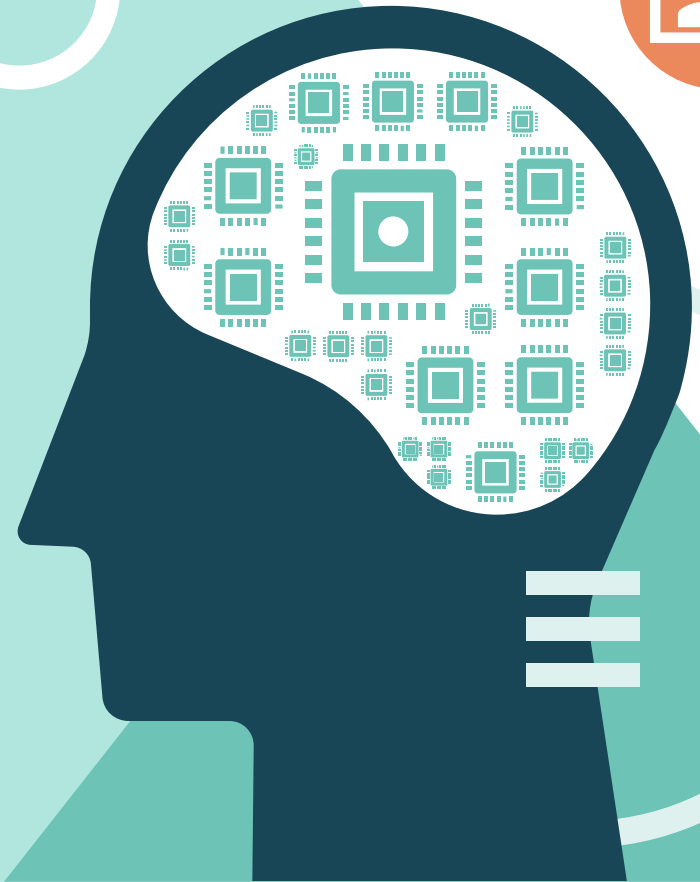


Feasibility of detecting depression from free text published in online forums

Ignacio Montes Álvarez
University of Oviedo, 2020

Table of Contents

Introduction

Depression is a serious disorder and the largest contributor to global dissability

+264 million

PEOPLE LIVE WITH THIS DISORDER

What is depression?

SYMPTOMS

Great sadness, reduced energy, loss of interest

OTHER FEELINGS

Guilt, worthlessness, irritability

CAUSES

Changes in life: accidents, drugs abuse, traumas, divorce…

RELATED DISORDERS

Anxiety, disturbed sleep, eating and concentration disorders

How can we help?

01
Motivation & Objectives

Social media can be used as a “behavioral health assessment tool” [1]

[1] *Munmun De Choudhury et al. “Social media as a measurement tool of depression in populations”*

Why social media?

SAFE ENVIRONMENT

Sufferers can express themselves better. It's like a "diary" [2]

LOTS OF DATA

We have acess to large amounts of data about users' beliefs

[2] *Munmun De Choudhury. "Role of social media in tackling challenges in mental health"*

Which platform to analyze?

REDDIT

Large group of forums or communities

STATISTICS

17th globally and 7th USA (Alexa Rank, October 2020)

STRUCTURE

Several forums also called subreddit with their own rules managed by moderators

ANONIMITY

No need to provide personal information, only a username

Subreddit subject to study

r/depression

- The biggest Reddit community about this disorder
- More than 695k users
- Created in 2009

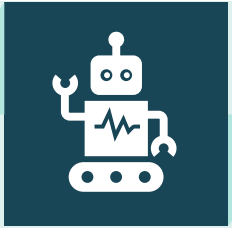
Goals

Detection of depression-related texts

Detection of depression-prone users in non depression-related forums

Potential fields of application

Tool for professionals



Automatic detection bot

Self-help web page

(*) Caveats: ethical implications should be carefully analyzed before deployment of any of these tools

02
Methodology

Datasets creation, text handling
and machine learning

Technologies used

Project structure

COMMON MACHINE LEARNING PIPELINE

PHASE A

Detection of depression-related texts

PHASE B

Detection of depression-prone users in non depression-related forums

How was it done? (phase A)

1. Subreddit corpus extraction (depression-prone dataset)
2. Control collection generation (non-depression-prone dataset)
3. **Machine learning pipeline**

(*) All posts extracted are older than the 1st of January of 2020

How was it done? (phase B)

1. Username extraction from depression-prone dataset (phase A, step 1)
2. Obtain users information
3. Obtain random users' sample (presumably depressed users)
4. Obtain control users' sample (presumably non-depressed users)
5. Corpus generation for both samples
6. **Machine learning pipeline**

(*) All posts extracted are older than the 1st of January of 2020

In depth: obtain users information

Creation date

Karma punctuations

+69 million users

Elasticsearch indexing

In depth: obtain users' samples

Systematic sampling

Similarity constraints

Not in r/depression

Pareto's Law and fine-tuning

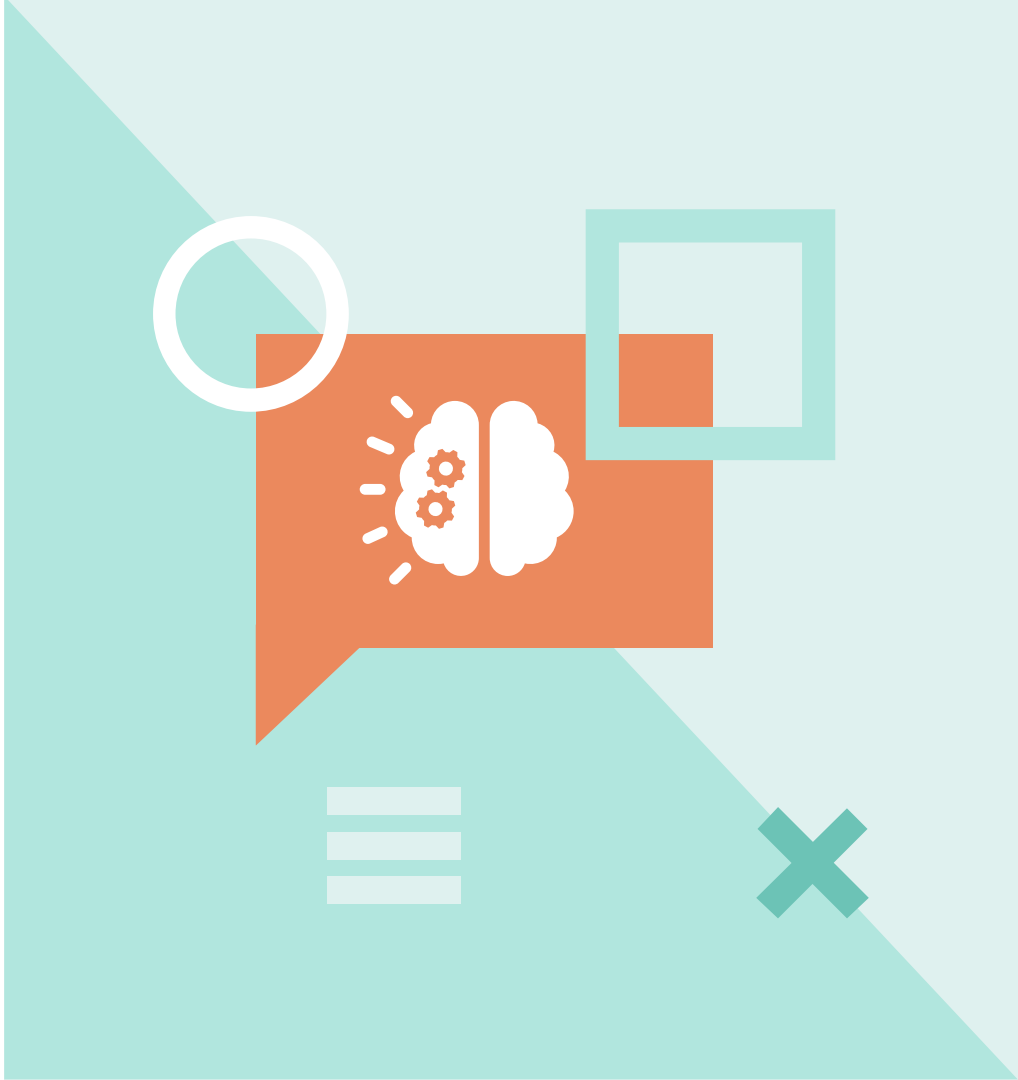
Additional steps: subreddits removal

To ensure we don’t distort the results we have to remove possible comorbidities of depression

- Remove subreddits that can be directly related to depression (i.e, r/Anxiety, r/SuicideWatch, r/mentalhealth...)
- Remove subreddits that can be indirectly related to depression (i.e, r/lgtbi, r/Alcoholism...)
- Remove subreddits that contain “-depress-” in their name

Machine Learning

Supervised binary classification using text features



Datasets cleaning

Lowercase

Remove punctuation and URLs

Stemming

Whitespace normalization

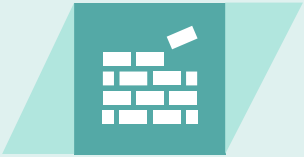
Stopwords removal

Unicode symbols conversion

Datasets vectorization

Bag of words

TF-IDF



1-2 n-grams

10,000 features

Train and test

Classifiers chosen		
Classifier	**Type**	**Alpha (α) values**
Multinomial Naïve Bayes	Bayesian	0.1, 0.2, 0.4, 0.6, 0.8, 1.0
Complement Naïve Bayes	Bayesian	0.1, 0.2, 0.4, 0.6, 0.8, 1.0
Stochastic Gradient Descent	Linear	$[10^{-1}, 10^{-6}]$ (step 10^{-1})

The alpha parameter is the one to be tuned for each classifier. In the Bayesian classifiers controls the smoothing and in the linear classifier controls the regularization strength

03
Results

Metrics presentation and interpretation

Runs

3 TYPES OF CLASSIFIERS

6 VECTORIZATION MODELS

540 ITERATIONS

Performance metrics

SENSITIVITY

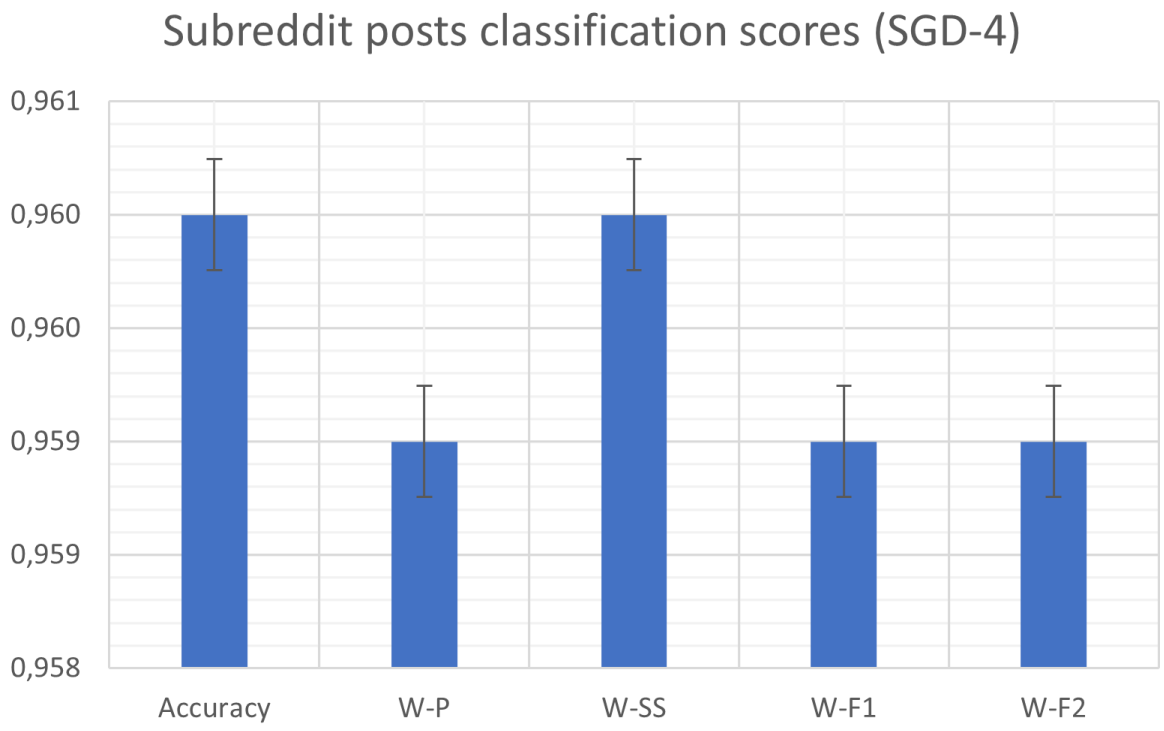
WEIGHTED SCORING

F2

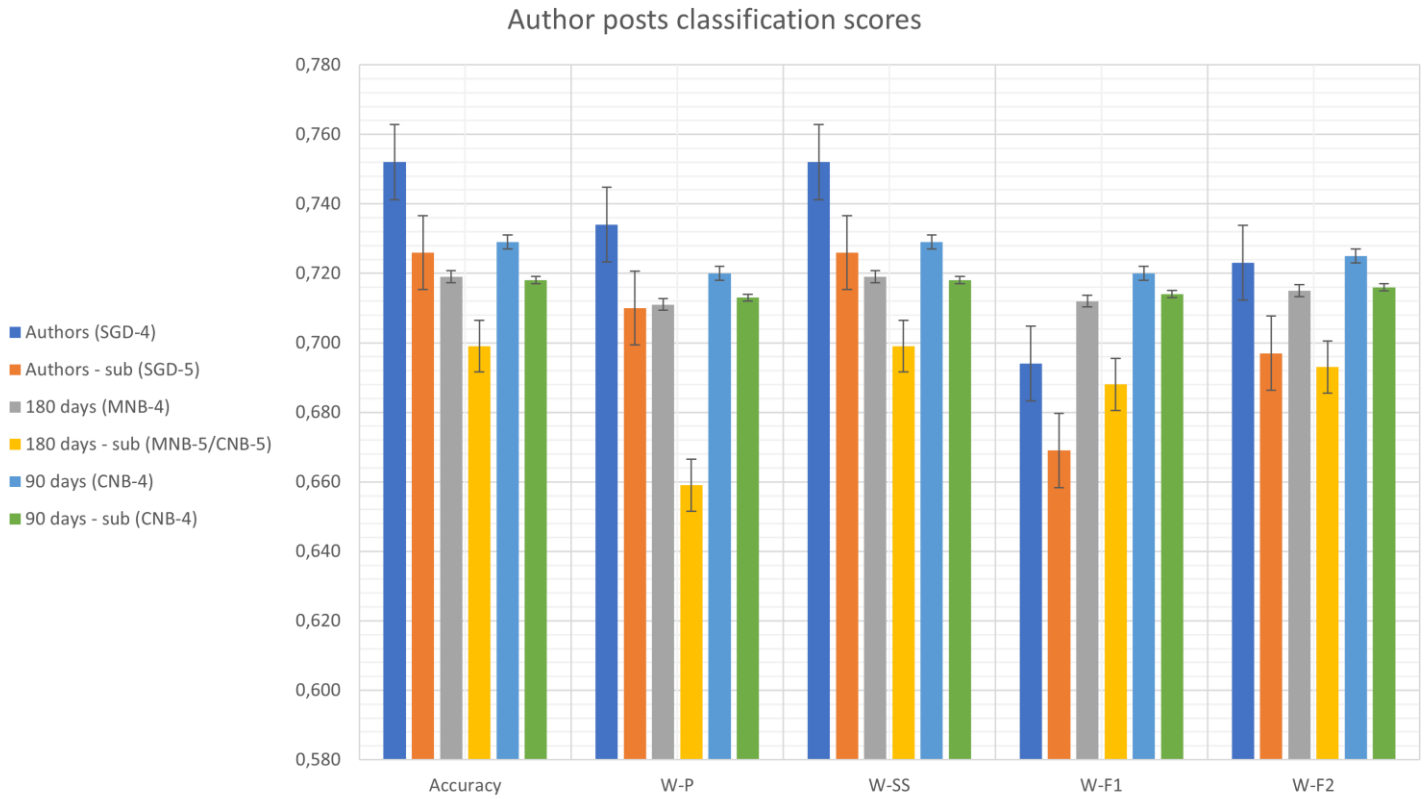
Best classifiers - I

	CLF	α	TYPE	Acc	W-P	W-SS	W-F1	W-F2
Posts	SGD-4	10^{-6}	TF-IDF 1-1	0.960	0.959	0.960	0.959	0.960
Authors	SGD-5	10^{-5}	TF-IDF 1-2	0.752	0.734	0.752	0.694	0.723
Authors - sub	SGD-5	10^{-5}	TF-IDF 1-2	0.726	0.710	0.726	0.669	0.697
180 days	MNB-4	0.1	TF-IDF 1-1	0.719	0.711	0.719	0.712	0.715
180 days - sub	MNB-5/CNB-5	0.1	TF-IDF 1-2	0.699	0.659	0.699	0.688	0.693
90 days	CNB-4	0.6	TF-IDF 1-1	0.729	0.720	0.729	0.720	0.725
90 days - sub	CNB-4	0.4	TF-IDF 1-1	0.718	0.713	0.718	0.714	0.716

Best classifiers - II (phase A)



Best classifiers – III (phase B)

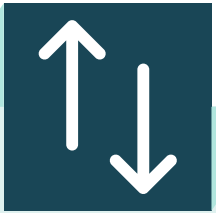


04
Conclusions

What could we “take home” from all of this?

Aside from numbers...

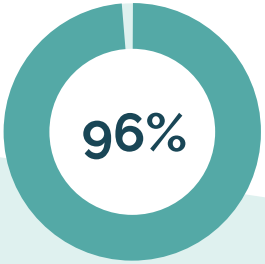
TF-IDF over BoW



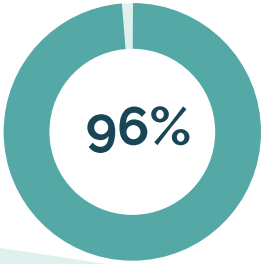
Bayesian for small datasets
and SGD for the largest

Phases' results

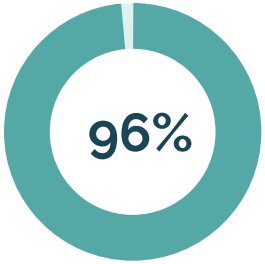
Accuracy



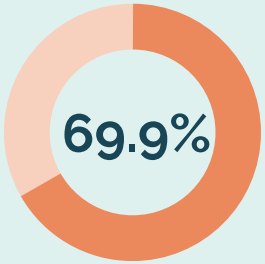
W-SS



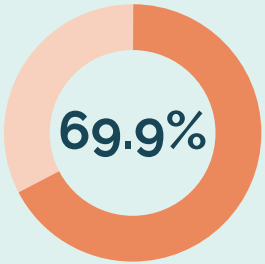
W-F2



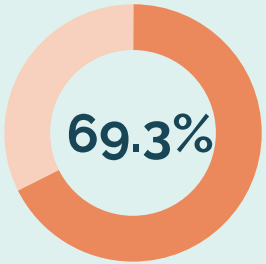
Accuracy



W-SS



W-F2



Future work...

WORD EMBEDDINGS

DEEP LEARNING

Questions?

Ignacio Montes Álvarez
uo257907@uniovi.es