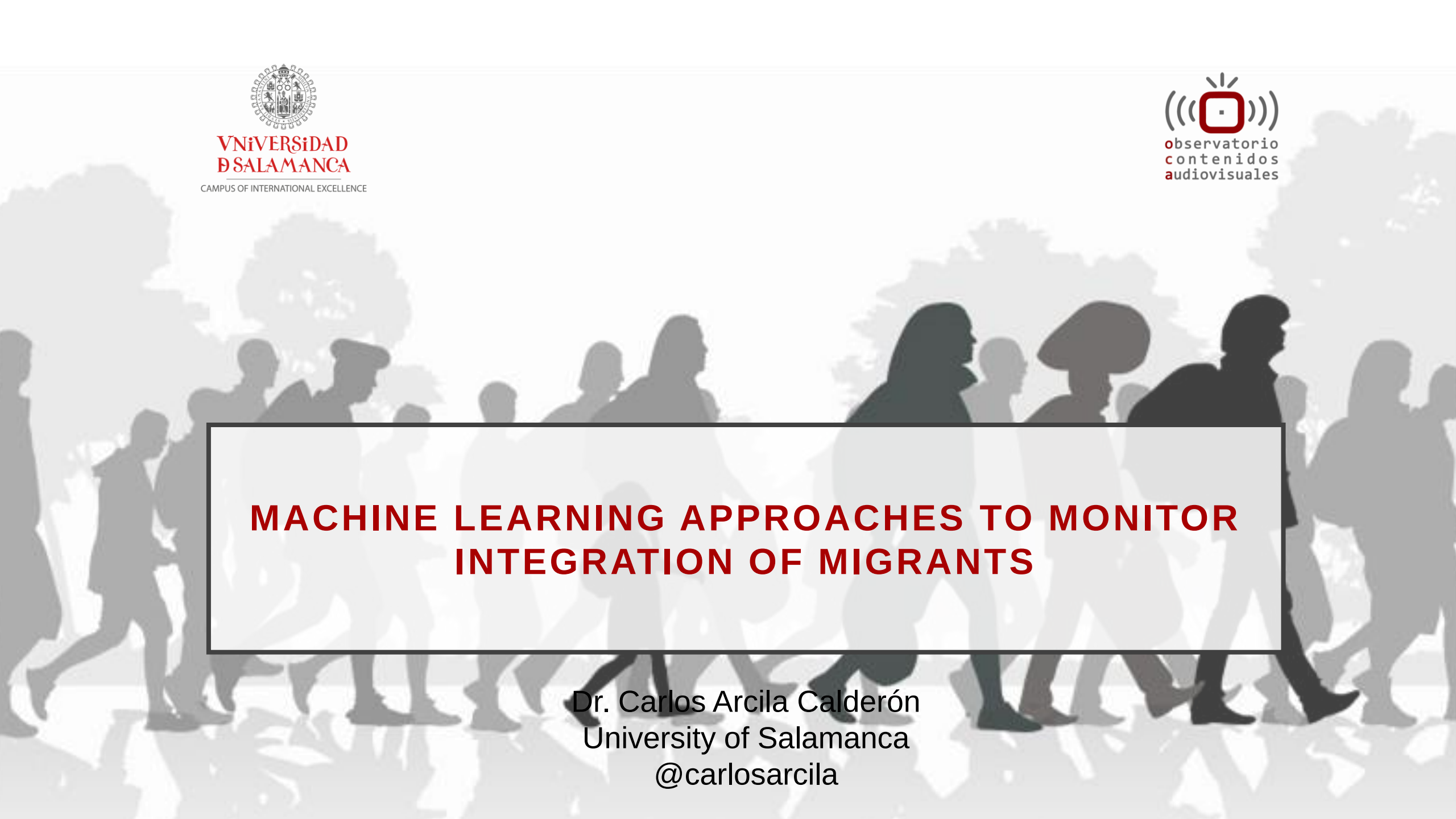


MACHINE LEARNING APPROACHES TO MONITOR INTEGRATION OF MIGRANTS

Dr. Carlos Arcila Calderón
University of Salamanca
@carlosarcila

BIG DATA ANALYSIS
FOR MIGRATION STUDIES

- Collecting and modelling **unstructured data** (e.g. social media contents, parliamentary speeches, video records, etc.).
  - Case of study: Large-scale detection of hate speech against migrants and refugees.
- Collecting and modelling **structured data** (e.g. databases, surveys, official statistics, etc.).
  - Case of study: Using machine learning and synthetic populations to predict social acceptance of asylum seekers in European regions

INTRODUCTION

- The mass arrival of forced displaced people in Europe is generating an unprecedented political and demographic impact
- In this context, **migration management has become one of the main challenges** for politicians and competent organizations
- Related to this topic, some of the main programs of the European Union and the UN Refugee Agency (UNHCR) in recent years was intended to resettle new refugees arriving in Europe and to relocate refugees and asylum seekers who remained in first-reception countries

CONTEXT

- One of the most important variables for resettlement processes to occur in a positive way is the **probability of acceptance of the receiving societies**
- Until now, the criteria taken into account to manage migration and refuge have been fundamentally economic and political (Jones, Teytelboym 2017), but the **contexts of reception by the local population have not been taken into account**

- **The prediction of attitudes at the regional level would mitigate the risks of exclusion** through the planning of integration policies, anticipating which populations could accept or oppose more resistance to the arrival of refugees

CONTEXT

- Previous studies indicate that **the perception of migration is strongly influenced by individual sociodemographic characteristics** (Finney and Peach, 2004)
- Some of these variables may be ***nationality***, ***age***, ***education***, ***income*** or ***political ideology***
- This gives enormous importance to sociodemographic factors to predict refugee acceptance, since **they can condition the degree of acceptance or rejection of refugees** (Finney and Peach, 2004; Schweitzer et al., 2005; Cameron et al., 2006; Song, 1992)

CONTEXT

- It is also known that **the level of acceptance of migrants and refugees varies between countries, but also between regions** (Crawley, Drinkwater and Kauser, 2013)

- However, **most refugee support surveys** (such as the Eurobarometer or the European Social Survey) **do not have enough geo-localized data** to make estimates at the regional or local level

BIG DATA SOLUTIONS

- We can **estimate the probability of acceptance of refugees in each European region based on:**
  - **data from social media using supervised text classification to detect hate speech against migrants**
  - **data from existing surveys using computational approaches that combine machine learning and synthetic populations**

DATA-BASED MIGRATION PLANNING

- Administrations continue to base their decisions on traditional studies and data
- However, **the academy has already begun to understand the great potential of big data and computational methods to study social processes and movements, including migrant and refugee flows**
- These studies based on large amounts of data should serve as a basis for making decisions and managing migration flows

DATA-BASED MIGRATION PLANNING

Some examples of these types of studies are:

- Hawelka et al (2014) and Zagheni et al (2014): They use geographically placed Twitter messages to identify global mobility patterns and predict migratory flows

- Lamana et al (2018): They use data extracted from Twitter, such as language and spatial-temporal communication patterns of individuals, to estimate the power of social integration of different cities in the world

DATA-BASED MIGRATION PLANNING

Bansak et al (2018): These authors **develop a model that uses multiple data to assign a specific place to refugees according to their profile**, and thus improve integration

They use information about the basic characteristics of refugees (country of origin, language skills, gender, age, etc.), time of arrival, assigned location and average employment success

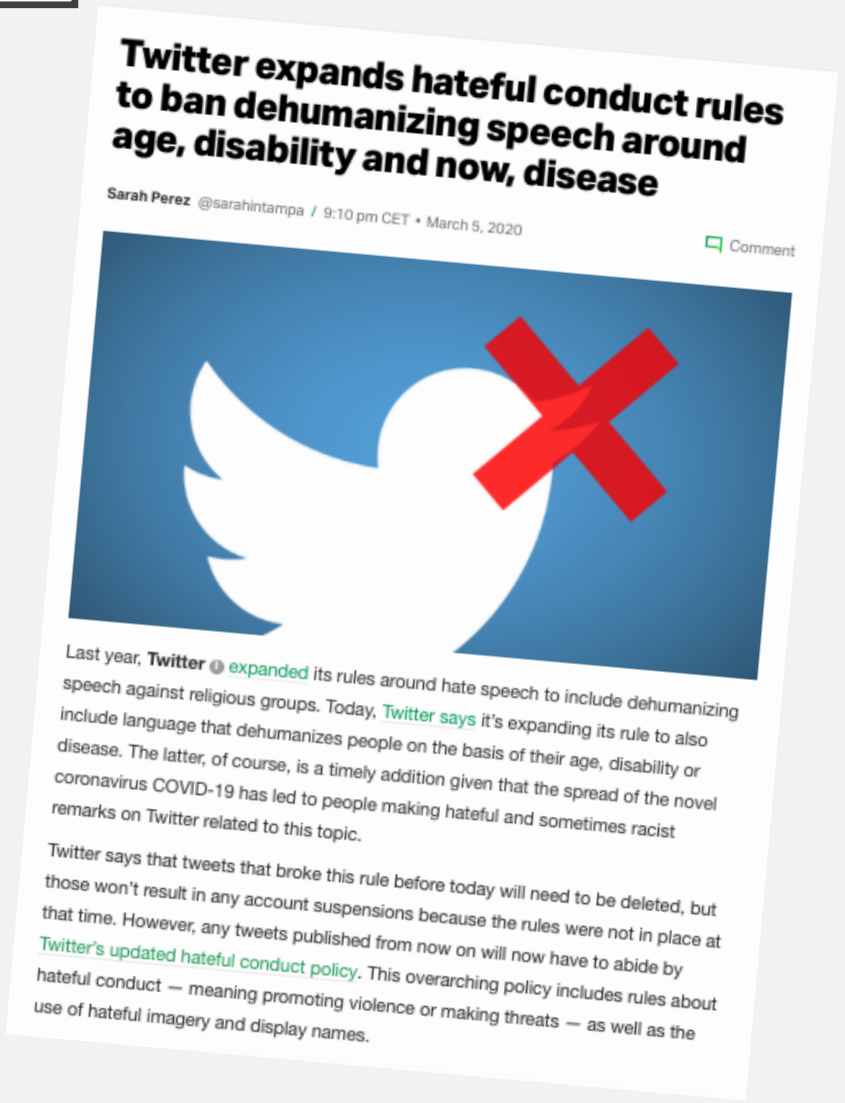
So **they create a set of supervised machine learning models to predict the expected job success of refugees based on their background characteristics**

SOLUTION 1

- **Big data from social media using supervised text classification to detect hate speech against migrants**

INTRODUCTION

- Online hate speech has increased exponentially in recent years
- This increase is especially worrying on social media like Twitter (Bartlett et al., 2014)
- Twitter has updated its Hate Speech Policy to counter the hateful conduct



Twitter expands hateful conduct rules to ban dehumanizing speech around age, disability and now, disease

Sarah Perez @sarahintampa / 9:10 pm CET • March 5, 2020

Comment

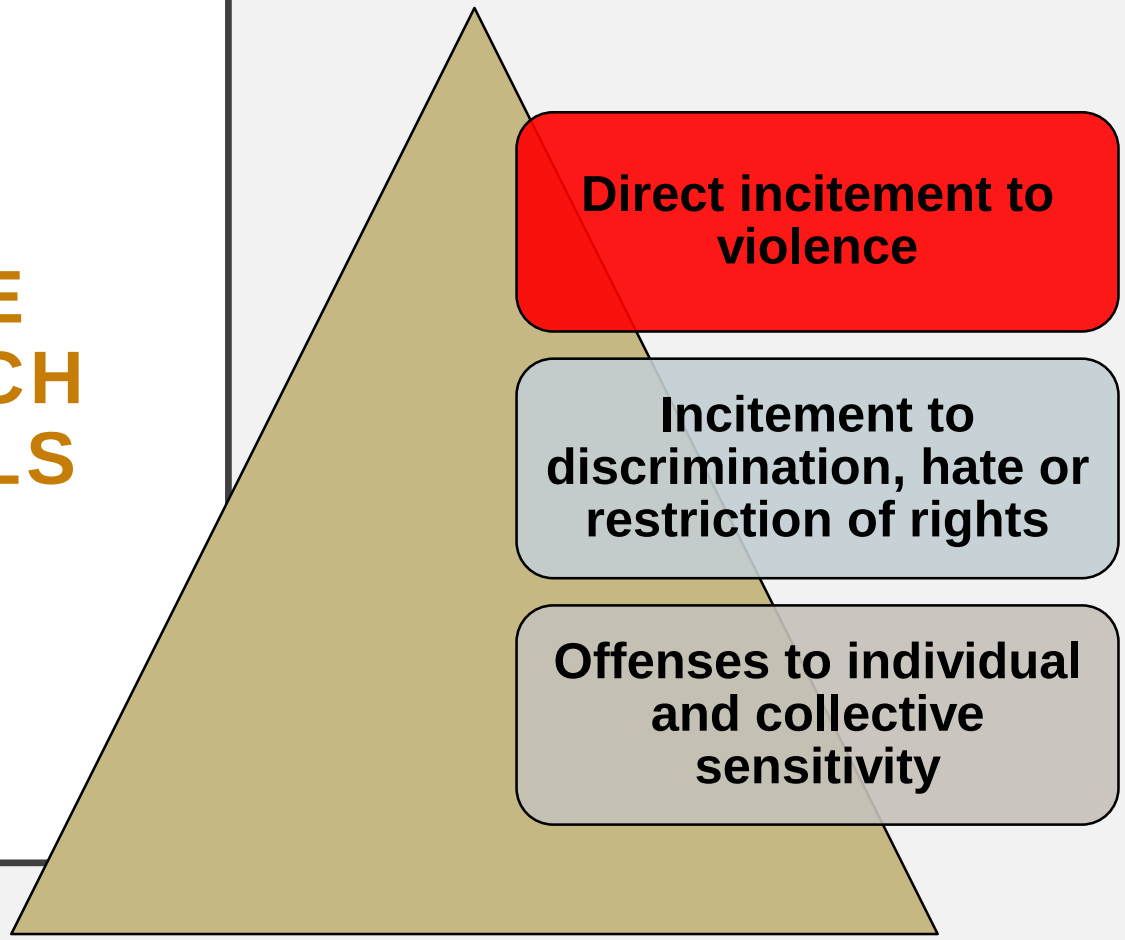
Last year, **Twitter** expanded its rules around hate speech to include dehumanizing speech against religious groups. Today, Twitter says it's expanding its rule to also include language that dehumanizes people on the basis of their age, disability or disease. The latter, of course, is a timely addition given that the spread of the novel coronavirus COVID-19 has led to people making hateful and sometimes racist remarks on Twitter related to this topic.

Twitter says that tweets that broke this rule before today will need to be deleted, but those won't result in any account suspensions because the rules were not in place at that time. However, any tweets published from now on will now have to abide by Twitter's updated hateful conduct policy. This overarching policy includes rules about hateful conduct — meaning promoting violence or making threats — as well as the use of hateful imagery and display names.

- Hate speech implies promoting messages that encourage **rejection, humiliation, harassment, discrediting and stigmatization of individuals or social groups**
- Any type of expression that **propagates, promotes or justifies hate** and intolerance towards the otherness

(Council of Europe, Recommendation Nº 97, 1997)

HATE SPEECH LEVELS



Source: Miró (2016)

JUSTIFICATION

- Recent studies establish **correlations** between **hate speech** spread through social media such as Twitter and **hate crimes** that take place in particular locations

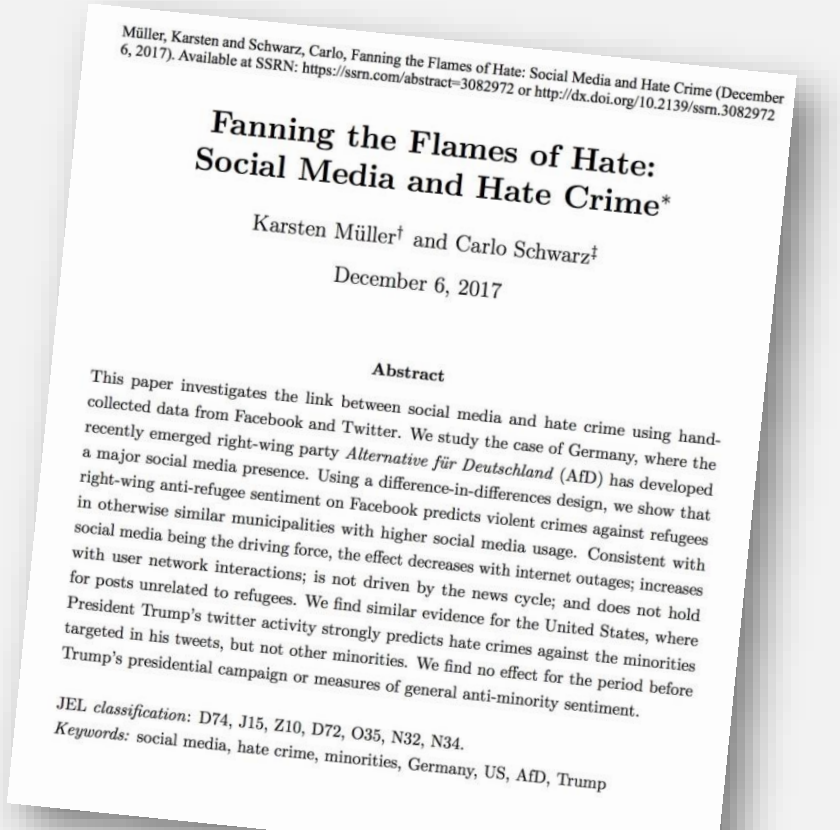
Prejudices

↓

Speech

↓

Crimes



Müller, Karsten and Schwarz, Carlo, Fanning the Flames of Hate: Social Media and Hate Crime (December 6, 2017). Available at SSRN: https://ssrn.com/abstract=3082972 or http://dx.doi.org/10.2139/ssrn.3082972

Fanning the Flames of Hate: Social Media and Hate Crime*

Karsten Müller† and Carlo Schwarz‡

December 6, 2017

Abstract

This paper investigates the link between social media and hate crime using hand-collected data from Facebook and Twitter. We study the case of Germany, where the recently emerged right-wing party *Alternative für Deutschland* (AfD) has developed a major social media presence. Using a difference-in-differences design, we show that right-wing anti-refugee sentiment on Facebook predicts violent crimes against refugees in otherwise similar municipalities with higher social media usage. Consistent with social media being the driving force, the effect decreases with internet outages; increases with user network interactions; is not driven by the news cycle; and does not hold for posts unrelated to refugees. We find similar evidence for the United States, where President Trump's twitter activity strongly predicts hate crimes against the minorities targeted in his tweets, but not other minorities. We find no effect for the period before Trump's presidential campaign or measures of general anti-minority sentiment.

JEL *classification*: D74, J15, Z10, D72, O35, N32, N34.

Keywords: social media, hate crime, minorities, Germany, US, AfD, Trump

JUSTIFICATION

- Recently **several prototypes of online hate speech detectors have been developed**

- All of them have significant limitations

Salminen et al. Hum. Cent. Comput. Inf. Sci. (2020) 10:1
https://doi.org/10.1186/s13673-019-0205-6

RESEARCH

Human-centric Computing and Information Sciences

Open Access

Developing an online hate classifier for multiple social media platforms

Joni Salminen[1,2*], Maximilian Hopf[3], Shammur A. Chowdhury[1], Soon-gyo Jung[1], Hind Almerekhi[4] and Bernard J. Jansen[1]

*Correspondence: jooleag@utu.fi
[2] Qatar Computing Research Institute, Hamad Bin Khalifa University, Doha, Qatar
Full list of author information is available at the end of the article

Abstract

The proliferation of social media enables people to express their opinions widely online. However, at the same time, this has resulted in the emergence of conflict and hate, making online environments uninviting for users. Although researchers have found that hate is a problem across multiple platforms, there is a lack of models for online hate detection using multi-platform data. To address this research gap, we collect a total of 197,566 comments from four platforms: YouTube, Reddit, Wikipedia, and Twitter, with 80% of the comments labeled as non-hateful and the remaining 20% labeled as hateful. We then experiment with several classification algorithms (Logistic Regression, Naïve Bayes, Support Vector Machines, XGBoost, and Neural Networks) and feature representations (Bag-of-Words, TF-IDF, Word2Vec, BERT, and their combination). While all the models significantly outperform the keyword-based baseline classifier, XGBoost using all features performs the best (F1 = 0.92). Feature importance analysis indicates that BERT features are the most impactful for the predictions. Findings support the generalizability of the best model, as the platform-specific results from Twitter and Wikipedia are comparable to their respective source papers. We make our code publicly available for application in real software systems as well as for further development by online hate researchers.

Keywords: Online hate, Toxicity, Social media, Machine learning

Introduction

Online hate, described as abusive language [1], aggression [2], cyberbullying [3, 4], hatefulness [5], insults [6], personal attacks [7], provocation [8], racism [9], sexism [10], threats [11], or toxicity [12], has been identified as a major threat on online social media platforms. Pew Research Center [13] reports that among 4248 adults in the United States, 41% have personally experienced harassing behavior online, whereas 66% witnessed harassment directed towards others. Around 22% of adults have experienced offensive name-calling, purposeful embarrassment (22%), physical threats (10%), and sexual harassment (6%), among other types of harassment. Social media platforms are the most prominent grounds for such toxic behavior. Even though they often provide ways of flagging offensive and hateful content, only 17% of all adults have flagged harassing conversation, whereas only 12% of adults have reported someone for such acts [13].

MDPI

sensors

Article

Detecting and Monitoring Hate Speech in Twitter

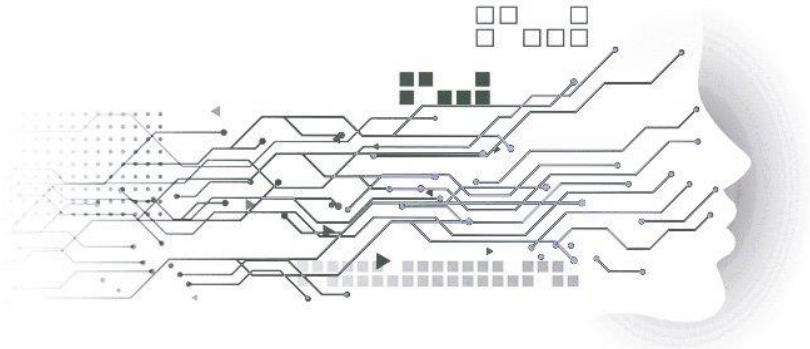
Juan Carlos Pereira-Kohatsu [1], Lara Quijano-Sánchez [1,2,*], Federico Liberatore [2,3] and Miguel Camacho-Collados [4,5]

1 Engineering School, Autonomous University of Madrid, 28049 Madrid, Spain; jc.pereira.kohatsu@gmail.com
2 UC3M-BS Institute of Financial Big Data, Charles III University of Madrid, 28903 Madrid, Spain
3 School of Computer Science & Informatics, Cardiff University, Cardiff CF24 3AA, UK; LiberatoreF@cardiff.ac.uk
4 Artificial Intelligence Area Coordinator, Cabinet of the Secretary of State for Digital Advancement, 28020 Madrid, Spain; mcamachoc@mineco.es
5 State Secretariat for Security, Interior Ministry, 28010 Madrid, Spain
* Correspondence: lara.quijano@uam.es

Received: 13 September 2019; Accepted: 21 October 2019; Published: 26 October 2019

Abstract: Social Media are sensors in the real world that can be used to measure the pulse of societies. However, the massive and unfiltered feed of messages posted in social media is a phenomenon that nowadays raises social alarms, especially when these messages contain hate speech targeted to a specific individual or group. In this context, governments and non-governmental organizations (NGOs) are concerned about the possible negative impact that these messages can have on individuals or on the society. In this paper, we present *HaterNet*, an intelligent system currently being used by the Spanish National Office Against Hate Crimes of the Spanish State Secretariat for Security that identifies and monitors the evolution of hate speech in Twitter. The contributions of this research are many-fold: (1) It introduces the first intelligent system that monitors and visualizes, using social network analysis techniques, hate speech in Social Media. (2) It introduces a novel public dataset on hate speech in Spanish consisting of 6000 expert-labeled tweets. (3) It compares several classification approaches based on different document representation strategies and text classification models. (4) The best approach consists of a combination of a LTSM+MLP neural network that takes as input the word, emoji, and expression tokens' embeddings enriched by the tf-idf, and obtains an area [illegible] on our dataset, outperforming previous methods presented in the

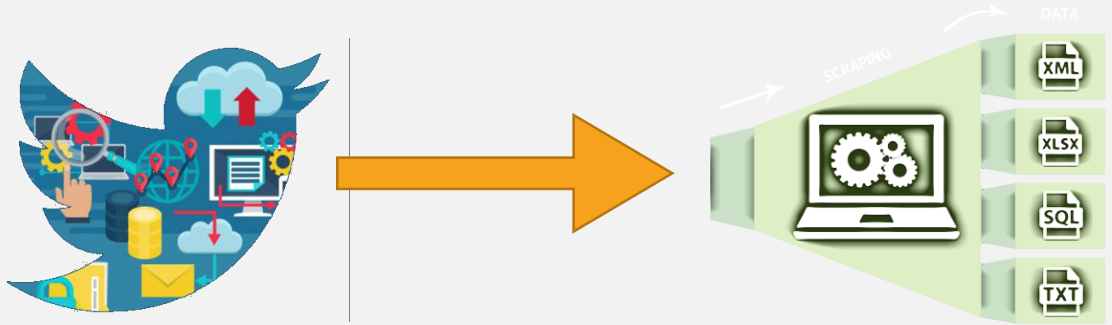
[illegible] policing; social network



Stop HATE PROJECT

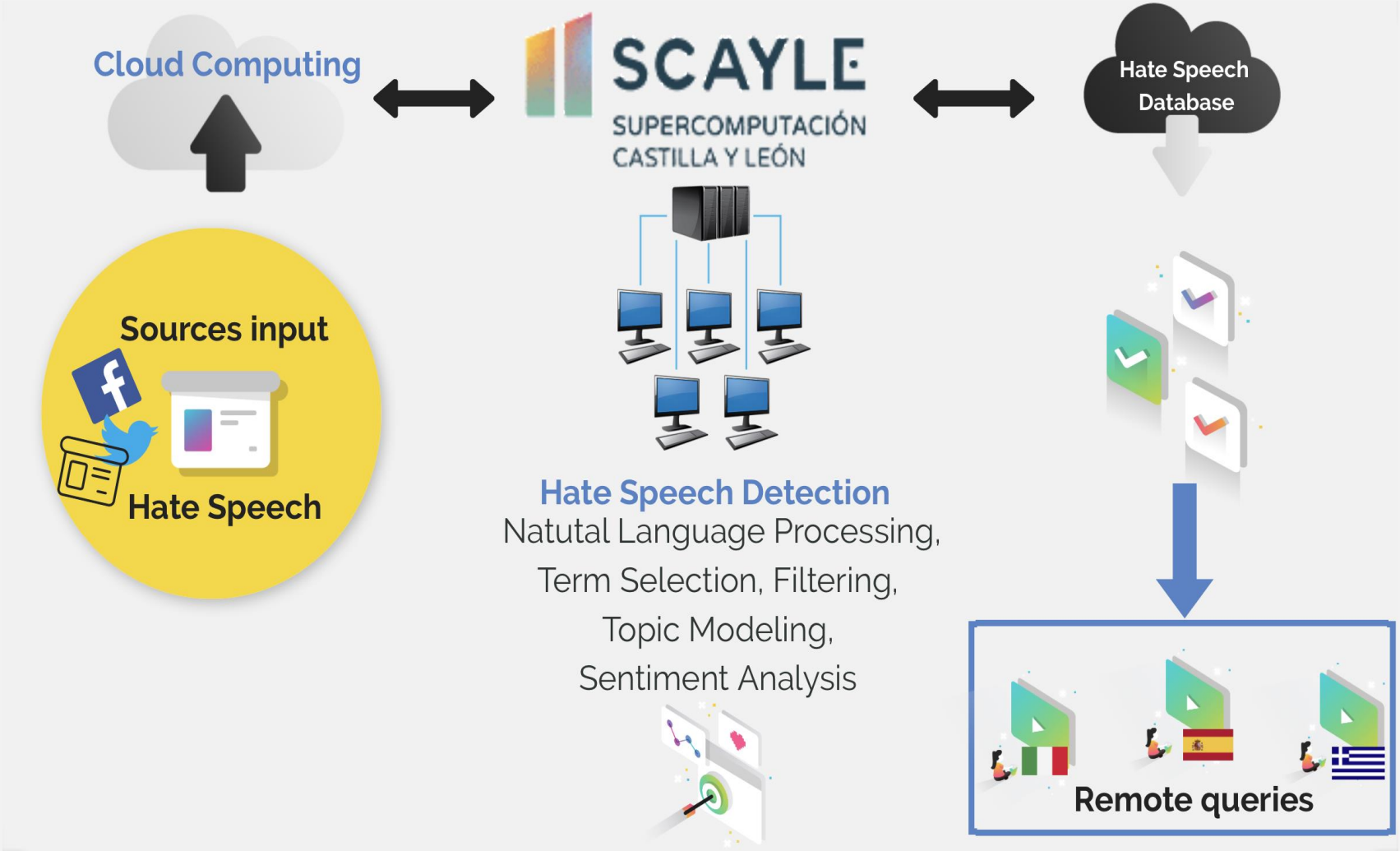
OBJECTIVE

- Development and evaluation of a **detector of online hate speech in Spanish** transmitted on Twitter, and for 3 specific types of prejudice:



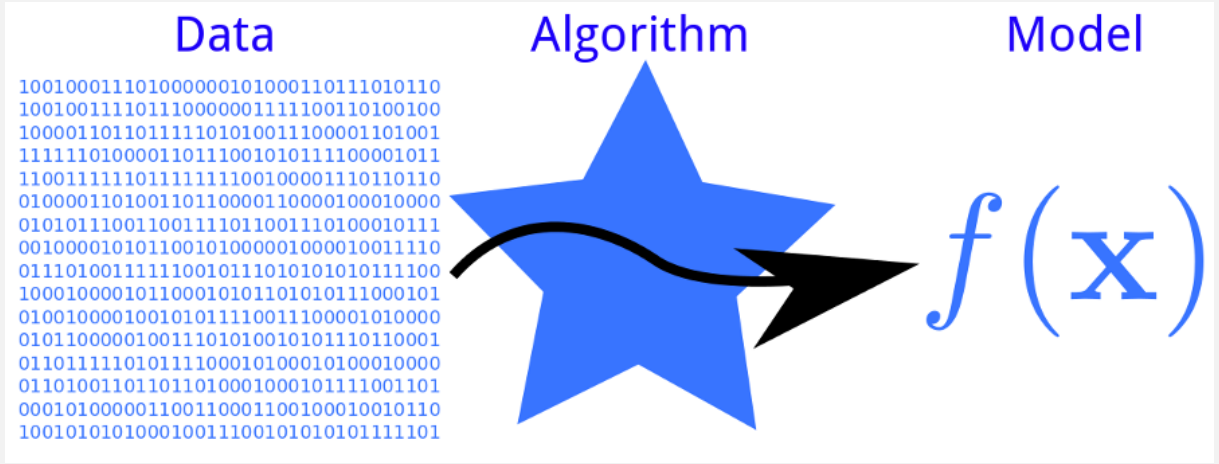
1. **Racism/xenophobia**
2. Sexual orientation
3. Politic ideology

LARGE-SCALE DETECTION STRATEGY

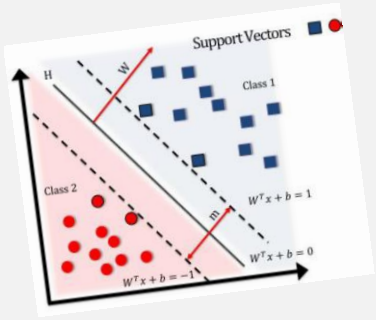


METHOD

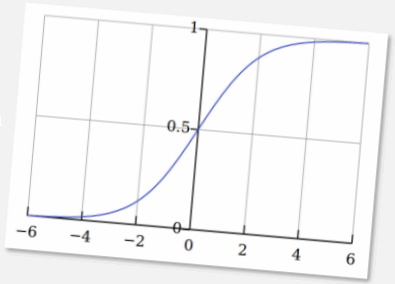
SUPERVISED CLASSIFICATION



- The predictive model is trained and **generated from the given examples**, manually classified for it



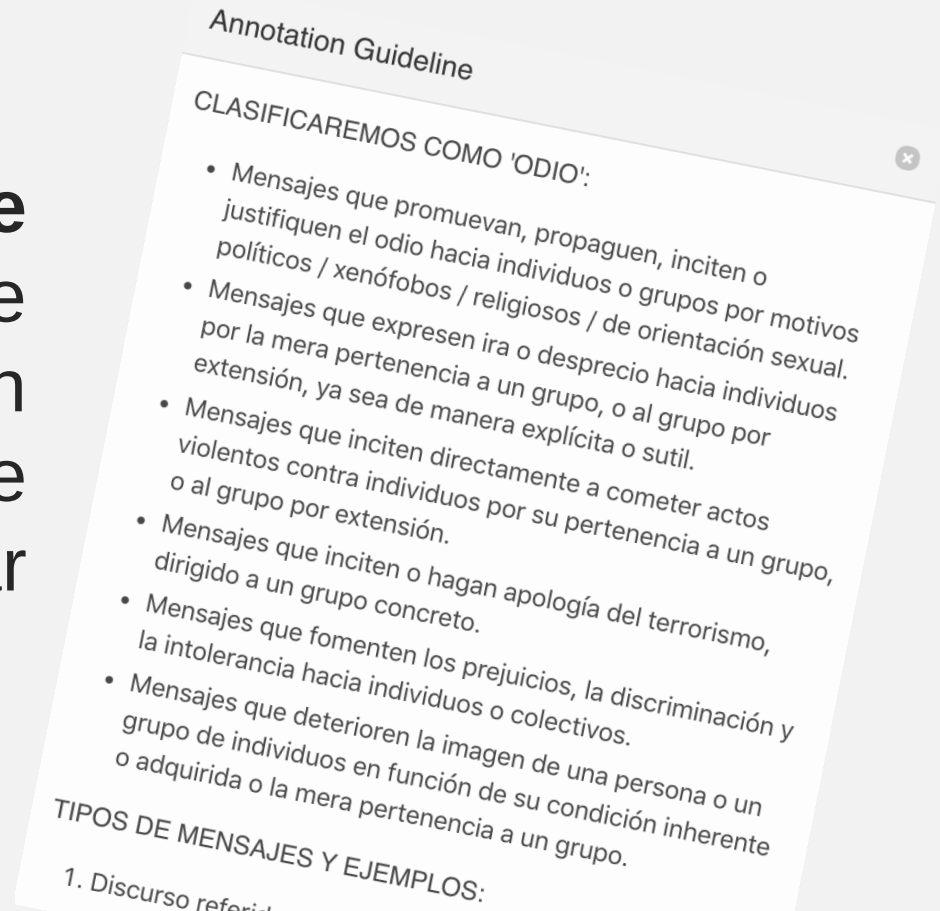
$$p(C_k \mid \mathbf{x}) = \frac{p(C_k)\, p(\mathbf{x} \mid C_k)}{p(\mathbf{x})}$$



METHOD

GENERATING THE TRAINING CORPUS FOR STOP-HATE

1. To **define the type of hate speech to detect**, and generate a manual or codebook in which those types of hate to be identified are explained with clear examples



Annotation Guideline

CLASIFICAREMOS COMO 'ODIO':

- Mensajes que promuevan, propaguen, inciten o justifiquen el odio hacia individuos o grupos por motivos políticos / xenófobos / religiosos / de orientación sexual.
- Mensajes que expresen ira o desprecio hacia individuos por la mera pertenencia a un grupo, o al grupo por extensión, ya sea de manera explícita o sutil.
- Mensajes que inciten directamente a cometer actos violentos contra individuos por su pertenencia a un grupo, o al grupo por extensión.
- Mensajes que inciten o hagan apología del terrorismo, dirigido a un grupo concreto.
- Mensajes que fomenten los prejuicios, la discriminación y la intolerancia hacia individuos o colectivos.
- Mensajes que deterioren la imagen de una persona o un grupo de individuos en función de su condición inherente o adquirida o la mera pertenencia a un grupo.

TIPOS DE MENSAJES Y EJEMPLOS:

1. Discurso referi

METHOD

GENERATING THE TRAINING CORPUS FOR STOP-HATE

2. To **generate filters**, identifying and selecting terms and combinations of words that could indicate and represent hate motivated by the 3 predefined types of prejudice:

   - Racism/xenophobia, - sexual orientation and - political ideology

LINGUISTIC FILTERS

Hate reason	**0.** Fixed terms or combinations	**1.** Primary words (descriptive within the hate group)
Racism / Xenophobia	panchito/a/o/os/as guachupin/o/a/os/as tiraflechas sudaca machupichu chino/a/os/as AND mandarin/o/a/os/as sin AND papeles, cierre AND frontera/s invasion AND silenciosa efecto AND llamada migracion/inmigracion AND masiva cerrar/cierre AND frontera/s	migrante/s, inmigrante/s, refugiado/a/os/as, desplazado/a/os/as, africano/a/os/as, asiatico/a/os/as, chino/a/os/as, latino/a/os/as, negro/a/os/as, negrato/a/os/as, moreno/a/os/as, mulato/a/os/as, moro/a/os/as, musulman/a/es/as, arabe/s, islamista/s, islamico/a/os/as

2. Secondary words (accompanying and expressing hatred)
carga, amenaza, inseguridad, peste, mierda, lacra, gentuza, basura, escoria, fuera, largo, culo, puto/a/os/as, jodido/a/os/as, maldito/a/os/as, sucio/a/os/as, peligro, peligroso/a/os/as, radical/es, terror, terrorista/s, asco/asqueroso/a/os/as, ilegal/es, indocumentado/a/os/as, invasion, invasor/a/es/as, invadir, invade, invaden, criminal/es, robo/s, robar, roba, roban, delincuencia, delincuente/s, delinquir, delinque, delinquen, vago/a/os/as, zangano/a/os/as, violencia, violento/a/os/as, violacion, violador/es, violar, viola, violan, trafico, traficar, trafica, trafican, ocupacion, ocupar, ocupa, ocupan, saqueo/s, saquear, saquean, asalto, asalta, asaltan, paga/paguita/a, subvencion/es, ayuda/s, carterista/s, hurto/s, deportacion/es, deportar, deportaria, deportaba, deporte/deporten, deportarlo/a/os/as, repatriacion/es, repatriar, repatriaria, repatrie, repatrien, repatriaba, repatriarlo/a/os/as, destrozo, destrozar, destroza, destrozan, destrozaria, destrozaba, destrozarlo/a/os/as, reventar, revienta, revientan, reviento, reventaria, reventaba, reventarlo/a/os/as, matanza, matar, mato, mata, mataria, mataba, matarlo/a/os/as

METHOD

Generating the training corpus for STOP-HATE

3. To **download the filtered tweets** from the Twitter API using the pre-set combinations as filters

word = [‘inmigrantes’, ‘cierre\nfrontera’, ‘cierre\nfronteras’, ‘anti\ninmigracion’, u‘anti\ninmigración’, u‘invasión\neuropea’, ‘invasion\neuropea’, u‘invasión\nsilenciosa’, ‘invasion\nsilenciosa’, ‘efecto\nllamada’, u’inmigración\nmasiva’, u’migración\nmasiva’, u’inmigracion\nmasiva’, u’migración\nmasa’ u’inmigración\nmasa’, u’inmigracion\nmasa’, ‘panchito’, ‘panchita’, ‘panchitos’, ’panchitas’, ‘machupichu’, ‘machu\npichu’, ‘machupicchu’, ‘machu\npicchu’, ‘guachupin’, u‘guachupín’, ‘guachupino’, ‘guachupina’, ‘sudaca’, u’tráfico\nmigrante’, u’tráfico\nmigrantes’, ‘u’tráfico\ninmigrante’, u’tráfico\ninmigrantes’, ‘repatriar’, ‘repatriarlo’, ‘repatriarlos’, u‘repatriación’, ‘repatriacion’, ‘deportar’, ‘deportarlos’, u‘deportación’, ‘deportacion’,

‘desplazado\nputo’, ‘desplazado\njodido’, ‘desplazado\nmaldito’, ‘desplazado\nsucio’, ‘desplazado\nasqueroso’, ‘desplazado\nasco’, ‘desplazado\nmierda’, ‘desplazados\nputos’, ‘desplazados\njodidos’, ‘desplazados\nmalditos’, ‘desplazados\nsucios’, ‘desplazados\nasquerosos’, ‘desplazados\nasco’, ‘desplazados\nmierda’, ‘refugiado\nputo’, ‘refugiado\njodido’, ‘refugiado\nmaldito’, ‘refugiado\nsucio’, ‘refugiado\nasqueroso’, ‘refugiado\nasco’, ‘refugiado\nmierda’, ‘refugiados\nputos’, ‘refugiados\njodidos’, ‘refugiados\nmalditos’, ‘refugiados\nsucios’, ‘refugiados\nasquerosos’, ‘refugiados\nasco’, ‘refugiados\nmierda’,

‘moro\nputo’, ‘moro\nsucio’, ‘moro\nasco’, ‘moro\nmierda’, ‘moro\njodido’, ‘moro\nmaldito’, ‘moro\nasqueroso’, ‘moros\nputos’, ‘moros\nmierda’, ‘moros\njodidos’, ‘moros\nmalditos’, ‘moros\nasquerosos’, ‘moros\nsucios’, ‘moros\nasco’, ‘mora\nputa’, ‘mora\nmaldita’, ‘mora\njodida’, ‘mora\nasquerosa’, ‘mora\nmierda’, ‘mora\nsucia’, ‘mora\nasco’, ‘moras\nputas’, ‘moras\nmalditas’, ‘moras\njodidas’, ‘moras\nasquerosas’, ‘moras\nmierda’, ‘moras\nsucias’, ‘moras\nasco’,

‘chino\mandarin’, u‘chino\mandarín’, ‘chino\nmandarino’, ‘china\nmandarina’, ‘chino\namarillo’, ‘china\namarilla’, ‘chino\nputo’, ‘china\nputa’, ‘chino\njodido’, ‘chino\nmierda’, ‘chino\nasqueroso’, ‘chino\nsucio’, ‘chino\nasco’, ‘chino\nmaldito’, ‘chinos\nputos’, ‘chinos\njodidos’, ‘chinos\nmalditos’, ‘chinos\nsucios’, ‘chinos\nasquerosos’, ‘chinos\nasco’, ‘chinos\nmierda’, ‘china\namarilla’, ‘china\njodida’, ‘china\nmierda’, ‘china\nasquerosa’, ‘china\nsucia’, ‘china\nmaldita’, ‘chinas\nputas’, ‘chinas\njodidas’, ‘chinas\nmalditas’, ‘chinas\nsucias’, ‘chinas\nasquerosas’, ‘chinas\nasco’, ‘chinas\nmierda’,

u‘asiático\nputo’, u‘asiático\njodido’, u‘asiático\nmaldito’, u‘asiático\nsucio’, u‘asiático\nasqueroso’, u‘asiático\nasco’, u‘asiático\nmierda’, u‘asiática\nputa’, u‘asiática\njodida’, u‘asiática\nmaldita’, u‘asiática\nsucia’, u‘asiática\nasquerosa’, u‘asiática\nasco’, u‘asiática\nmierda’, u‘asiáticos\nputos’, u‘asiáticos\njodidos’, u‘asiáticos\nmalditos’, u‘asiáticos\nsucios’, u‘asiáticos\nasquerosos’, u‘asiáticos\nasco’, u‘asiáticos\nmierda’, u‘asiáticos\nputas’, u‘asiáticos\njodidas’, u‘asiáticos\nmalditas’, u‘asiáticos\nsucias’, u‘asiáticos\nasquerosas’, u‘asiáticos\nasco’, u‘asiáticos\nmierda’,

‘latino\nputo’, ‘latino\njodido’, ‘latino\nmaldito’, ‘latino\nsucio’, ‘latino\nasqueroso’, ‘latino\nasco’, ‘latino\nmierda’, ‘latinos\nputos’, ‘latina\nputa’, ‘latina\njodida’, ‘latina\nmaldita’, ‘latina\nsucia’, ‘latina\nasquerosa’, ‘latina\nasco’, ‘latina\nmierda’, ‘latinos\njodidos’, ‘latinos\nmalditos’, ‘latinos\nsucios’, ‘latinos\nasquerosos’, ‘latinos\nasco’, ‘latinos\nmierda’, ‘latinas\nputas’, ‘latinas\njodidas’, ‘latinas\nmalditas’, ‘latinas\nsucias’, ‘latinas\nasquerosas’, ‘latinas\nasco’, ‘latinas\nmierda’,

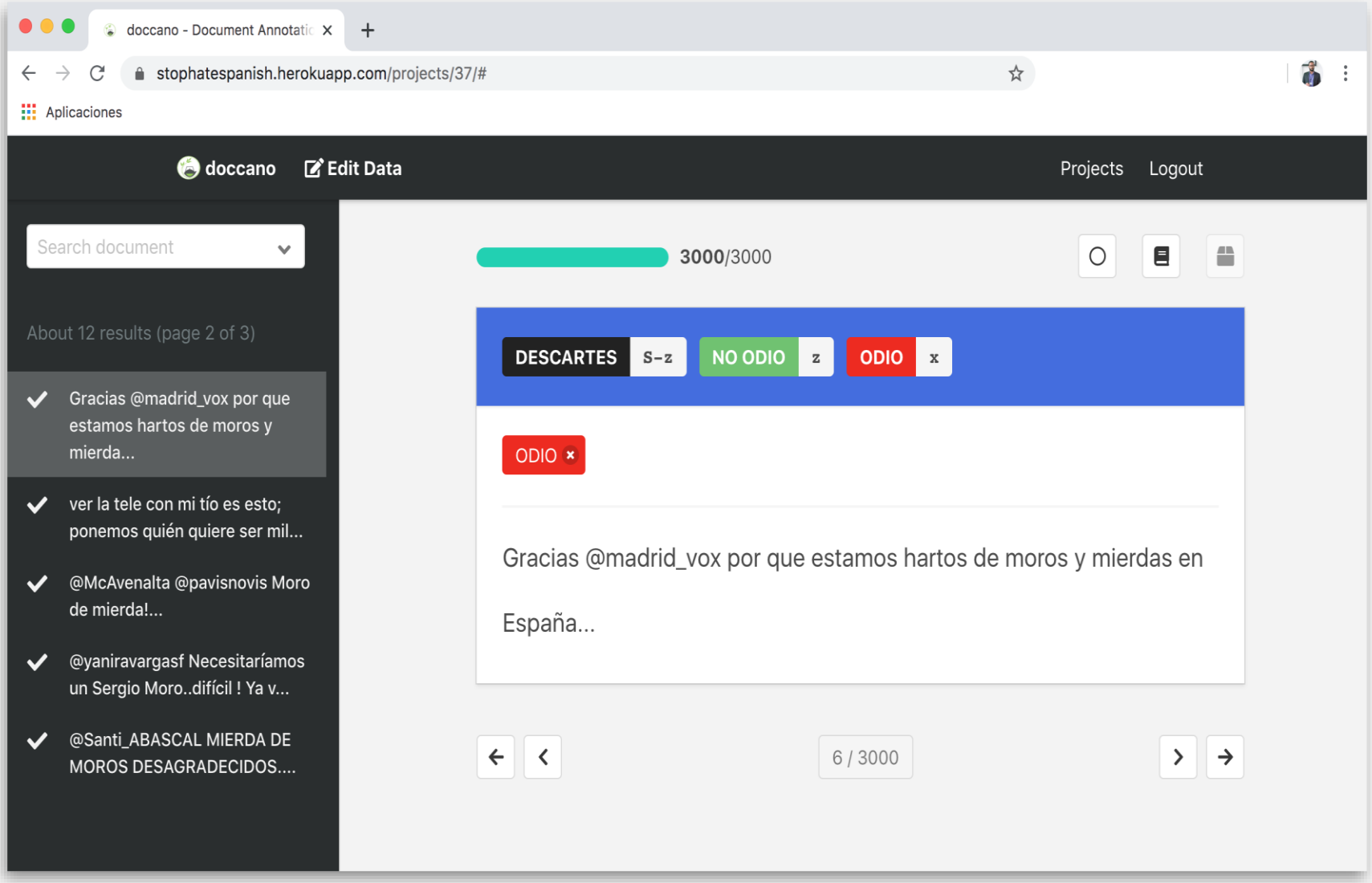
METHOD

Generating the training corpus for STOP-HATE

- Compilation of a total set of 72,000 filtered tweets:
  - **24,000 tweets from the racism / xenophobia group**
  - 24,000 tweets from the political ideology group
  - 24,000 tweets from the sexual orientation group

4. To **annotate and classify the downloaded filtered tweets** with 2 coders per message on the ***Doccano*** platform:
   - All the downloaded tweets from each group were classified by a project researcher and various secondary coders

ANNOTATION ON THE *DOCCANO* PLATFORM



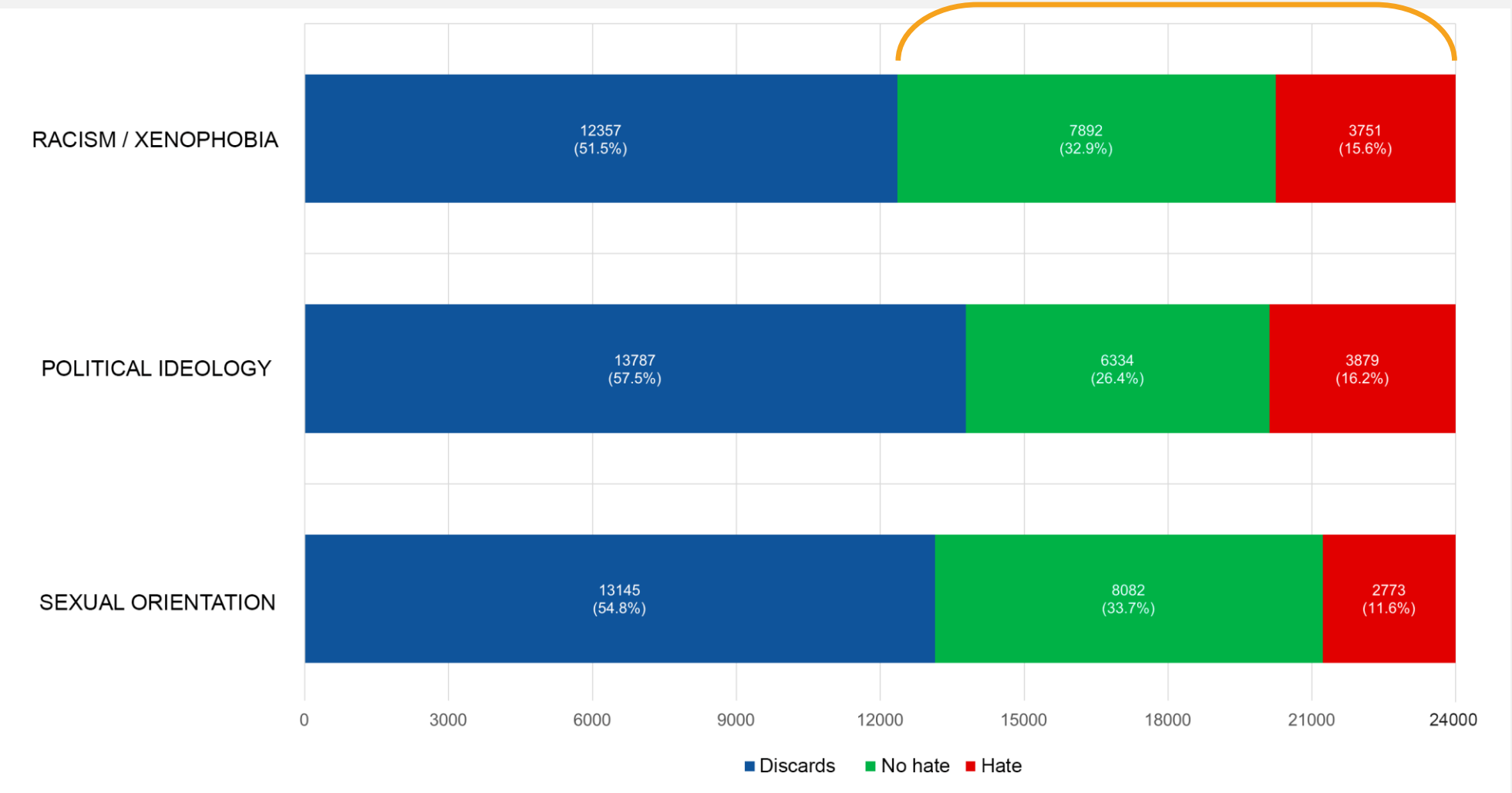
METHOD

Generating the training corpus for STOP-HATE

5. To **check the inter-coder reliability**, crossing all classified tweets to reach **TOTAL inter-coder agreement**

6. To **clean the resultant dataset**, discarding the previously rejected tweets and tweets without agreement

Distribution of frequencies and percentages of tweets coded for each of the hate reasons

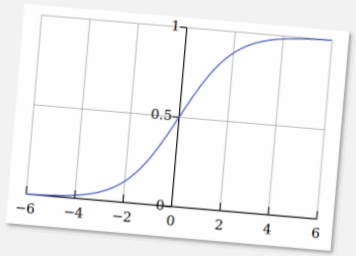


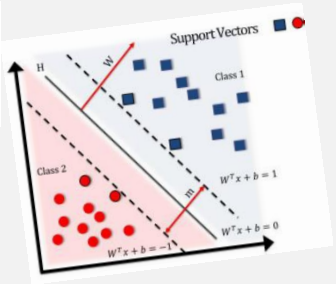
* The label discards corresponds to tweets without agreement plus those discarded in the classification process

METHOD

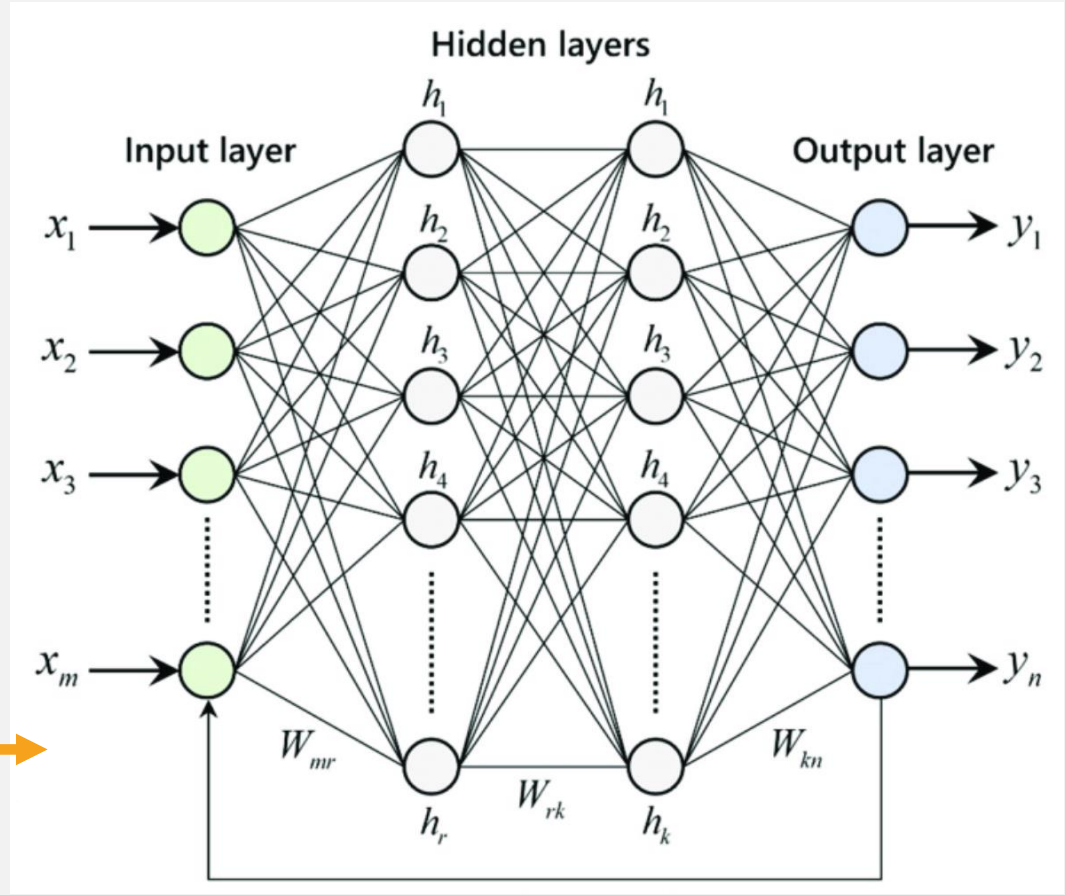
Generating the predictive models

$$p(C_k \mid \mathbf{x}) = \frac{p(C_k)\, p(\mathbf{x} \mid C_k)}{p(\mathbf{x})}$$





SHALLOW LEARNING
VS
DEEP LEARNING



Recurrent
Neural
Networks

scikits learn

TensorFlow + K Keras

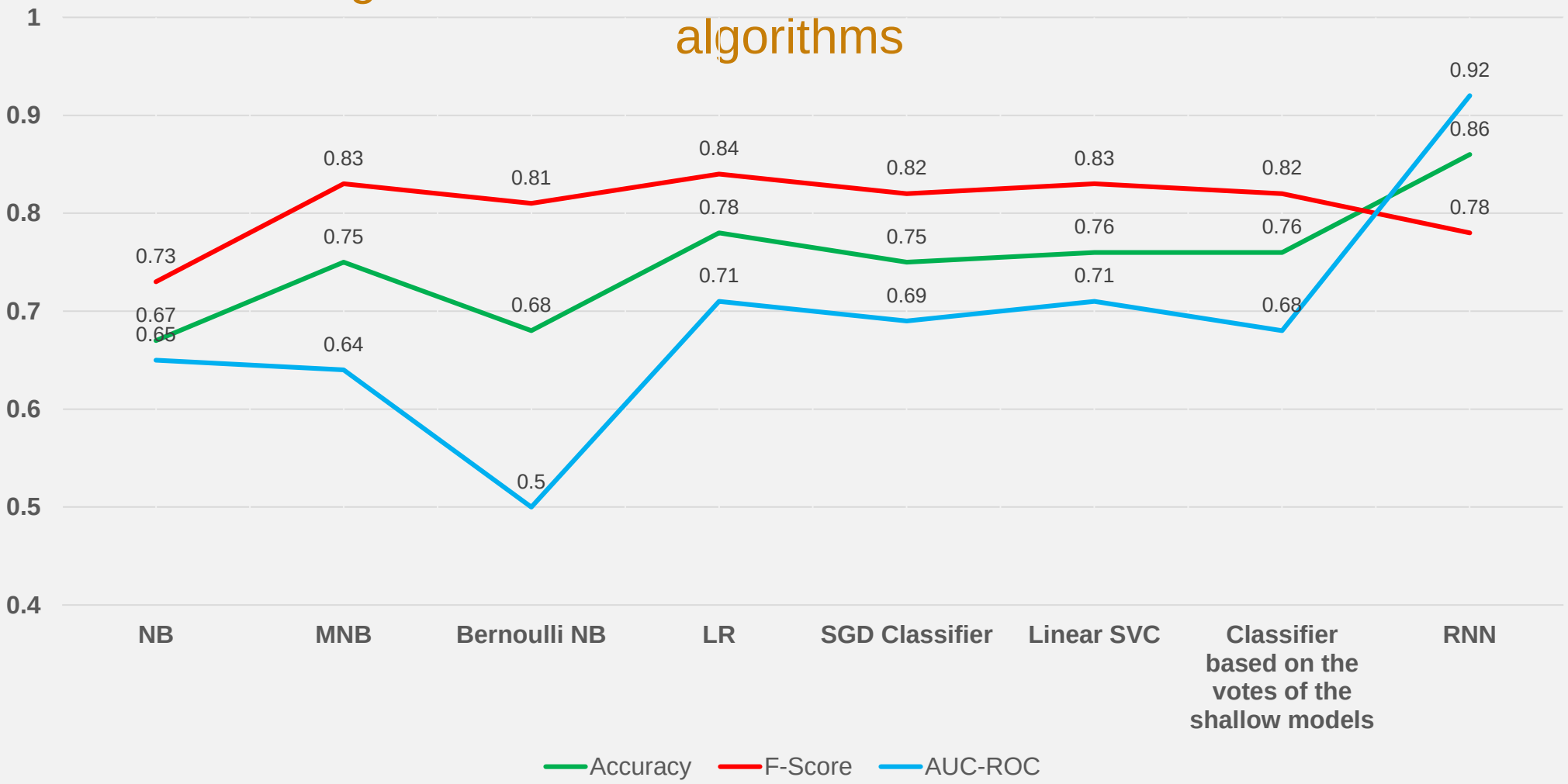
RESULTS

- We generated 6 predictive models with shallow algorithms, one based on the votes of those models; and a model generated with Recurrent Neural Networks
- In summary, we developed **8 models for each of the 3 groups, a total of 24 classification models**

Evaluation measures of the clasification algorithms

Shallow Algorithms	Accuracy	F-Score	AUC-ROC
Original Naive Bayes	.67	.73	.65
Naïve Bayes for Multinomial Models	.75	**.83**	.64
Naïve Bayes for Bernoulli's multivariate Models	.68	.81	.50
Logistic Regression	**.78**	**.84**	**.71**
Lineal classifiers with SGD training	.75	**.82**	.69
Support Vector Machines	**.76**	**.83**	**.71**
Classifier based on the votes of the other models	**.76**	**.82**	.68
Deep modelling	Accuracy	F-Score	AUC-ROC
Recurrent Neural Networks	**.86**	**.78**	**.92**

Evaluation measures of the models generated with each of the classification algorithms



NEXT STEP

- **PHARM project** aims to **develop a detector of hate speech** towards migrants and refugees in **Spain, Italy and Greece**, in more online sources, and with the added implementation of **early warning systems** that will allow predicting possible increases in geo-located hate crimes

PHARM

PREVENTING HATE AGAINST REFUGEES AND MIGRANTS

SOLUTION 1I

- **Big data from existing surveys using computational approaches that combine machine learning and synthetic population**

METHOD
1. Predicting individual acceptance

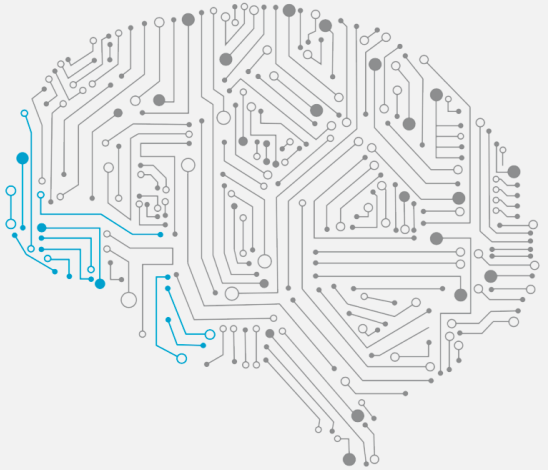
- In the first stage, **data from the available Eurobarometers that included the objective question "–Our country– should help refugees?"** (objective characteristic) **were extracted, along with 8 other variables that could explain this attitude** (descriptive characteristics)
- In total, **data were collected from 165,089 people distributed in 5 periods between 2015 and 2017**
- Of the 47 descriptive variables, **only sociodemographic characteristics that could later be identified in a census were used:**
  - *Country, origin, gender, age, occupation, marital status, education type of community* and *household composition*

METHOD
1. Predicting individual acceptance

- To facilitate the application of the algorithms, all measurements were standardized, turning them into dichotomous variables

- After eliminating the remaining values, a final matrix of 112,837 cases x 48 characteristics was obtained

With this data, supervised machine learning models were built to predict the individual support of Europeans to refugees



Variables taken from the Eurobarometer to model the level of support for refugees of a European citizen

	Variable	Description
0	*Support for refugees (target variable)*	1 = Not Support, 0 = Support
1	*Country*	Dummy (0-1): 'country_BALGARIJA', 'country_BELGIQUE', 'country_CESKA REPUBLIKA', 'country_DANMARK', 'country_DEUTSCHLAND OST', 'country_DEUTSCHLAND WEST', 'country_EESTI', 'country_ELLADA', 'country_ESPANA', 'country_FRANCE', 'country_GREAT BRITAIN', 'country_HRVATSKA', 'country_IRELAND', 'country_ITALIA', 'country_KYPROS', 'country_LATVIA', 'country_LIETUVA', 'country_LUXEMBOURG', 'country_MAGYARORSZAG', 'country_MALTA', 'country_NEDERLAND', 'country_POLSKA', 'country_PORTUGAL', 'country_ROMANIA', 'country_SLOVENIJA', 'country_SLOVENSKA REPUBLIC', 'country_SUOMI', 'country_SVERIGE', 'country_ÖSTERREICH'
2	*Origin*	0= Foreigner, 1= Native
3	*Marital status*	Dummy (0-1): 'Married', 'Partnership', 'Single', 'Divorced/Separated', 'Widow'
4	*Education*	0= No formal education <10, 1= ISCED Level 1. Primary education 10-12, 2= ISCED Level 2. Lower secondary education 13-15, 3= ISCED Level 3. Upper secondary education 16-18, 4= ISCED Level 4. Post secondary non-tertiary education, ISCED Level 5, First stage of tertiary education, ISCED Level
5	*Gender*	Dummy (0-1): 'Woman', 'Man'
6	*Age*	0= under 15 years, 1= 15 to 29 years, 2= 30 to 49 years, 3= 50 to 64 years, 4= 65 to 84 years, 5= 85 years and over
7	*Occupation*	Dummy (0-1): 'Employed', 'Not active', 'Unemployed'
8	*Type of community*	Dummy (0-1): 'Small/middle town', 'Rural area or village', 'Large town'
9	*Household composition*	0= 1 person, 1= 2 persons, 2= 3 to 5 persons, 3= 6 and more persons

METHOD
1. Predicting individual acceptance

- **Social acceptance of refugees was modeled** using *logistic regression* (LR), *decision trees* (DT), *random forests* (RF), *support vector machines* (SVM) and *nearest k-neighbors* (K-NN) **on the data Longitudinal Eurobarometer, obtaining good prediction accuracy in all algorithms**

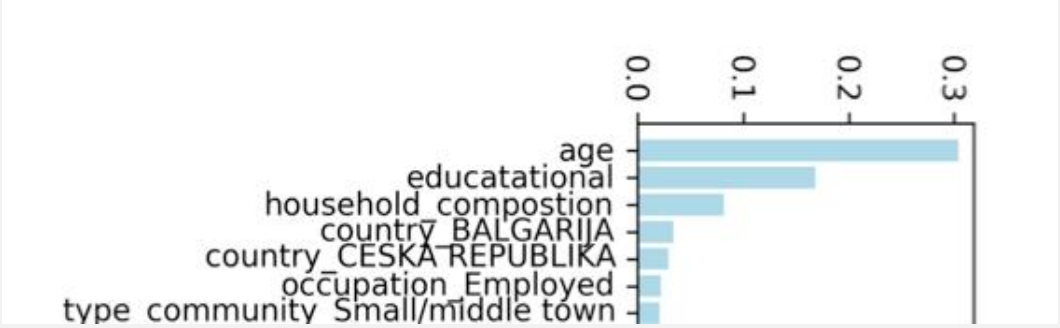
- **With these models the probability (0-1) of refugee support of any European citizen could be predicted**

- Some of the factors with the most significant coefficients were the *country* and *education* in logistic regression and *age* and *education* in random forests

Accuracy of the models generated with each of the algorithms

Algorithm	Accuracy
Logistic Regression (LR)	0.73
Decision Tree (DT)	0.70
Random Forest (RF)	0.69
Support Vector Machine (SVM)	0.72
SVM with kernel and gamma	0.65
K Nearest Neighbour (K-NN)	0.64

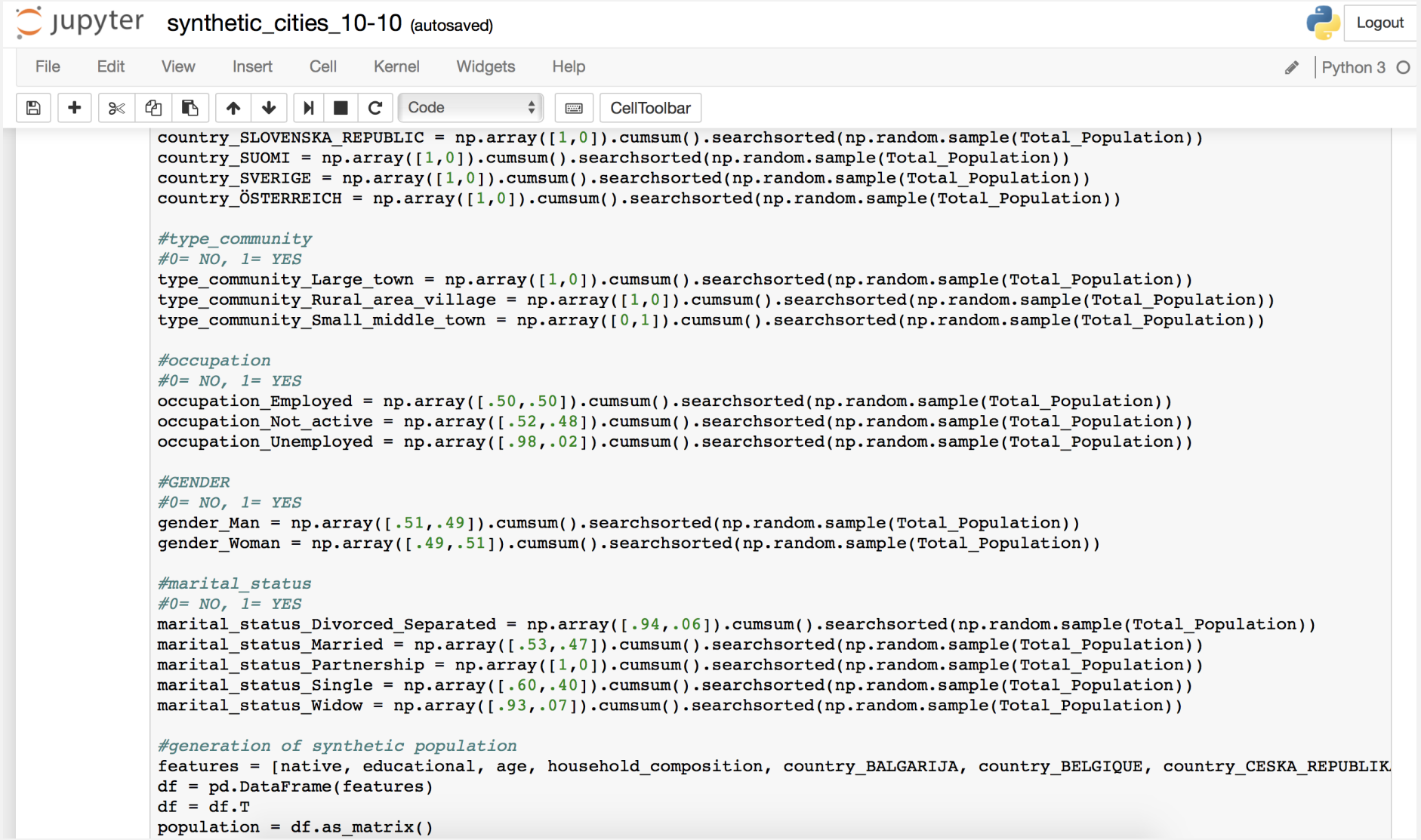
Weights of the variables in the random forest algorithm



METHOD
2. Predicting social acceptance

- To overcome the limitations of individual predictions and extend predictive models to large populations, we simulate the sociodemographic characteristics of 2,710,000 European citizens, corresponding to 10,000 in each of the 271 geographic regions (NUTS2).

Specifically, **we generate synthetic populations for each region using distributions of the sociodemographic characteristics obtained in the Eurostat census data**



Jupyter notebook used to generate synthetic populations

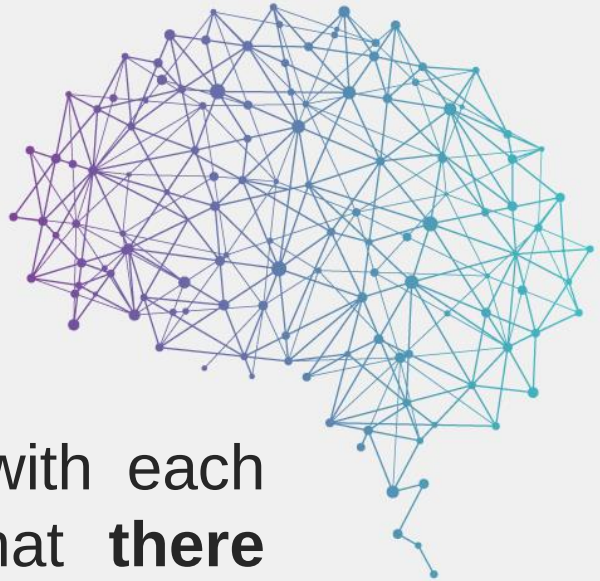
METHOD
2. Predicting social acceptance

- **Extrapolation at the regional level was carried out by calculating the probability of acceptance in each individual using the 6 predictive models, and for each of the 5 periods**

- The average probability was then estimated for all 271 regions, and an average variable was created to summarize the 6 predictive models

- In summary, **a database was generated with 271 cases** (NUTS2 in Europe) **and 42 (7 x 6) estimated probabilities of acceptance of refugees**

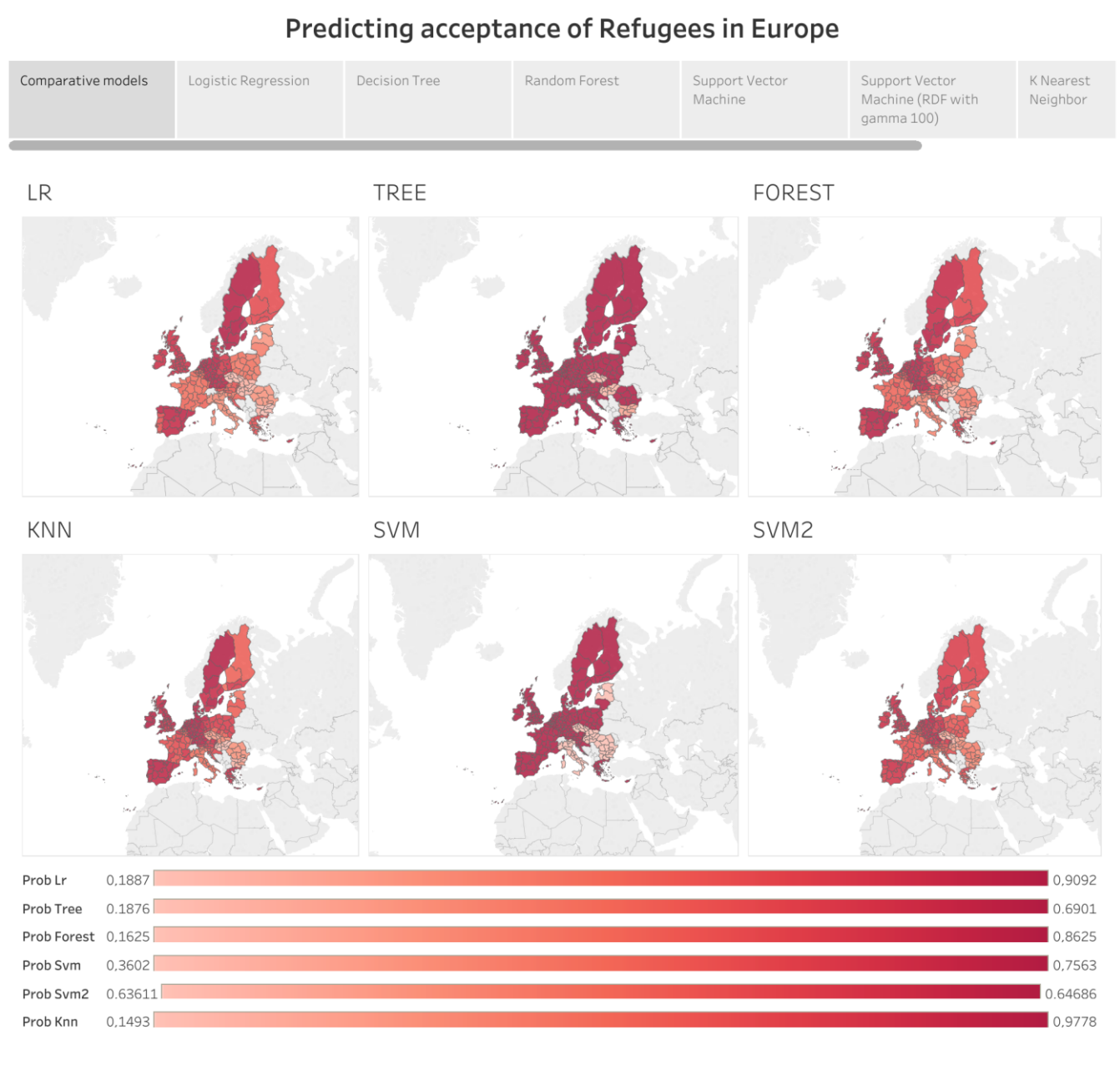
RESULTS



- When visualizing the regional predictions with each model for aggregate data, it is perceived that **there are certain differences between the models**, from more homogeneous predictions in the *decision trees* to more heterogeneous in the *logistic regression* or *K closest neighbors*

- **In all the maps we clearly observe that the country is a strong predictor of social support and acceptance of refugees**

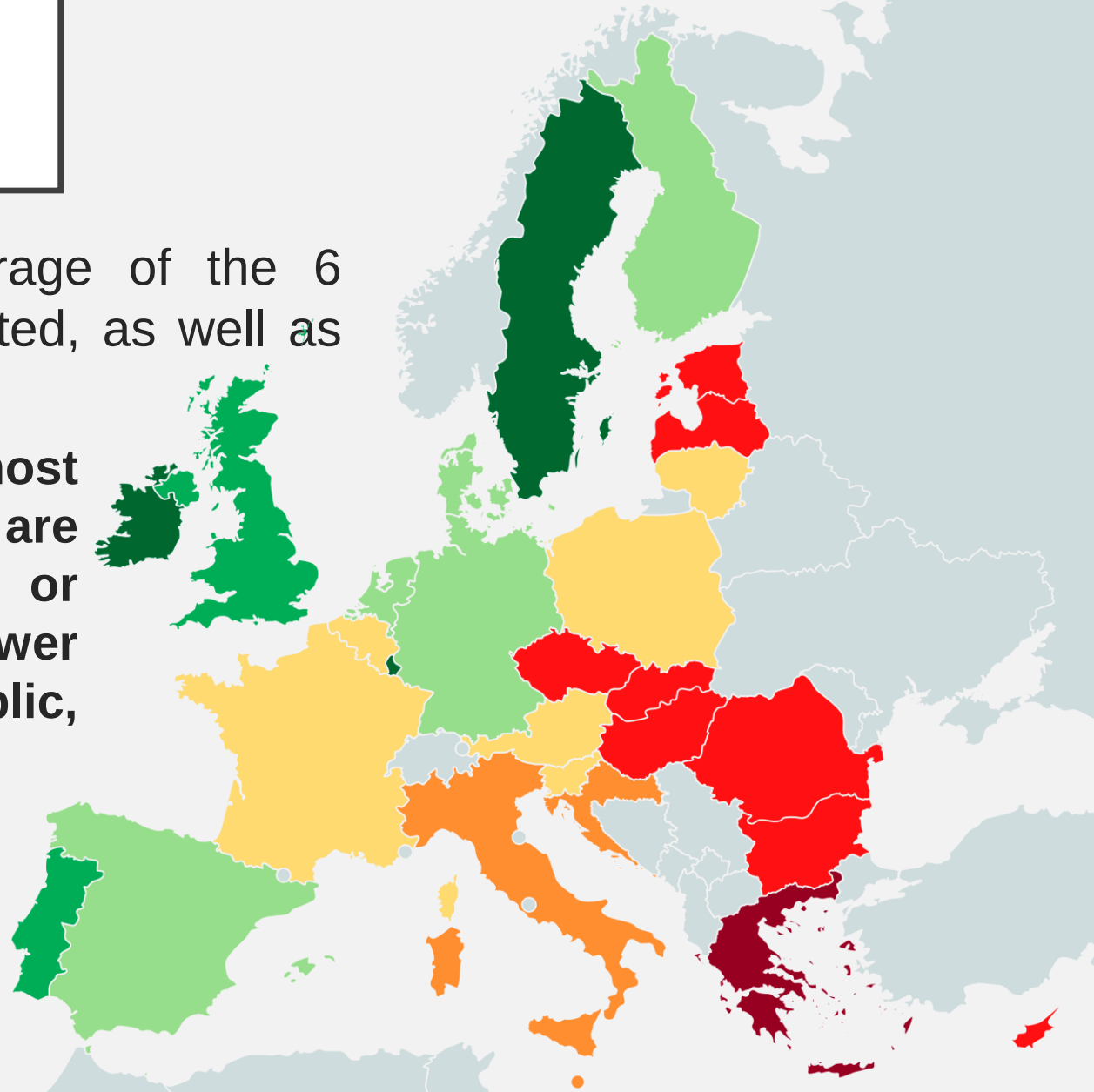
Comparative visualization of the 6 predictive models for social acceptance of refugees in NUTS2 regions of Europe



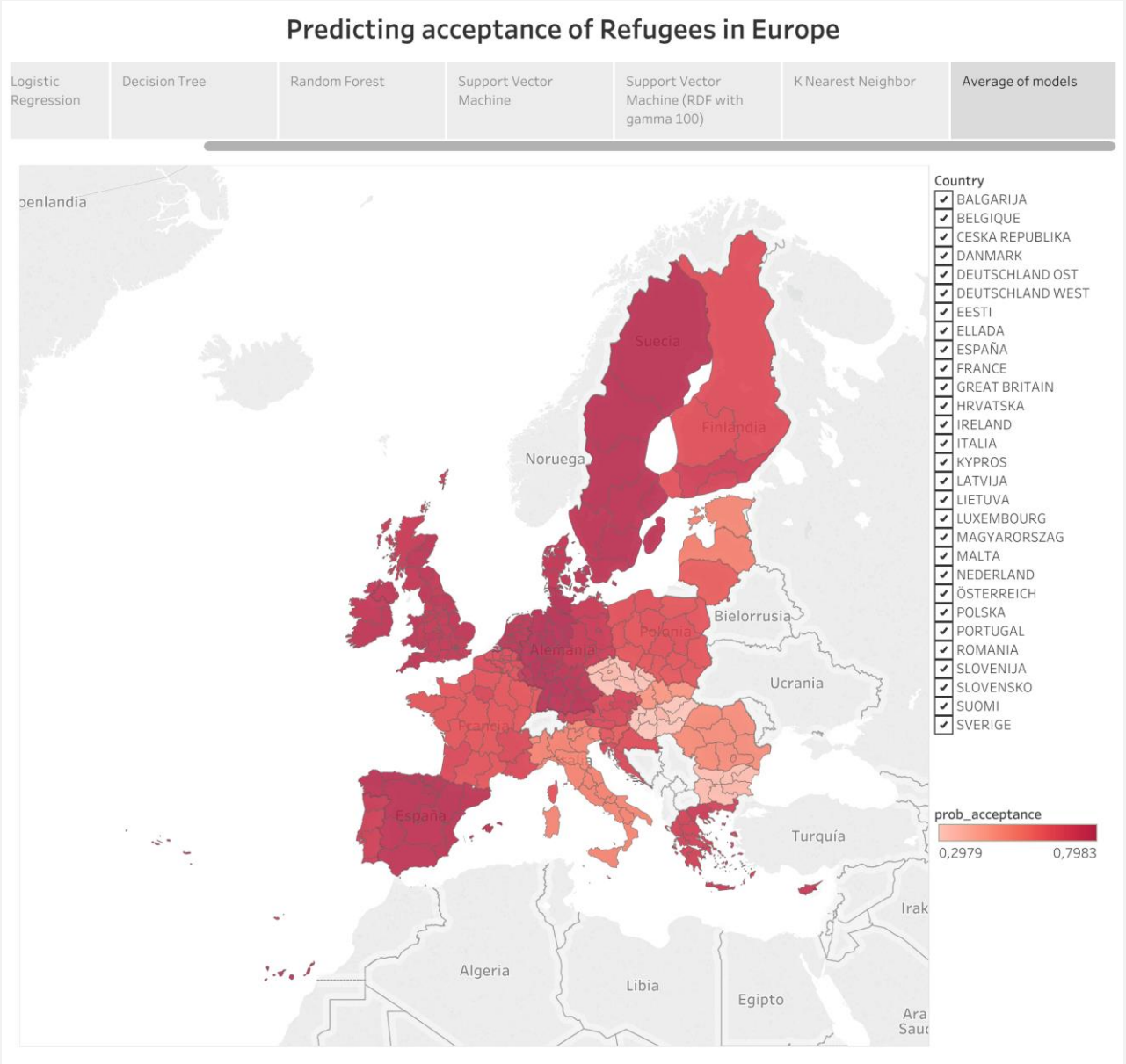
RESULTS

- To summarize the models, the average of the 6 estimates for each region was calculated, as well as for all periods analysed
- According to estimates, **the countries most likely to socially integrate refugees are Norway, Spain, United Kingdom or Germany, and those with lower probabilities are the Czech Republic, Hungary or Romania**

These estimates fully coincide with the comparative European surveys

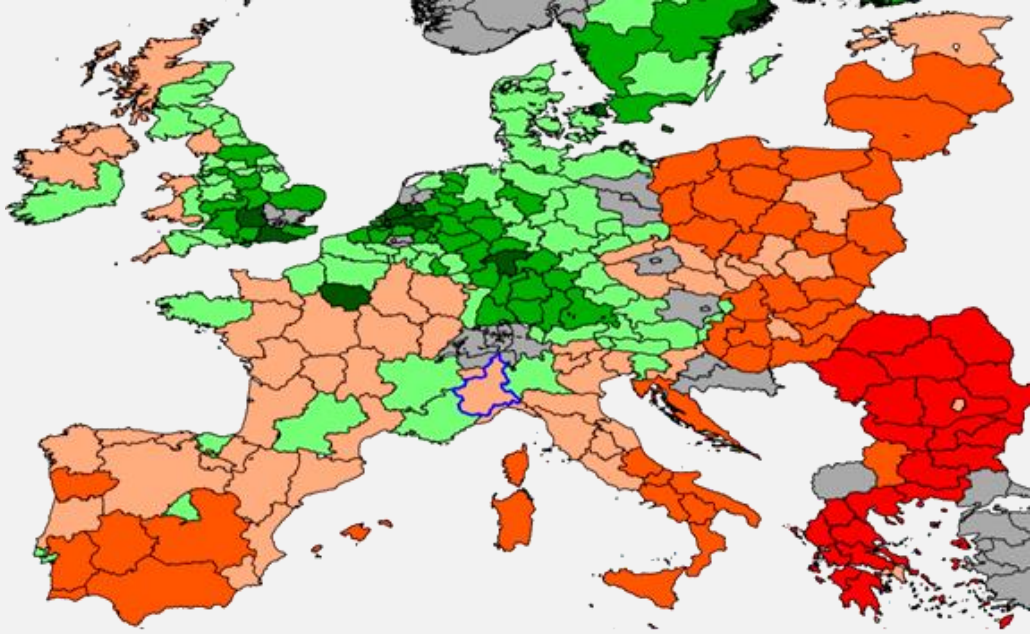


Summary display with the average acceptance probability for the 6 models and all periods analyzed

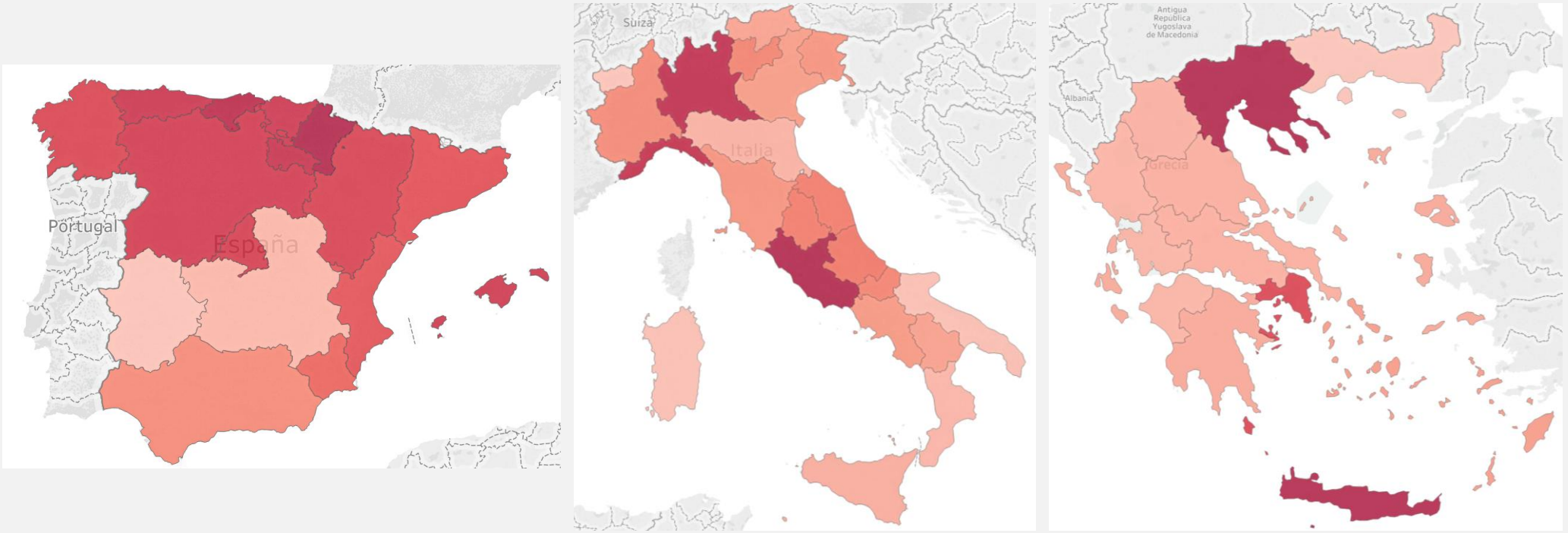


https://afly.co/ggw2

RESULTS



- **If the estimate is focused on each country, more interesting internal differences are perceived**
- For example, in the case of Italy, Greece and Spain it can be seen that the most populated regions (Athens, Thessaloniki, Rome, Milan, Madrid or Catalonia) are more likely to accept refugees
- **We create an interactive online map** (available for reproducibility) **to explore each country, with each model at each time point**

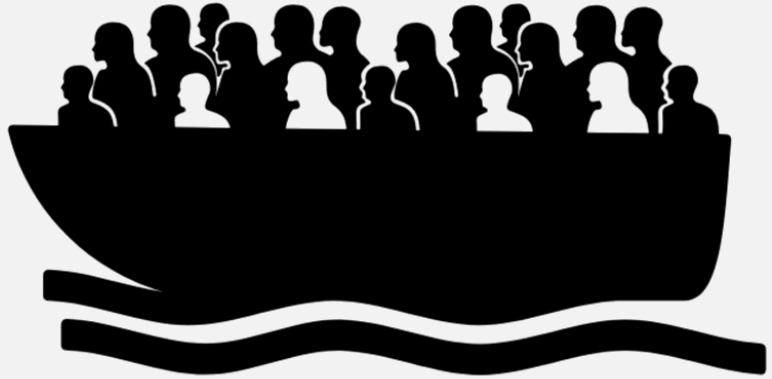


Estimation of the probability of acceptance of refugees in the regions of Spain, Italy and Greece

RESULTS

- **Regarding the longitudinal changes, the statistical analysis revealed that 2016 was the worst year for social integration**

- There is a significant decrease in the probability of accepting refugees by Europeans in May (M = 0.66, SD = 0.13) and November 2016 (M = 0.67, SD = 0.14), compared with November 2015 (M = 0.69, SD = 0.12) or May (M = 0.68, SD = 0.14) and November 2017 (M = 0.69, SD = 0.12)



REMARKS

- In this work, **the first articulated data that estimate the possible social integration of refugees in all regions of Europe have been generated**
  - The use of synthetic populations is an innovation that provides validated computational methods from other fields to the social and communication sciences, where they have rarely been implemented

REMARKS

- **This predictive analytical approach based on data can contribute to generating new relocation and resettlement schemes**

- **The prediction of the probability of acceptance of refugees at the regional level will allow mitigating the risks of exclusion of these groups**, through the planning of integration policies, anticipating which populations could accept or oppose more resistance to the arrival of refugees

Machine learning approaches to monitor integration of migrants

Carlos Arcila Calderón

THANK YOU

VNiVERSiDAD D SALAMANCA
CAMPUS OF INTERNATIONAL EXCELLENCE

observatorio contenidos audiovisuales