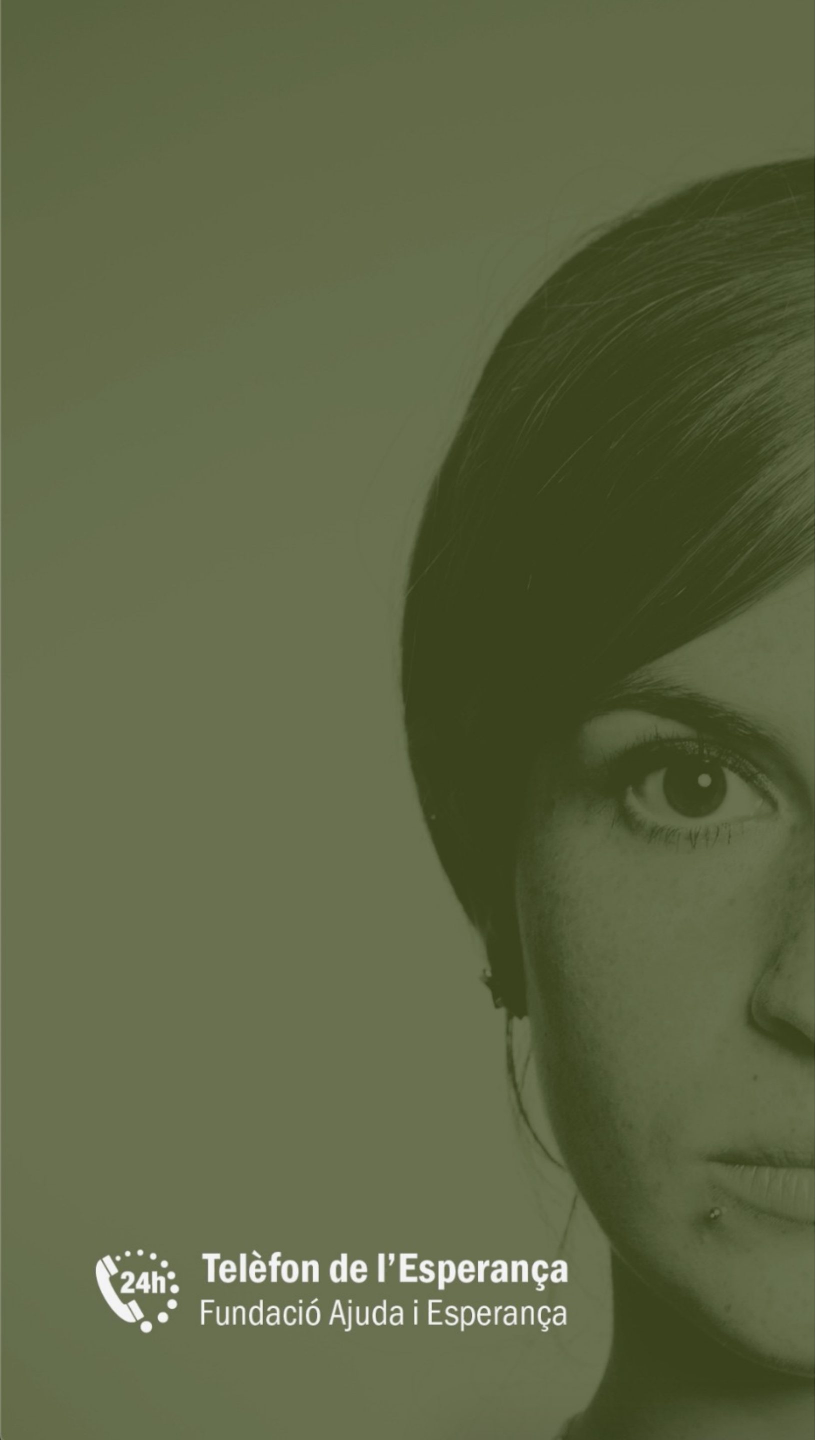


STOP PROJECT

STOP: Suicide prevenTion in sOcial Platforms

Ana Freire
Head of the Technology Unit
Senior Lecturer

upf. BARCELONA SCHOOL OF MANAGEMENT

@ana_ _freire

Ana Freire PhD Computer Science

STOP PROJECT

- Head of the Technology Unit - UPF Barcelona School of Management
- Senior Lecturer

upf. BARCELONA SCHOOL OF MANAGEMENT

- Founder

WISIBILÍZALAS

- Invited Researcher

YAHOO! LABS

National Research Council of Italy

University of Glasgow | School of Computing Science

- Sales Engineer

denodo

Premio Avanzadoras (1st Finalist) – Intermón Oxfam & 20minutos - 2021

Premio Mujeres a Seguir (Finalist) - 2020

23 tech leaders under 35- Business Insider - 2019

Premio al Joven Talento Científico Femenino - Real Academia de las Ciencias & Mastercard - 2019

Premio Ada Byron Jr – Universidad de Deusto - 2019 http://stop-project.upf.edu/

STOP PROJECT

STOP: Suicide prevenTion in sOcial Platforms

Studying mental health problems in social platforms
(suicide behaviour, depression, eating disorders)

upf. Universitat Pompeu Fabra Barcelona

Parc Taulí Sabadell Hospital Universitari

CVC Centre de Visió per Computador

fita. fundación instituto de trastornos alimentarios

24h Telèfon de l'Esperança Fundació Ajuda i Esperança

UNIVERSITÉ DE LYON

Università della Svizzera italiana

800.000

people die due to suicide every year

40

seconds per suicide

2nd

leading cause of death among 15-29-year-olds

World Health Organization

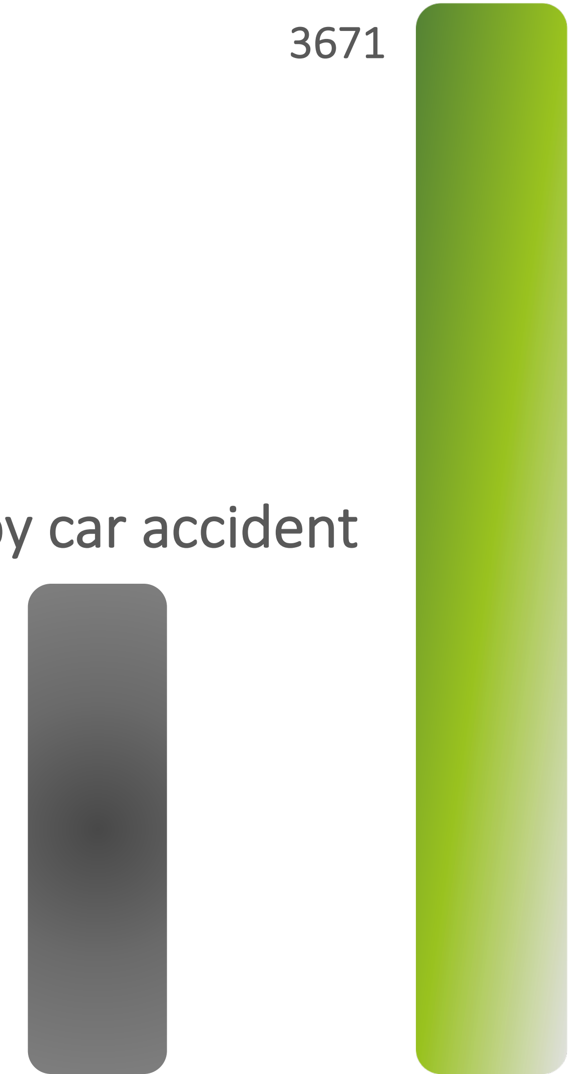
2019

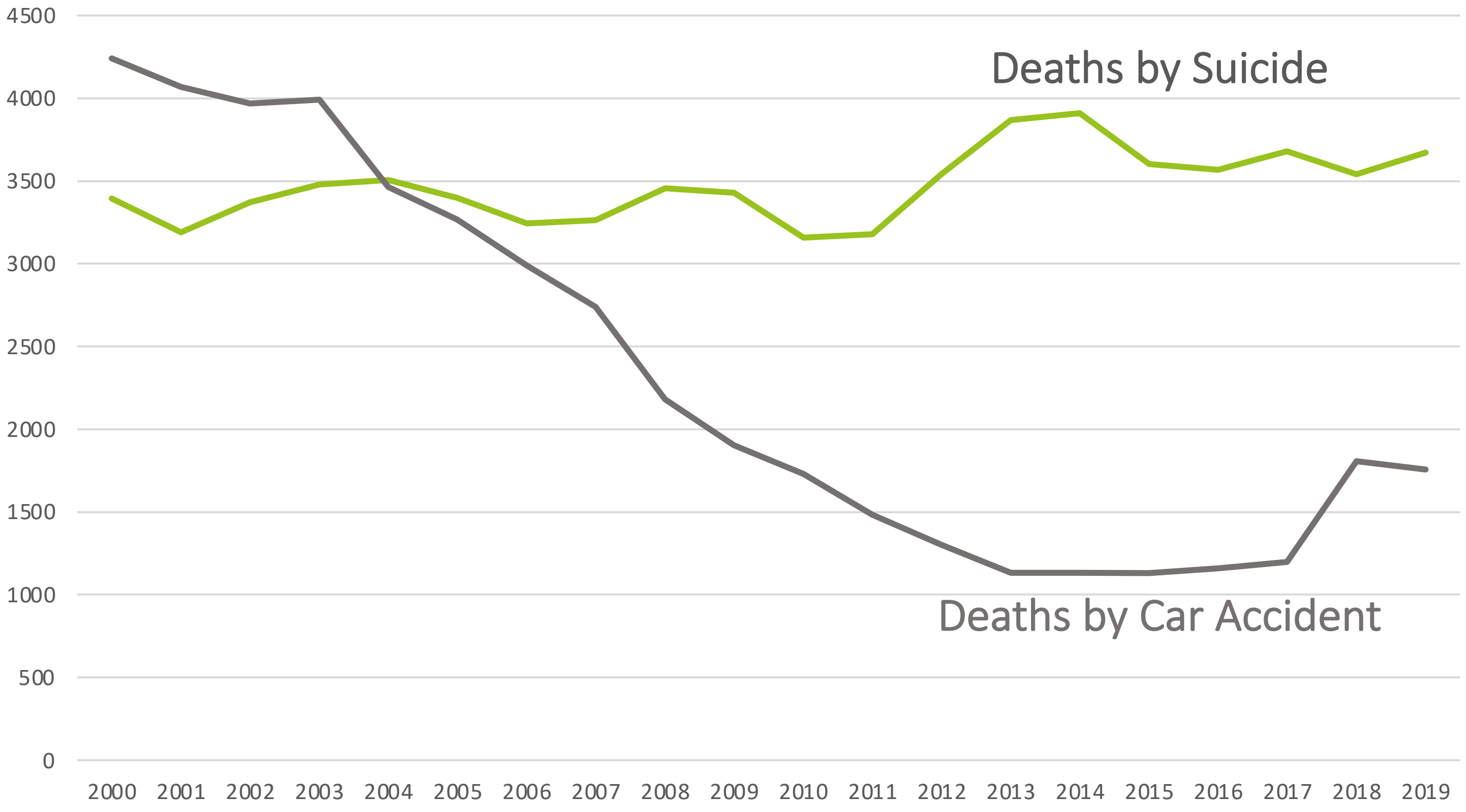
Deaths by Suicide

3671

Deaths by car accident

1755





Instituto Nacional de Estadística (INE)

$58 billion

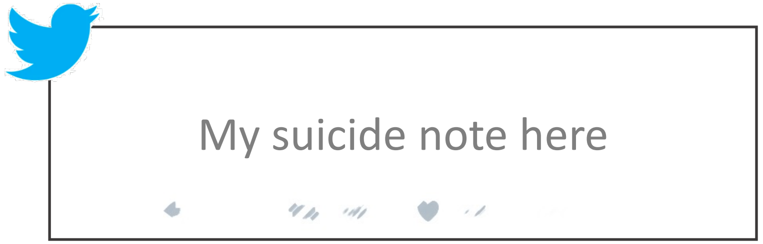
yearly **economic loss** of
suicide in USA

AU$22 billion

yearly **economic loss** of
youth suicide in Australia

Department of Health and Human Services of US. NIH Num. STR 18–6389

Kinchin, I., & Doran, C. M. (2018). The cost of youth suicide in Australia. International journal of environmental research and public health, 15(4), 672.



"Twitter is used by individuals to express suicidality and it is possible to distinguish the level of concern among suicide-related tweets, using machine learning."

O'Dea, Bridianne, et al. "Detecting suicidality on Twitter." *Internet Interventions* 2.2 (2015): 183-188.

Suicide ideation

Depression

Eating Disorders

Contributions

1. First methodology to generate a reliable Twitter dataset for suicide risk assessment
2. First user-level dataset for suicide risk assessment in Spanish
3. Method to obtain a subset of the user tweets related to a specific topic
4. Mapping between psychological consultations and social networks
5. First work in combining text-based, social networks and image—based features with data coming from 3 different social networks

Suicide Risk Detection on Social Media: A multi-modal approach

1. Methodology to create a reliable dataset for suicide risk assessment
2. Characterization of Twitter users with suicide ideation
3. Machine Learning based detection of users with suicide ideation
4. Main findings

Suicide Risk Detection on Social Media: A multi-modal approach

1. Methodology to create a reliable dataset for suicide risk assessment

2. Characterization of Twitter users with suicide ideation

3. Machine and Deep Learning based detection of users with suicide ideation

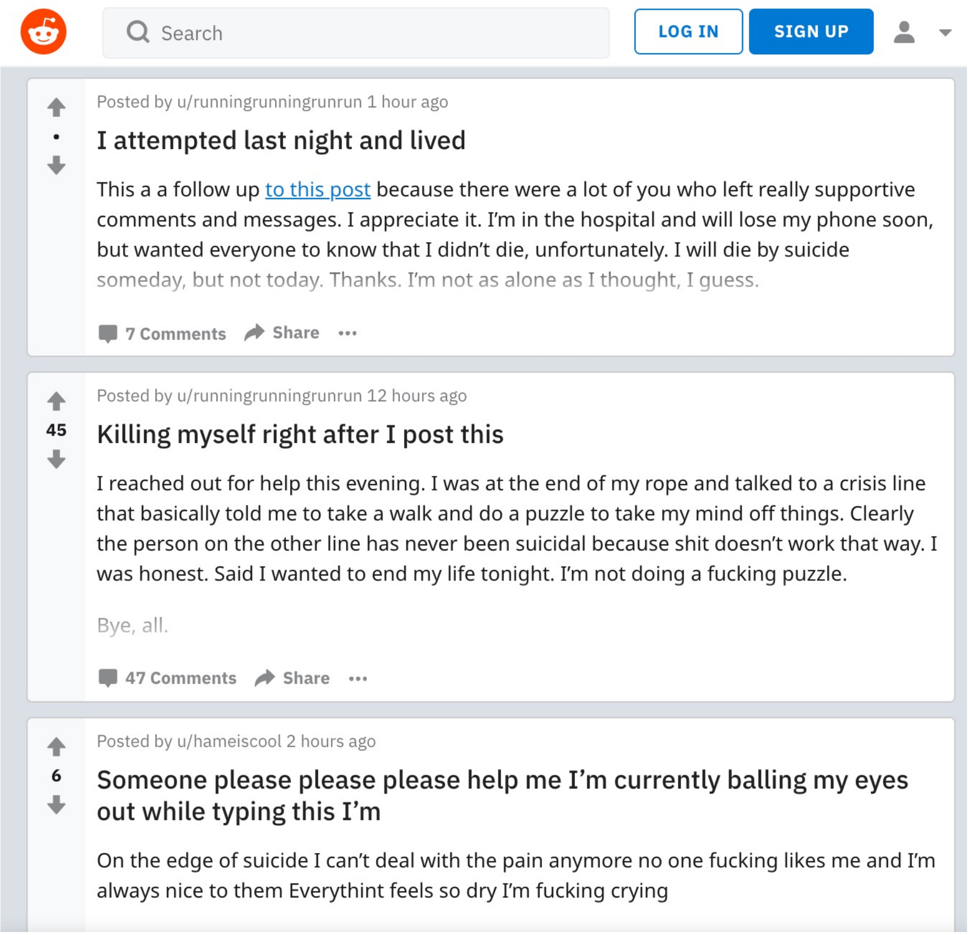
4. Main findings

Methodology to create a reliable dataset for suicide risk assessment

About 8000 tweets are published **by second** worldwide

Methodology to create a reliable dataset for suicide risk assessment

Extracting common expressions from Reddit /r/SuicideWatch/



500 titles

↓ Psychologists review

454 (110) sentences

Methodology to create a reliable dataset for suicide risk assessment

Twitter Crawling (Twitter API)

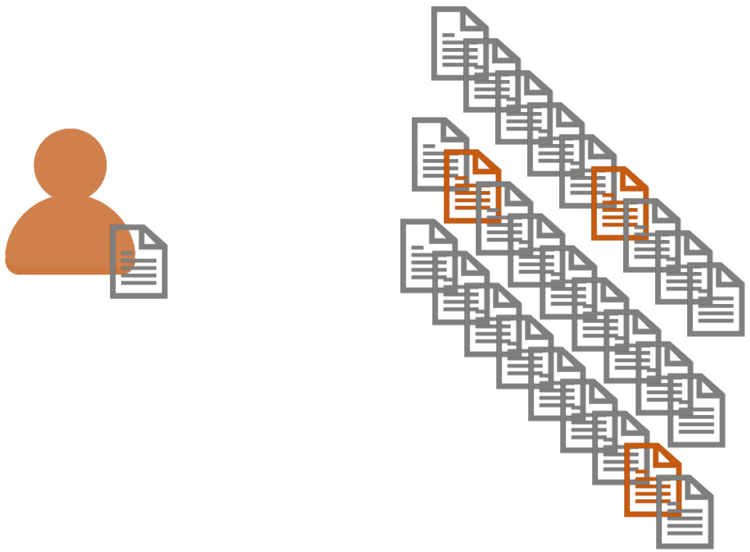
1 year period: Dec 21, 2017- Dec 21, 2018

98,619 tweets – 81,572 users

2-Level Annotation Process

Methodology to create a reliable dataset for suicide risk assessment

Short Profile Version



Short Profile Version Classifier (SPVC)

513 tweets (Suicide Risk) + 346 sentences
859 random tweets (Twitter API)

BoW (1-5 grams) + PCA + Logistic Regression
10-fold CV

F-measure: 0.90
Precision = 0.91
Recall = 0.89

Methodology to create a reliable dataset for suicide risk assessment

Twitter Crawling (Scrapping + Twitter API)

1 year period: Dec 21, 2017- Dec 21, 2018

98,619 tweets – 81,572 users

2-Level Annotation Process

Suicide Risk Detection on Social Media: A multi-modal approach

1. Methodology to create a reliable dataset for suicide risk assessment

2. Characterization of Twitter users with sucide ideation

3. Machine Learning based detection of users with suicide ideation

4. Main findings

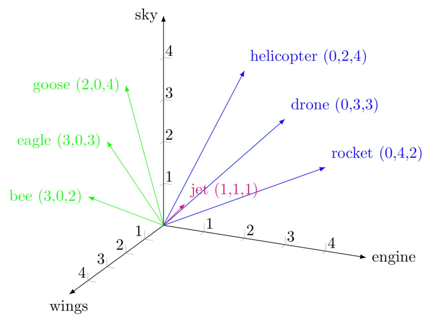
Characterization of Twitter users with suicide ideation

Features

- Textual features
- SNPSY: Social Network and Psychological features
- Image-based features

Characterization of Twitter users with suicide ideation

Features: Textual Features



- Bag of Words (BoW) and n-grams
- Word Embeddings
- Lexicons and suicide risk factors vocabulary - **LIWC: Linguistic Inquiry and Word Count LIWC 2007 Spanish dictionary:**
  - All LIWC Categories
  - Other categories added: self-injuries, explicit suicide ideation references, self-loathing terms, insomnia, fear, previous suicide attempts, racial or sexual discrimination, eating disorders, substances abuse, bullying, lack of social support, family and money issues, lack of spiritual beliefs, etc.
- Sentiment Analysis - **senti-py**: polarity of Spanish texts.

Characterization of Twitter users with suicide ideation

Features: Social Network and Psychological features

Behavioral features

Feature	Description	Source
Working week tweets count ratio	Total number of tweets on week days (Monday to Friday) normalized by the total amount of tweets.	SPV
Weekend tweets count ratio	Total number of tweets on weekend days (Saturday and Sunday) normalized by the total amount of tweets	SPV
Median time between tweets	Median of the time (in seconds) that passes between the publication of each tweet.	SPV
Sleep time tweets ratio	Ratio of tweets posted during the inferred sleep period of the user.	Full profile
Daytime tweets ratio	Ratio of tweets posted during the period the user is usually awake	Full profile
Normalized tweet count per quarter (4 features)	Number of tweets posted by the user within each quarter of the year, normalized by the total amount of tweets generated by the user during the year.	SPV

Characterization of Twitter users with suicide ideation

Features: Social Network and Psychological features

Behavioral features

Sleep time tweets ratio	Ratio of tweets posted during the inferred sleep period of the user.

$$STTR = \frac{\min_{i=0..7}\{t_i + t_{(i+1) mod\ 8}\}}{T}$$

Characterization of Twitter users with suicide ideation

Features: Social Network and Psychological features

Tweet Statistics

Feature	Description	Source
Suicide related tweets ratio	Ratio of tweets retained by the SPVC over all the tweets of the full profile	SPV and full profile
Median SPVC score	Median of the scores obtained by the tweets that are part of the SPV after applying the SPVC	SPV
Median tweet length	Median length of all the user tweets (word level)	SPV
Number of SPV tweets	Number of tweets	SPV
Number of user tweets	Number of tweets posted by the user since the creation of the account	User metadata

Characterization of Twitter users with suicide ideation

Features: Social Network and Psychological features

Relational Features

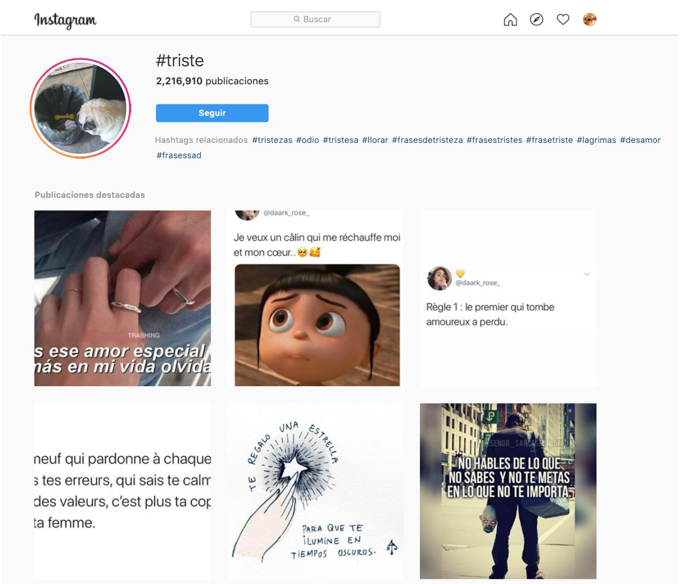
Feature	Description	Source
Followers number	Number of followers	User metadata
Friends number	Number of people followed by the user	User metadata
Favorites (favs) given	Total number of favs given by the user	User metadata
Median favs count	Median of all the favorites received by the user	SPV
Median retweets count	Median of all the retweets received by the user	SPV

Characterization of Twitter users with suicide ideation

Features: Image-based features

Suicide related keywords → Suicide images

Random keywords → Control images



90,000 images for training + 60,000 for validation

Characterization of Twitter users with suicide ideation

Features: Image-based features

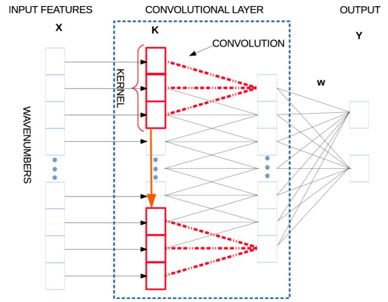


Image user score 1

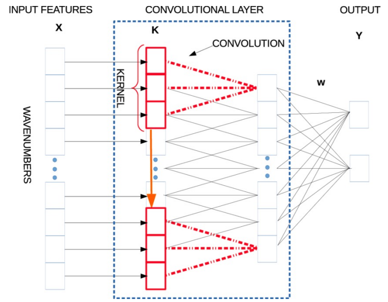


Image user score 2

AVG

User's aggregated score

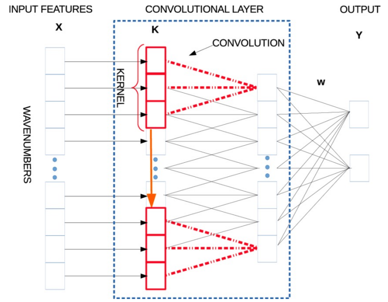


Image user score 3

Convolutional Neural Networks

Suicide Risk Detection on Social Media: A multi-modal approach

1. Methodology to create a reliable dataset for suicide risk assessment
2. Characterization of Twitter users with sucide ideation
3. Machine Learning based detection of users with suicide ideation
4. Main findings

ML based detection of users with suicide ideation

Experimental Setup

CONTROL GROUPS

1. Focused control group: users writing suicide-related keywords in a non-suicide risk context: users that either trivialize about suicide, news reports, information regarding the topic or simply manifest their support or opinions to people at risk. **Identifying these users is challenging for the classification systems but is key in reducing false positives**. These users were chosen at random among the users labeled as control cases during the first annotation process.

2. Generic control group: a set of Twitter users, which might not necessarily use terms related with suicide. These users were **selected randomly** using the Sample tweets API [31].

For both control groups, a second annotation process was followed as well to discard possible cases of users at risk within these samples.

ML based detection of users with suicide ideation

Experimental Setup

BASELINES

1. Baseline 1: BoW model trained with (1-5)-grams;
2. Baseline 2: the second baseline consists of a deep learning model defined by a Convolutional Neural Network (CNN) architecture

ML METHODS

Random Forest, Multilayer Perceptron, Logistic Regression and Support Vector Machines (SVM)

CNN (Embeddings model)

ML based detection of users with suicide ideation

	Suicide vs. Focused Control Group						Suicide vs. Generic Control Group					
Model	P	R	F1	A	AUC	Classifier	P	R	F1	A	AUC	Classifier
BoW model (baseline 1)	0.78	0.81	0.79	0.78	0.81	MLP	0.79	0.85	0.81	0.80	0.91	MLP
Embeddings model (baseline 2)	0.76	0.81	0.79	0.77	0.82	CNN	0.78	0.87	0.82	0.80	0.84	CNN
BoW model (SPV)	0.81	0.85	0.83	0.82	0.85	LR	0.80	0.92	0.86	0.84	0.89	MLP
Embeddings model (SPV)	0.79	0.85	0.82	0.80	0.83	CNN	0.77	0.87	0.82	0.80	0.82	CNN
SNPSY model	0.85	0.85	0.85	0.84	0.86	SVM	0.85	0.88	0.87	0.86	0.94	LR
(BoW + SNPSY) model	0.82	0.88	0.85	0.84	0.89	RF	0.85	0.88	0.87	0.86	0.94	LR
(Images + BoW) model	0.79	0.88	0.84	0.82	0.86	MLP	0.82	0.88	0.85	0.84	0.90	LR
(Images + SNPSY) model	0.88	0.85	0.86	0.86	**0.91**	**SVM**	0.88	0.88	0.88	0.88	0.94	LR
(Images + BoW + SNPSY) model 1	0.85	0.85	0.85	0.83	0.87	LR	0.85	0.92	0.88	0.88	0.92	MLP
(Images + BoW + SNPSY) model 2	0.88	0.81	0.84	0.84	**0.92**	**SVM**	0.85	0.88	0.87	0.86	0.94	LR
Selected features model 1 (P<0.05)	0.85	0.85	0.85	0.84	0.90	MLP	0.91	0.77	0.83	0.84	0.94	SVM
Selected features model 2 (P<0.001)	0.83	0.77	0.80	0.80	**0.92**	**SVM**	0.91	0.81	0.86	0.86	**0.95**	SVM

Suicide Risk Detection on Social Media: A multi-modal approach

1. Methodology to generate a reliable Twitter dataset for suicide risk assessment
2. Characterization of Twitter users with sucide ideation
3. Machine Learning based detection of users with suicide ideation
4. Main findings

Main findings (suicide behaviour)

- ML method for detecting users in risk of suicide: AUC = 0.92%
- SPVC (Short Profile Version Classifier) is successful as a first filter.
- Considering the different feature types individually, SNPSY (Social Network and Psychological features) gets the best results.
- Image + BoW + SNPSY model 2: 11% increasing AUC regarding baseline.
- Better results are achieved in terms of AUC when using generic control users instead of users that make use of suicidal vocabulary.

Main findings (suicide behaviour)

- We found multiple significant features that can be used to detect users at risk:
  - Use of first person (self-references)
  - More use of negations
  - They tend to show depression and anxiety.
  - **Topics**: usage of suicide explicit terms, depression related terms, self-loathing, substance abuse, self-injuries, and terms expressing lack of social support
  - More **images** related to suicide
  - Shorter tweets
  - Less number of friends
  - More **activity** on weekends

JMIR Publications
Advancing Digital Health & Open Science

Articles | Search articles

Journal of Medical Internet Research | Journal Information | Browse Journal | Sub

Published on 7.7.2020 in Vol 22, No 7 (2020): July

Preprints (earlier versions) of this paper are available at https://preprints.jmir.org/preprint/17758, first published January 16, 2020.

Detection of Suicidal Ideation on Social Media: Multimodal, Relational, and Behavioral Analysis

Diana Ramírez-Cifuentes [1]; Ana Freire [1]; Ricardo Baeza-Yates [1]; Joaquim Puntí [2, 3]; Pilar Medina-Bravo [4]; Diego Alejandro Velazquez [5]; Josep Maria Gonfaus [6]; Jordi Gonzàlez [5]

Article | Authors | Cited by (4) | Tweetations (47) | Metrics

- Abstract
- Introduction
- Methods
- Results
- Discussion
- References
- Abbreviations
- Copyright

Abstract

Background:
Suicide risk assessment usually involves an interaction between doctors and patients. However, a significant number of people with mental disorders receive no treatment for their condition due to the limited access to mental health care facilities; the reduced availability of clinicians; the lack of awareness; and stigma, neglect, and discrimination surrounding mental disorders. In contrast, internet access and social media usage have increased significantly, providing experts and patients with a means of communication that may contribute to the development of methods to detect mental health issues among social media users.

Objective:
This paper aimed to describe an approach for the suicide risk assessment of Spanish-speaking users on social media. We aimed to explore behavioral, relational, and multimodal data extracted from multiple social platforms and develop machine learning models to detect users at risk.

JMIR Publications
Advancing Digital Health & Open Science

Articles | Search articles

Journal of Medical Internet Research | Journal Information | Browse Journal

Published on 20.7.2021 in Vol 23, No 7 (2021): July

Preprints (earlier versions) of this paper are available at https://preprints.jmir.org/preprint/25925, first published November 21, 2020.

Characterization of Anorexia Nervosa on Social Media: Textual, Visual, Relational, Behavioral, and Demographical Analysis

Diana Ramírez-Cifuentes [1]; Ana Freire [1, 2]; Ricardo Baeza-Yates [1, 3]; Nadia Sanz Lamora [4]; Aida Álvarez [4, 5, 6]; Alexandre González-Rodríguez [4, 5, 6]; Meritxell Lozano Rochel [7]; Roger Llobet Vives [7]; Diego Alejandro Velazquez [8]; Josep Maria Gonfaus [9]; Jordi Gonzàlez [8]

Article | Authors | Cited by | Tweetations (25) | Metrics

- Abstract
- Introduction
- Methods
- Results
- Discussion
- Abbreviations
- Copyright

Abstract

Background:
Eating disorders are psychological conditions characterized by unhealthy eating habits. Anorexia nervosa (AN) is defined as the belief of being overweight despite being dangerously underweight. The psychological signs involve emotional and behavioral issues. There is evidence that signs and symptoms can manifest on social media, wherein both harmful and beneficial content is shared daily.

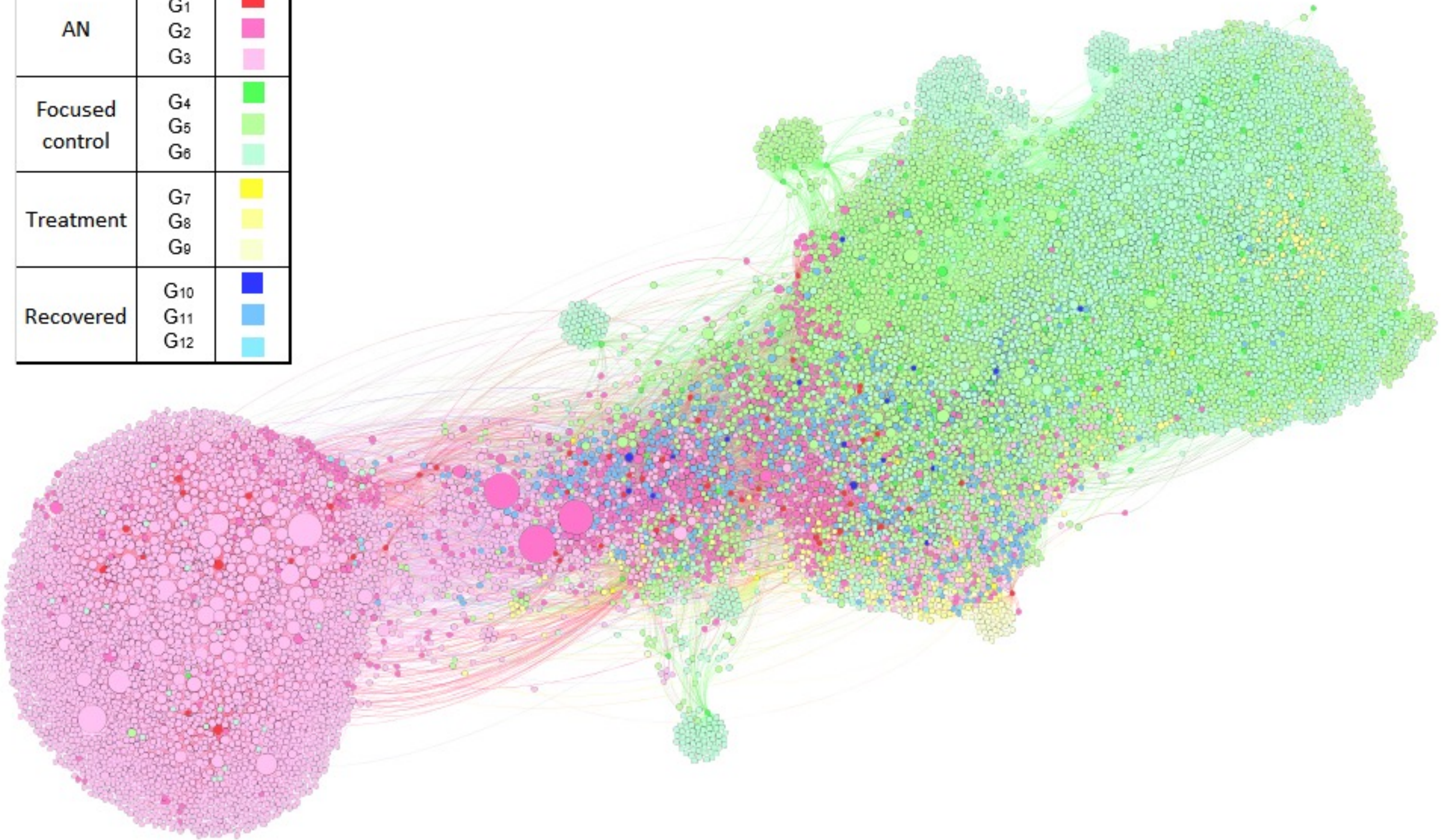
Objective:
This study aims to characterize Spanish-speaking users showing anorexia signs on Twitter through the extraction and inference of behavioral, demographical, relational, and multimodal data. By using the transtheoretical model of health behavior change, we focus on characterizing and comparing users at the different stages of the model for overcoming AN, including treatment and full recovery periods.

STOP PROJECT

Main findings (eating disorders)

- We found multiple significant features that can be used to detect users at risk:
  - Use of first person (self-references)
  - More use of negations
  - They tend to show sadness, disgust and anger.
  - **Topics**: laxatives, suicide, death, bullying (at later stages of the disorder)
  - They tend to use darker **images**, with fewer letters, selfies, images of body parts and altered images of idealized bodies
  - Few **interaction** with users
  - More **activity** on weekends and during sleeping periods

Class	Group	Color
AN	G1 G2 G3	
Focused control	G4 G5 G6	
Treatment	G7 G8 G9	
Recovered	G10 G11 G12	



Suicide Ideation

<39 years old
55.6% Female

Topics: usage of suicide explicit terms, depression related terms, self-loathing, substance abuse, self-injuries, and terms expressing lack of social support

Depression

<29 years old
58% Female

Topics: loneliness, insomnia, low self-confident, divorce, marital problems, bullying, antidepressants...

Eating Disorders

<29 years old
84% Female

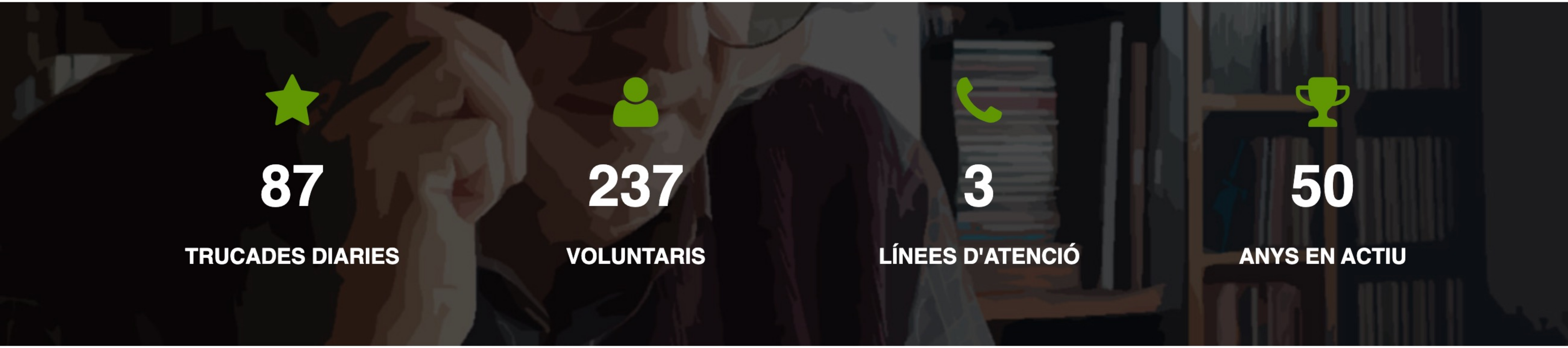
Topics : veganism, vegetarian diets, extreme exercise, laxatives, calories, low self-esteem, shame, self-harm, family ...

Telèfon de l'Esperança de Barcelona
Fundació Ajuda i Esperança

Sempre que et sentis EN CRISI t'escoltem i ajudem a reduir el patiment.

Acompanyem en situacions emocionals crítiques cercant conjuntament camins alternatius per atenuar el patiment.

Truca | Et truquem




Si estás pasando por un momento difícil, expresarlo te ayudará.
Llámanos, te escuchamos
682 900 500
24h
Telèfon de l'Esperança
Fundació Ajuda i Esperança

Te sientes sola?
Sientes que la vida te supera?

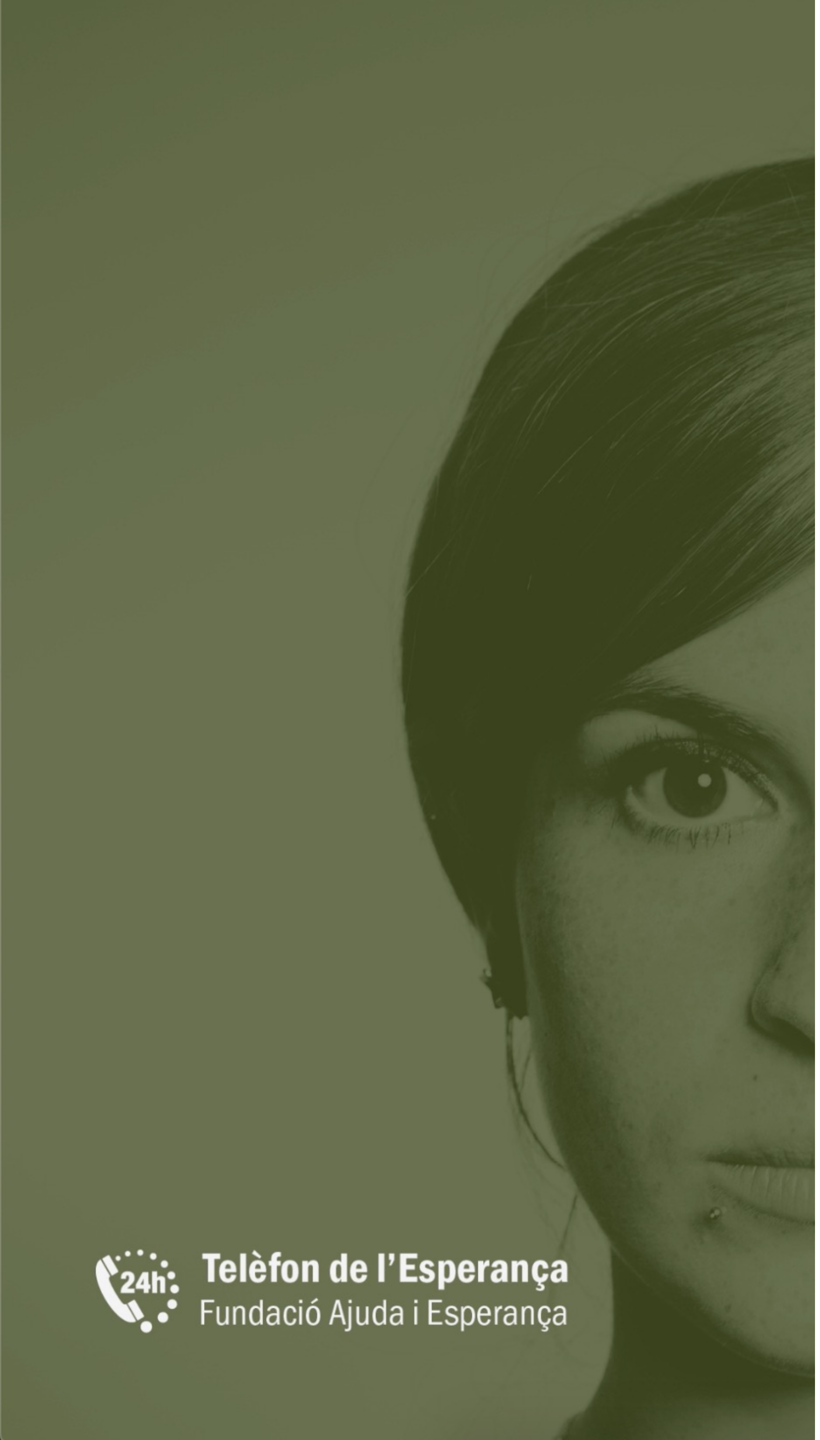
Hablar puede ayudarte,

Llámanos.

682 900 500

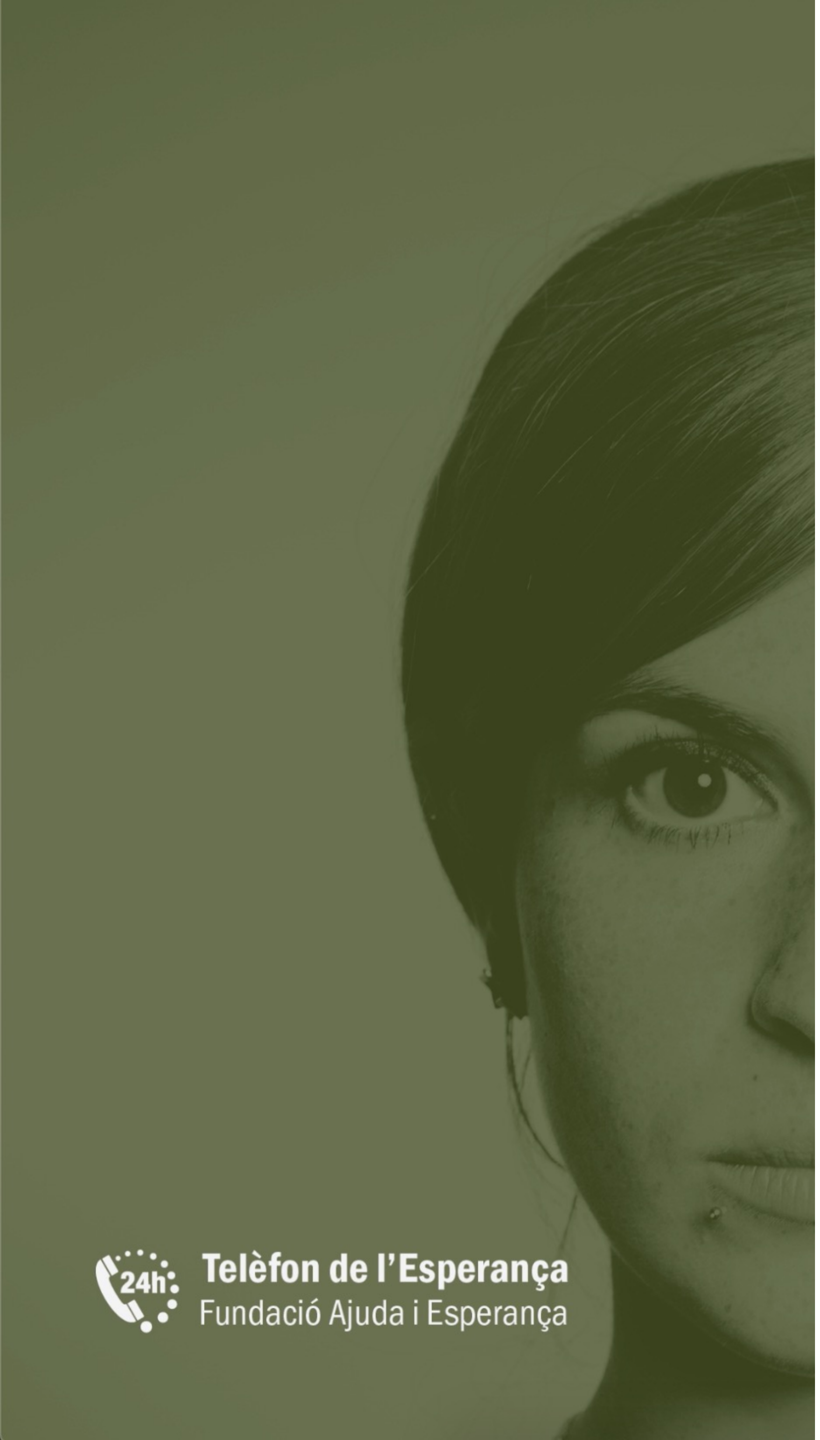
Telèfon de l'Esperança
Fundació Ajuda i Esperança


Fundació Ajuda i Esperança
Ajuntament de Barcelona
Si estás pasando por un momento difícil, expresarlo puede ayudarte.
Llámanos, te escuchamos
900 92 55 55
TELÉFONO DE PREVENCIÓN DEL SUICIDIO



STOP PROJECT

60%



STOP PROJECT

STOP: Suicide prevenTion in sOcial Platforms

Ana Freire
Head of the Technology Unit
Senior Lecturer
upf. BARCELONA SCHOOL OF MANAGEMENT
@ana_ _freire