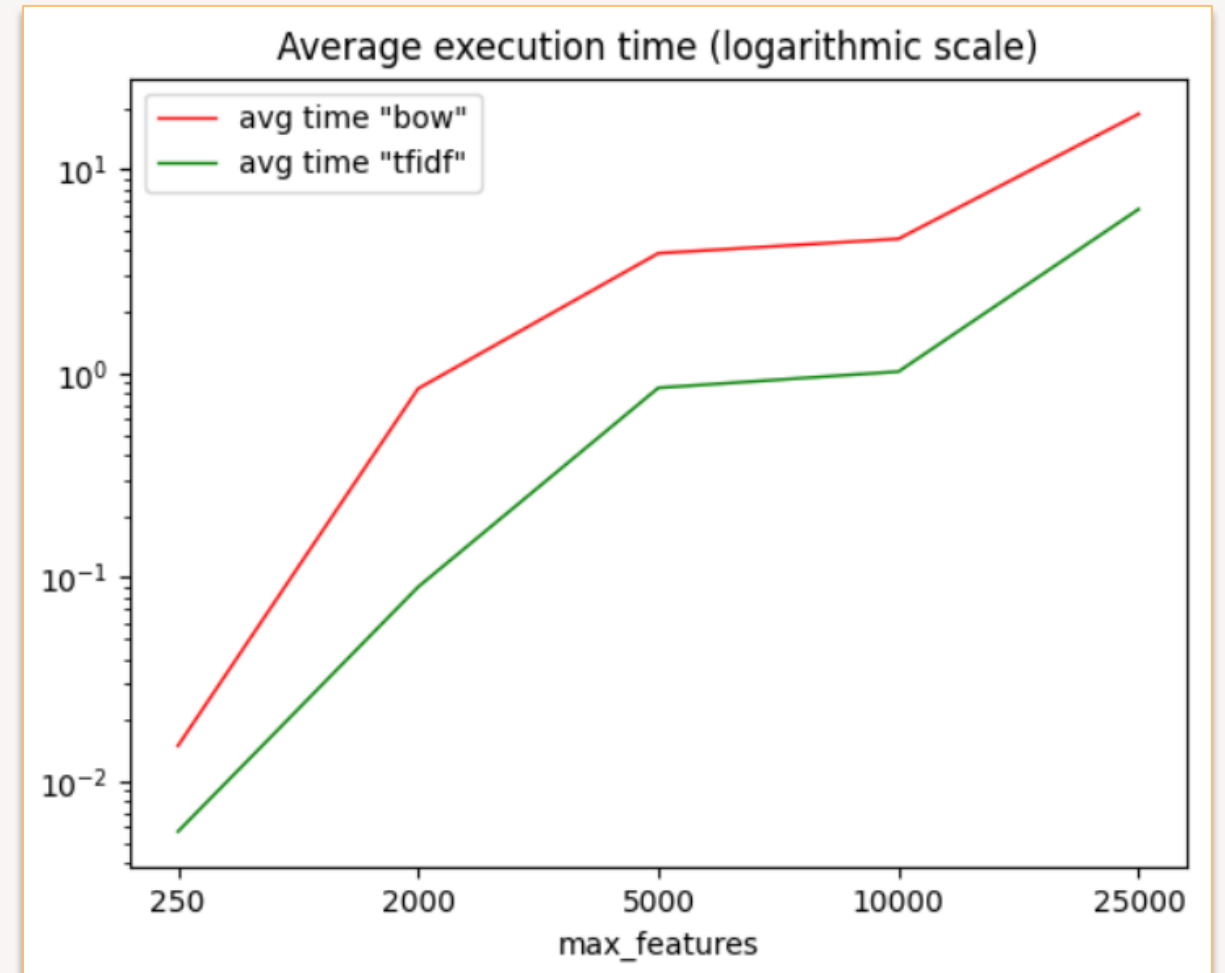


Twitter3 dataset (zoom)

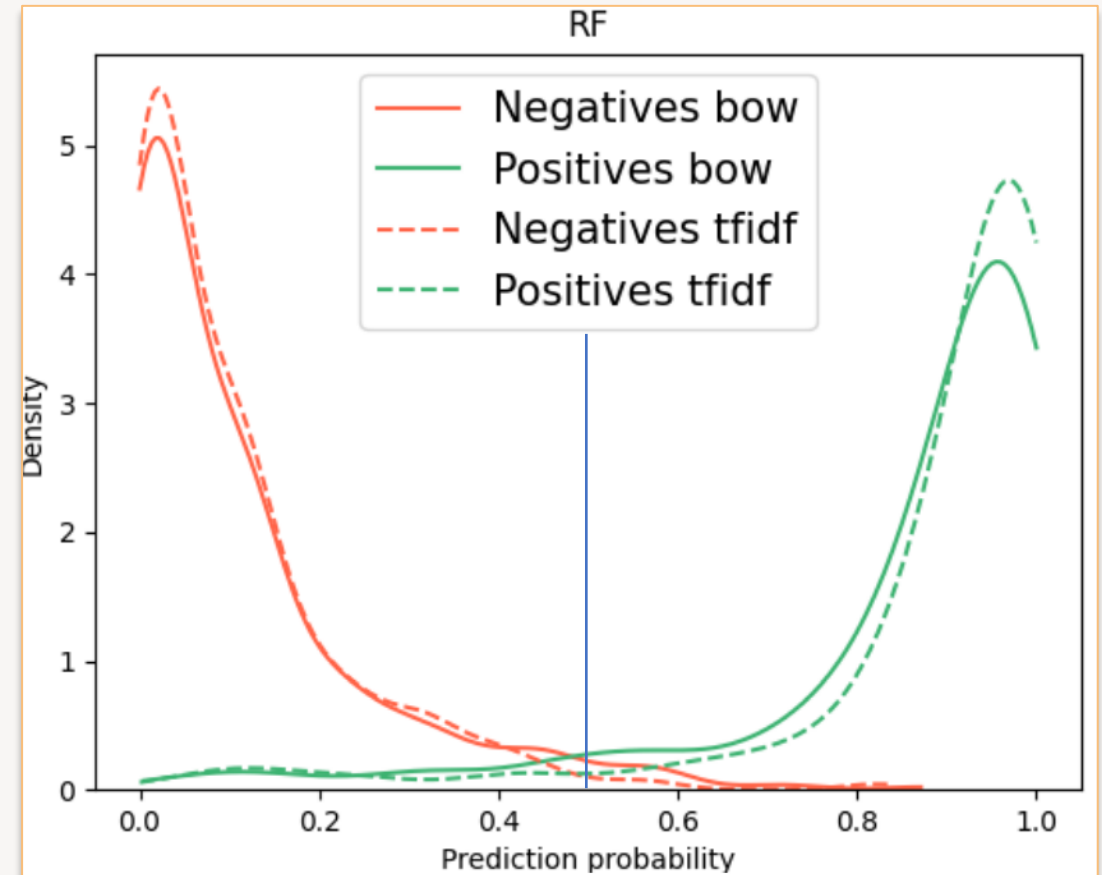
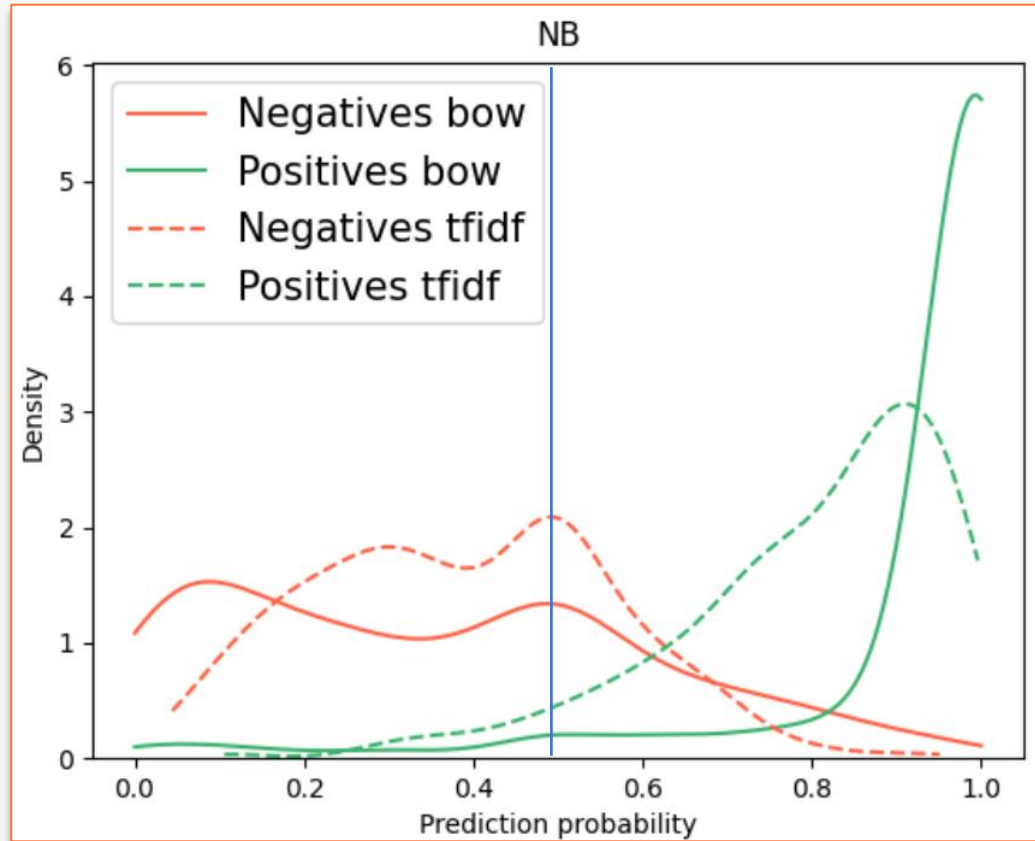


Execution time

- ✓ TF-IDF slightly better
- ✓ n^0 max_features improves results
- ✓ Better execution time



Confidence in predictions



Hyperparameter search

Done with Optuna:

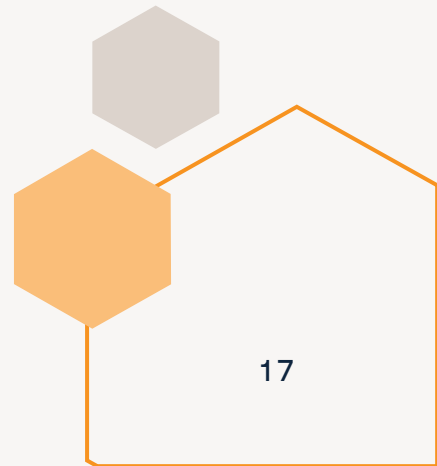
- Python library
- Optimized search
- Parallelization

Model	N° parameters	executions
SVM	3	10^3
KNN	3	10^3
DT	4	8^4
RF	4	8^4
NB	1	100^1



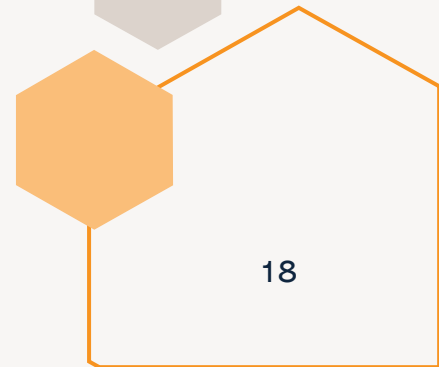
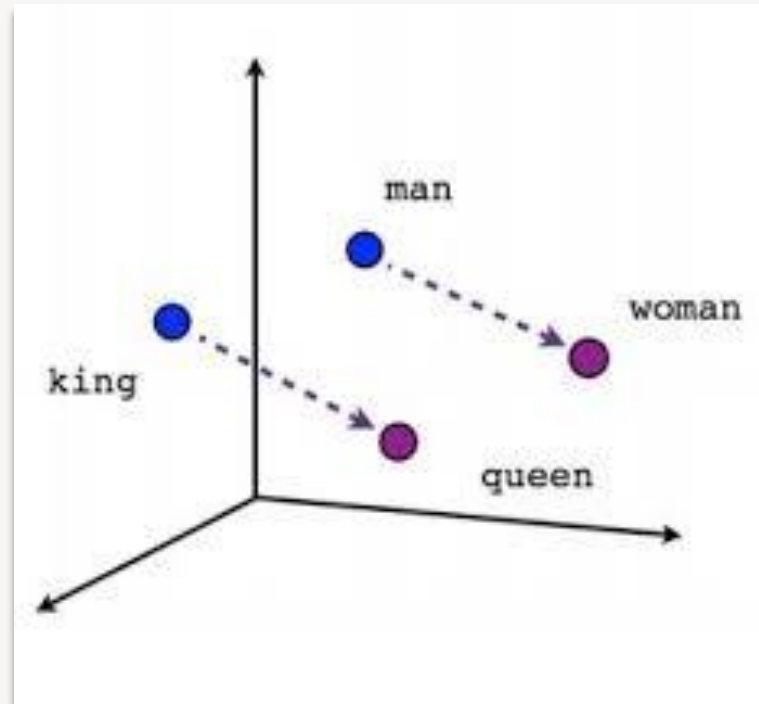
O P T U N A

× Results do not improve





Word Vectoring





Deep learning results

- Keras library
- RTX 3070 Ti

- RNN
- RNN GRU
- RNN LSTM
- BERT



RNN

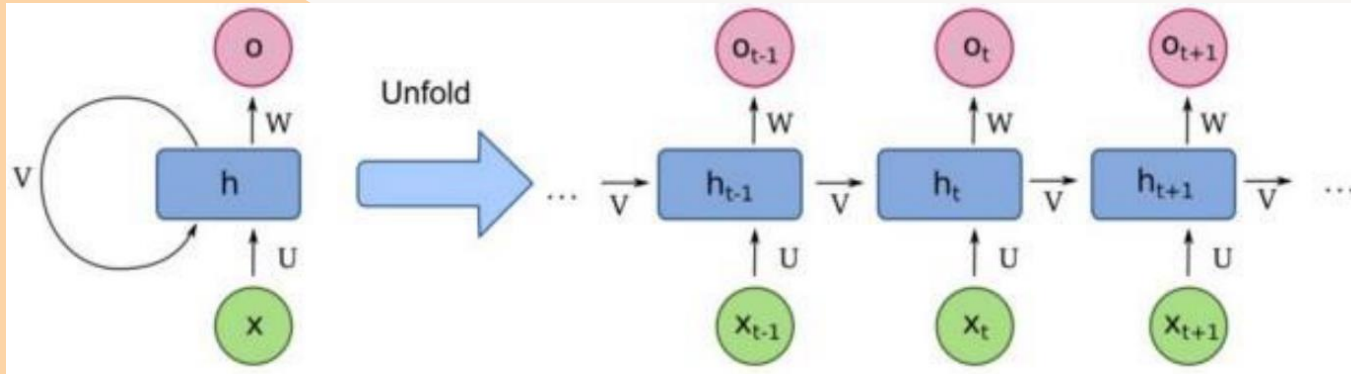
- Sequence of layers
- Input, activation function, output
- No memory

RNN LSTM

- 3 gates
- Memory

RNN GRU

- 2 gates
- Memory
- Simplified LSTM



RNN

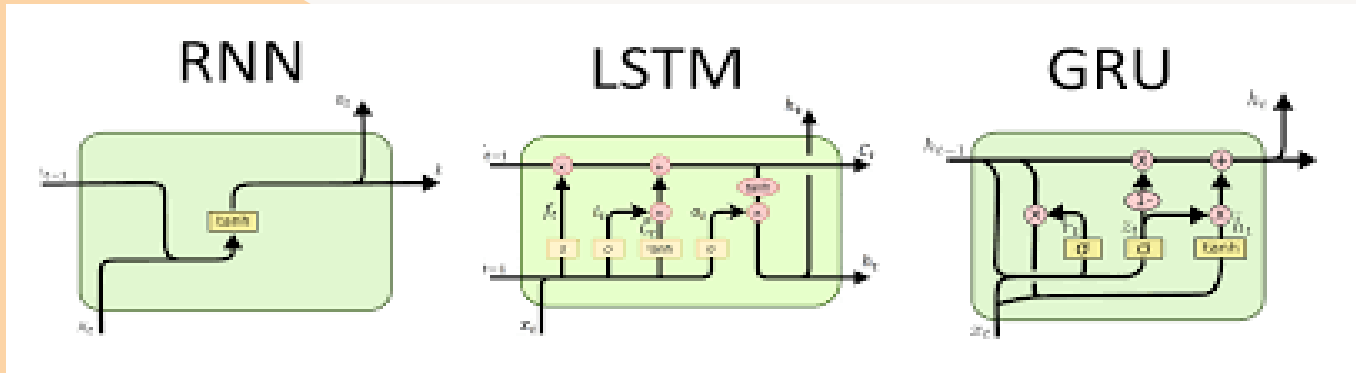
- Sequence of layers
- Input, activation function, output
- No memory

RNN LSTM

- 3 gates
- Memory

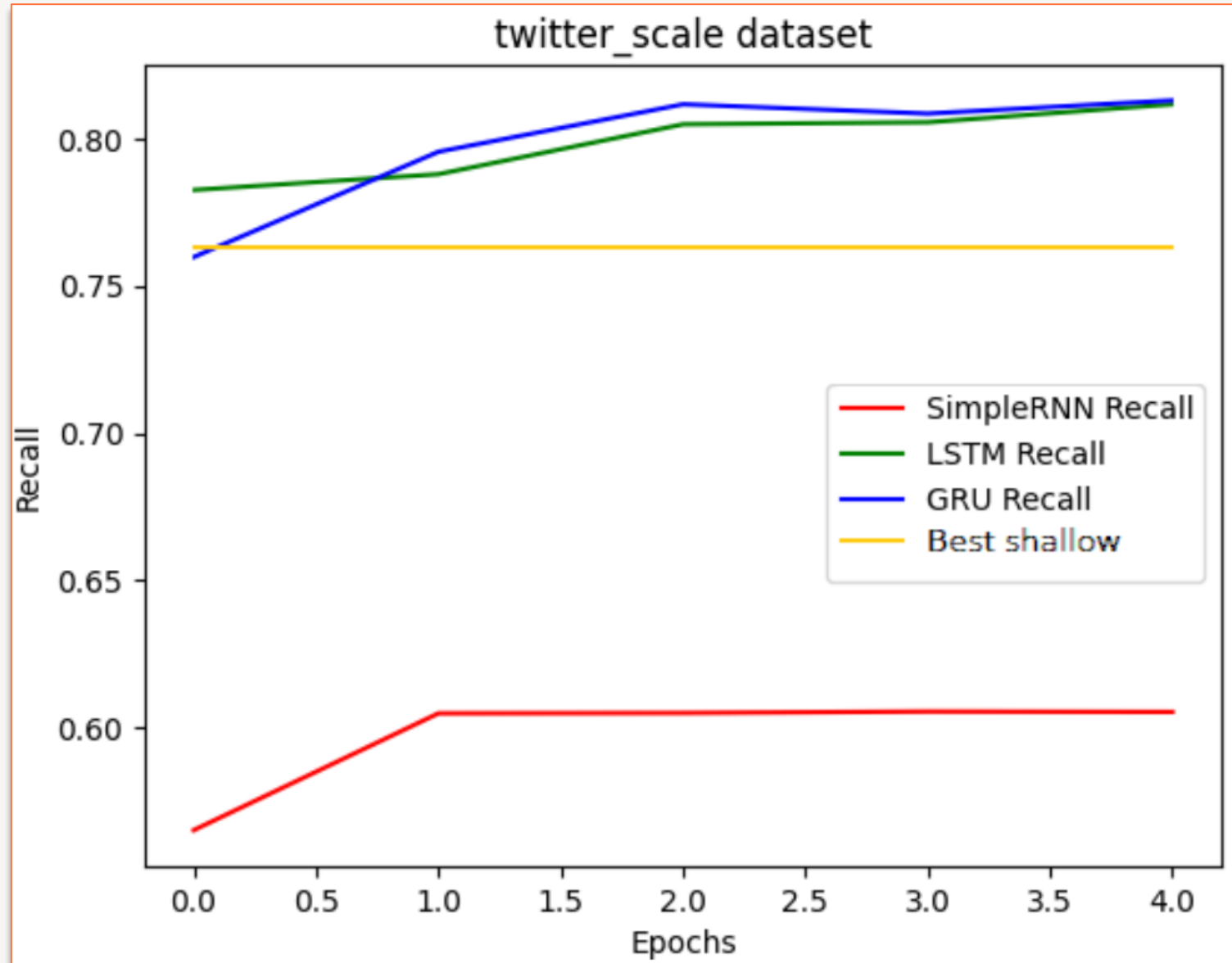
RNN GRU

- 2 gates
- Memory
- Simplified LSTM





Results

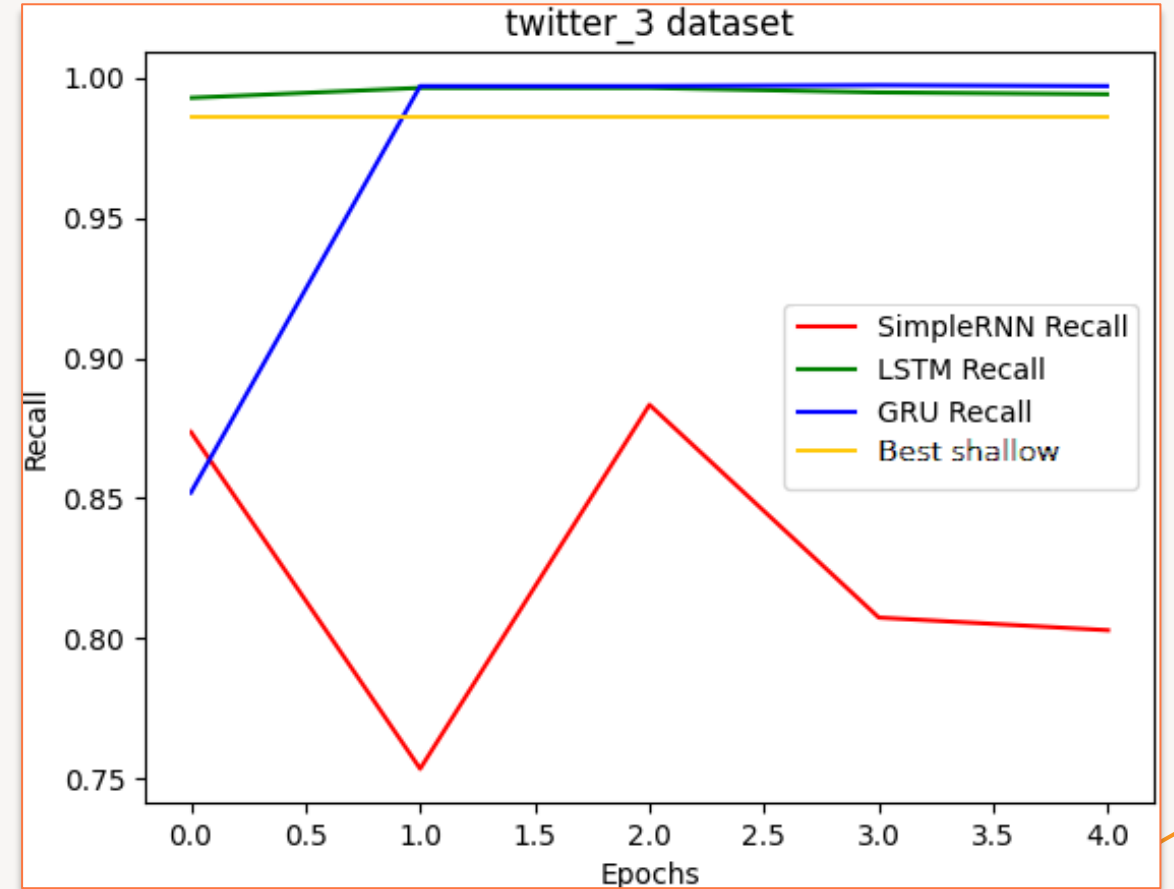
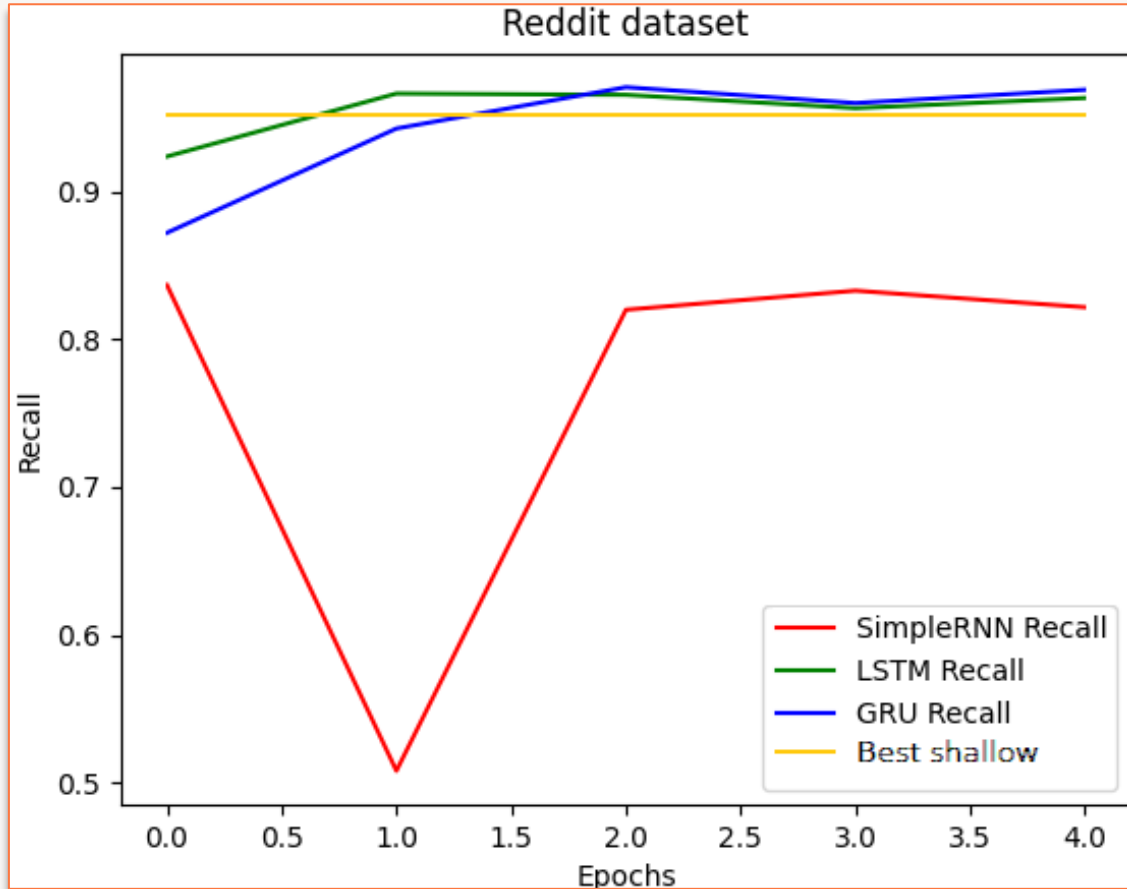


RNN

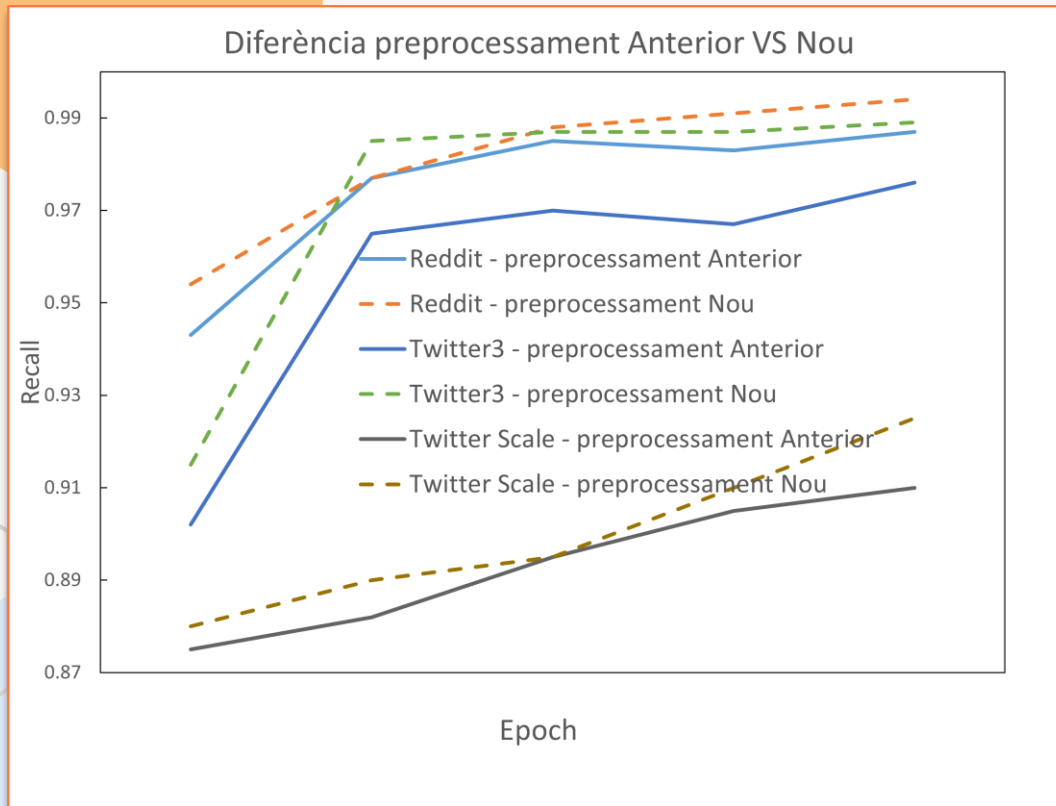




Results



New preprocessing



Delete usernames

~~Delete Stop Words~~

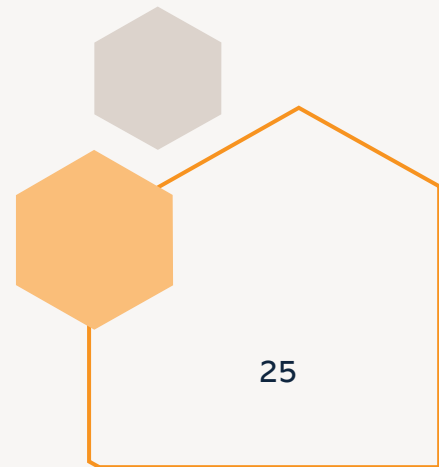
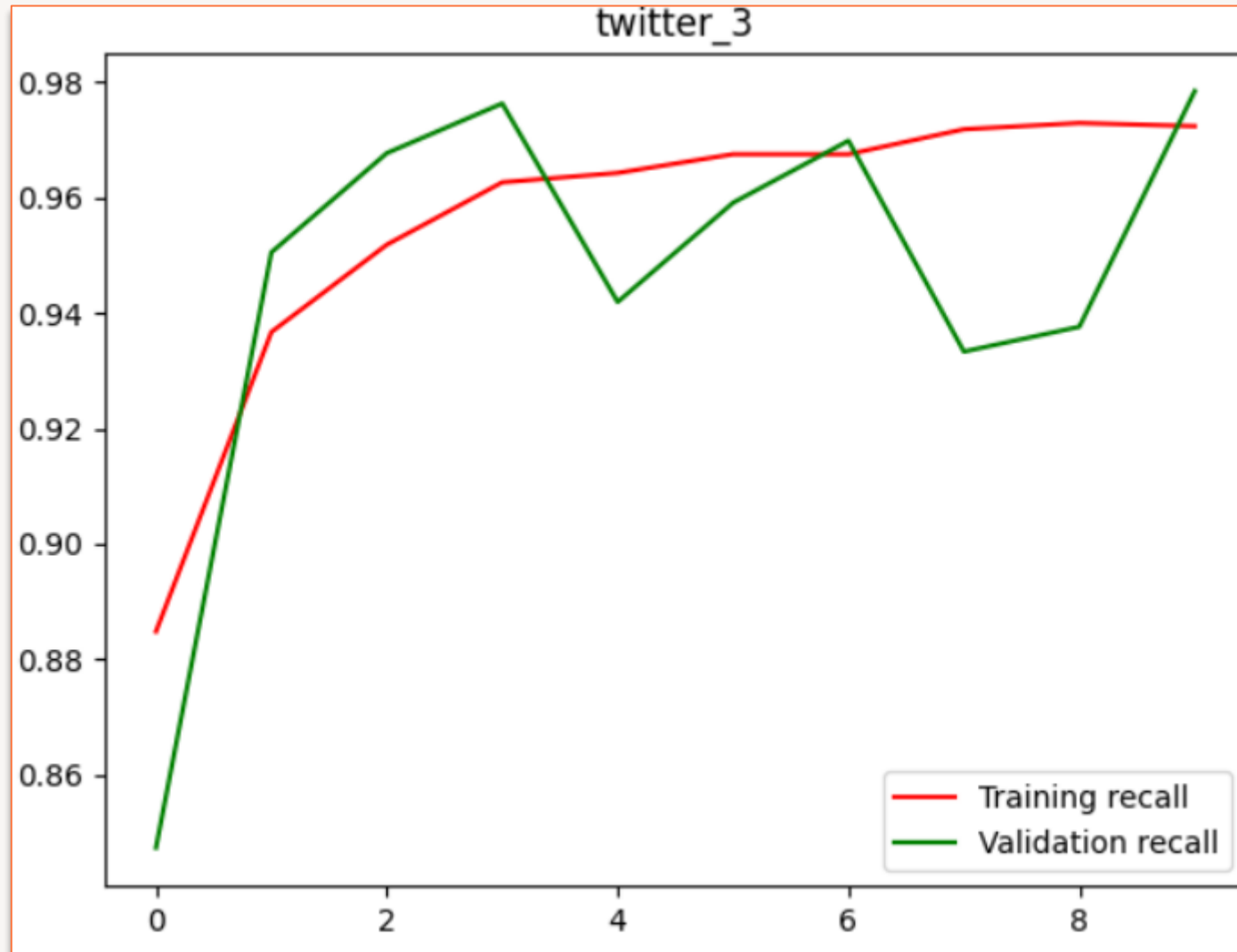
Delete numbers

~~Lemmanization~~

Delete punctuation

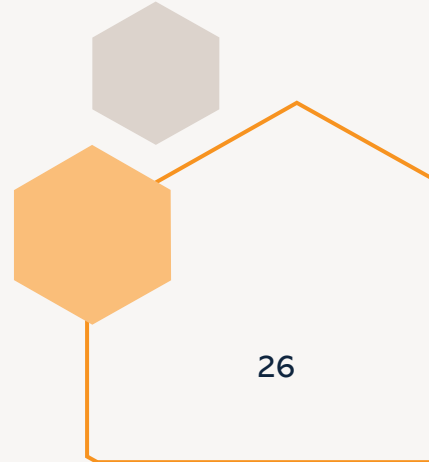


BERT (transformers)



Differences in predictions

	Deep Learning	Shallow Learning
“study finds no casual relationship between cannabis and depression”	✓	✗
“dailytonic exposure to the bacteria in soil can be good for mental hearlth and could treat depression and prevent ptsd”	✓	✗
“don’t be sad, armys are here for you we will always suport you btstwt be strong”	✓	✗



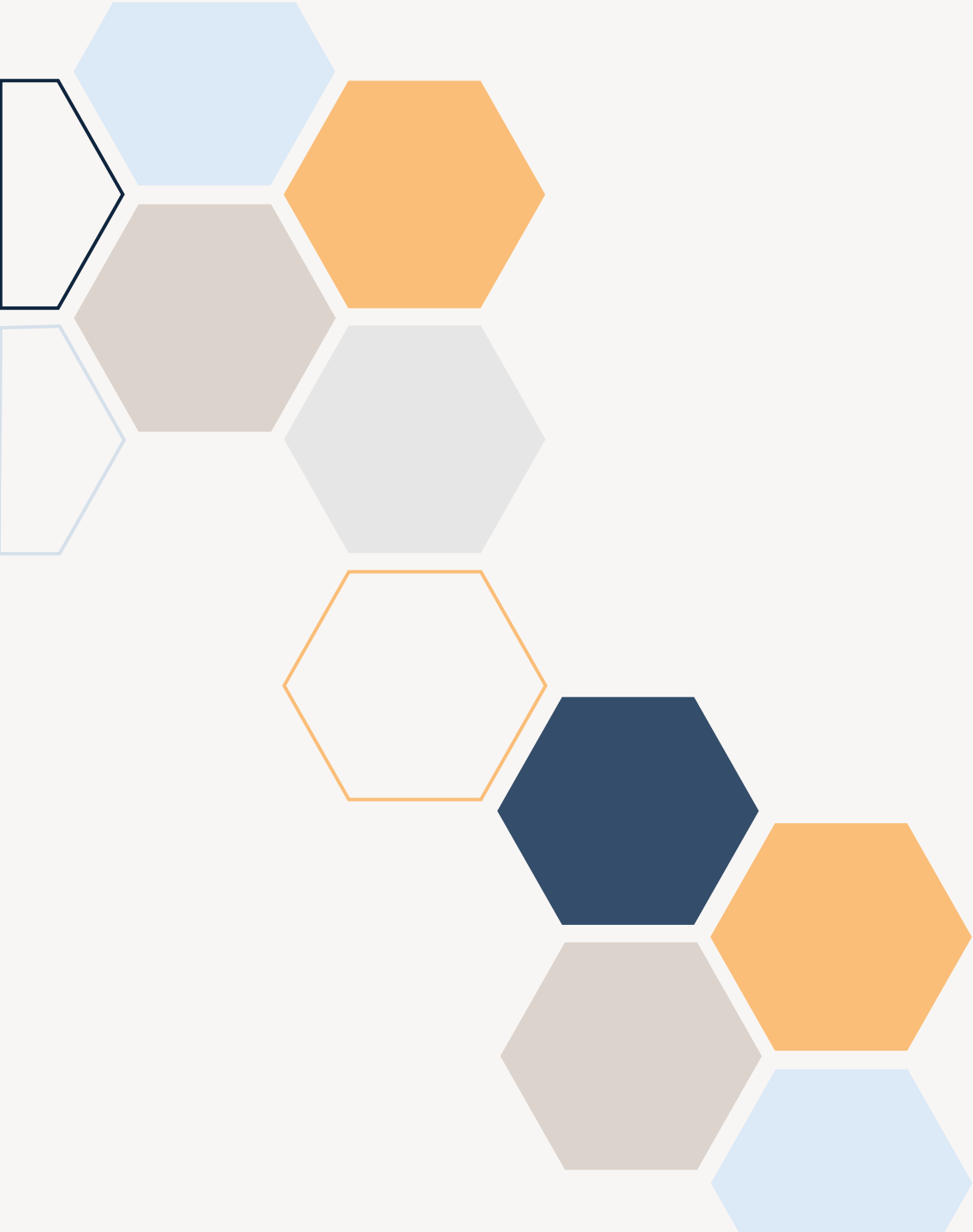


Shallow learning

- Best: SVM and RF (relative to confidence)
- Preprocessing highly affects on metrics
- Feature extraction highly affects on execution time
- Parameters are not decisive

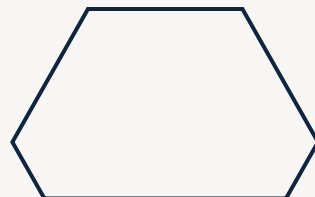
Deep learning

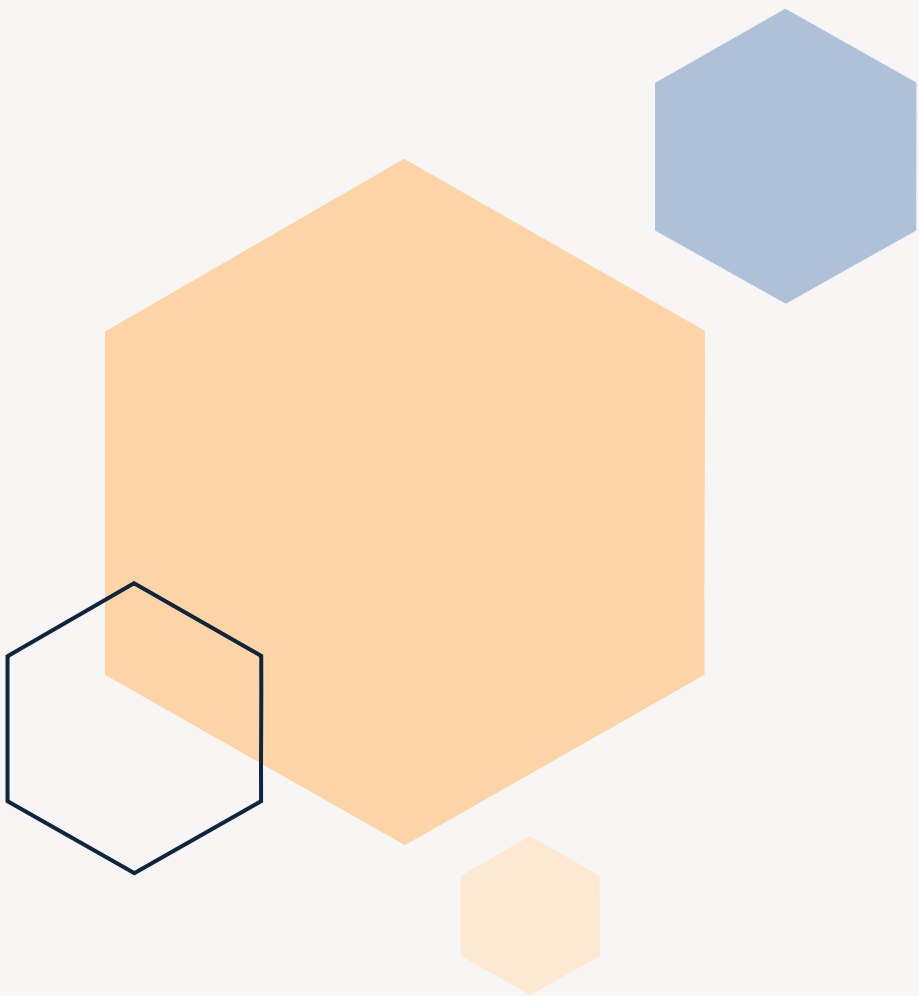
- Results about 10% better
- Simple RNN not good at all, GRU and LSTM are needed
- LSTM better than GRU with long messages
- Gets the semantics instead of the relations
- BERT needs more data and computing resources

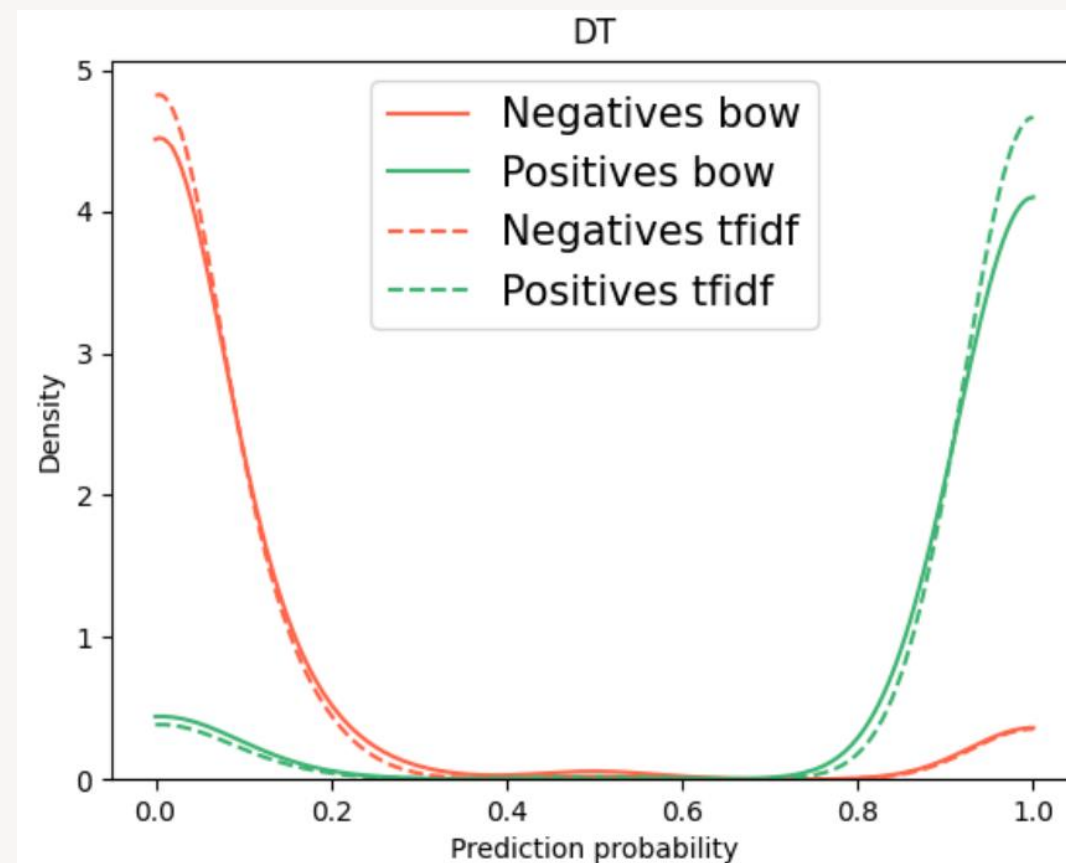
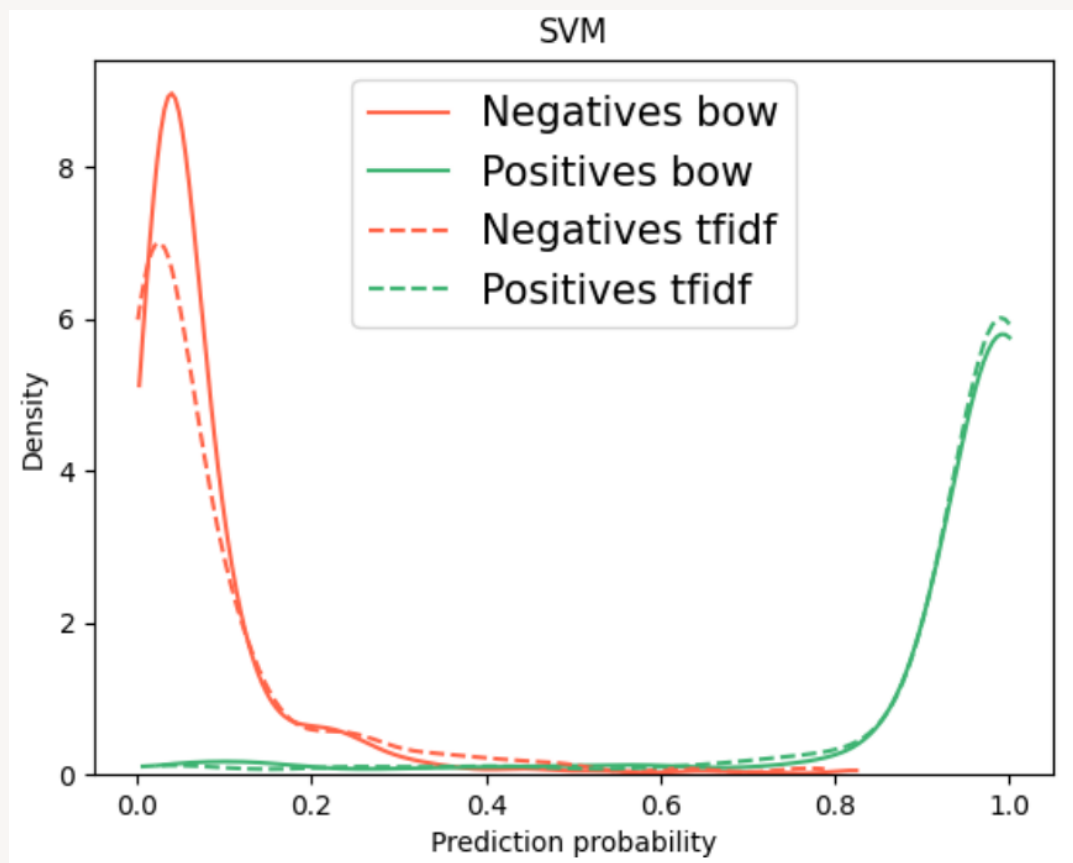


Thank you

Martí Caixal i Joaniquet







Execution time

