

UNIVERSITÀ CATTOLICA DEL SACRO CUORE –
MILANO

Interfacoltà Scienze Bancarie, Finanziarie e Assicurative -
Economia

Master of Science in Statistical and Actuarial Science

Profile: Data Analytics for Business and Economics



HATE SPEECH DETECTION:
AN ITALIAN EXAMPLE OF HATE AGAINST
IMMIGRANTS AND MINORITIES

Supervisor:
Prof. Marco L. DELLA VEDOVA

Candidate:
Federica DANI
ID: 4906750

TABLE OF CONTENTS

1	Introduction to Hate Speech	3
1.1	Hate Speech Definition	3
1.2	Hate Speech Detection	4
1.2.1	Automatic approaches for Hate Speech Detection	7
2	Introduction to hate against immigrants and minorities in Italy	9
2.1	Immigration in Italy: reading of the statistics	9
2.2	Immigration through media communication	12
2.2.1	How immigrants are perceived by the Italian society	19
2.3	Hate speech against immigrants: reading of the statistics	25
2.4	How Covid-19 has affected hatred and hate speech towards minorities	28
3	Hate Speech Detection: The Analysis	37
3.1	Goal of the analysis	37
3.2	Related work	37
3.3	Dataset Creation and Description	39
3.4	Pre-processing analysis	42
3.4.1	Data Cleaning	43
3.4.2	Explorative Analysis	45
3.4.3	NLP: A deeper analysis	50

4	Hate Speech: Implementation of the models	54
4.1	Models' description	54
	BIBLIOGRAPHY	65

1 Introduction to Hate Speech

1.1 Hate Speech Definition

Hate Speech can be defined as any expression *“that is abusive, insulting, intimidating, harassing, and/or incites to violence, hatred, or discrimination. It is directed against people on the basis of their race, ethnic origin, religion, gender, age, physical condition, disability, sexual orientation, political conviction, and so forth”* (Erjavec and Kovacic, 2012)

Hate is all but a new phenomenon, yet the global spread of Internet and social network services has provided it with new means and forms of dissemination. This spread allows users to create, access and share knowledge more easily and with a wider public than before, but most importantly barely any skill and cost is required. The result has been the so-called “democratisation” of the web that refers to the process by which access to technology rapidly continues to become more accessible to more people. The phenomenon reflects the up-to-date view of freedom of expression, thus freedom of speech. Yet, this freedom can result also in the publication of content that is abusive and harmful toward the rights of some groups of people – namely hate speech.

Indeed, online hateful content can be potentially more dangerous than offline communication because of specific characteristics such as the presumed anonymity that usually leads the user to be more open with his thoughts and direct with his words, confident of his non-traceability and impossibility of being personally attacked. Also, high virality is a main characteristic of the web, that allows information to spread quickly, and usually in case of hateful content, it causes harm before it is possible to prevent it.

Even if there is more awareness than before, today the Internet is still commonly considered reliable information, regardless of the real source; therefore, hateful content can be persuasive and can rapidly spread also starting from some tweets. As

such, many online forums such as Facebook, YouTube, and Twitter consider hate speech harmful, and have policies to remove hate speech content.

When talking about Hate Speech, it is important to consider the trade-off between freedom of speech and the protection of dignity and rights of minority groups. That is why, despite the numerous efforts, there is no universally accepted definition of HS. While the US is more oriented to granting freedom of speech, other countries such as Europe tend to be less tolerant. In fact, several European treaties and conventions ban HS: to mention but one, the Council of European Union condemns publicly inciting violence or hatred towards persons or groups defined by reference to race, colour, religion, descent or national or ethnic origin. (Poletto et al., 2017). It is also worth mentioning the *No Hate Speech Movement*¹, which is promoted by the Council of Europe and whose main pursuit is endorsing responsible behaviours and preventing HS among European citizens.

1.2 Hate Speech Detection

Due to the societal concern that online hate is becoming, there is strong motivation to study automatic detection of hate speech. Universities and centres of research have been trying to develop new techniques to identify hate language and its impact. By automating its detection, the spread of hateful content can be reduced. Most social media platforms have established rules that prohibit hate speech. Enforcing these rules requires much manual effort. Some platforms, like Facebook, have recently increased the number of content moderators. Automatic tools could accelerate the reviewing process and help allocate the human resources to the specific posts that require a deeper examination.

Detecting hate speech is a challenging task. One of the main reasons is related to the specific definition of the concept. As we anticipated, there is no universally accepted definition of hate speech: according to some, it is necessary that the target is a

¹ <https://www.nohatespeechmovement.org>

group, while in other cases, also the attack to an individual can be considered hate speech (e.g., Encyclopaedia of the American Constitution). A common theme among the definitions is that the attack is based on some aspect of the group or peoples' identity, while sometimes the identity's characteristics are what is valued. At last, certain definitions – e.g., Fortuna et al. – specifically call out variations in language style and subtleties, which can be extremely challenging for the conventional text classification tools.

A particular problem not covered by many definitions relates to factual statements. Some sentences are more evident expressions of hate speech, for example “Jews are swine ”, which is a statement of inferiority. However, it is also common to have sentences with less clear hate reference, for example “Many Jews are lawyers”, which can either be a factual statement or a characterization of the group of people that hides a hateful comment. Understanding this type of speech is hard because it relates to real-world fact. More so, to evaluate validity, we would initially need to define precise word interpretations, namely, is “many” an absolute number or by relative percentage of the population, further complicating the verification (MacAvaney et al., 2019).

Another issue that is worth mentioning is praising a hateful group. For example, the KKK does not contain per se hateful reference, but it represents a group who express hate against another group.

Despite the mentioned differences, some of the developed approaches found promising results for detecting hate speech in textual content. The proposed solutions employ machine learning techniques to classify text as hate speech. The problem with the implemented techniques is that their decisions can be difficult to interpret, and manual intervention is still needed to check the context in which the sentences are used and validate the result.

Because of the difficulty of defining hate speech another issue arises, namely, the importance of obtaining the right dataset. Collecting and annotating data for the training of automatic classifiers to detect hate speech is challenging, specifically, identifying and agreeing whether specific text is hateful is difficult.

Social media platforms are a hotbed for hate speech, yet many have very strict data usage and distribution policies. Due to privacy's issues not all the social networks available allow to extract data, making it difficult to make a thorough study. Twitter is the most common platform used to extract sample texts having a more lenient data usage policy. The unique genre of Twitter posts - direct, brief, usually expression of one's thoughts and beliefs - if on one hand, can be considered a valuable resource, on the other, the character limitation results in terse, short-form text. In contrast, posts from other platforms are typically longer and can be part of a larger discussion on a specific topic. Longer posts could help provide additional context and interpreting the meaning of the text.

Some researchers have worked on collecting posts and building curated datasets to identify hateful context. To mention but a few: Hatabase Twitter², WaasemA and WaasemB³, Stromfront⁴, TRAC⁵, HatEval⁶, Kaggle⁷. A representative sampling of available training and evaluation public datasets is shown in Table 1.

Most of these datasets come from Twitter, but they all vary considerably in their size, scope, characteristics of the data and of the hate speech considered. As mentioned before, the Twitter datasets do capture a wide variety of aspects in many different languages such as attacking different groups; however, the construction process including the filtering and sampling methods introduce uncontrolled factors for analysing the corpora. Also, on a platform such as Twitter, hate speech occurs at a very low rate compared to non-hate speech. Although datasets reflect this imbalance to an extent, they do not map the actual percentage due to training needs. For example, in the WaseemA dataset, 20% of the tweets were labelled sexist, 11.7%

² Davidson T, Warmley D, Macy MW, Weber I. Automated Hate Speech Detection and the Problem of Offensive Language. ICWSM. 2017.

³ Waseem Z, Hovy D. Hateful Symbols or Hateful People? Predictive Features for Hate Speech Detection on Twitter. In: SRW@HLT-NAACL; 2016.

⁴ de Gibert O, Perez N, Garcia-Pablos A, Cuadros M. Hate Speech Dataset from a White Supremacy Forum. In: 2nd Workshop on Abusive Language Online @ EMNLP; 2018.

⁵ Kumar R, Ojha AK, Malmasi S, Zampieri M. Benchmarking Aggression Identification in Social Media. In: Proceedings of the First Workshop on Trolling, Aggression and Cyberbullying (TRAC-2018). ACL; 2018. p. 1–11

⁶ CodaLab—Competition;. Available from: <https://competitions.codalab.org/competitions/19935>.

⁷ Detecting Insults in Social Commentary;. Available from: <https://kaggle.com/c/detecting-insults-in-social-commentary>

racist, and 68.3% neither. In this case, there is still an imbalance in the number of sexist, racist, or neither tweets, but it may not be as imbalanced as expected on Twitter (MacAvaney et al., 2019).

Table 1. Hate-related dataset characteristics

Dataset	Labels and percents in dataset	Origin Source	Language
Hatebase/Twitter [9]	Hate 5% Offensive 76% Neither 17%	Twitter	English
WaseemA [17]	Racism 12% Sexism 20% Neither 68%	Twitter	English
WaseemB [18]	Racism 1% Sexism 13% Neither 84% Both 1%	Twitter	English
Stormfront [14]	Hate 11% Not Hate 86% Relation 2% Skip 1%	Online Forum	English
TRAC (Facebook) [19]	Non-aggressive 69% Overtly agg. 16% Covertly agg. 16%	Facebook	English & Hindi
TRAC (Twitter) [19]	Non-aggressive 38% Overtly agg. 29% Covertly agg. 33%	Twitter	English & Hindi
HatEval [20]	Hate 43% / Not Hate 57% Agg. / Not agg. roup / Individual	Twitter	English & Spanish
Kaggle [21]	Insulting 26% Not Insulting 74%	Twitter	English
GermanTwitter (Expert 1 annotation) [11]	Hate 23% Not Hate 77%	Twitter	German

Source: <https://doi.org/10.1371/journal.pone.0221152.t001>

1.2.1 Automatic approaches for Hate Speech Detection

As we mentioned before, several automatic approaches have been developed for hate speech detection from text. A brief description of the logic behind the most relevant techniques will follow in this section.

Keyword-based approach

It is a basic approach that relies on using an ontology or dictionary to identify text that contain potentially hateful keywords. Keyword-based approaches are fast and straightforward to understand. However, they have severe limitations. As we observed in our study of the definition of hate speech, simply using a hateful term is not enough to constitute hate speech. Also, a system that mainly relies on keywords would be very precise in identifying the speech that has those words but would not be able to identify hateful content that does not contain them. In contrast, including

terms that are not always hateful (e.g., “swine”, “trash”) could lead to excessive false alarms. Lastly, this approach cannot identify hate speech with figurative or nuanced language, namely, speech where no hate word is present.

Source metadata

The approach considers using additional information from social media to categorize the users and have a better understanding of the posts. However, external researchers hardly ever have full access to the user information due to privacy’s issues. Also, using user information potentially rises ethical problems. Indeed, models or systems might be biased against certain users and frequently flag their posts as hateful even if they are not. Similarly, relying too much on the user’s statistics could lead to missing hateful posts from users who do not typically post hateful content. Thus, the approach needs further studies.

Machine learning classifiers

According to MacAvaney et al. (2019), “Machine learning models take samples of labelled text to produce a classifier that is able to detect the hate speech based on labels annotated by content reviewers” . Numerous models have been proven successful in the past. The most common ones are Naïve Bayes, Support Vector Machine and Logistic Regression, Natural Ensemble, BERT, FastText, C-GRU.

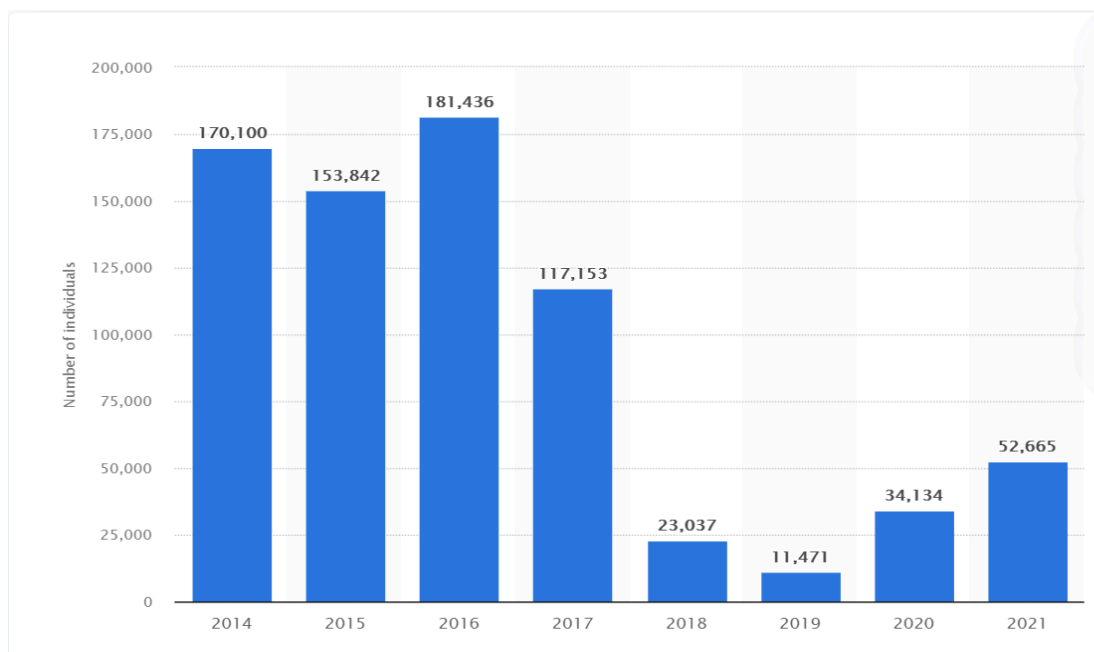
2 Introduction to hate against immigrants and minorities in Italy

2.1 Immigration in Italy: reading of the statistics

This study is focused on detecting hate speech against immigrants and minorities in the specific context of Italy. To better interpret the results of the study it is necessary to get a general insight on the immigration status and how immigrants are perceived in Italy.

Although landings occur also in Malta, Cyprus and Spain, Italy remains the country where most ships come ashore after a rescue. In 2020, the number of migrants arriving by sea in Italy exceeded 34000, mainly on the islands of Lampedusa and Sicily and in Calabria. The table below (fig 1) shows the number of immigrants who arrived by sea in Italy between the year 2014 and 2021.

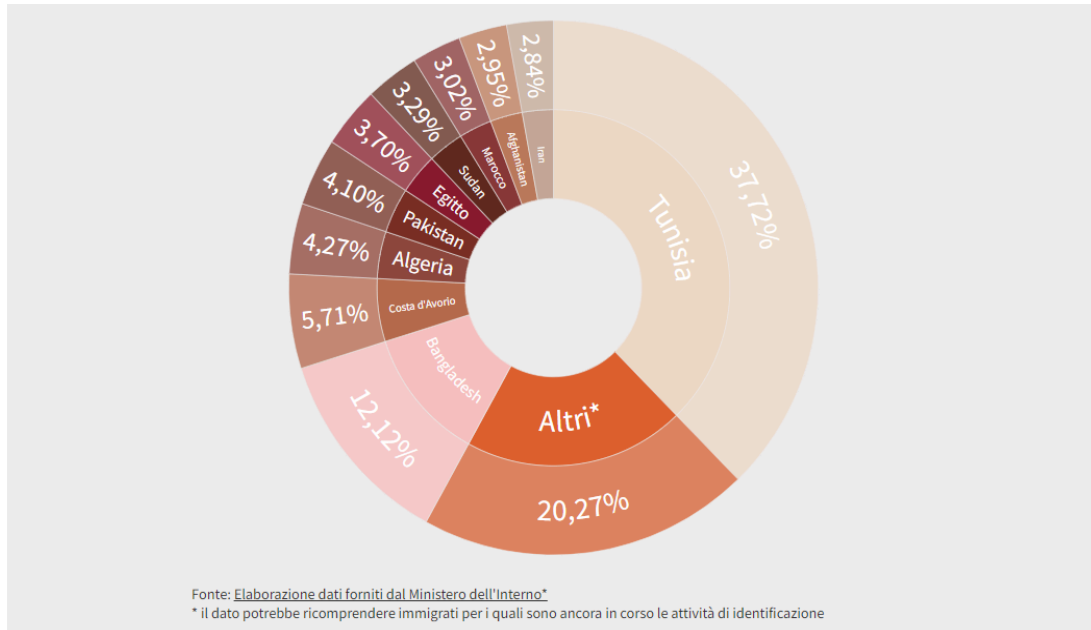
Fig 1. Number of immigrants who arrived by sea in Italy between 2014 - 2021



Source: www.statista.com

In the following table (Fig 2), we can see the declared countries of origin and the percentage of immigrants coming from each of it for the year 2020.

Fig 2. Country of origin of immigrant in Italy (2020)



Source: [Sbarchi e immigrazione in Italia: i dati degli ultimi 5 anni - YouTrend](#)

The migration phenomenon has started taking place in Italy quite recently, for forty years, a shorter period compared to the other western countries who were leading actors in the colonisation. However late, immigration in Italy has become a stable and structured phenomenon, making the country not a transitory one but a destination. In 2011, the regular migrants were 7.5% of the Italian population, a percentage higher than the European average (6.2%). As already shown, their origin is quite varied: 49% is European and mainly comes from East Europe (37%); the 26% comes from Africa; the 17,1% comes from Asia and the 11,3% from Latin America. Among the national groups the Romanians stand out with around 887.763 individuals, the Albanians (466.684), the Moroccans (431.529), followed by the Chinese (188.352), the Ukrainians (174.129) and the Filipinos (123.584). This results in a cultural and religious pluralism: the immigrants of Christian religion, mainly

orthodox, count for 43.7% of the total, the Muslims for 33,2%, the animists are about 19% of the total, while there are very few Hindus (2.5%) and Buddhists (1.9%).

The median age of the immigrants is 32,3 compared to 43,9 of the Italians. Italy is the second country to host the youngest group of foreigners. Also, it has been proven that foreigners help increase the birth rate of the country. In 2009, their births were about 16% of the total number of new-borns in Italy and their contribution to the increase in the Italian population is about 92%. Demographic studies show that in the long term, they could help evolve the population that, as we know, has been dropping for the past decades. According to Istat, the contribution of the immigrant women to the capacity of the Italian population is crucial: about a fifth of the total births is of immigrant women, also about 63 thousand new-borns are conceived with a foreign partner. The migratory dynamic lets us foresee an Italian population that is ever more variegated.

Furthermore, immigrants represent a non-negligible segment of the job market. The foreign workers are about 2 million and are active mainly in the sectors forsaken by the autochthonous population such as the agricultural, construction, personal services among others. The foreigners account for 10% of the total workers in Italy. Even though they often have less prestigious and remunerative jobs, the immigrants generally produce 9% of the national PIL and they are holders of 3,5% of economic activities of the country.

The advantages of having immigrant workers can be certified by looking at different indicators: according to the Caritas report, in a one-year period, they paid contributions to the *Istituto per la Previdenza Sociale* equal to EUR 7,5 billion and declared to the tax authorities a taxable amount of EUR 33 billion. Moreover, already in 2005, the 57% of immigrants had an Italian bank account and represented the 15% of those who had access to property through the purchase of a house. These data suggest that the immigrants, apart from being necessary workers, contribute to the development of the country as taxpayers, consumers, and even as savers.

To confirm the structural nature of immigration and its rooting in Italian society, it is worth mentioning that even if the overwhelming majority has a job that can be

qualified as generic, a quarter of them is evolving towards a specialised and qualified job. The progression of the career, albeit slow, is a crucial step of the integration because it facilitates the social mobility, a better collocation in the society and hopefully the breaking of the stereotype of the precarious and excluded immigrant. Also, one out of two regular immigrants are registered to a trade union organisation showing willingness to participate and be integrated. Additional indicators of establishment can be observed in the mixed marriages which represent 14% of the civil union celebrated in Italy and in the increasing number of students attending Italian schools equal to 7% of the scholastic population in 2019.

The other face of the migratory phenomenon, which is abused by the media, is the irregular immigration and the deviant acts committed by and often only attributed to foreigners. According to Istat, the crime rate of immigrants in Italy is slightly higher than that of Italians: 1.3% compared to 0.75%. It is however necessary to mention that 87% of crimes relate to the absence of residency permit, which according to the law Bossi-Fini is considered an offence punishable with imprisonment for 3 to 12 months.

This is the demographic, economic and social profile of the migratory phenomenon. In the light of the above statistical reading, it is worth mentioning what Italians think and really know about their immigrants. Experts state that the knowledge of the phenomenon only rarely comes from direct experience; more often Italians have instead an abstract vision, filtered by the stereotypes that are linked to the different ethnicities and influenced by the image portrayed by the media. The media have a significant role in shaping and interpreting reality, therefore they have a huge responsibility in the identification of the Us and Other, identification that has an impact on the foreigner's condition, social status and on the legitimization of inclusion policies. The sociologist Mario Marcellini says that "*the mass communication can in fact become the pivot itself and the key element in the construction of a diffident and intolerant society or, on the contrary, of a more pluralist and multicultural one*"⁸.

⁸ Mario MARCELLINI, « Alle porte della cittadella mediale, preludio alla lettura », in *Fuoriluogo. L'Immigrazione e i media italiani*, a cura di Marco Binotto e Valentina Martino, Cosenza, Pellegrini Editore, 2004, p.7.

Given the importance that means of mass communication have in the construction of collective representations and in a society always more ethnically and culturally pluralists, the following paragraph will quote the results of a research done on the media contents to understand if the disclosed information mirrors the real status of immigration in Italy and if it favours inclusion and respect of the diversity.

2.2 Immigration through media communication

Italy is a country with growing ethnic and cultural diversity. Currently, immigrants constitute 7% of the resident population. The media has a major role in the complex process of integration into Italian society. But how is the media supporting this process? To answer this question, we propose the results and insights of the analysis on media content carried out by C.Maltone. She is a professor of social and political history of contemporary Italy and a researcher at CEMMC (*Centre d'études des mondes modern et contemporaine*) at the University Montaigne (Bordeaux, France). Her main field of research is the study of immigration and its representations.

The following statements are based on the “analysis of content of public and private television channels as well as national and regional news journals of diverse political leanings that exposes the recurring connection between immigration and deviance, criminal tendencies and a propensity for producing stereotypes” , as she affirms in the study.

In general, it has been proven that the media somehow tends to participate in the reiteration of these prejudices and thereby transforming the reality of the immigrants. This causes, especially among the youths, feelings of intolerance and hostility towards certain ethnic groups that are perceived as dangerous regarding national security and a threat to their identity.

The complex thematic of defining how television, social platforms and newspapers talk about migrants, in which contexts and what image they spread has been a hot topic for studies and surveys since the beginning of 2000. One of the first study, *L'immagine degli immigrati e delle minoranze etiche nei media* (ADD

REFERENCE), was carried out by Censis in 2002 related to the European project *Tuning into diversity*⁹. The study was followed in 2003 by the survey *L'immigrazione e i media italiani*, promoted by the European project *Etnequal Social Communication*¹⁰ and carried out jointly by the Sociology and Communication department of the La Sapienza University (Rome), Amnesty International (Italian Section), Caritas (Rome) and Rai (i.e., Radiotelevisione italiana).

More recently, further studies have been promoted, among which: *Immigrazione nei media. Rete di monitoraggio sui media locali bolognesi* (2005) and *Osservatorio Carta di Roma. Il tempo delle rivolte* (2010). The mentioned studies, however, are local and often referred to a precise event. That is why C. Maltone took the analysis to a higher level, with the aim of having a wider and updated vision of the phenomenon. Together with her students, she carried out a survey on 15 newspapers both national and local with different political orientation, that are: *Il Corriere della Sera, La Stampa, Il Giornale, L'Unità, Liberazione, Il Secolo XIX, Libero, Il Foglio, Il Messaggero, Il Resto del Carlino, Romagna Oggi, Il Tirreno, La Gazzetta del Sud, Il Mattino e Il Giornale di Sicilia*.

In Italy, 90% of adults learn the news through the television that is still widely common. Therefore, it is worth mentioning its impact. Findings of the analysis of tv content (both private and public television) show that immigration is discussed for the 90% by news broadcasts, more specifically the 80% is included in the general current events' section and the 56,7% among the crime events. Overall, 78% of news concerning immigrants are negative and referred to criminal acts, drug dealing, thefts, homicides, prostitution, clandestinity, illegality, as well as social degradation caused by immigrants. News of illegal landings on the southern coasts constantly fill the broadcasts. News about positive occurrences is less common: work and social activities of immigrants count for the 3,5% of narrated facts, whereas the solidarity acts and support cover the 13% of the content.

⁹ Cf. the *Rapporto finale* of Fondazione Censis 2002, Roma, Fondazione Censis, 2002.

¹⁰ The project was financed by the Ministero del lavoro e delle politiche sociali in the scope of the EQUAL project with the aim of contrasting the prejudice and intolerance toward immigrants through media channels to facilitate their insertion in the job market.

In around 75% of told news, the immigrants are depicted by their nationality, their ethnicity, and their country of origin. Such a practice tends to produce dangerous stereotypes that consider the immigrant more as the representative of a category than an individual with its specificities. More generally, the theme of immigration is rarely discussed with debates, news columns or inquiries as it usually happens with aspects of Italian society. Furthermore, the narration of current events tends to involve the emotional side of users: in 81% of the cases, news arouses compassion, pity, or preoccupation. Also, even if immigrants are often protagonists of the broadcasts, they rarely get a chance to defend or talk for themselves, expressing their point of view on the happenings in which they are directly involved: in 65% of cases they are simply mentioned, and only for the 25% they are interviewed. In 85% of television news, men are involved whereas women are under-represented even if they constitute 49% of the total immigrated population.

For what concerns newspapers, the painted image is not much different than the television one. From the study of the headings (1058 articles) emerges the tendency of speaking about immigrants when they are the protagonists of negative news mostly concerning crime and marginally also terrorism. More specifically, data shows that in national daily newspapers, general news counts for 44,8%, terrorism facts count for 7,4%; the 35,6% of articles cover landings, clandestineness, debates about the regulation of the migratory phenomenon, usually in a negative light. News about immigrants' crime is even more present in local papers (70% of articles about immigrants).

Regarding the crimes reported, those against the person stand out - 42.3% of the news articles compared to the 29.7% against property and 10.4% against the economy. This media coverage is in contrast with the data of the Ministry of Justice according to which in the typology of crimes committed by immigrants, thefts clearly prevail over violence against people; it is also in contrast with the Istat investigations according to which the most widespread crime is clandestinity. Like television, 85% of the articles focus on male immigration. 80% of the articles published by national newspapers and 95% by local newspapers are short, 47.5% unsigned and only 11%

of the news concerning immigration is updated or subjected to a thematic discussion. It is quite rare that the same fact is followed in its development over time or deepened by several articles in the same edition.

The computerised treatment of the lexicon used by the fifteen newspapers confirms the tendency to associate immigration with crime and deviance. The terms that most frequently accompany the words *immigrant/immigration* are *illegal, traffickers, crime, reporting police checks, arrests, law enforcement, border re-entry, expulsion, repatriation, insecurity*. The use of such a linguistic register helps to create the stereotype – Immigration equals danger – and inevitably to fuel mistrust as well as to raise ever higher barriers between Us and Them.

It is also customary in the press to use nationality as the only reference to the protagonists or victims of the narrated events. The attribution to an ethnic category appears since the title, just to name a few:

Original text	English translation
<i>Forzano un posto di blocco. Presi due rumeni e un albanese</i>	<i>Forced a roadblock. Taken two Romanians and an Albanian.</i>
<i>Ferisce un carabiniere. Albanese in manette</i>	<i>Wounds a policeman. Albanian in cuffs</i>
<i>Non sono terroristi islamici. Assolti due operai marocchini</i>	<i>They are not Islamic terrorists. Acquitted two Moroccan workers</i>
<i>Tratta di uomini, arrestato Pakistano¹¹</i>	<i>Human trafficking, arrested a Pakistani man</i>

Romanian and Albanian are the most common stereotypes. Through this generalised reference to nationality, the press tends to associate entire ethnic groups with crime and illegality. “It would proceed”, as the Romanian writer Mihai Butcovan argues “to a collective condemnation of all people who belong to the same people or nation”.¹² The labelling through geo-cultural belonging risks attributing an offensive value to nationality and becoming a discriminating factor. However, immigration as a danger is also present in articles that go beyond crime, and which refer to landings or

¹¹ *Il Resto del Carlino* 2010, *Gazzetta del Sud* 28 agosto 2010 e 23 agosto 2010, *Corriere della Sera* 7 luglio 2010, *Il Secolo XIX* giugno 2010, *Il Tirreno* 10 e 14 agosto 2010.

¹² Interview to Mihai Butcovan titled *Butcovan, immigrazione e lingua del cuore*, 13 October 2009 published on the website Osservatorio Balcani e Caucaso.

refugees. The mainstreams tend to describe this phenomenon as ‘a problem’, an ‘alarming situation’, often as ‘an invasion’. According to the Italian Northern League newspapers, like *La Padania*, the immigrants that land on the southern coasts are ‘armies of illegal immigrants’ and ‘hordes of barbarians’. Using these metaphors can affect the reader and have a long-term impression in his mind.

Conversely, when the alarmist tones subside, immigration becomes instrumental to the Italian labour market; the focus is not so much on the needs of the migrants in search of a better life or the economic initiatives they have undertaken, as on the needs of the Italian economy. The immigrant is depicted as the one who must fill the vacancy left by Italians, that is do the jobs that Italians tend to avoid.

In the printed media, the share of articles, 11.6%, relating to the economic, working and health conditions of the migrant is quite small; even lower is the attention paid to its culture, religion, and integration; this news represents 8.8% of the total articles on immigration. Furthermore, it considers immigrants as a block of indistinct individuals.

The immigration – social dangerousness paradigm is particularly rampant in the conservative and populist press which tends to systematically present foreigners as an element of disturbance, disorder, degradation. These papers lean towards the idea that the Italians, especially the northern ones, are the real victims of immigration because, and I quote, ‘*submerged by individuals who invade our spaces, who do not respect our rules, who violate our values*’. The immigrant becomes the aggressor because he is an invader, filthy and uncivilised, a stranger to the Italian community and values.

A different image is painted by the progressivist and leftist press (such as *L’Unità* and *Liberazione*). Generally, the immigrant is described as a human being, owner of rights and obligations, to be integrated in Italian society and defended from racism. In this case, the most common lemmas referred to the immigrants are *rights, citizenship, equality, tolerance, solidarity, integration, anti-racism, anti-discrimination*. Going against the grain, these newspapers cover the social aspects of immigration deepening the study of their life history, their discomfort, and

the suffered injustices. However, this vision, which some define as good-natured, is a minority in the Italian media; most of the mainstreams return an image of a foreigner who is involved in illegal activities, report stories of forced coexistence, a climate of tension or exasperated with the host community.

As in a game of mirrors, both printed and television media, on one hand, think they are interpreting feelings of fear and mistrust inherent in the society, on the other, they are feeding them. The result is that the media coverage of immigration raises a discrepancy between real immigration and media immigration. The distorted representation of the immigrants tends to create confusion between the everyday migrant life and the exceptional, favouring a collective perception of the foreigner as dangerous. All of this leads to the idea of the immigrant as non-integrable. The disinterest in his ordinary life, in knowing the culture, in recognizing the contribution of immigrants to the country is decisive in preventing their integration.

The reference to nationality as the only element of identification is another restraining element of this long and delicate process. The constant reference to the place of origin in fact brings the immigrant back to his dimension as a foreigner, relocates him to his country of origin, isolates him from the host society, excludes him and almost seems to suggest the existence of two opposing worlds, ours, characterised by a positive self-presentation and, their world, charged with negativity. The absence of comments, problematization and in-depth analysis, only at a first reading can it appear an expression of a neutral or objective journalism, in truth we are faced with information that is profoundly simplified and flattened to result in a stereotyped representation of immigrants.

We could try to give an interpretation to the reasons behind the behaviour of the media. The communication expert Marco Binotto suggests that reporters often get the information from the Minister of Interior and the Police (*Forze dell'Ordine* and *Questure*). Frequently, articles are strongly affected by the police's idea of immigrants centred on criminal activity and collisions with the society. At the same time, it is well known that scandalous news makes more audience, therefore the

immigrant that breaks the law makes the headlines more easily than the one who follows it.

It appears that the reasons to negatively stereotype the immigrants are also cultural and political. In the media alarmism we can read the fear that Italy, in welcoming illegal immigrants or being an easily reachable side, could be perceived by Europe as the soft underbelly of the Union and therefore as an unreliable partner. There is the fear of being perceived as a country that is poor and going toward uncivilization. In the Italian context of pauperization of the middle class and of a development without growth, we tend to reflect into the immigrant the anxiety for the social regression and the loss of the wealth fare and the fear of being left behind by the other European countries. Furthermore, showing the immigrant in a negative light can become a means to blame for the social problems of the country, such as the unemployment, the precariousness and, the economic stalemate. It becomes the politicians' means to gain consent from Italians, by proving to them that they will be safe and will always be the priority.

The anthropologist Annamaria Rivera suggests as a motive the fear of losing the identity of the nation. The immigrant is seen as different and bearer of divergent ideals and customs into Italian culture. The diversity of the foreigner forces us to interrogate ourselves about our nature, history, and values. This contrast can be harder on Italy which is a country with a fragile identity.

The real problem that follows this disinformation is the hindering of integration, placing Italian society in front of greater difficulties so that it fosters attitudes of intolerance, hostility as well as the spread of a neo-racist culture, that is, a contemporary racism very different from the past. The sociologists Hall Stuart and Dorothy Hobson argue that "racism is a historically contextualised ideological construction that changes according to the economic, political and socio-cultural condition". Also, the scholar Martin Barker in his essay *New Racism in the United Kingdom* examined the changes that occurred after the war: "the alleged differences of yesterday are now replaced by differences between cultures, religions, or nations. In Western societies, the new racism takes on a new form, the defence of our values,

our way of life, our traditions against strangers not because they are inferior but because they belong to other cultures. Racist practises develop due to the spread of stereotypes, prejudices, and commonly shared representations. This cultural racism ends up legitimising a different economic, juridical, and social treatment of aliens to one's own culture. The differentiation of statutes expressed by a limited number of rights and opportunities leads to the legalization of the discriminatory principle" [C. Maltone].

The new racism, as said by the linguist Tuen Van Dijk, is exercised moderately by people who even though they declare themselves as democratic, tolerant, respectful of multiculturalism, when communicating they take their distance from the ethnical minorities. Media have accelerated this subtle form of racism making, precisely, hinging on stereotypes, on the Us/Them polarisation, on the minimization or concealment of the virtues of others and the emphasis on one's own.

2.2.1 How immigrants are perceived by the Italian society

While the scientific community recognizes that collective reception depends on various factors – such as experience, individual culture, social status, habitat – they believe that there is a very strong link between media and collective representation to such an extent that public opinion tends to perceive as real what appears on television or what is read in the newspaper. Therefore, in our society, the migratory phenomenon in the collective imagination is shaped by the media. To understand the idea that Italians gave about immigrants and to measure the impact that media have on it we can refer to the research *Io e gli altri. I giovani italiani nel vortice dei cambiamenti*¹³. On one hand, the answers of the young reveal a wide openness toward diversity, universalism – probably thanks to globalisation – on the other hand, they reflect the same fears and mistrusts told by the media.

¹³ The research was carried out on a sample of 2085 young Italians aged between 18-29 years. It was promoted by the Conference of Regional Presidents and conducted by the Iard Institute (Institute of socio-economic research). The report is available on the website: www.parlamentiregionali.it (2019)

72% of the young interviewed believe that immigrants, provided they are legal, must be able to have the same social rights as natives; 63% say they are in favour of extending citizenship to immigrants: 52% would grant them the right to vote. In general, these answers were given by young people with a medium-high level of education and who are non-practicing Catholics. Two-thirds of the young people interviewed, in declaring themselves willing to share their rights with non-nationals, express a rather positive image of immigrants.

This positivity is confirmed by the responses inherent to the role of foreigners in the workplace. 68% of young people declare that they do not fear their competition and 47% emphasise the benefits that Italy derives from their activity in terms of support for the pension system and the welfare state. Only 24% believe that immigrants take jobs away from Italians. These data, according to the researchers of the Iard research centre, show that, after thirty years of immigrant presence and for young people who have known and lived in an ethnically diverse society, living alongside people from other countries is part of normality.

This attitude of openness, however, begins to fray when the variable economic crisis/shortage of work takes over; 49% believe Italians should be given priority in hiring; 26%, mainly the less wealthy ones, find it unfair to help immigrants before their own situation is resolved. Also, the feeling of economic solidarity towards immigrants prevails only in the 37% of cases compared to the 36% that is placed in the middle, undecided. These findings indicate a wavering tolerance and a tendency to believe in the Italians' supremacy. The positive image is even more lessened when considering the cultural appreciation; 43% considers the presence of immigrants as non-threatening to the Italian identity, while 31% has an opposite opinion and 26% hesitates. The distress is more common among northern young people.

C. Maltone affirms in his work: "from this picture the researchers do not draw optimistic conclusions. It seems that young Italians would also be permeable to stereotyped images and crossed by impulses of rejection or reticence towards immigrants, as shown by the responses inherent to crime and the level of satisfaction of the various ethnic groups. Among the young interviewed, while 47% do not share

the idea of immigration equals crime, 54% believe that some ethnic groups are more prone to violence. In the ranking of the most aggressive there are the Rom, the Romanians, and the Albanians followed by the Arabs, the Maghribs and the Balkans. Among the less dangerous groups: the Chinese, the Filipinos, the Indians, and the Africans. In this hit-parade of ethnic violence, all the conditioning of the media unfolds. It is clear that the recurrent combination of nationality and crime may have produced this type of stereotype in young people. The immigration-crime binomial is emphasised above all by less educated young people and more convinced Catholics". It is presumed that media in mentioning the ethnicity concur in creating prejudices that become feelings of frustration and aversion. 69% of young people dislikes the Roman, 55% the Romanians, 52% the Albanians; 47% the Arabs, 43% the Balkans, 40% the Muslim Turks, 39% the Chinese, 36% the Maghrebi, 31% dislikes Russians and Ukrainians, while Africans, South Americans, Indian, and other Asian populations have a dislike index less than 30%.

This and other studies demonstrate how the media are often places of rumination of stereotypes, of distortion of immigrant reality as well as of hidden xenophobia. However, these journalistic practises are against the Italian and European legislation and with the code of ethics of the Order of Journalists. Moreover, the "Carta dei Doveri: etica e deontologia" (1993) compiled by the *Ordine dei Giornalisti* and *Federazione Nazionale della Stampa* indicates as fundamental obligation the following:

"The respect for the person, their dignity, their right to privacy and non-discrimination based on race, religion, sex, political opinion, physical or mental condition"

Other recommendations aimed at promoting information that is free, truthful, and non-racist are included in the *Dichiarazione d'impegno per un'informazione a colori* (1994) and the *Carta per un'informazione non razzista* (1996).

Given that these ethic codes have been disregarded, lately many journalists have started reflecting on their way of operating and in 2007, the National Order of Journalists (tr. *Ordine Nazionale dei Giornalisti*) compiled the *Carta di Roma*,

namely a deontological protocol, in which the media are invited to a more correct and in-depth treatment of news on immigration, both in terms of language and themes. Among the aims was the modification of the information about the immigration through the use of the right lexicon. Words of media are relevant because they are neither neutral nor without history; they could sound as a salvation, or a condemnation and they leave an impact on the listeners. “Words”, as the American philosopher Judith Butler says, “they produce social and personal subjectivity; once pronounced they cannot be reset; they can create collective lacerations that are difficult to mend”. A reformulation of the journalistic language could favour social interactions and lead to a high-profile reflection on citizenship, on the multicultural models of society that we intend to imagine.

The awareness campaign on the right use of words includes the definition of a new glossary to be used when talking about immigrants, and the exclusion of some disrespectful and prejudiced words. An example of the words to avoid is “*clandestino*” (tr. clandestine). This term is highly used by Italian media to indicate foreign people that do not have a residency permit; however, in the collective imagination, it evokes secrecy, lives spent in the shadows and associated with crimes. Therefore, besides being used inappropriately, this term arouses negative images. Some proposed alternatives are *irregular*, *illegal*, *undocumented*, *asylum seeker* or *refugee*.

Another example is “*extracomunitario*”, which literally refers to the non-EU citizens, but it is never referred to Americans or Helvetic, and only used to designate people coming from poor countries. This term emphasises the extraneity from Italy, the being different and excluded from Italian culture. Other words to be eliminated are *badante*, *vu cumprà*, *zingaro*, *nomade*. Good alternatives are *collaboratore familiare* (tr. family worker), *venditore ambulante* (tr. peddler), *Rom*.

A support to this campaign came from the civilian society that in 2008 opened the website “*occhioaimedia*”¹⁴. Its main role was to observe the activity of mass-media

¹⁴ www.occhioaimedia.org

and collect titles and articles offensive towards ethnical minorities or somewhat racist.

A similar initiative was undertaken by the group *Giornalisti contro il Razzismo* (i.e., Journalists against Racism) as a response to a wave of articles and TV broadcasts that pointed the Rom as “..., dangerous, violent, linked to criminality, source of problems to our society”. This group of publicists invited readers to report to the Order of Journalists the articles that seem discriminatory, xenophobic, inciting racial hatred, etc. Lastly, another helping action was the ratification of the UNESCO Treaty on the Protection of Cultural Diversity by the media.

These initiatives, however, are considered as insufficient by some journalists and intellectuals, so that some have asked for the rather problematic institution of supervisory bodies on the correctness of information and on the cultural pluralism of the media, with sanctioning powers (such as fines or subtraction of public funding) against those mainstreams that favour an oligo-cultural or monocultural representation of Italy.

Finally, as mentioned above, from the research on media and immigrations it emerges that immigrants hardly ever are given the possibility to express or defend themselves, or simply to talk about their needs, expectations, their everyday life, their perception of Italy and Italians. Italian society asks its immigrants for rapid social and cultural integration without the media providing them with information that favours this process. Paradoxically, they are not the recipients of information, even though in the Censis it is said that “information is a basic necessity for immigrants, essential for knowing the realities and codes of the host country as well as for creating social relations and integrating”¹⁵.

From this paradox, the multicultural media was born. Multicultural media include television or radio broadcasts called socially useful aimed at turning the spotlight on the hardships, loneliness, difficulties encountered by immigrants without pietism and stigmatisation and to provide practical information about their rights, duties, administrative procedures, regulations relating to work, health, or residence card. In

¹⁵ Rapporto Censis, *L'immagine degli immigrati e delle minoranze etniche nei media*

open contrast with the mainstreams, these service media try to satisfy the real needs of the immigrant citizen by providing him with concrete answers. By distributing this type of information, they tend to promote, on the one hand, a gradual integration of immigrants and, on the other, to develop greater sensibility towards migrants in Italian society. They are part of the very early media to consider immigrants as new citizens and not as new citizens and not as unwanted guests.

Intercultural journalism is expressed both in the form of inserts or special pages in the traditional press or as a completely independent newspaper. One of the earliest intercultural experiences found in Italian newspapers is *MondoInsieme*, an insert combined with the Reggio Emilia Gazette; *Incroci. Vicenza crocevia delle culture*, monthly supplement hosted by the newspaper *Il Giornale di Vicenza*; *Bergamondo*, weekly insert of the newspaper *L'Eco di Bergamo*; the Monday page of the *Sole 24 Ore* dedicated to the world of immigration. The experience of *Yalla Italia*, a magazine produced by young second generation Muslim women distributed together with the weekly *Vita*, was unusual. In some cities, autonomous initiatives have been launched such as *Città Meticcias* in Ravenna, *Métissage* in Aosta and among the latest creations, *Il Tamburo*, born in 2007 in Bologna.

The weekly supplement of *La Repubblica*, *Metropoli*, had a national circulation and was entirely conceived and dedicated to foreigners living in Italy. *Metropoli* distinguished itself from the dominant press because it introduced profound changes: immigrants were the recipients of the information, they expressed themselves with their voices, immigration was treated with greater sensitivity and preparation and finally, a fundamental novelty, the reporters have given a positive and enhancing representation. This journalistic narrative unambiguously translated the integrationist and pluricultural position of the newspaper.

The use of the Italian language was the clear sign of its intercultural and dialoguing approach; *Metropoli* was in fact a newspaper that intended to act as a bridge between migrants and Italians and between different communities. Italian as a language of communication was an anti-discrimination and unifying choice.

However, this journal did not talk about the other side of immigration made up of prostitution, drugs, illegal human traffic, etc. The omission of controversial themes and a univocal vision would have made *Metropoli* fall into the trap of simplification and the reader into the illusion that Italy is a multicultural country, respectful of diversity. In short, *Metropoli* would have represented an idealised reality, a society without barriers and at peace.

This newspaper was also innovative for having encouraged foreign journalists or journalists of immigrant origin, especially women, to enter the editorial office, even if only Italians were really in charge.

The intercultural media is flanked by ethnic ones, a vast aggregate of information consisting of newspapers and radio and television broadcasts in the language of origin intended and dedicated to specific national, linguistic, or geographical communities. The audience to which this type of media is addressed is made up of 190 nationalities or ethnic groups and has over 3 million users of which one and a half million readers.

Ethnic newspapers range from news of political and cultural events in the country of origin to news and comments on what is happening in Italy, with particular attention to immigrant issues and service information. These media are therefore spaces self-managed by the individual linguistic communities with the function of imparting, especially to newcomers, essential information on the reality and the Italian bureaucratic rules.

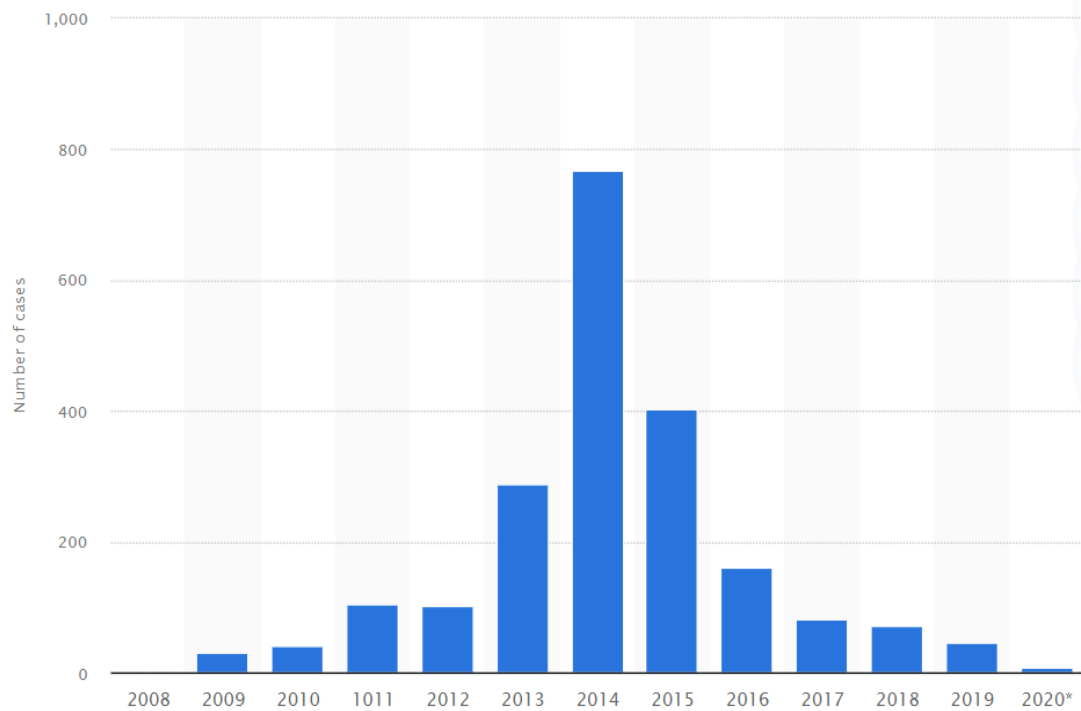
These identity spaces are a fundamental part of the process of integration of different cultures. Nowadays, as C. Maltone says, “Immigrants are a structural component of the Italian population and in the current multicultural and multi-ethnic context, less ethnocentric information, free from generalisations, phobias, and prejudices, is becoming increasingly indispensable. Information that respects diversity, which considers difference as an asset to be protected, which looks at diversity as a resource, can play a key role both in the process of inclusion of new Italian citizens and in giving them one more reason to reflect with trust and hope in the country of Italy and to feel a constructive part in the adopted society”.

2.3 Hate speech against immigrants: reading of the statistics

In Italy, the episodes of public hate speech and incitement to hatred motivated by racism have been increasing in the last years. In the first three months of 2020, eight cases were recorded, while in the previous year, there were 45 episodes. The data only reports the cases collected by source. Thus, the actual cases of racial hate speech and incitement to hatred might be much higher.

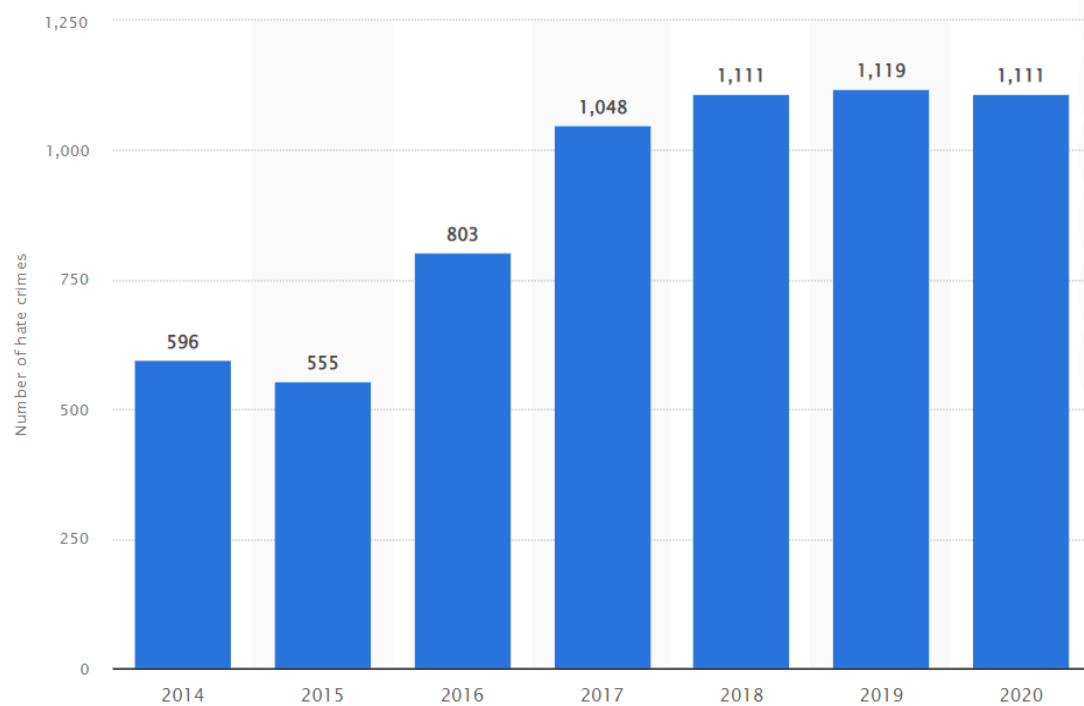
Hate speech and incitement to ethnic or racial hatred represent a major issue in Italy. In politics, for instance, this phenomenon is quite widespread. For example, one of the largest political parties in the country, *Lega*, has been making a significant number of statements based on xenophobia and incitement to hatred.

Fig 3. Number of hate speech incidents motivated by racism in Italy 2008 - 2020



Source: www.statista.com

Fig 4. Number of hate crimes reported by the police in Italy 2014 - 2020



Source: www.statista.com

2.4 How Covid-19 has affected hatred and hate speech towards minorities

The spread of the COVID-19 epidemic throughout the whole world during the past two years has been cause of economic and social crisis in several countries and has increased the mistrust of other populations, powered by the constant news about countries in lockdown, new virus' variants (often identified by the country in which it first spread - e.g., Omicron, the African variant) and the closure of borders. Since its beginning people have tried to find someone to blame, the infector. At first, they referred to it as the "*virus giallo*" (tr. yellow virus), implying that it was caused by the Chinese; hashtag like *#kungflu*, *#chinesevirus* were trending on Twitter in the United States in February and March 2020, which caused the increase of online hat and bullying by the 900%. However, overtime and with the progression of the virus, the scope of targets has broadened.

Already at the end of March, Fernand de Varennes, the *special rapporteur* of the United Nations on minorities said "Covid-19 is not just a question of health but a virus capable of exacerbating xenophobia, hatred and exclusion". And just over a month later, on the 8th of May, it was the turn of the Secretary General of the United Nations Antonio Guterres to reaffirm the concept by appealing to all states to counteract the hate speech tsunami linked to the spread of the virus.

Therefore, it makes sense to understand how this general feeling of distress and mistrust towards others has impacted minorities and migrants in Italy. To do so, we avail ourselves of the survey carried out by Amnesty International Italy¹⁶, that has monitored content published in social networks and journals' headlines. The posts were collected starting from June 15, 2020, through September 15, 2020. Comments were collected from the same date up to 30 September 2020 (15 days after the closure of the post collection, to ensure that most of the comments were collected.

¹⁶ Amnesty International Italy, *Barometro dell'odio 2021 - Intolleranza pandemica*
[Layout 1 \(d21zrvtkxtd6ae.cloudfront.net\)](https://www.amnesty.it/Barometro%20dell%27odio%202021%20-%20Intolleranza%20pandemica)

During such a period they collected 177000 posts and tweets published by the authors of public/pages profiles (44,000 from Facebook, 133,000 from Twitter) and 22 million comments (13 million from Facebook, 9 million from Twitter).

The problem can be taken back to the polarisation of *us*, to be helped, and *them*, those who do not have the right to be helped, with the government to be blamed if it does not follow this ‘order’. In fact, around this polarisation, not only new social fractures are built (e.g., the “*professori*”, the professors are overly protected against other categories of workers), but also the concept of “parasitism” is consolidated, e.g., the words “*parassita*” (parasite) and “*zecca*” (tick) are frequent epithets. Due to the circumstances, long-standing hostilities are rearticulated such as that between “immigrants” and “poor autochthonous people”. Thus, for example, the idea around the “*clandestino*” is radicalised both as factors being “bastards”, “criminals”, “terrorists”, “infected”, and as subjects of plots to “favour the clandestine immigration” and of aid denied “to Italians”.

A clear example is what happened around May 10 and 11 2020 to Silvia Aisha Romano¹⁷, “guilty”, according to her detractors, not only of having been freed (with what money?), but also of being so ungrateful that she converted and became Muslim. Instead of expressing satisfaction for the happy ending of a dramatic situation, both sexists and haters stormed adopting a colonial and Islamophobic look to judge the clothes and conversion of the woman and an increasingly ravenous, aggressive, sensationalist press, which should have protected the person instead of feeding to public opinion confidential information, inferences, and summary judgements.

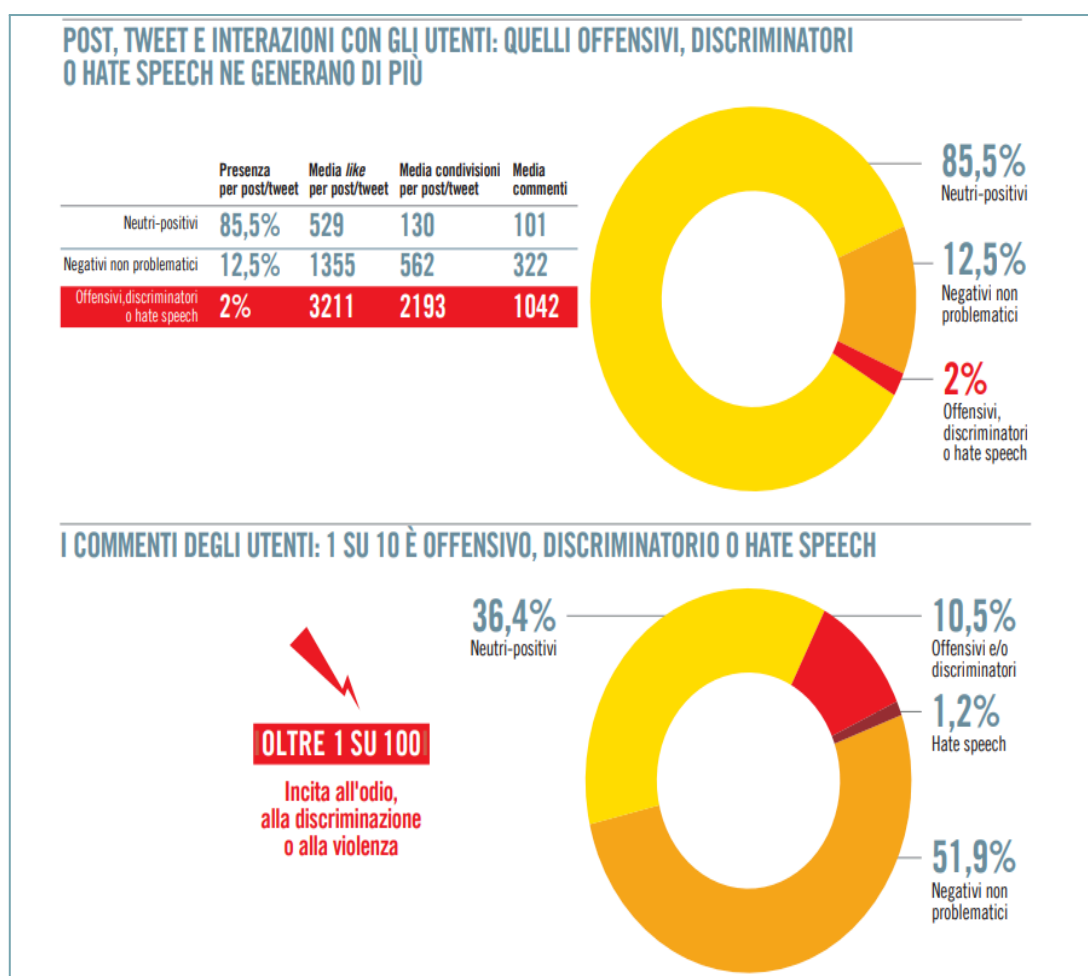
The combination of hate speech and propaganda is not surprising, but still cause for worries. In a hard time, it hit two vulnerable subjects: the ‘victims’ and the public opinion made fragile and insecure by the circumstances. In this case, hate speech had at least three interrelated objectives: trying to oppose the central government and the political forces that supported it cashing in on growing discontent; divert attention

¹⁷ Silvia Aisha Roman has a bachelor degree in linguistic mediation for social security and defence. She was kidnapped in 2018 in Kenya during her second mission as a volunteer of the “Africa Milele Onlus” association. May 9 2019, Italian intelligence freed her. Her return in Italy was object of interest because of her conversion to Islam.

from deficiencies of many regional governments, often governed or supported by the same political leaders who tried to stir up crowds on social media; indicate minorities and migrants as an easy scapegoat to be pointed out if necessary (the “illegal immigrants” as “infected”, “affected parasites from Covid”; foreigners as a pandemic “outbreak”, “asymptomatic than yes make them unavailable for checks”).

The Amnesty International work studies the magnitude of what the New Yorker called *public shaming pandemic*. Looking at the results on the comments, it turns out that 10.5% is offensive and/or discriminatory and 1.2 % is hate speech.

Fig 5. General course of the debate



It has been noticed a growth of 0.5% of hate speech content. A variation which, if referred to the size of the analysed percentage, represents a growth of 40% in the incidence itself. Together with other elements useful to read the phenomenon, such

as the recorded peaks of hatred and the analysis themes, targets and terms, this change could be indicative of a radicalization of hatred on the net. The anxiety and fear generated by the social and economic crisis find expression in the radicalisation of online hate. People feel the need to annihilate those who are the alleged cause of this problem, especially where the perpetration of the pandemic is perceived because of the lack of control of the situation by the authorities. However, the targeted groups are always the most vulnerable ones, that cannot defend their human rights.

Economic, social, and cultural rights are the main themes of the posts/comments. Immigration remains the most common topic, followed by women and gender rights, LGBTI, disability, religious minorities, and Roma.

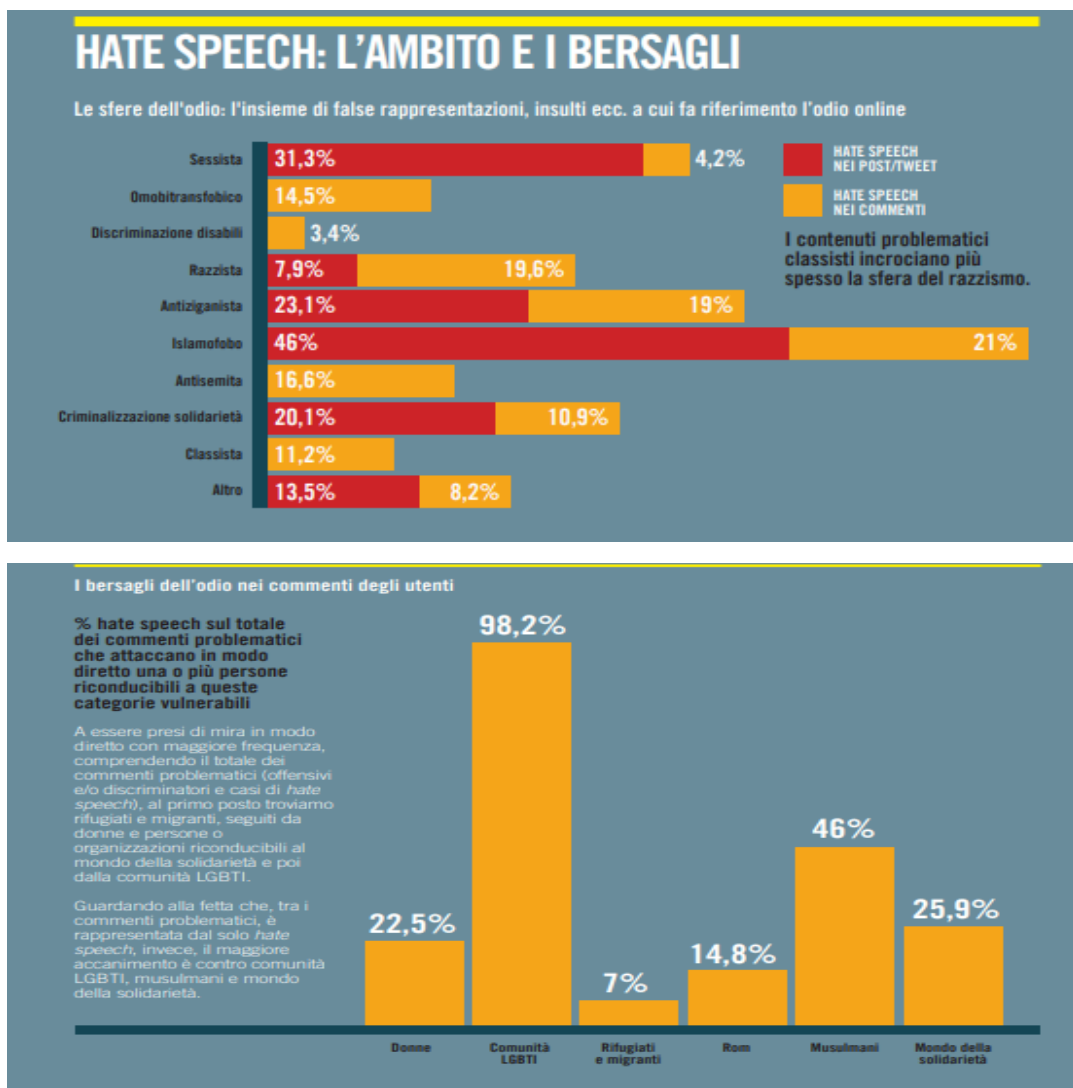
Not all the comments are negative, but the positive ones generate less likes, comments, and shares. On the podium of problematic comments, we find “religious minorities” (55.6%), “Roma” (47.6%) and “immigration” (42.1%); on that of hate speech, only the order varies: “Roma” (14.1%), “religious minorities” (12.7%) and “immigration” (7.9%).

Fig 6. Main topics: how much and how they are talked about

NEI POST/TWEET							
TEMA	PRESENZA %	ACCEZIONE NEGATIVA %	PROBLEMATICI %*	HATE SPEECH %	MEDIA LIKE	MEDIA CONDIVISIONI	MEDIA COMMENTI
Donne e diritti di genere	3,5	8,3	4,3	1	864,1	272,3	125,4
LGBTI	1,3	7,8	7,1	7,1	2186,9	805,3	259,6
Disabilità	0,5	4,1	1,8	0	11167,9	1685,6	493
Immigrazione	7,1	32,5	17	1,8	1320,1	662,5	374,8
Minoranze religiose	0,4	23,2	10,8	6,1	913,1	773,1	441,9
Rom	0,1	74,1	40,2	8,3	2312,7	1218,4	581
Solidarietà	1,5	28,9	14,1	3	4033,9	1113,8	337,1
Diritti economici sociali e culturali	29,7	16,4	1,8	0,1	669,3	205,2	126
Altro	56,8	12,6	0,6	0	640	156,2	112,6






NEI COMMENTI DEGLI UTENTI				
TEMA	PRESENZA %	ACCEZIONE NEGATIVA %	PROBLEMATICI %*	HATE SPEECH %
Donne e diritti di genere	1,9	59,5	26,7	2,5
LGBTI	0,4	62,9	25,2	5
Disabilità	0,5	33,3	7,2	1,3
Immigrazione	8	79,3	42,1	7,9
Minoranze religiose	0,4	75,9	55,6	12,7
Rom	0,2	92	47,6	14,1
Solidarietà	1,1	50,9	28,4	2,3
Diritti economici sociali e culturali	27,4	75,2	9,9	0,7
Altro	54,2	62	19	1

Fig 7. Hate Speech: scope and targets



In a further analysis, Amnesty International collects pages/profiles of newspapers, politicians, trade union organisations and bodies related to welfare. They considered a sample of 38: 22 politicians (half men and half women), 8 newspapers, 4 representatives of workers, 4 relating to welfare. The biggest weight in the observed debate is that of politicians and newspapers. Also, considering the public reactions on Facebook, the politicians arouse more anger, whereas newspapers arouse laughter but are often meant as a mockery.

Fig 8. Users' reaction by type of page/public profile

Le reazioni predilette dagli utenti per tipologia di pagina/profilo pubblico					
					
	Haha	Wow	Love	Sad	Angry
Politici	11,80%	2,79%	39,28%	8,55%	37,56%
Testate	15,08%	3,93%	29,33%	28,88%	22,78%
Rappresentanti del mondo del lavoro	1,05%	0,62%	74,67%	19,73%	3,92%
Enti welfare	5,97%	18,62%	30,31%	35,56%	9,55%

Furthermore, regarding hate speech, it has been noticed that:

- The 5 posts that generated the highest incidence of hate speech are all focused on the themes of “immigration” and “religious minorities”; they are problematic, and all published by right-wing politicians on Facebook.
- The top 5 in generating the highest incidence of offensive and/or discriminating comments and hate speech, are once again all regarding the topics of “immigration” and “religious minorities”; among them, one is published on Twitter by a right-winged newspaper and 4 are published on Facebook by right-wing politicians.
- Also, regarding the economic, social, and cultural rights, the main topic is immigration; however, in these cases, there is a higher transversality both in terms of categories of public profiles /pages and with respect to the problematic nature of the comment itself.

Some interesting indicators with respect to the way the online debate on the social, economic, and cultural rights is configured, come from the headwords found more frequently within the corpus of analysed contents. Starting with those that, regardless of the meaning of the content itself, are more present, the first three are: “oggi”,

“lavoro”, *“governo”* (i.e., today, work, government). However, when considering only the offensive ones, the order changes, and the word *“governo”* is the most common one. That is because, not surprisingly, the government is considered responsible for the difficulties experienced by Italians.

Moreover, it is interesting observing that if we look at the most common words cross-referencing them with the reading of the contents in which they are found, in the witch hunt on the web unleashed by the health and social crisis that impacts on the people’s emotions, the government is often accused in combination with those groups, that are the most fragile and vulnerable social identities- already mentioned in the previous pages. In fact, the government is considered “guilty of privileging them, the scapegoats (the infectors, those who steal the jobs and benefit from economic support), while ignoring and harassing us”, as most Italians think.

Going back to the list of most common words, on the third and fourth place we find *“migranti”* (migrants) and *“clandestini”* (clandestine), and also *“stranieri”* (foreigners) at the twentieth position, followed by *“immigrati”* (immigrants), *“risorse”* (resources), *“sbarchi”* (landings) and *“#rom”*.

To sum up, the first blame that is attempted to attribute is that of the health crisis itself. Who spread Covid-19 in Italy? If at the beginning, the Chinese were the main target, now migrants and refugees are, indeed, the favourite gourmets of the hater online, facilitated by the fact that their target does not have a voice to defend himself. Not too rarely, they are incited by politicians blaming not only the migrants, but also those who allow them to arrive, asking to put an end to these arrivals.

Furthermore, today among the alleged benefits reserved for them by the Italian government they are blamed for being those for whom the limitations are not valid, not being subjects to controls.

Some examples are shown in the following images (Fig 9.1 and Fig 9.2).

Fig 9.1 Example of hate comments towards immigrants

“basta sbarchi ci infettano tutti lo volete capire???”

“Si rischia un'altra epidemia! Blocco navale”

“Stop importazione clandestini malati di Covid e altro”

“Il giorno che li vedrò soccombere sarò la donna più felice del mondo. Li voglio vedere in carcere. Perché sono dei criminali. Non so se si rendano conto che tutti questi negri malati non bisogna più farli entrare. Ma cosa c'è sotto per questo accanimento”

“C'e' solo una cosa da fare, chi li ha fatti sbarcate, vuole la MORTE DEGLI ITALIANI. LA PRECEDENZA E' X I POLITICI PRO IMMIGRATI. TUTTI...”

“Un garante che, NON GARANTISCE IL PROPRIO POPOLO FACENDO SBARCAR MIGLIAIA DI CLANDESTINI INFETTI VA RIMOSSO CON OGNI MEZZO”

Quest'ultimo commento introduce un altro elemento che notiamo in numerosi commenti: il richiamo all'esigenza di *fare qualcosa*, imbracciare le *armi*, i *forconi* contro chi è considerato al vertice del processo decisionale.

“ci vorrebbe una rivoluzione contro questi figli di.....”

“Rivoluzione!!!! Li mandiamo a casa a calci in culo”

“L unica giustizia ..che potrai avere.. sarà quella che ti farai tu!!!”

“ITALIANI METTIAMO MANO ALLE ARMI E INCOMINCIAMO A FARCI GIUSTIZIA DA NOI”

Fig 9.2 Example of hate comments towards immigrants

“Mi raccomando controllate bene se sono Italiani, se sono clandestini e, possibilmente infetti, voltatevi dall'altra parte e fateli scappare”

“Qui a Vicenza continuano ad arrivare badanti/lavoratori dalla Serbia. Viaggiano indisturbati ed evitano i controlli. ??????”

“Non solo quelli del Bangladesh sono infetti ,che sono a zonzo..e gli africani arrivati in tutti questi mesi sono infetti pure loro ,e sono a vagabondare evinfettare noi ? Eh ? Quelli non li fermate no ? Quelli servono alle coop rosse ,scafisti,centri di accoglienza , la chiesa ecc., Soldi, solo bisness. Credo che tra non molto chiuderanno noi nuovamente questi fancazisti altroché !”

“La multa di mille € chi paga? Gli immigrati? Che non portano mai la mascherina e scappano dai centri di accoglienza?”

To support the results obtained by the Amnesty International survey, the study *Mappa dell'Intolleranza*¹⁸ of the Vox- Osservatorio Italiano sui Diritti, done by the co-founder Silvia Brena, confirms the above-described tendencies. In particular, it is found that even if the positive comments are more than the negative ones, there is a radicalisation of online hate: the peaks of intolerant comments are very high and focused on specific targets.

Going into detail, the Map highlights a redistribution of the negative tweets; indeed, in 2019 the most affected clusters were migrants (32.74%), followed by women (26.27%), Muslims (14.84%), people with disabilities (10.99%), Jews (10.01%) and homosexuals (5.14%). In 2020, the first two places are women (49.91%) and Jews (18.45%), followed by migrants (14.40%), Muslims (12,02%), homosexuals (3.28%) and people with disabilities (1.95%). Therefore, women remain the preferred target.

Finally, it is worth mentioning the growing phenomenon of *zoombooming*. The word refers to the phenomenon whereby some unwanted and uninvited people, often organised in small groups, take part in video conferences or meetings on Zoom, Google Meet, or others to disturb, offend, and prevent participants from talking. At worst, they share materials that are sexist, homobi-transphobic or praising racism, denial, misogynists.

2.5 Contrasting Hate Speech

During the last years, Italy has started taking action against online hate and hate crimes. An example is “*Contro l’odio*”, a platform for detecting, monitoring and visualising Hate Speech against Immigrants in Italian social media. It applies a combination of computational linguistics techniques for hate speech detection and data visualisation tools on data drawn from Twitter.

¹⁸ Born in 2015, the Mappa dell'intolleranza is a project focused on Twitter and aimed at the extraction and geolocation of tweets that contain sensible words, trying to detect the sentiment that animates online communities.

In 2020, the survey covered the months from March to September; 1.304.537 tweets were analysed (43% of which were negative comments and 57% were positive ones).

3 Hate Speech Detection: The Analysis

3.1 Goal of the analysis

The idea behind this study is to expand the scope of the hate speech detection to the newspapers' articles. The goal is to analyse the text of articles from different Italian newspapers to detect hate speech and more generally their sentiment.

The challenge of this analysis relies on two aspects: on one hand, the study is done on the Italian language, for which the sentiment analysis algorithms are less developed, and the research is thinner; on the other hand, newspapers' articles, even if collected from websites, are longer than tweets or posts, requiring more trained techniques. Above all, the greatest challenge linked to articles is that they are part of official sources of information, therefore they tend to have cleaner language and to maintain an objective point of view when reporting facts and narrating events.

We will try to adapt some of the existing models for the Italian language (mentioned later in detail), which were pre-trained mainly on tweets/comments/posts and will test their performance. The analysis is performed using the Python programming language.

3.2 Related work

As already mentioned, the research for the Italian language is thinner than for the English one but several progresses have been made.

An important contribute is given by MediaVox¹⁹, an online information and communication laboratory against hate. It is led by Milena Santerini and supported by different organizations among which there is the Università Cattolica del Sacro Cuore. The aim of the observatory is to promote information of quality on the web

¹⁹ <http://mediavox.network/>

spreading a different way of narrating facts. They carry out advanced research on Twitter contents, analysing online hate and categorizing it and its authors.

Another big impact is given by the initiative called EVALITA, which is a periodic evaluation campaign of Natural Language Processing (NLP) and speech tools for the Italian language. As written in their homepage²⁰, “*the general objective of EVALITA is to promote the development of language and speech technologies for the Italian language, providing a shared framework where different systems and approaches can be evaluated in a consistent manner*” and “*The good response obtained by EVALITA, both in the number of participants and in the quality of results, showed that it is worth pursuing such goals for the Italian language*”. EVALITA is an initiative of the Italian Association for Computational Linguistics (AILC) and it is endorsed by the Italian Association for Artificial Intelligence (AI*IA) and the Italian Association for Speech Sciences (AISV)²¹.

Finally, the best functioning NLP models have been developed by the University of Bari A.Moro – Department of Computer Science and the University of Turin – Dept. Computer Science. They successfully trained a language understanding model for the Italian language (AIBERTO²²) that uses the so-called BERT classifier. AIBERTO is focused on the language used in social networks, specifically on Twitter.

The University of Turin is also active in the detecting Hate Speech against immigrants in Italy. The paper “*An Italian Twitter Corpus of Hate Speech against Immigrants*” (2018) by M.Sanguinetti, F.Poletto, et al., is one of the most important studies in this field. Their paper describes a recently created Twitter corpus of about 6,000 tweets, annotated for hate speech against immigrants, and developed to be a reference dataset for an automatic system of hate speech monitoring.

²⁰ [evalita – Evaluation of NLP and Speech Tools for Italian](#)

²¹ [AILC – Associazione Italiana di Linguistica Computazionale \(ai-lc.it\)](#); [Aixia |](#) ; [Benvenuti sul sito AISV](#)

²² <https://github.com/marcopoli/AIBERTO-it>

As reference for the analysis on newspapers websites, we consider the paper *“You Don't Understand, This is a New War!” Analysis of Hate Speech in News Web Sites' Comments* by Karmen Erjavec a & Melita Poler Kovačič. Even though, their analysis is focused on the comments published on newspaper websites, their work gives insights about the perception people get from the articles and how it affects their thinking. As Erjavec and Kovačič affirm, their work *“tries to contribute to uncovering the characteristics of Internet hate speech by combining discourse analyses of comments on Slovenian news websites with online in-depth interviews with producers of hate speech comments, researching their values, beliefs, and motives for production. Producers of hate speech use different strategies, mostly rearticulating the meaning of news items”*.

3.3 Dataset Creation and Description

To create the dataset, 60 articles were collected from several sources. The considered newspapers are *Il Primato Nazionale, Il Gazzettino, Libero, Il Corriere della Sera, La Repubblica, Il Messaggero, La Gazzetta del Sud, Il Quotidiano Net, Il Giornale, Il Foglio, La Stampa, Manifesto* as well as other local smaller papers (such as *Il Piacenza Sera, Il Corriere del Veneto*).

In selecting the articles, the guideline was to be as exhaustive as possible. Therefore, the considered newspapers are from different political ideology, and they include both local and national news. They also vary in length and cover different aspects related to immigrants and minorities in Italy such as the narration of current events, news related to legislation and politics as well as statistics on the perceptions and feelings of the Italian society on the matter.

The articles were collected considering a period between the year 2020 and 2022.

The extraction of the text from the websites was made using the Python library *Newspaper3k* . After selecting the desired language, the library allows the download

of the article's text. It also recognizes the date, title, authors, and additional elements like, for example, images or graphs as well as publicity.

For most of the cases, the download of the texts worked well, and the system was able to extract only the relevant text from the website page along with its title. However, it often happened that a piece of additional information was misplaced: for example, when downloading articles from *Il Primato Nazionale*, *Il Giornale* and *Il Quotidiano Net*, the authors were not correctly identified, and were included in the text; or for the *Libero Quotidiano* articles, the date of publication resulted null.

The worst text extraction resulted from *Il Gazzettino* and *Libero Quotidiano*, probably due to the high presence of extra content on the website's pages. In particular, the download included text from the links to other articles of related topics which are placed between paragraphs and quote the name of the newspaper.

Conversely, the download from more official websites like *Repubblica* and *Il Corriere* worked very efficiently.

Overall, the performance of the *Newspaper3k* library on the Italian language can be considered positive. In our specific case, the problems related to authors and publication date could be overlooked, as they are of no relevance to the scope of the analysis. However, the problematic linked to the non-pertinent text had to be resolved manually. That was not computationally expensive considered the limited dimension of the selected dataset.

Annotation

Once the articles were gathered, the subsequent necessary step was to annotate them. The implemented technique was manual annotation. Two classes were identified: *Hateful* and *Non-Hateful*, where the former indicates an article, whose sentiment is not hostile towards immigrants and minorities or whose author is promoting positive behaviour towards them. While the latter refers to articles whose sentiment appears as hostile towards immigrants and minorities, narrating extremely aggressive events or using hateful words, such as unkind nicknames and similes.

The resulting dataset was made up of 60 observations and 5 variables: the id, the link, the title, the text, and the tag of the article.

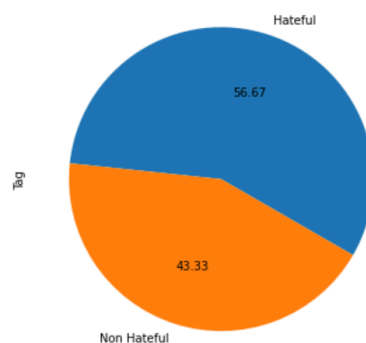
The title was maintained because it often mirrors the content and message of the article and sometimes it can be rather expressive - e.g., “*Immigrazione? No grazie! Un sondaggio rivela cosa pensano veramente gli italiani*” (tr. “Immigration? No thanks! A survey reveals what Italians really think”)²³; “*Quarta Repubblica, Toni Capuozzo: "Stupri, secondo voi è un caso?". Islam, la più scomoda delle verità*” (tr. “Fourth Republic, Toni Capuozzo: "Rape, do you think it is a coincidence?". Islam, the most inconvenient of truths”)²⁴; “*Paura dei migranti: metà degli europei è a favore dei muri*” (tr. “Fear of migrants: half of Europeans are in favor of walls”)²⁵.

Fig 10. Extract of the dataset

ID	Testo	Link	Titolo	Tag
0 1	Roma, 17 gen – Dopo le violenze perpetrate da ...	https://www.ilprimatonazionale.it/cronaca/alla...	La propensione al crimine degli immigrati è 4,...	Hateful
1 2	Roma, 12 ago – Con alcuni semplici grafici, a ...	https://www.ilprimatonazionale.it/politica/fer...	Fermare l'immigrazione clandestina è possibile...	Non Hateful
2 3	Roma, 21 lug. – Nonostante la martellante prop...	https://www.ilprimatonazionale.it/cultura/immi...	Immigrazione? No grazie! Un sondaggio rivela c...	Hateful
3 4	Roma, 16 mar – “Sul fronte dell’immigrazione e...	https://www.ilprimatonazionale.it/approfondime...	“Immigrazione? Ecco come tornare sovrani nel r...	Non Hateful
4 5	Roma, 1 dic – Per meglio comprendere l’increme...	https://www.ilprimatonazionale.it/inferno-spa/...	Immigrati: aumento degli sbarchi e zero ricoll...	Non Hateful

To be truthful to the goal of the analysis, the dataset is well balanced between the two categories: the “Hateful” observations are 34, whereas the “Non-Hateful” ones are 26.

Fig 11. Target variable distribution



²³ <https://www.ilprimatonazionale.it/cultura/immigrazione-sondaggio-italiani-202067/>

²⁴

<https://www.liberoquotidiano.it/news/terra-promessa/30051845/quarta-repubblica-toni-capuozzo-stupri-gruppo-parola-arabo-islam-milano-capodanno.html>

²⁵ https://www.repubblica.it/esteri/2021/12/23/news/paura_dei_migranti_meta_degli_europei_e_a_favore_dei_muri-331267289/

3.4 Pre-processing analysis

Before implementing the models, it is useful to perform an explorative analysis of the text to better understand the characteristics of the dataset.

When dealing with text, the mentioned analysis is carried out through the so-called Natural Language Processing (NLP). NLP is referred to as the large subfield of machine learning that deals with data in the form of natural language, both in terms of written text and speech. NLP allows the computer to read, understand and derive meaning from the human languages. It is one of the most actively researched fields within machine learning at the moment.

NLP works by transforming text into numerical data so that the machine learning models can recognize it. There are several techniques to do so, such as one-hot-encoding, bag of words, TF-IDF of all the words in the data. An alternative approach that is usually better is to use word embeddings. This means that the words in a text are represented as vectors in a large vector space. Word embeddings models have been trained on very large word corpuses to create the vector space such that the words that share common contexts in the training corpus will be close to each other in the vector space.



“Queen” + “Man” - “Woman” = “King”. (Image by author)

Source: *NLP: Gaining insights from text reviews* | by Fredrik Olsson | *Towards Data Science*

3.4.1 Data Cleaning

As for numerical analyses, the first step is to clean the data. The tool used in this case is the library NLTK or Natural Language Toolkit, a leading platform for building Python programs with human language data.

Removing punctuation and stop words

The cleaning of the data consists in homologating the text and removing the non-significant components. Therefore, all the text is set to lowercase, and the punctuation is removed.

In addition to the punctuation, the so-called “stop words” need to be removed. Stop words are the most common words that many search engines avoid, for the purposes of saving space and time in processing large data because they do not contain any meaning, strictly speaking. Among stop words, conjunctions, articles, pronouns are included.

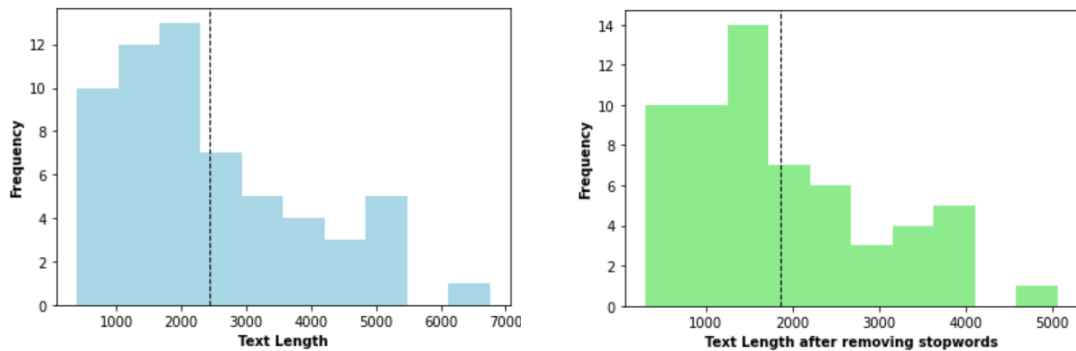
There is no single universal list of stop words used by all natural language processing tools, nor any agreed upon rules for identifying stop words, and indeed not all tools even use such a list. Therefore, any group of words can be chosen as the stop words for a given purpose.

For each language in NLTK there is a default list of stop words that can be modified. Figure 12 shows an extract of the list for the Italian language.

Fig 12. Extract of default stop words in Italian

```
{'fai', 'finche', 'subito', 'cento', 'ha', 'glielle', 'oggi', 'siete', 'sareste', 'spesso', 'giacche', 'facevate', 'facesse',  
'generale', 'negli', 'starò', 'quasi', 'quest', 'nell', 'averlo', 'alcuno', 'stemmo', 'quale', 'dai', 'sembri', 'avrà', 'poic  
he', 'non', 'avevi', 'avrò', 'glieli', 'dentro', 'infatti', 'successivamente', 'sull', 'allora', 'dove', 'ma', 'ognuna', 'de  
gl', 'seguito', 'avemmo', 'anticipo', 'hai', 'vale', 'quindi', 'basta', 'avrei', 'ho', 'modo', 'vita', 'fosse', 'nostri', 's  
ui', 'avessi', 'sarete', 'qui', 'va', 'ansa', 'la', 'preferibilmente', 'nuovo', 'del', 'momento', 'siamo', 'stai', 'sue', 't  
orino', 'fine', 'anno', 'glielo', 'starà', 'mesi', 'relativo', 'fui', 'cos', 'riecco', 'ti', 'gia', 'sarò', 'dagl', 'foste',  
'sto', 'certo', 'soltanto', 'insieme', 'come', 'che', 'sulla', 'se', 'stato', 'qualcosa', 'siano', 'più', 'dappertutto', 'tr  
a', 'milioni', 'adesso', 'avevamo', 'diventare', 'stati', 'avessimo', 'diventa', 'realmente', 'fossimo', 'così', 'piu', 'qua  
li', 'stessi', 'chiunque', 'miei', 'stavate', 'dagli', 'fummo', 'diventato', 'gruppo', 'a', 'abbastanza', 'su', 'stata', 'ha
```

It is interesting to see how much the length of the text changes after removing stop words, also to get an idea of their incidence over the whole dataset. In our case, the average number of stop words per article is equal to 46.



Lemmatization

Lemmatization is a crucial step in the data cleaning process. Often words appear in several inflected forms, to facilitate the analysis, the process of stemming and lemmatization are introduced. The goal of both stemming and lemmatization is to reduce inflectional forms and sometimes derivationally related forms of a word to a common base form. *Stemming* usually refers to a crude heuristic process that eliminates the ends of words in the hope of achieving this goal correctly most of the time, and often includes the removal of derivational affixes. *Lemmatization* is the algorithmic process of determining the lemma of a word based on its intended meaning. Unlike stemming, lemmatisation depends on correctly identifying the intended part of speech and meaning of a word in a sentence, as well as within the larger context surrounding that sentence, such as neighbouring sentences or even an entire document.

In our dataset we tried to apply both the techniques. As expected, the lemmatization produced a better performing result, whereas the stemming techniques had some confusing results. This is probably also due to the complexity of the Italian language.

Fig 13. Example of text after lemmatization

'roma 17 gen dopo violenza perpetrare banda immigrato cosiddetto italiano secondo generazione milano durare notte capodanno analizzare dato dell'istat riguardare arresto denuncia 2020 quadrare uscire fortemente allarmare nonostante straniero censito italia 5 milione 1845 popolazione residente commettere 30 cento delitto propensione crimine 47 volto superiore rispettare it aliano 41 cento violenza sessuale commettere immigrato analizzare specificare delitto evidenziare forte pericolosità sociale straniero esempio 2020 immigrato stato arrestatidenunciati 4059 cento caso omicidio preterintenzionale 3622 cento caso seque strare persona 4109 cento caso violenza sessuale ben 6753 cento caso sfruttamento favoreggiamento prostituzione straniero st ato arrestatidenunciati 4248 cento furto 4252 cento rapina 4024 cento ricettazione 3211 cento danneggiamento 3695 cento reato riguardare trafficare stupefare 2555 cento crimine connettere all'associazione delinquere propensione crimine straniero tipo delitto propensione determinato tipo delitto straniero allarmare immigrato propensione sfruttamento favoreggiamento prostituzione 225 volto superiore italiano omicidio preterintenzionale furto rapina 8 volto superiore violenza sessuale 76 volto superiore nazionalità commettere reato cittadino nove nazionalità venire arrestatidenunciati 53 cento crimine totale commette

3.4.2 Explorative Analysis

Given the goal of the analysis, it is intriguing to explore the typology of the words in the article and how they are used.

Firstly, we check the most common words in the whole dataset, afterwards we perform the same check by dividing the data according to their tag, so either “hateful” or “non-hateful”. By comparing them, we aim at seeing whether the most common words change, thus reflecting the sentiment of the text.

Fig 14. Top 40 more used words in the dataset

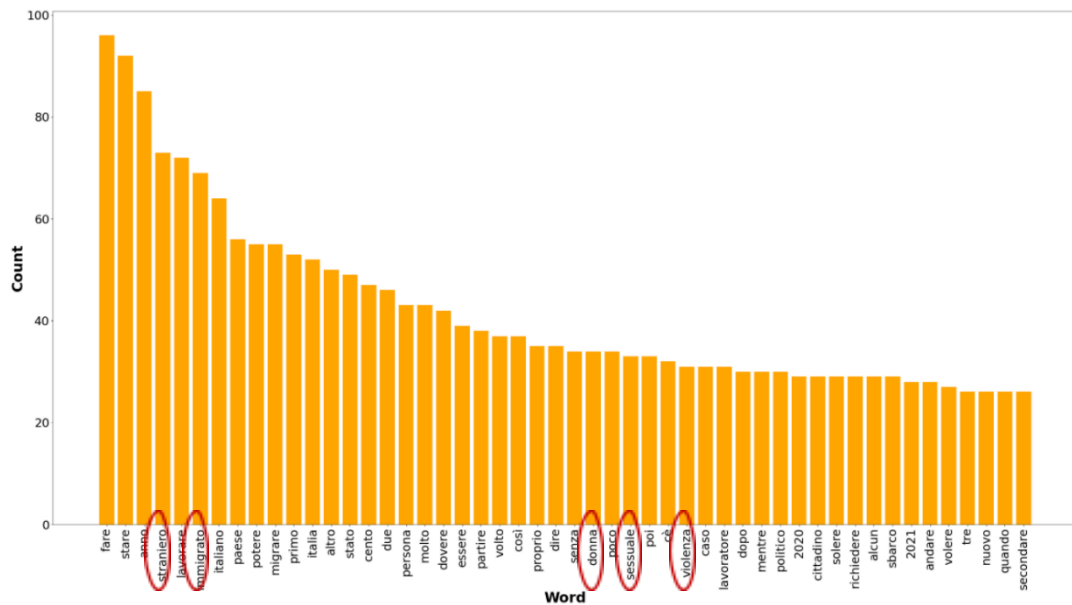


Fig 15. Top 40 more used words in the non-Hateful dataset

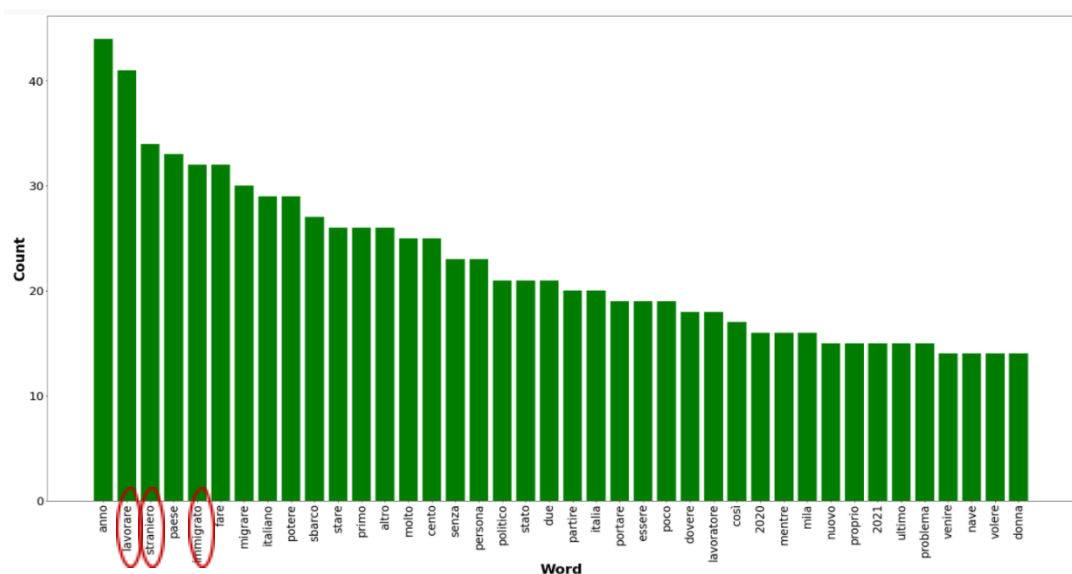
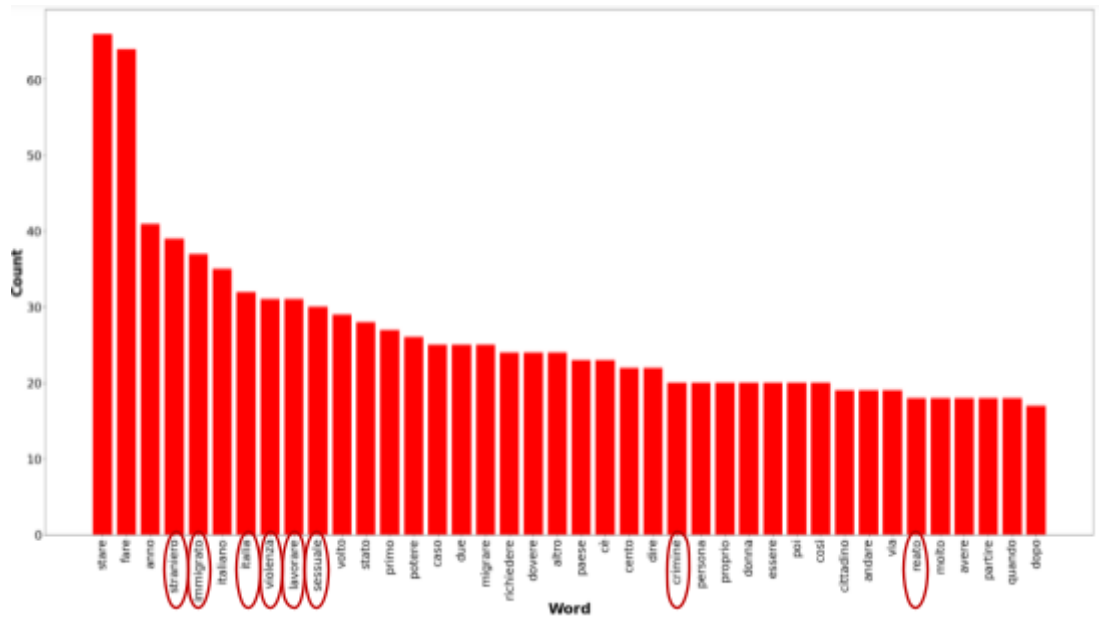


Fig 16. Top 40 most used words in the Hateful dataset



Simply by glancing at the graphs, it can be stated that there is some noticeable difference in the words and/or in their order of appearance.

Overall, the most frequent words are the ones related to the topic at hand, that can be considered “neutral”. Examples of these words are *immigrato* (immigrate), *migrare* (migrate), *italiano* (Italian), *Italia* (Italy), *paese* (country), *lavorare/lavoratore* (work), *sbarco* (landing). The most common term is *straniero* (foreigner), which is not a malicious adjective, but its high popularity in the articles suggests that there may be a tendency to underline the extraneity of the migrants, thus their being different from us.

The biggest difference in the usage of the words can be seen between the graphs representing the two groups. As expected, the most common words of the non-hateful dataset are associated with a more positive view of the migrants. In general, we find more verbs indicating action and feelings such as *lavorare* (work), *fare* (to do), *stare* (to stay), *potere* (can), *partire* (to leave), *portare* (to bring), *venire* (to come) and *volere* (to want). Furthermore, there are no words linked to negative

situations except for *problema* (problem), which however is among the last ones, and it is probably used when raising issues in defence of the immigrants.

It is also interesting to see how in this case the word *Italia* (Italy) is less used than the word *paese* (country), suggesting a narrative that is less Italy-centred.

The graph from the “Hateful” dataset offers a quite different scenario. Among the most recurrent words we find, in sequence, *straniero* (foreign), *immigrato* (immigrate), *violenza* (violence), *lavorare* (to work), *sessuale* (sexual) and the words *crimine* (crime) and *reato* (offense/felony) appear, whereas they were not included in the two previous graphs. Moreover, even the verbs are slightly different, we find: *richiedere* (request), *dovere* (duty/must), *andare* (to go), *avere* (to have). These verbs arouse an idea of obligations versus needs, probably related to the management of the migrants by the Italian government and its duty to fulfil the Italian demands.

Also, in this case, the word *Italia* (Italy) is among the first ones, suggesting that it is the focus of this type of narration.

It is interesting to see how the word *volto* (face) is widely used. Given that it does not appear in the “non-hateful” dataset, it may be linked to episodes of violence, or it could be used in its metaphoric meaning, that is the “face” of a nation/population or of a situation.

Lastly, a term worth mentioning is *donna* (woman) which is included in all three top lists and reaches the highest position in the general dataset. It can be noticed that, conversely, the word *uomo* (man) is not present. The dominance of the word *donna* could be related to several aspects. To cite but a few: in several statistics it is shown that most of the immigrants in Italy are women; foreign women are contributing to increasing the birth rate. But more often women are the victims of stories of violence and crime and more than often the ones that are narrated in papers are the action of immigrants.

The same kind of analysis can be performed on the titles of the articles. In newspapers’ articles the titles play an important role in captivating the reader’s

attention, therefore, the choice of the words is peculiar and tends to sum up the concept expressed in the article. However, it often happens that being that the title is formulated to catch the reader's attention, it turns out to be deceiving.

In our case, it is interesting to see what kind of words are chosen and if there is any glaring difference between the titles of the "hateful" articles and those of the "non-hateful" ones.

The top 15 most common words in the titles are shown in the graphs below.

Fig 17. Most common words in the titles (total dataset)

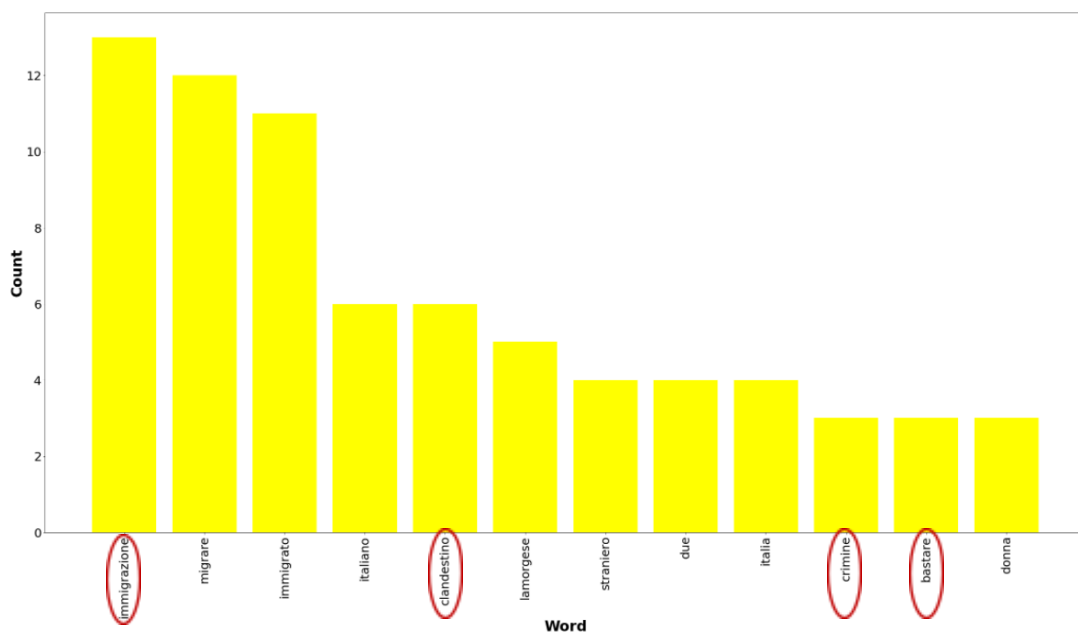


Fig 18. Most common words in the titles (Non-Hateful dataset)

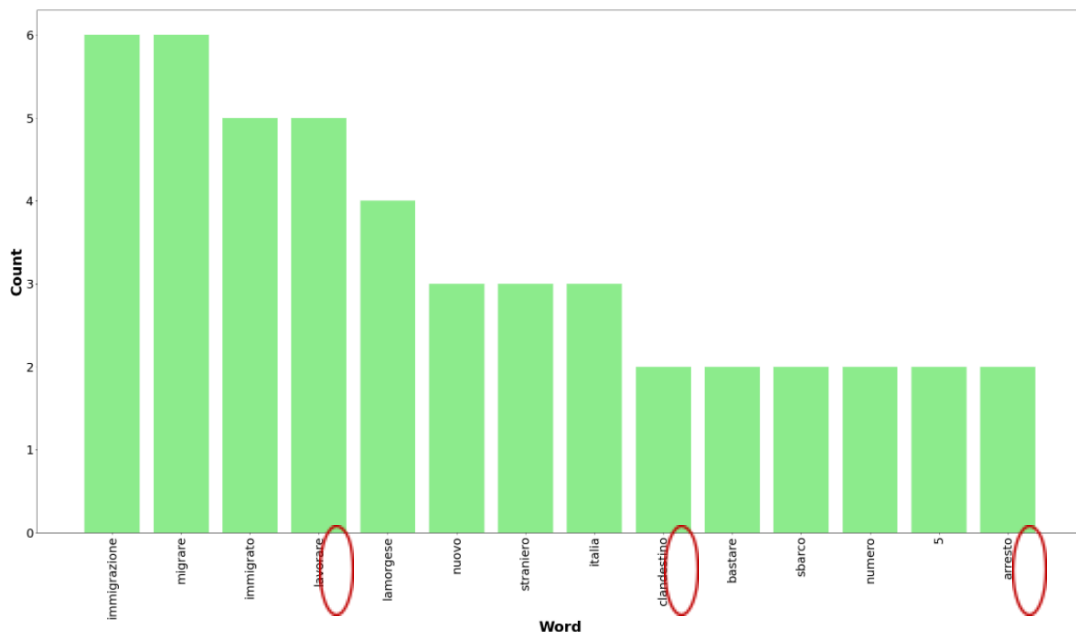
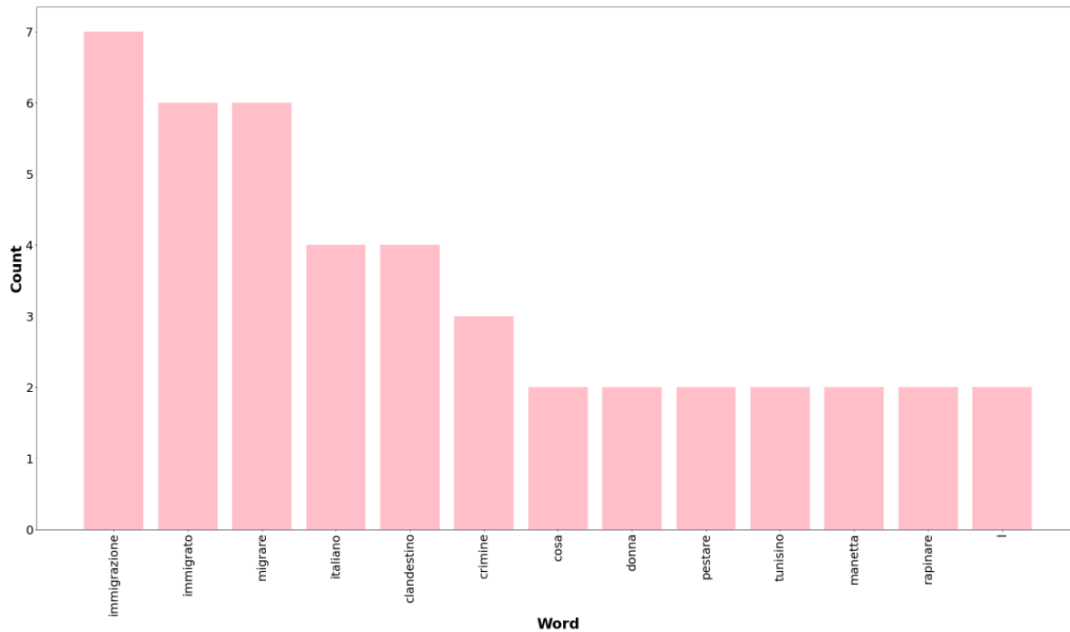


Fig 19. Most common words in the titles (Hateful dataset)



All three graphs show that the most common words are the ones related to the concept of immigration, i.e., *immigrazione* (immigration), *immigrato* (immigrate), *migrare* (to migrate). As for the rest, also in the titles, we find choices of words that could be reflecting the sentiment of the writer. Indeed, the “hateful” titles have a recurrency of words like *clandestino* (clandestine), *crimine* (crime), *pestare* (to beat someone up), *rapinare* (to rob someone/someplace), whereas the “non-hateful” titles often include words such as *lavorare* (to work), *nuovo* (new), *arresto* (arrest) and Lamorgese which is the name of the Italian Minister of Internal Affairs. Therefore, it seems that the non-hateful articles are more focused on practical matters related to the immigrants’ conditions in Italy. In fact, among the common words, we find *bastare*, the naturalized form of the term *basta* which means “enough”, and it is used to express the need of stopping a certain thing/action/behaviour.

It also must be noticed how in the non-hateful titles we find both the words *straniero* and *clandestino*, with the latter being more common, whereas in the hateful titles we only find the word *clandestino*, in a much higher position than in the other dataset. The word *clandestino*, as we mentioned earlier, expresses a more hostile feeling. Moreover, it is interesting to notice how the word *tunisino* is widely used in the hateful titles.

Lastly, it is worth mentioning that in the titles of the hateful articles there is a large use of the word *donna*, which supports the theory mentioned before.

3.4.3 NLP: A deeper analysis

Aside from the cleaning steps, the NLP allows to understand texts in a more specific way, recognizing the structure of the sentences, the dependency between words and the type of entity that a word represents. To show the functioning of these methods, an example of their application to the first article in the dataset is shown below. The analysis is done on the raw dataset (pre-cleaning) because it is necessary that the original structure is maintained.

Part of Speech Tagging

In corpus linguistics, part-of-speech tagging (also called POS tagging or POST), is the process of marking up a word in a text as corresponding to a particular part of speech, based on both its definition and its context—i.e., its relationship with adjacent and related words in a phrase, sentence, or paragraph. The tool identifies the words as nouns, adjectives, verbs, adverbs, etc.

Fig 20. Exhibit of POST application

Il	DET	RD__Definite=Def Gender=Masc Number=Sing Prontype=Art	None
quadro	NOUN	S__Gender=Masc Number=Sing	None
che	PRON	PR__Prontype=Rel	None
ne	PRON	PC__Clitic=Yes Prontype=Prs	None
è	AUX	VA__Mood=Ind Number=Sing Person=3 Tense=Pres VerbForm=Fin	None
uscito	VERB	V__Gender=Masc Number=Sing Tense=Past VerbForm=Part	None
è	VERB	V__Mood=Ind Number=Sing Person=3 Tense=Pres VerbForm=Fin	None
fortemente	ADV	B	None
allarmante	X	A__Number=Sing	None
:	PUNCT	FC	None
nonostante	ADP	E	None
gli	DET	RD__Definite=Def Gender=Masc Number=Plur Prontype=Art	None
stranieri	NOUN	S__Gender=Masc Number=Plur	None
censiti	VERB	V__Gender=Masc Number=Plur Tense=Past VerbForm=Part	None
in	ADP	E	None
Italia	PROPN	SP	space
siano	VERB	V__Mood=Sub Number=Plur Person=3 Tense=Pres VerbForm=Fin	None
5	NUM	N__NumType=Card	None
milioni	NOUN	S__Gender=Masc Number=Plur	None
,	PUNCT	FF	None
1'	DET	RD__Definite=Def Gender=Masc Number=Sing Prontype=Art	None
8,45	NUM	N__NumType=Card	None
%	SYM	SYM	symbol
della	ADP	E_RD__Definite=Def Gender=Fem Number=Sing Prontype=Art	None
popolazione	NOUN	S__Gender=Fem Number=Sing	None
residente	X	A__Number=Sing	None
,	PUNCT	FF	None
questi	PRON	PD__Gender=Masc Number=Plur Prontype=Dem	None
commettono	AUX	V__Mood=Ind Number=Plur Person=3 Tense=Pres VerbForm=Fin	None
il	DET	RD__Definite=Def Gender=Masc Number=Sing Prontype=Art	None
30	NUM	N__NumType=Card	None
per	ADP	E	None
cento	NUM	N__NumType=Card	None
dei	DET	E_RD__Definite=Def Gender=Masc Number=Plur Prontype=Art	None
delitti	NOUN	S__Gender=Masc Number=Plur	None
,	PUNCT	FF	None
con	ADP	E	None

una	DET	RI__Definite=Ind Gender=Fem Number=Sing PronType=Art	None
propensione	NOUN	S__Gender=Fem Number=Sing	None
al	DET	E_RD__Definite=Def Gender=Masc Number=Sing PronType=Art	None
crimine	NOUN	S__Gender=Masc Number=Sing	None
4,7	NUM	N__NumType=Card	None
volte	NOUN	S__Gender=Fem Number=Plur	None
superiore	ADJ	A__Degree=Cmp Number=Sing	None
rispetto	ADP	E	None
a	ADP	E	None
quella	PRON	PD__Gender=Fem Number=Sing PronType=Dem	None
degli	DET	E_RD__Definite=Def Gender=Masc Number=Plur PronType=Art	None
italiani	NOUN	S__Gender=Masc Number=Plur	None
.	PUNCT	FS	None

Dependency Parsing

Dependency parsing is the process of analysing the grammatical structure in a sentence and find the related words and the type of relationship between them.

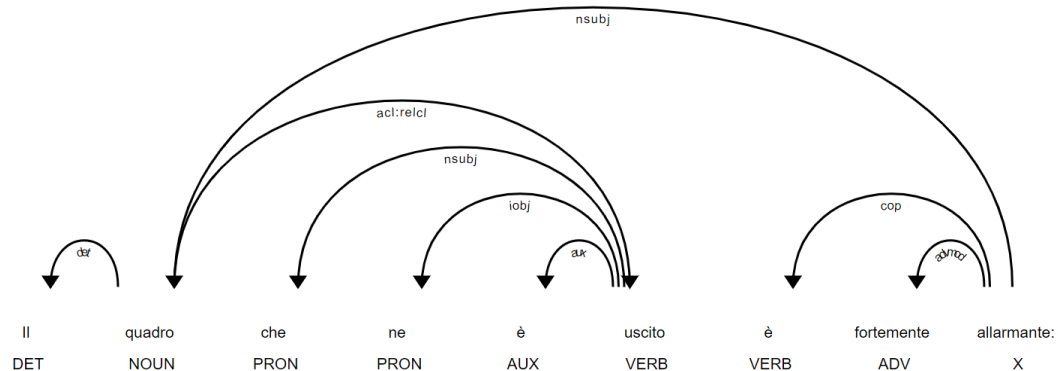
Each relationship has one head and a dependent that modifies the head; is labelled²⁶ according to the nature of the dependency between the head and the dependent.

Fig 21. Exhibit of Dependency Parsing

Il	DET	det	determiner
quadro	NOUN	nsbj	nominal subject
che	PRON	nsbj	nominal subject
ne	PRON	iobj	indirect object
è	AUX	aux	auxiliary
uscito	VERB	acl:relcl	None
è	VERB	cop	copula
fortemente	ADV	advmod	adverbial modifier
allarmante	X	ROOT	None
:	PUNCT	punct	punctuation
nonostante	ADP	case	case marking
gli	DET	det	determiner
stranieri	NOUN	obl	oblique nominal
censiti	VERB	acl	clausal modifier of noun (adjectival clause)
in	ADP	case	case marking
Italia	PROPN	obl	oblique nominal
siano	VERB	cop	copula
5	NUM	nummod	numeric modifier
milioni	NOUN	ROOT	None
,	PUNCT	punct	punctuation
1'	DET	det	determiner
8,45	NUM	nummod	numeric modifier
%	SYM	conj	conjunct
della	ADP	case	case marking
popolazione	NOUN	nmod	modifier of nominal
residente	X	amod	adjectival modifier
,	PUNCT	punct	punctuation
questi	PRON	nsbj	nominal subject
commettono	AUX	conj	conjunct
il	DET	det	determiner
30	NUM	obj	object
per	ADP	case	case marking
cento	NUM	nummod	numeric modifier
dei	DET	case	case marking
delitti	NOUN	nmod	modifier of nominal
,	PUNCT	punct	punctuation
con	ADP	case	case marking
una	DET	det	determiner
propensione	NOUN	obl	oblique nominal
al	DET	case	case marking
crimine	NOUN	nmod	modifier of nominal
4,7	NUM	nummod	numeric modifier
volte	NOUN	obl	oblique nominal
superiore	ADJ	amod	adjectival modifier
rispetto	ADP	case	case marking
a	ADP	fixed	fixed multiword expression
quella	PRON	obl	oblique nominal
degli	DET	case	case marking
italiani	NOUN	nmod	modifier of nominal
.	PUNCT	punct	punctuation

²⁶ The labels can be found at [Universal Dependency Relations \(universaldependencies.org\)](http://universaldependencies.org)

Fig 22. Visualization of Dependency Parsing



Named Entity Recognition

Named entity recognition (NER) is “a process where a sentence or a chunk of text is parsed through to find entities that can be put under categories like names, organizations, locations, quantities, monetary values, percentages, etc. [...] NER is a simple but effective approach to reduce searching a state space by directing the algorithm to weigh the sentences more if a chunk of entities are found.”

Fig 23. Exhibit of NER

La moglie del capo del **Dipartimento LOC** per le libertà civili e immigrazione del **Viminale LOC**, **Michele di Bari PER**, è tra le 16 persone indagate in un'inchiesta per caporalato dei **Carabinieri ORG** e della procura di **Foggia LOC** che ha portato all'arresto di cinque persone, due delle quali in carcere.

Dipartimento - LOC - Non-GPE locations, mountain ranges, bodies of water
 Viminale - LOC - Non-GPE locations, mountain ranges, bodies of water
 Michele di Bari - PER - Named person or family.
 Carabinieri - ORG - Companies, agencies, institutions, etc.
 Foggia - LOC - Non-GPE locations, mountain ranges, bodies of water

More in depth, the tool offers different annotations among which those that can tell the position at which the entity is in the text, its hash value (the numerical value associated to the word in Python) and the string description. An example follows below.

Fig 24. Exhibit of entity's annotations

Dipartimento 5 6 23 35 LOC
Viminale 13 14 77 85 LOC
Michele di Bari 15 18 87 102 PER
Carabinieri 31 32 168 179 ORG
Foggia 36 37 199 205 LOC

Overall, the NLP exploratory tools work efficiently for the Italian language. However, it can be affirmed that the cleaning steps works more poorly. While tools like POS, Dependency parsing, and NER seem to correctly recognize the entities and their role in a sentence, with the performed analysis we found that the operation of removing the punctuation does not produce a complete clean output and often makes mistakes when considering the apostrophe. Moreover, even if the stop words removing process works correctly, the list of default stop words needs to be updated to avoid the need of manually adding terms to it and allow more efficiency.

Lastly, the techniques of Stemming and Lemmatization are the ones with the lowest performance. The stemming operation seems to work poorly with the Italian language and to produce confusing results, whereas the lemmatization is more precise, therefore is the one chosen in this analysis.

The following images shows examples of stemming and lemmatization applied to the first article of the dataset. It is noticeable how in Figure 25 the stemming does not correctly truncate the desinences of all the words.

Fig 25. Exhibit of text after stemming

"roma , 17 gen - dopo le violenz perpetr da band di immigrati e dai cosiddetti " italiani di seconda generazione " a milano du rant la nott di capodanno , abbiamo analizzato i dati dell ' istat riguardanti gli arresti e le denunc del 2020 . il quadro c he ne è uscito è fortemet allarmant : nonostant gli stranieri censiti in italia siano 5 milioni , l ' 8,45 % della popolazio n resident , questi commettono il 30 per cento dei delitti , con una propension al crimin 4,7 volt superior rispetto a quella degli italiani . il 41 per cento dell violenz sessuali sono commess da immigrati analizzando nello specifico i delitti , si e videnzia la fort pericolosità social degli stranieri . ad esempio , nel 2020 , gli immigrati sono stati arrestati/denunciati nel 40,59 per cento dei casi di omicidio preterintenzional , nel 36,22 per cento dei casi di sequestro di person , nel 41,09 per cento dei casi di violenza sexual , e ben nel 67,53 per cento dei casi di sfruttamento e favoreggiamento della prostituz ion . gli stranieri sono stati arrestati/denunciati per il 42,48 per cento dei furti , per il 42,52 per cento dell rapin , pe r il 40,24 per cento dell ricettazioni , per il 32,11 per cento dei danneggiamenti , per il 36,95 per cento dei reati riguard anti il traffico di stupefacenti , e per il 25,55 per cento dei crimini connessi all ' associazioni per delinquer . la propen

Fig 26. Exhibit of text after lemmatization

'roma 17 gen dopo violenza perpetrare banda immigrato cosiddetto italiano secondo generazione milano durare notte capodanno analizzare dato dellistat riguardare arresto denuncia 2020 quadrare uscire fortemente allarmare nonostante straniero censito italia 5 milione 1845 popolazione residente commettere 30 cento delitto propensione crimine 47 volto superiora rispettare it aliano 41 cento violenza sessuale commettere immigrato analizzare specificare delitto evidenziare forte pericolosità sociale straniero esempio 2020 immigrato stato arrestatidenunciati 4059 cento caso omicidio preterintenzionale 3622 cento caso seque strare persona 4109 cento caso violenza sessuale ben 6753 cento caso sfruttamento favoreggiamento prostituzione straniero st ato arrestatidenunciati 4248 cento furto 4252 cento rapina 4024 cento ricettazione 3211 cento danneggiamento 3695 cento reat o riguardare trafficare stupefare 2555 cento crimine connettere allassociazioni delinquere propensione crimine straniero tip o delitto propensione determinato tipo delitto straniero allarmare immigrato propensione sfruttamento favoreggiamento prosti tuzione 225 volto superiora italiano omicidio preterintenzionale furto rapina 8 volto superiora violenza sessuale 76 volto s uperiora nazionalità commettere reato cittadino nove nazionalità venire arrestatidenunciati 53 cento crimine totale commette

4 Hate Speech: Implementation of the models

4.1 Models' description

The aim of the analysis is fine-tuning models that are pre-trained for text classification and/or sentiment analysis on the Italian language and then evaluating their performance on the constructed dataset.

As earlier mentioned, unfortunately there are not many models developed for Italian text and just a few of them are available to public downloading and usage.

The most used model is the **BERT Classifier**. BERT, which stands for Bidirectional Encoder Representations from Transformers, is an open-source machine learning framework for natural language processing (NLP). It is designed to help computers with the ambiguity of the language contained in text and it does so by taking into consideration the surrounding text and trying to determine the context. The BERT framework was pre-trained using text from Wikipedia.

BERT is based on Transformers, a deep learning model in which each output element is connected to every input element, and the weightings between them are dynamically calculated based upon their connection. (In NLP, this process is called *attention*.)

More specifically, BERT is based on the so called Bidirectional Recurrent Neural Network.

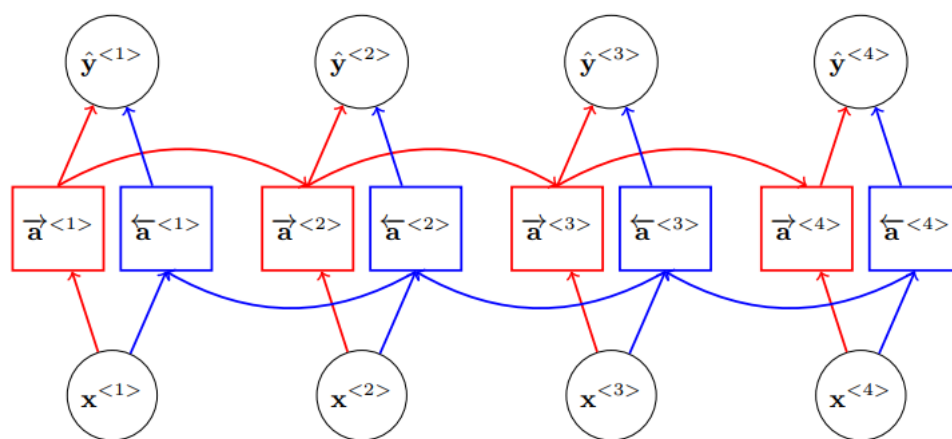
A Recurrent Neural Network (RNN) is based on the concept of neural networks. A neural network is a computational model that mimics the way nerve cells work in the human brain and use learning algorithms that can independently learn as they receive new input. The NN is made up of three or more layers that are interconnected. The first layer consists of input neurons. Those neurons send data on to the deeper layers, which in turn will send the final output data to the last output layer. In text classification, the inputs are words, and the output is the classification of the sentiment of the text.

The RNN, instead, is a class of artificial neural networks where connections between nodes form a directed or undirected graph along a temporal sequence. This allows it to exhibit temporal dynamic behaviour. Derived from feedforward neural networks, RNNs can use their internal state (memory) to process variable length sequences of inputs. This makes them applicable to tasks such as unsegmented, connected handwriting recognition or speech recognition. The key idea is that each unit of the output is a function of the previous members of the input and each unit of the output is produced using the same update rule applied to the previous outputs, also the parameters are shared across different time steps.

Bidirectional RNN takes a step forward and connects two hidden layers of opposite directions to the same output. With this form of generative deep learning, the output layer can get information from past (backwards) and future (forward) states simultaneously. This technique applied to text allows the model to understand the context of the text and make a more precise classification.

The bidirectionality, is introduced by the Transformer encoder which, as opposed to directional models, which read the text input sequentially (left-to-right or right-to-left), reads the entire sequence of words at once.

Fig 27. A Bidirectional Neural Network structure



Employing its bidirectional capability, BERT is pre-trained on two different, but related, NLP tasks: *Masked Language Modelling* and *Next Sentence Prediction*.

The aim of Masked Language Model (MLM) training is to hide a word in a sentence and then make the model predict which term has been hidden (or masked) based on the hidden word's context. The objective of Next Sentence Prediction training is to have the program predict whether two given sentences have a logical, sequential connection or whether their relationship is simply random.

The source of the models is *Hugging Face*. As the TensorFlow blog²⁷ says “*Hugging Face is an NLP-focused start-up with a large open-source community*”.

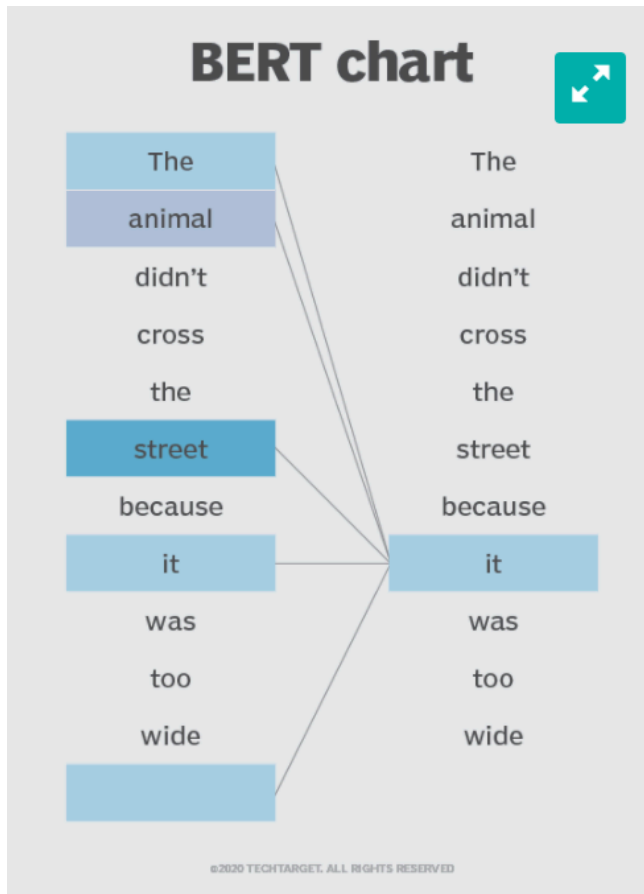
The special tool that raises our interest is the library called *Transformers*. “Transformers is a python-based library that exposes an API (Application Programming Interface) to use many well-known transformer architectures, such as BERT, RoBERTa, GPT-2 or DistilBERT, that obtain state-of-the-art results on a variety of NLP tasks like text classification, information extraction, question answering, and text generation. Those architectures come pre-trained with several sets of weights.” It is easy to download and implement and it is very flexible and adaptable to several systems. The library has seen super-fast growth in PyTorch and has recently been ported to TensorFlow 2.0, offering an API that now works with Keras’ fit API, TensorFlow Extended, and TPUs .

Transformers relies on the idea of pre-trained transformers models. These models vary in several aspects: they have different sizes, shapes, and architectures. Moreover, each of them has its precise way of receiving the input data.

Transformers was first introduced by Google in 2017. It is considered an improvement with respect to RNN because while RNN requires that the sequences of data are processed in a precise order, Transformers can process data in any order, and they enable training on larger amounts of data than ever was possible before their existence. This facilitated the creation of pre-trained models like BERT, which was trained on massive amounts of language data prior to its release.

²⁷ [Hugging Face: State-of-the-Art Natural Language Processing in ten lines of TensorFlow 2.0 — The TensorFlow Blog](#)

Fig 28. Example of BERT's understanding



Source: [What is BERT \(Language Model\) and How Does It Work? \(techtarget.com\)](https://techtarget.com/what-is-bert-language-model-and-how-does-it-work/)

As we can see from the image, BERT can understand the subject of the text and the context related to it.

How BERT works

The goal of NLP is to understand human language which is by definition ambiguous. BERT typically does so by predicting a word in a blank. It was pre-trained using only unlabelled, plain text corpus (from Wikipedia). Thus, it continues to learn unsupervised from unlabelled text and improve as it is being used in practical applications. The pre-training serves as a base layer to build from. BERT can adapt and be fine-tuned to the users' specifications.

The library Transformers allows this process to be intuitive and systematic.

The library builds on three main classes: a configuration class, a tokenizer class, and a model class.

- *The configuration class*: the configuration class, as the word says, is the one that stores all the necessary information related to the format of the model. It includes details about the number of layers, the hidden layers, the number of attention heads. An example of a BERT configuration file is shown below.

Fig 29. Example of a BERT configuration file

```
{
  "attention_probs_dropout_prob": 0.1,
  "hidden_act": "gelu",
  "hidden_dropout_prob": 0.1,
  "hidden_size": 768,
  "initializer_range": 0.02,
  "intermediate_size": 3072,
  "max_position_embeddings": 512,
  "num_attention_heads": 12,
  "num_hidden_layers": 12,
  "type_vocab_size": 2,
  "vocab_size": 28996
}
```

Source: [Hugging Face: State-of-the-Art Natural Language Processing in ten lines of TensorFlow 2.0 — The TensorFlow Blog](#)

- *The tokenizer class*: the tokenizer class is the part of the code responsible for the transformation of the text into a format that the model can understand. It relies on the rules of pre-processing described before, such as removing non-relevant words (the so-called stop words), removing punctuation, and cleaning from other symbols like emoticons, hashtags and urls. Luckily our model is made up of articles and does not include these aspects. Also, the tokenizer, as suggested by the name, divides the text into tokens, or sequences of tokens, and performs the vectorization, converting text into a specific numerical index. Each model has its own tokenizer, being that the tokenization varies according to the model.

- *The model class*: the model class contains the neural network modelling structure itself. When implementing a TensorFlow model, it inherits from *tf.keras.layers.Layer* which means that it can be applied easily by the Keras' fit API. Indeed, Keras is more flexible and user-friendly than TensorFlow.

The implemented models

For this analysis, three pertinent models were found on Hugging Face, all based on the BERT classifiers.

It is worth mentioning that all the following models are thought to be used to understand Twitter/Facebook comments.

Each one of the models has its specific vocabulary with whom the input text is compared and that allows the determination of the sentiment of the sentence.

A brief description of their functioning is displayed below.

AIBERTO-it

AIBERTO²⁸ is the first Italian BERT model for Twitter language understanding. The reference model is *a bert-based lower cased model*. The difference between an uncased and a cased model lies in the tokenization. BERT uncased and BERT cased are different in terms of BERT training using case of text in WordPiece tokenization step and presence of accent markers. In BERT uncased, the text has been lowercased before WordPiece tokenization step while in BERT cased, the text is same as the input text without any changes, moreover in BERT cased the accent markers are kept, which is appropriate for the Italian language that makes a large use of accents and apostrophe. AIBERTO is based on the AIBERT model, developed for the English language.

ALBERT²⁹ is a lighter version of BERT, it is a transformers model pretrained on a large corpus of English data in a self-supervised fashion, that is on raw text only

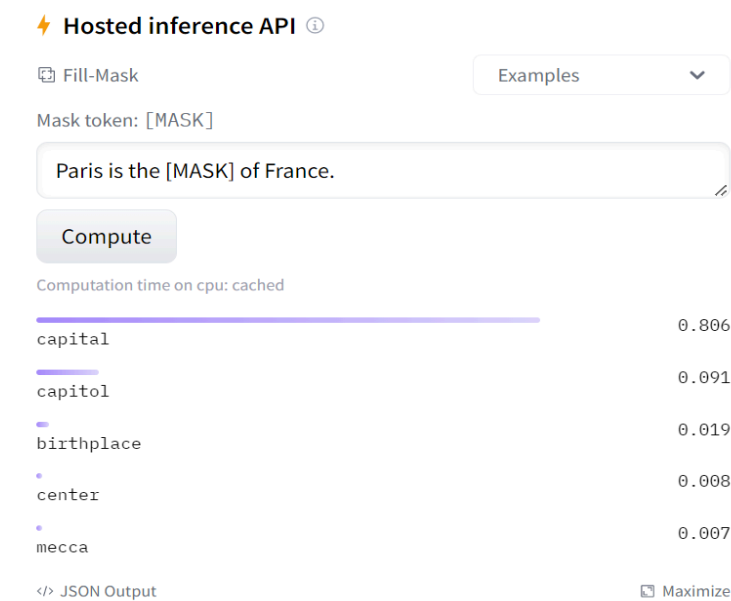
²⁸ [GitHub - marcopoli/AIBERTO-it: AIBERTO the first italian BERT model for Twitter language understanding](#)

²⁹ [albert-xxlarge-v2 · Hugging Face](#)

without any labelling. It was pre-trained with two objectives: Masked language modelling (MLM) and Sentence Ordering Prediction (SOP).

The model learns an inner representation of the English language. This representation can then be used to extract features useful for downstream tasks, like text classification.

Fig 30. Example of MLM with ALBERT model (English)



Source: [albert-xxlarge-v2 · Hugging Face](#)

The pre-trained ALBERT model's structure is the following:

- 12 layers
- 768 hidden
- 12 heads
- 110M parameters

It is implemented as a TensorFlow model and was pre-trained on a sample of tweets collected in 2018-2019.

The model was developed by M. Polignano, a researcher of the Department of Computer Science of the Università “Aldo Moro” of Bari in collaboration with the Department of Computer Science of the University of Turin.

Hate-speech-CNERG/dehatebert-mono-italian

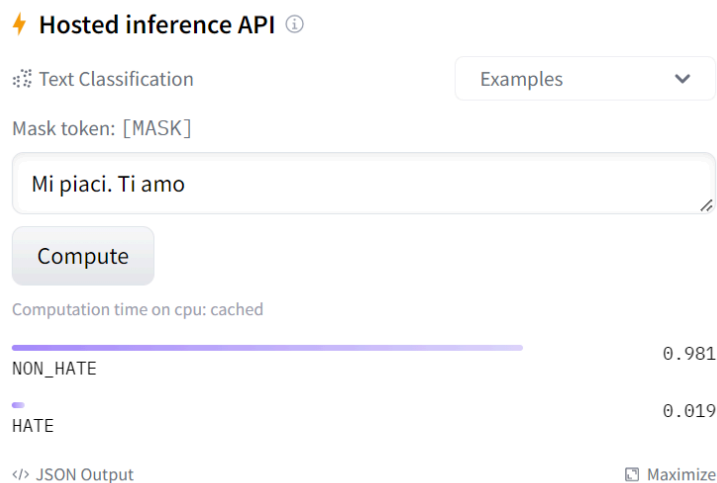
The *dehatebert-mono-italian model*³⁰ was developed by the Hate Speech CNERG, a research group that aims at building “Hate-ALERT (AnaLysis,dEtection and counteRing sysTem for hate speech) with the goal of peace in the online world.”

It is a multilingual model, adapted to be used in different languages (9 languages to be exact), in fact, the mono in the name refers to the monolingual setting, while the model is trained using only English text. It is then finetuned in multilingual languages.

The model is trained with different learning rates and the best validation score achieved is 0.837288 for a learning rate of 3e-5.

As for the parameters, in this case the best parameters are automatically selected through cross-validation. The only ones that is necessary to specify are the language and the directory .

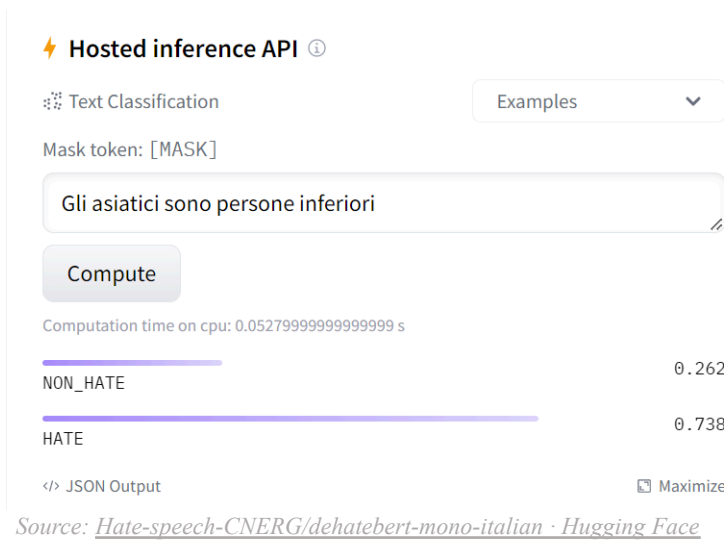
Fig 30.1 Example of Dehatebert-mono-italian classifier in practice



Source: [Hate-speech-CNERG/dehatebert-mono-italian · Hugging Face](#)

³⁰ [Hate-speech-CNERG/dehatebert-mono-italian · Hugging Face](#)

Fig 30.2 Example of Dehatebert-mono-italian classifier in practice



MilaNLPProc/feel-it-italian

The *FEEL-IT* can be fine-tuned to detect either sentiment or emotion.

It was developed by MilaNLP lab of Università Bocconi in Milan. It is a research group dedicated to the study, development, and the deployment of NLP algorithms to solve real-world problems.

This model represents a big upgrade in the Italian NLP field because it is the first one that can tackle both the task of sentiment and emotion detection.

FEEL-IT is a novel benchmark corpus of Italian Twitter posts annotated with four basic emotions: anger, fear, joy, sadness.

Sentiment

The sentiment analysis, instead, is obtained by collapsing the four emotions.

The `feel-it-italian-sentiment` model performs sentiment analysis on Italian, and it is based on the `UmBERTo`³¹ model. The data is collected by annotating a total of 2037 tweets from a broad range of topics.

³¹ [Musixmatch/umberto-commoncrawl-cased-v1 · Hugging Face](#)

Fig 31.1 Example of the FEEL-it-sentiment classifier in practice

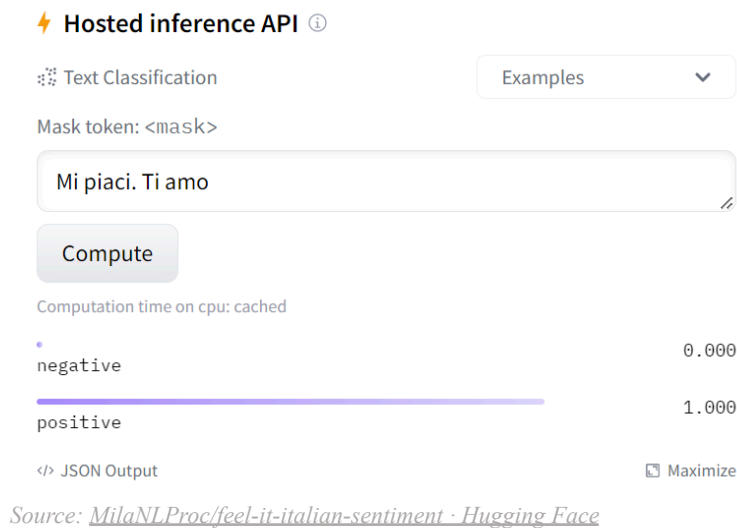
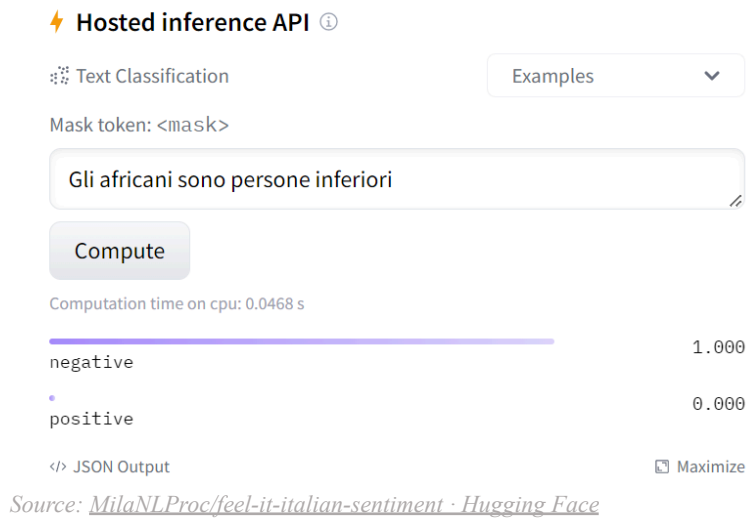


Fig 31.2 Example of the FEEL-it-sentiment classifier in practice



The structure of the model is the following:

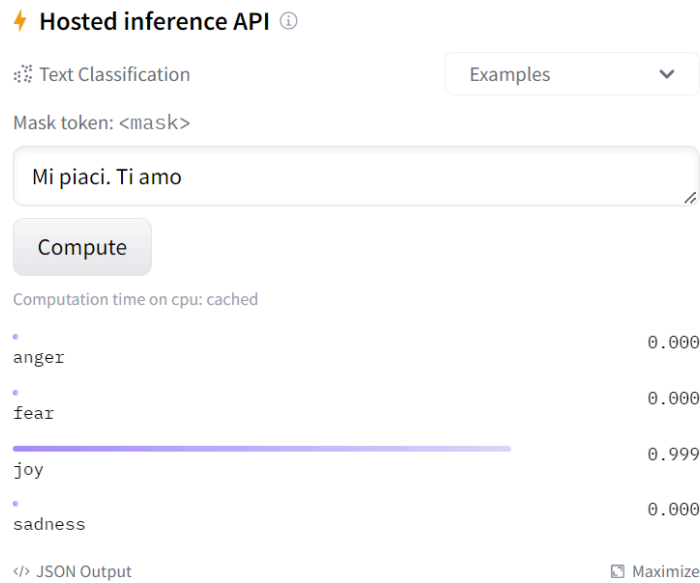
- 12 layers
- 12 attention heads
- 768 hidden
- *Gelu* activation function
- Output labels: positive/negative

- Vocabulary size: 32005

Emotion

The feel-it-italian-emotion model performs emotion classification on Italian text, and it is also based on the UmBERTo³² model. The 2037 tweets are annotated with an emotion label.

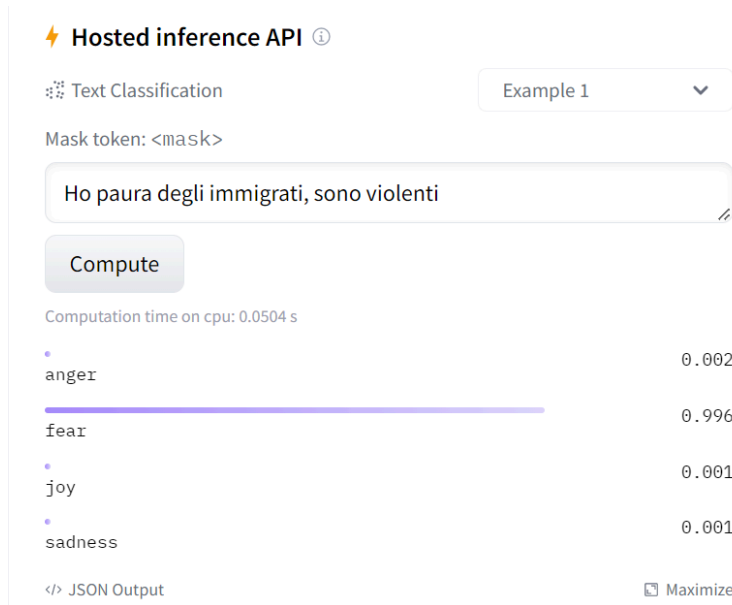
Fig 31.1 Example of the FEEL-it-emotion classifier in practice



Source: [MilaNLProc/feel-it-italian-emotion](#) · Hugging Face

Fig 31.2 Example of the FEEL-it-emotion classifier in practice

³² [Musixmatch/umberto-commoncrawl-cased-v1](#) · Hugging Face



4.2 Results sharing and evaluation

In this section, we share the results obtained from the evaluation of the implemented models and we offer an interpretation.

The dataset was split into train (60%) and test (40%). The former is used to fine tune and fit the model, and the latter is used to make predictions and evaluations.

For evaluating the performance of the models, we consider the so-called *confusion matrix*. The confusion matrix is “it is a performance measurement for machine learning classification problem where output can be two or more classes. It is a table with 4 different combinations of predicted and actual values.”

Fig 33. Confusion Matrix

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Where the True Positive (TP) are the positive values correctly identified; the True Negative (TN) are the negative values correctly identified, whereas the False Positive (FP) and False Negative (FN) are the values wrongly predicted. FP and FN are respectively called “Type 1 error” and “Type 2 error”.

The confusion matrix is a powerful tool because it allows to measure the following metrics: recall, precision, f1-score, accuracy. We will briefly describe them.

- **Recall**

$$\text{Recall} = \frac{TP}{TP + FN}$$

Recall indicates how many positive values were correctly predicted. This value should be as high as possible.

- **Precision**

$$\text{Precision} = \frac{TP}{TP + FP}$$

Precision tells us from all the classes that were predicted as positive, how many were effectively positive. This value also should be as high as possible.

- **F-measure**

$$F - \text{measure} = \frac{2 * \text{Recall} * \text{Precision}}{\text{Recall} + \text{Precision}}$$

The F-score helps to measure recall and precision at the same time. It is useful since it can be difficult to compare models that have high precision and low recall or vice

versa. It is the Harmonic Mean of the two metrics, more efficient than the arithmetic one since it punishes the extreme values more strongly. Consequently, the higher the F-score, the better.

- **Accuracy**

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + FP + TN + FN)}$$

Accuracy is the most used metric. It measures how many values were correctly predicted over the whole sample. Usually, higher values are better. Accuracy is more suitable for balanced datasets (as in our case).

AIBERTO-it (da caricare)

Based on our criteria, the

BIBLIOGRAPHY (in aggiornamento)

Karmen Erjavec & Melita Poler Kovačič (2012) “You Don't Understand, This is a New War!” Analysis of Hate Speech in News Web Sites' Comments, *Mass Communication and Society*, 15:6, 899-920, DOI: [10.1080/15205436.2011.619679](https://doi.org/10.1080/15205436.2011.619679)

Poletto, Fabio, Marco Stranisci, Manuela Sanguinetti, Viviana Patti, and Cristina Bosco. 2017. *Hate speech annotation: Analysis of an Italian twitter corpus*. In *Proceedings of the Fourth Italian Conference on Computational Linguistics (CLiC-it 2017)*, Rome, Italy, December. CEUR.

Sanguinetti, Manuela, Fabio Poletto, Cristina Bosco, Viviana Patti, and Marco Stranisci. 2018. *An Italian Twitter Corpus of Hate Speech against Immigrants*.

Sean MacAvaney, Hao-Ren Yao, Eugene Yang, Katina Russell, Nazli Goharian, Ophir Friede. 2019. *Hate speech detection: Challenges and solutions*

MacAvaney S, Yao H-R, Yang E, Russell K, Goharian N, Frieder O (2019) Hate speech detection: Challenges and solutions. *PLoS ONE* 14 (8): e0221152. <https://doi.org/10.1371/journal.pone.0221152>

Carmela MALTONE, «L’immigrazione nei media italiani. Disinformazione, stereotipi e innovazioni.», *Line@editoriale* [En ligne], N° 003 - 2011, Scritture italiane della migrazione, mis à jour le : 11/05/2017, URL : <https://revues.univ-tlse2.fr/443/pum/lineaeditoriale/index.php?id=314>.

“Barometro dell'odio – Intolleranza pandemica” - Amnesty International Italia, 2021

Arthur T. E. Capozzi, Mirko Lai, Valerio Basile, Fabio Poletto, Manuela Sanguinetti, Cristina Bosco, Viviana Patti, Giancarlo Ruffo, Cataldo Musto, Marco Polignano, Giovanni Semeraro and Marco Stranisci, ““Contro L’Odio”: A Platform for Detecting, Monitoring and Visualizing Hate Speech against Immigrants in Italian

Social Media”, IJCoL [Online], 6-1 | 2020, Online since 01 June 2020, connection on 10 April 2021. URL:<http://journals.openedition.org/ijcol/659> ; DOI: <https://doi.org/10.4000/ijcol.659>

W. Warner and J. Hirschberg, Columbia University, Department of Computer Science, New York, NY 10027, *Detecting Hate Speech on the World Wide Web*. 2012

https://ec.europa.eu/migrant-integration/library-document/ecri-third-report-italy_en
ECRI - Country monitoring in Italy (coe.int)

Catelli, R.; Pelosi, S.; Esposito, M. Lexicon-Based vs. Bert-Based Sentiment Analysis: A Comparative Study in Italian. *Electronics* 2022, 11, 374. <https://doi.org/10.3390/electronics11030374>

overview.pdf (ceur-ws.org)

Connecting on hate crime data in Italy | Facing Facts

paper49.pdf (ceur-ws.org)

Hate speech annotation: Analysis of an Italian twitter corpus (unito.it)

L'epidemia dell'odio online: l'hate speech ai tempi del Covid-19 (economiecomportamentale.it)

Measuring and Characterizing Hate Speech on News Websites | 12th ACM Conference on Web Science

omsa.pdf (cornell.edu)

<http://ceur-ws.org/Vol-2481/paper57.pdf>

XXIX Rapporto Immigrazione Caritas e Migrantes 2020. “Conoscere per comprendere” – Fondazione Migrantes

An Impossible Dialogue! Nominal Utterances and Populist Rhetoric in an Italian Twitter Corpus of Hate Speech against Immigrants - ACL Anthology

Hugging Face – The AI community building the future.

Natural Language Processing — Dependency Parsing | by Shivane Jaiswal |
Towards Data Science

Named-entity recognition Definition | DeepAI