FAST LISA E-Participation platform

15.09.2023

Chiara Catizone, Giulia Venditti, Monica Palmirani – CIRSFID
monica.palmirani@unibo.it
chiara.catizone2@unibo.it
giulia.venditti2@unibo.it

FASTLISA

Index

Index

- Software developement
  - Flow
  - Classificators selection and fine-tuning
- Dissemination
  - JSON format
  - Upload on the server
- **RESULTS VISUALIZATION**
  - Dashboard Access and UI (header, help button, info button)
  - Dashboard full visualization division in fore area
    - Left (»level of severity»)
    - central (“lexicon” and “hate speech classification”),
    - right (“purpose”) ,
    - bottom (“discussion monitoring”)
  - Report
    - Purpose
    - Structure

INTRODUCTION

FASTLISA

Objective

The creation of an **online portal** at http://fast-lisa.unibo.it (based on WordPress), designed to ensure the **protection of personal data** while **fostering open public discussions** led by the **public administration** on topics closely linked to the everyday concerns of **citizens**.

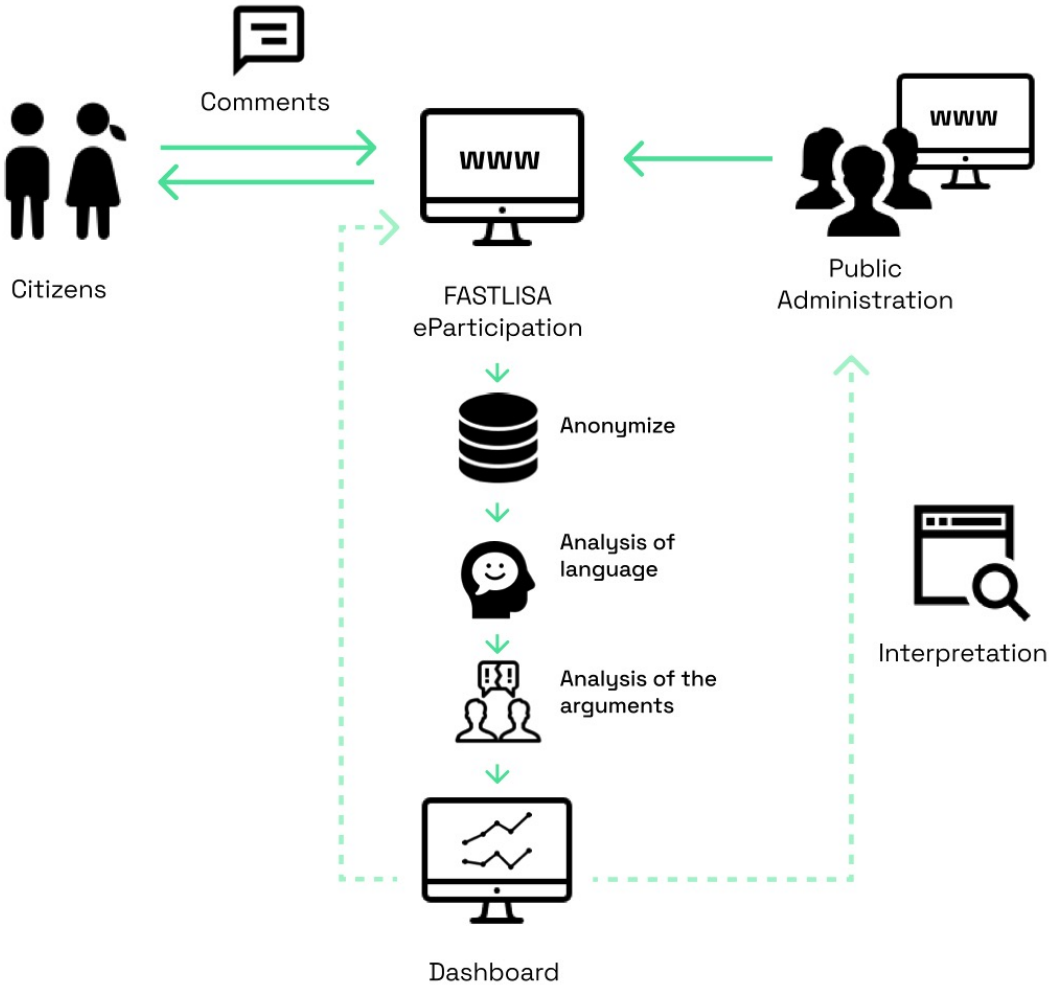
The main objective is to **assess the quality of these discussions, detect potential instances of hate speech**, and **provide** decision-makers with **a concise summary of prevailing opinions**.

This approach empowers public administrations to **evaluate policy effectiveness**, **monitor shifting citizen needs**, and **make more informed and impactful decisions**.

FASTLISA

System Architecture

1. **Public Administration Operators** initiate topic-specific consultations for citizen engagement.
2. **Citizens** search and participate in specific consultations leaving comments stored in anonymous form.
3. **Debates are moderated by FAST LISA ambassadors** for productive and respectful discussions.
4. **AI algorithm** analyse comments and messages for hate speech detection and cluster arguments into pros and cons.
5. **Interactive Dashboards** provide clear visualisations of aggregated hate speech data, promoting transparency and aid decision-making for public administration.



FASTLISA

CONSULTATION & DATA GATTERING

FASTLISA

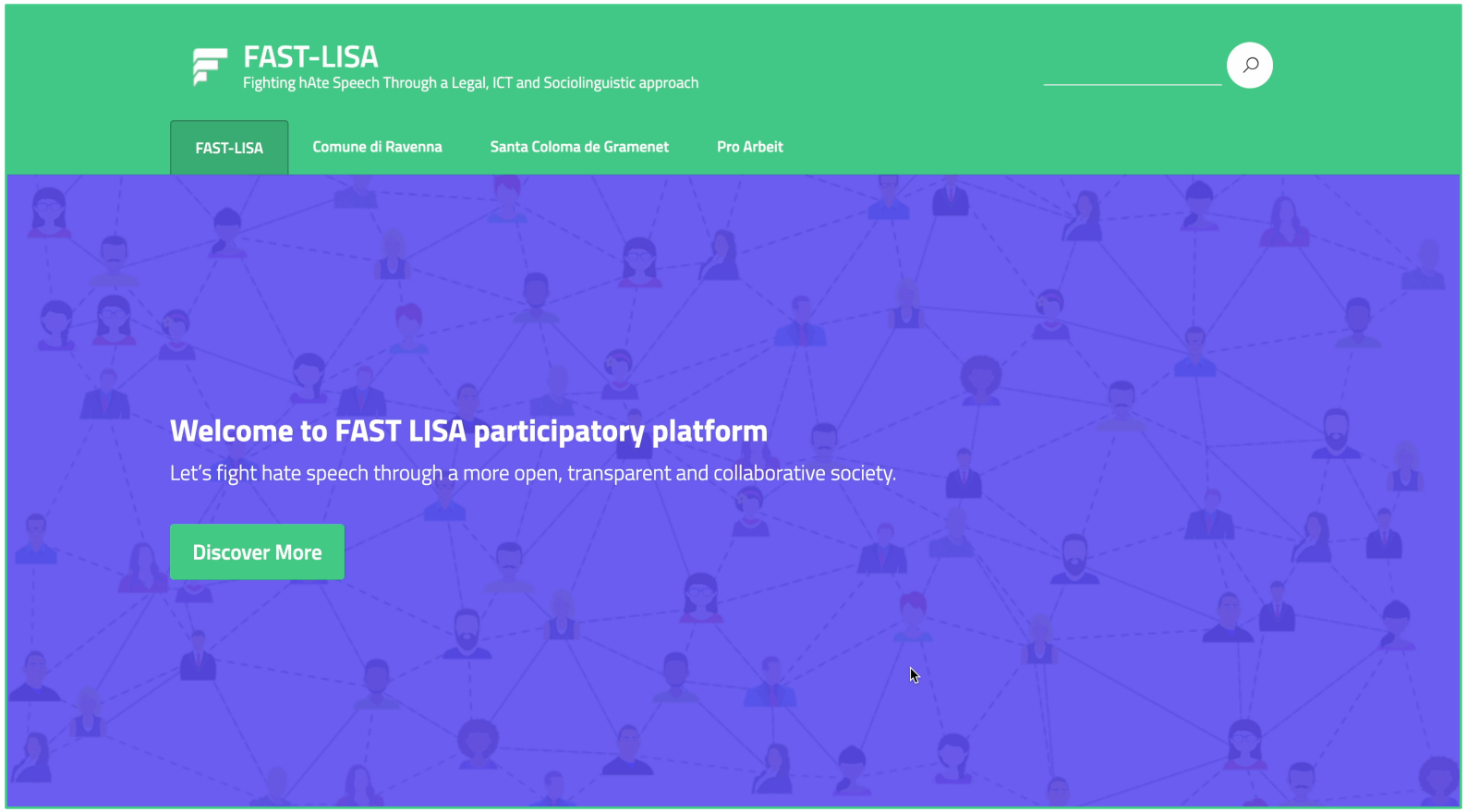
FAST LISA e-Participation platform

- CONSULTATION & DATA GATTERING

FAST LISA e-Participation platform works as a hub of participation activities across different partner organisation.

Home page
Primary Entry Point. It works as central hub of information

Partner Organisation
Customise in the respect of Public Organisation community cultural and linguistic expectation



FASTLISA

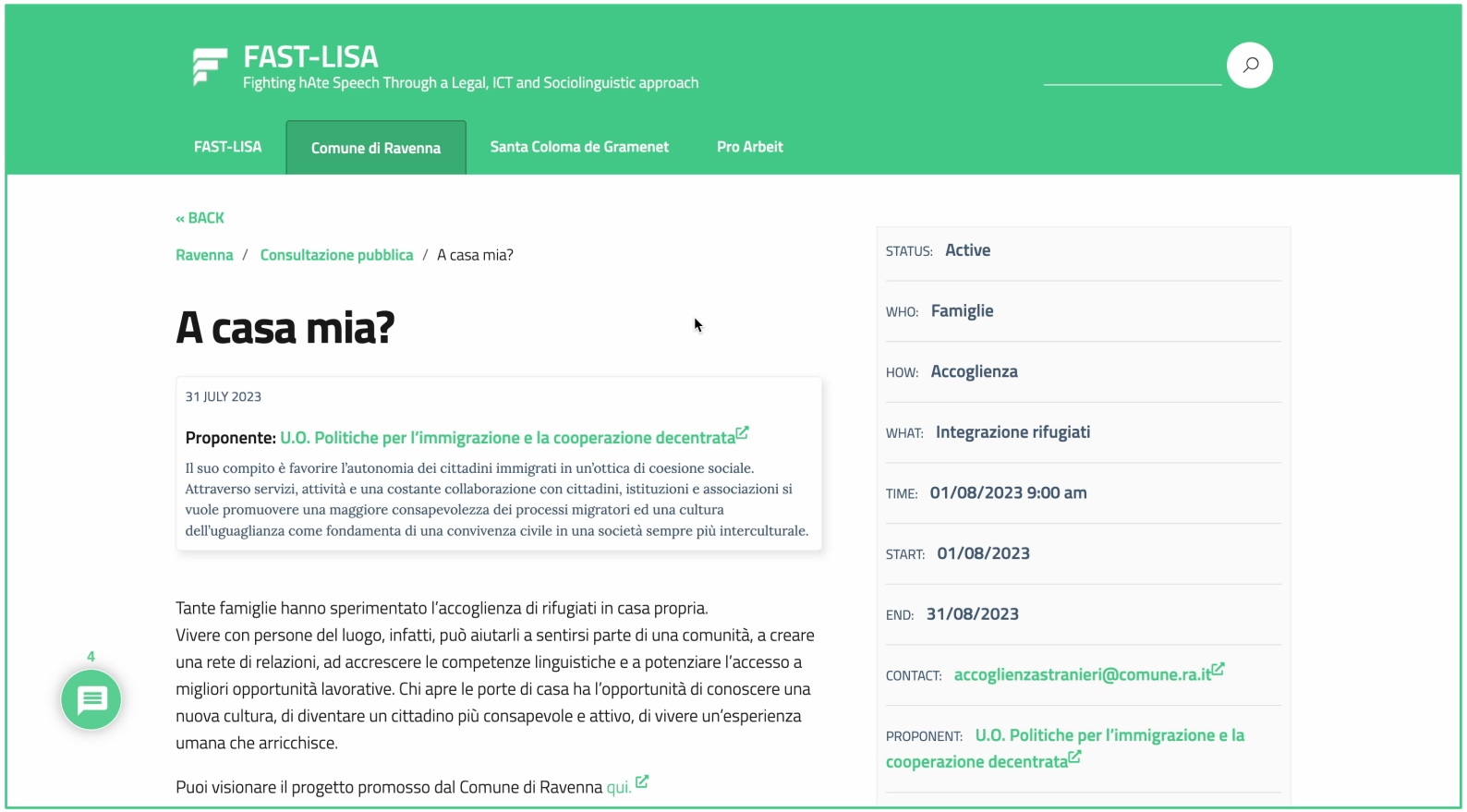
Public Consultation

- CONSULTATION & DATA GATTERING

Extensive explanation of the **discussion topic** and **participation rules** for informed and meaningful contributions.

Discussion forum for **anonymous commenting** allows for:

- **Filtering** options to arrange comments chronologically.
- **Navigate** to prioritize reacted ones, or focus on highlighted contributions.
- **Reply** to others comments and leave a reaction (like/dislike)

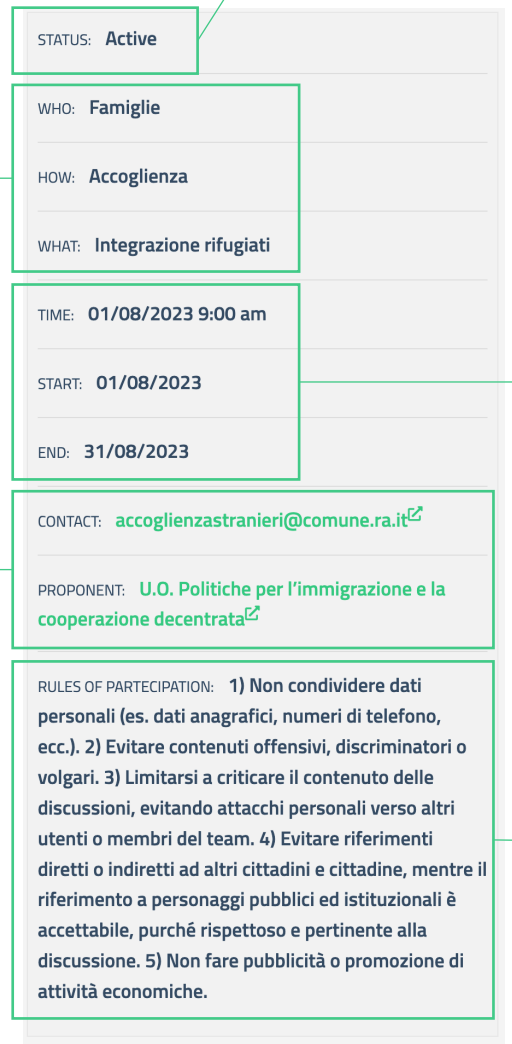


FASTLISA

Consultation Metadata

Activation status (active or closed)

Topic details.
(who, how, what)



Time-related information, such as **duration** and **start/end dates**, providing context for participants.

Proponent details with contact information and link to their official webpage for transparency and accessibility.

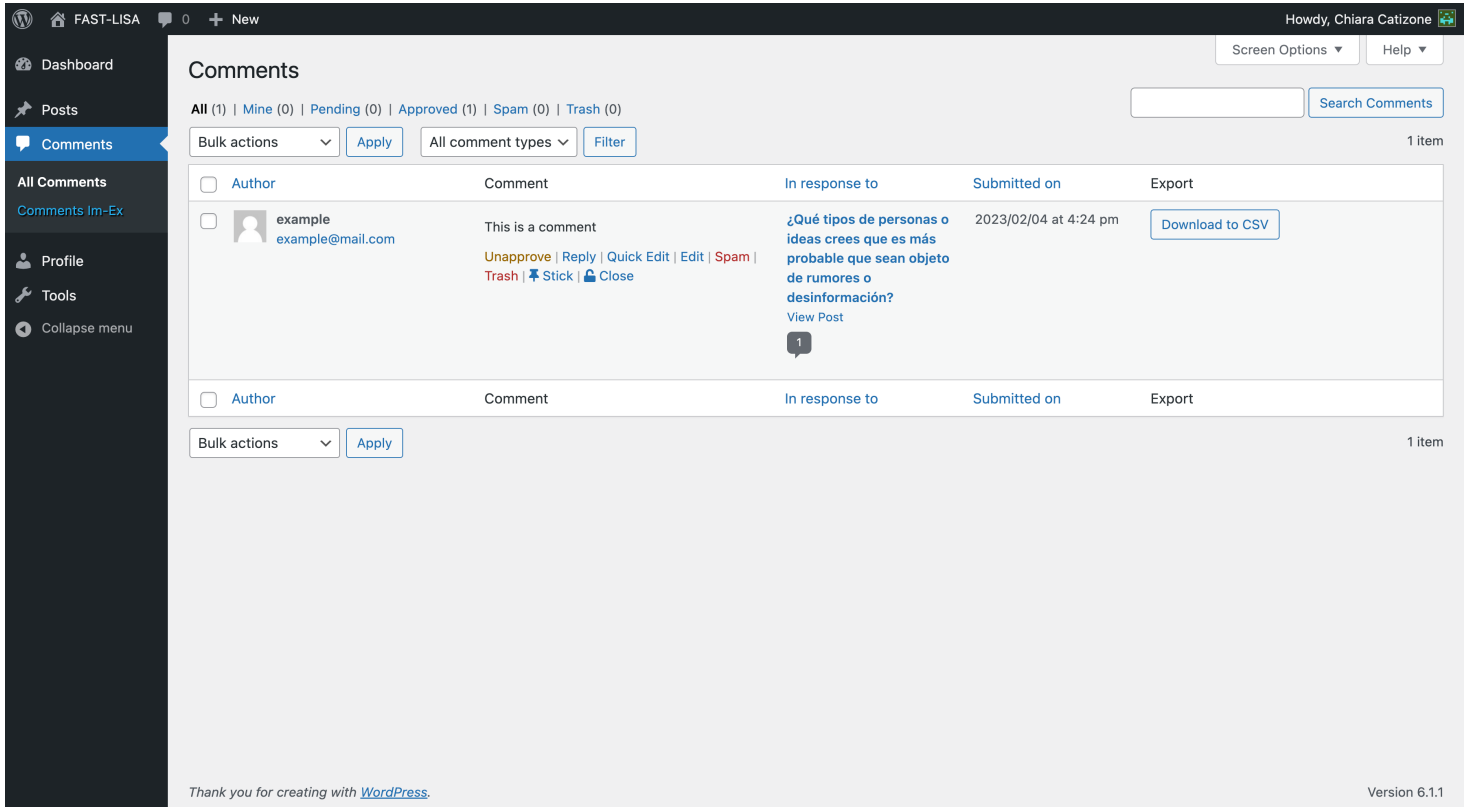
Schematic list of **participation rules** to ensure respectful and constructive engagement.

Debate Moderation

- CONSULTATION & DATA GATTERING

FAST LISA Ambassadors, thanks to provided credential, can access the WordPress Dashboard where they can perform a series of action on users comments:

- **Approve/Unapprove**
- **Reply** to comment to foster and guide dialog
- **Stick** at the top of the conversation
- **Edit** comment for example when personal data are shared
- **Close** reply to a particular comment
- **Mark as Spam**
- **Trash** the comment

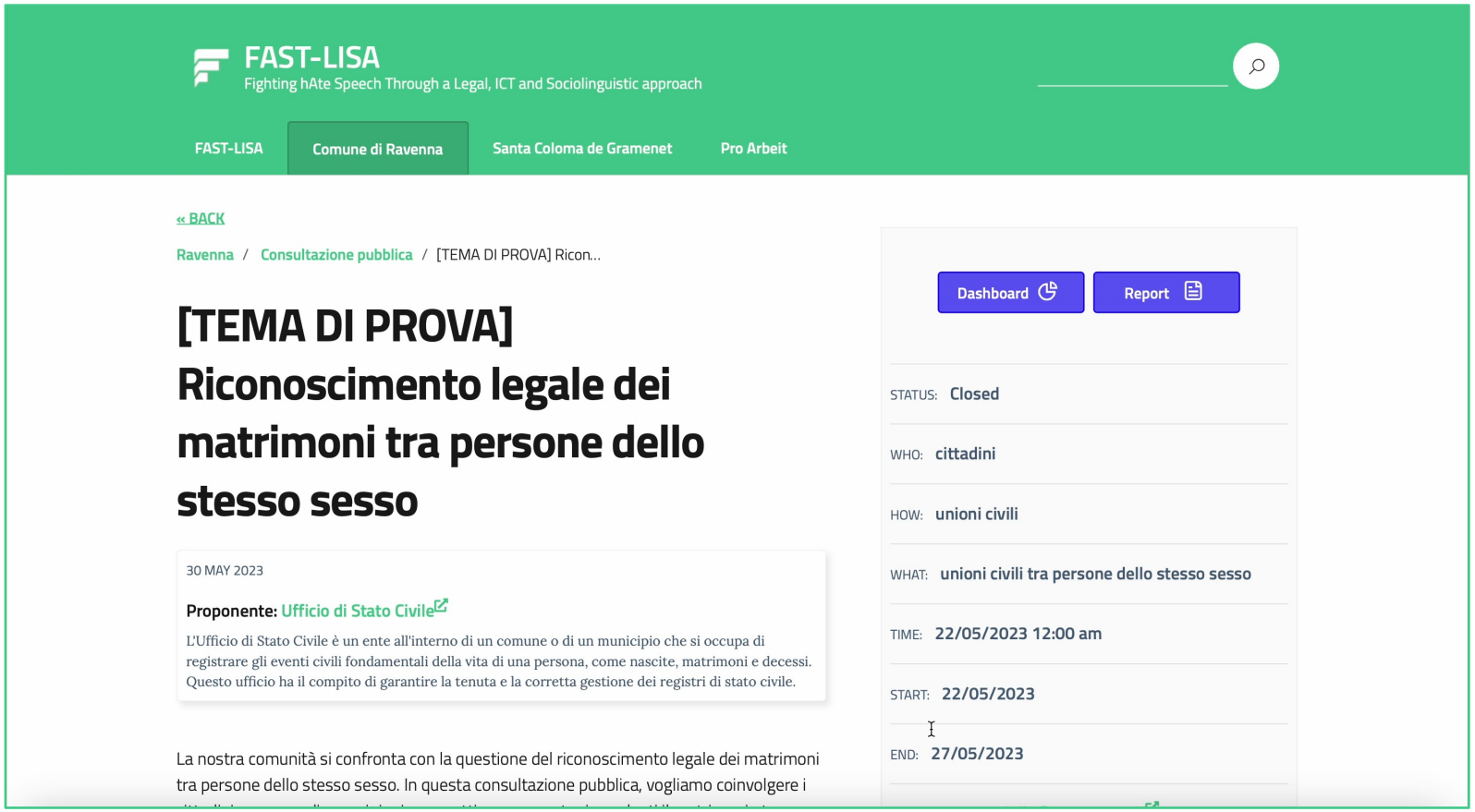


FASTLISA

Consultation Closure

- CONSULTATION & DATA GATTERING

1. **Consultation Status** is changed to **Closed.**
2. **Comment form** is **deactivated** but posted comments are steal visible and navigable.
3. **Data gattered** are **processed by the AI algorithm**
4. **Results** are **published** and accessible through the **“Dashboard”** and **“Report” buttons**.



FASTLISA

DATA ELABORATION:
attori e workfow

FASTLISA

Workflow

- DATA ELABORATION

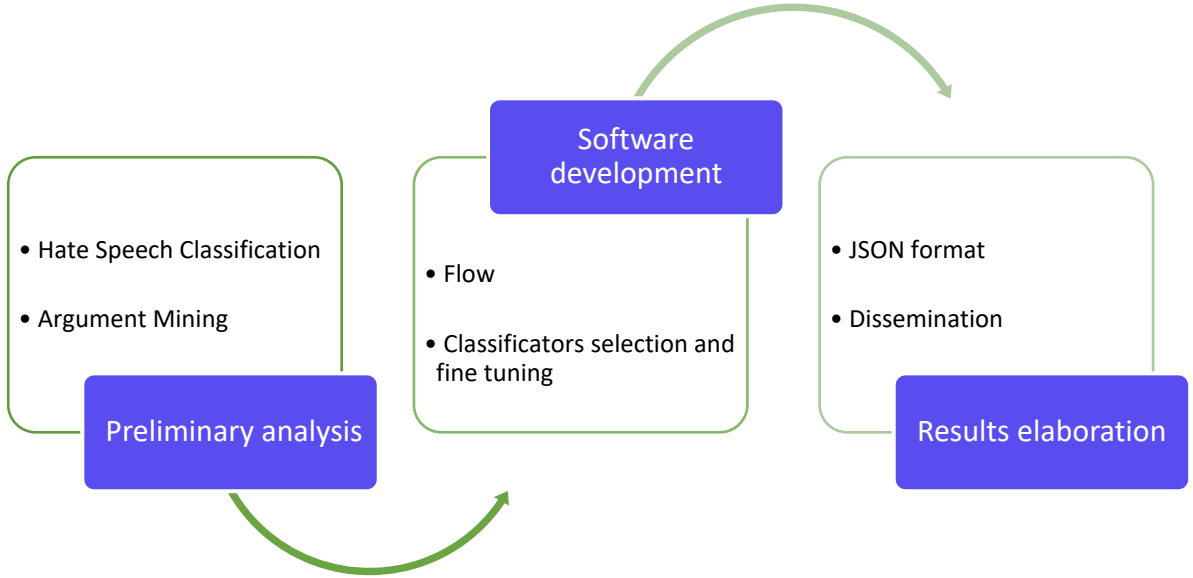
Premises

Focus on detecting and describing **hate speech** and the **flow of the discussion**

- **Legal, social and linguistic basis** for the taxonomy
- **Pragmatic aspect** of argumentation must be highlighted

Methodological pillars:

1. *Lexicon oriented to the legal domain*
   Classes extracted from legal documents to support the creation of a trans-border taxonomy for classifying online hate speech.
2. *Situations of the hate speech (pragmatic aspect)*
   Designing a methodology for extracting arguments from raw text to understand the users' inclinations when expressing their opinion in institutional context.



Preliminary studies

- DATA ELABORATION

Hate Speech Classification – I

Online Hate Speech can be automatcally descripted by AI tools trained in classification tasks.

To automate the descriptive process we first needed to set a trans-border taxonomy to classify data in an harmonyzed – yet legallly valid – way.

"A taxonomy (or taxonomical classification) is a scheme of classification, especially a hierarchical classification, in which things are organized into groups or types."

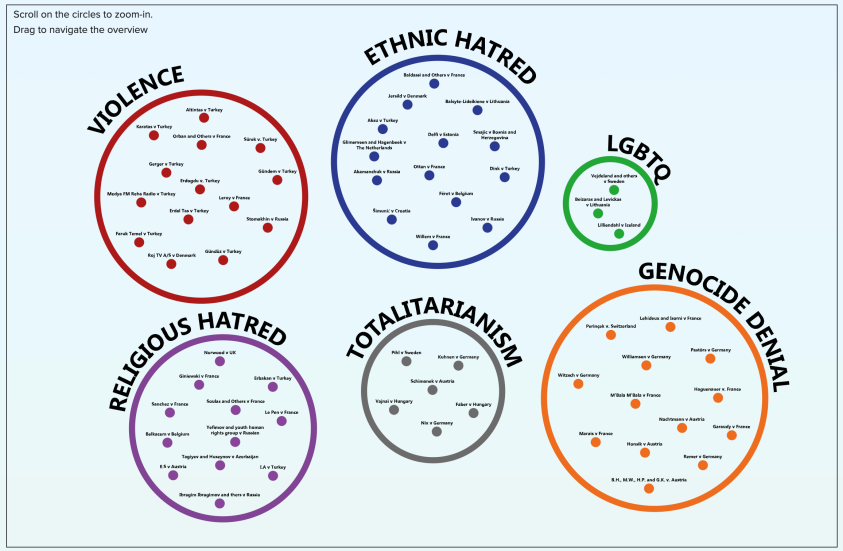
– Wikipedia

In literature research step we gathered existingtaxonomies used for describing hate speech.

Unfortunately they have no legal value, necessary for our project as specified in the premises.

The Future of free speech's taxonomy, unique with legal value, because it wasvbuilt starting from European Court of Human Rights (ECHR) documents.

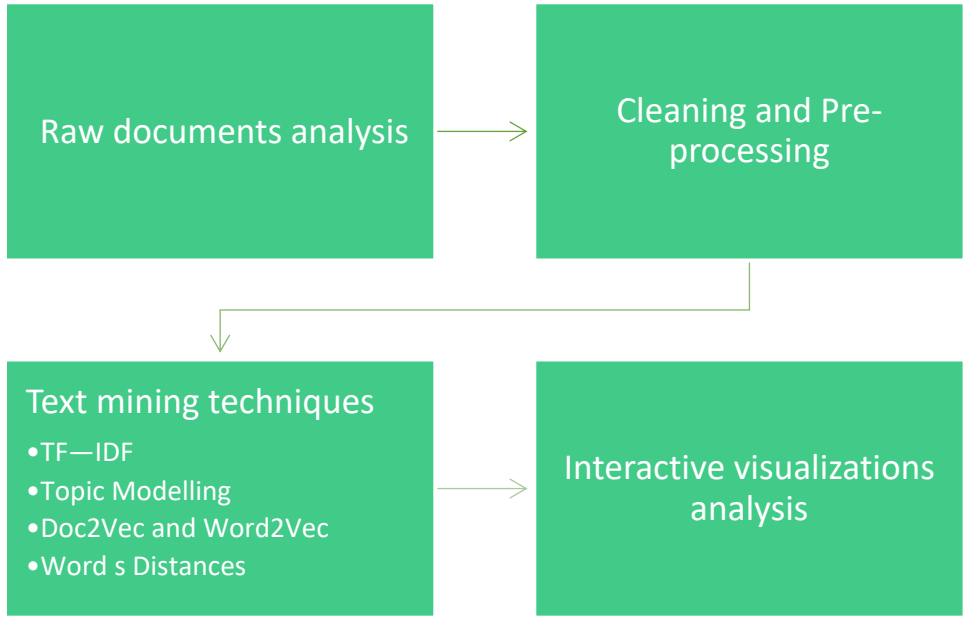
Hate Speech Case Database



FASTLISA

Preliminary studies

- DATA ELABORATION



Hate Speech Classification – II

All the legal documents have been gathered by domain experts.

Analysis tool:

- KNIME, an Open Source software for data analysis

Following (Salminen et al. 2018) this methodology is based on the **open coding technique**, in which classes emerge from the material.

Features:

- Information retrieval techniques like TF-IDF, which has been useful to retrieve important terms (Rajaraman and Ullman 2011; Luhn 1957; Robertson 2004; Vrysis et al. 2021)
- Latent Dirichlet Allocation was used to model topics and extract them atics proposed by (Blei 2003; Salminen et al. 2018)
- We used embeddings (Doc2Vec and Word2Vec) and word distances to support the linguistic categorization of terms related to hate speech (Lewandowska-Tomaszczyk et al. 2021).

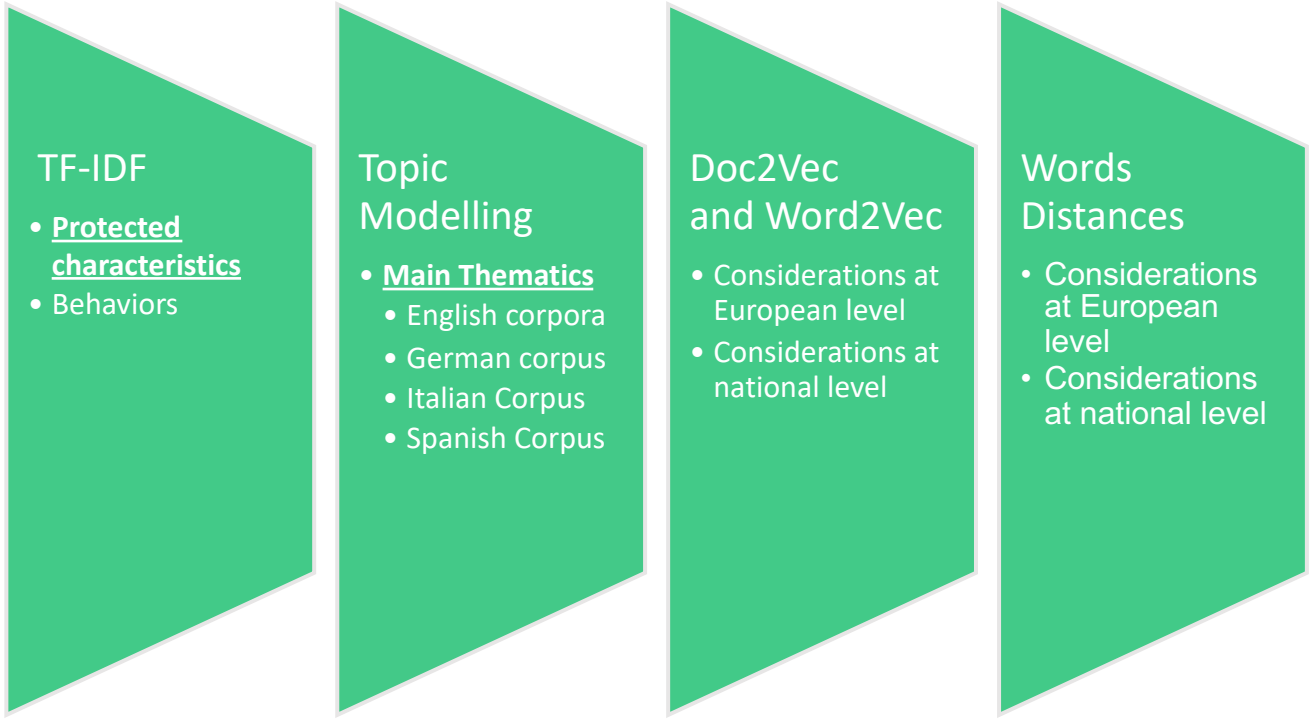
Preliminary studies

- DATA ELABORATION

Hate Speech Classification– III

In brief, for each investigated country we were able to retrieve the protected characteristics constituting criminal grounds –i.e., bias motivation– in the national legislation and a lexicon of abusive behaviors characterizing the crimes targeting minorities.

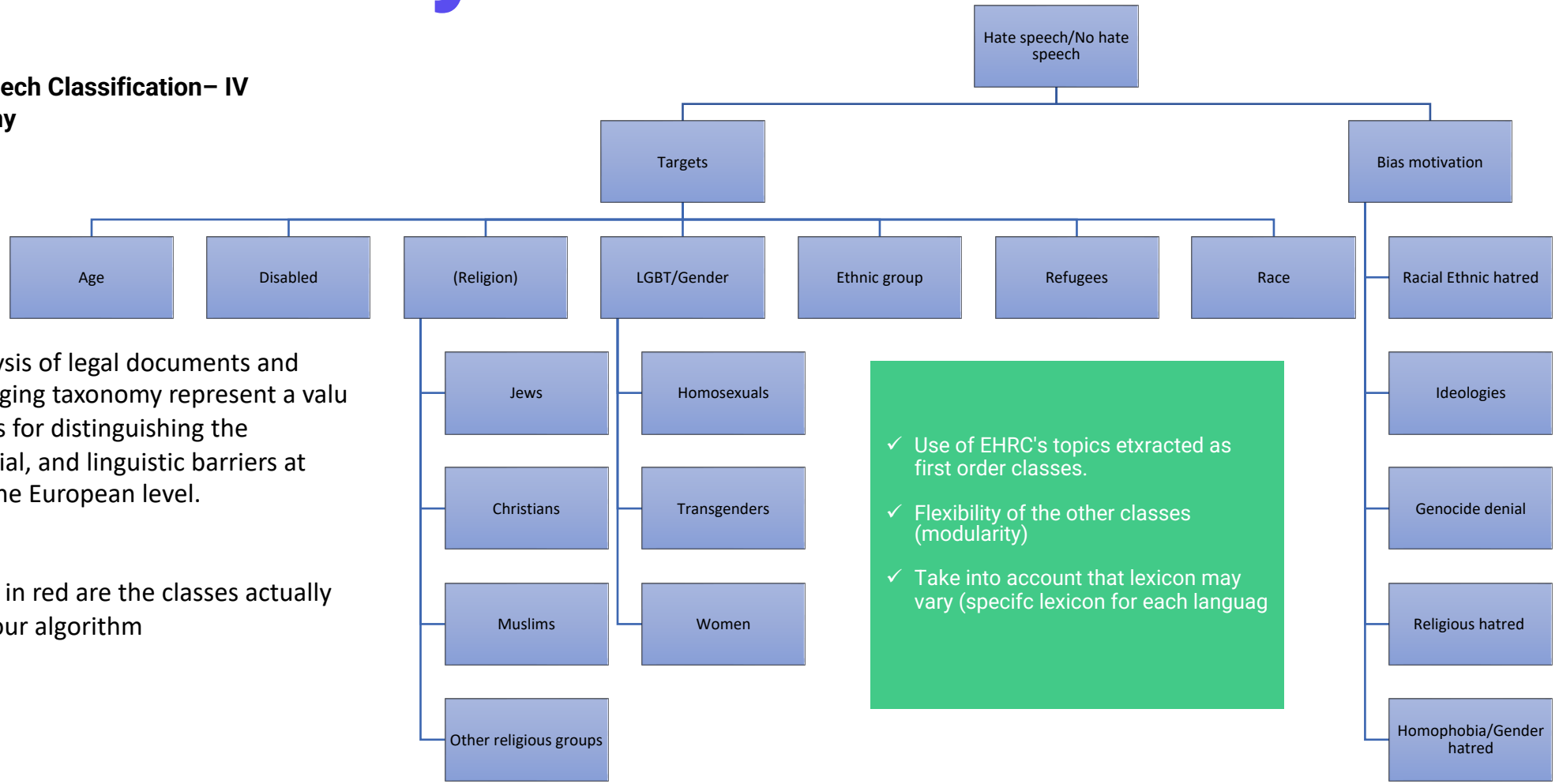
Racism and ethnic hatred are the most common bias motivation among the investigated countries



Preliminary studies

- DATA ELABORATION

Hate Speech Classification– IV Taxonomy



The analysis of legal documents and the emerging taxonomy represent a valu able tools for distinguishing the legal, social, and linguistic barriers at least at the European level.

NB: texts in red are the classes actually used by our algorithm

- ✓ Use of EHRC's topics etxracted as first order classes.
- ✓ Flexibility of the other classes (modularity)
- ✓ Take into account that lexicon may vary (specifc lexicon for each languag

FASTLISA

Preliminary studies

- DATA ELABORATION

Argument Mining -I

Objective:

How to effectively structure argumentations from an online institutional environment?

Following the principle by (C. Joshi 2019) "*bigger data doesn't always translate into better decisions*", we employed text mining techniques to explore and compare the structure of argumentations and the behaviour of users to understand how to develop the argumentation mining process.

Preliminary studies

- DATA ELABORATION

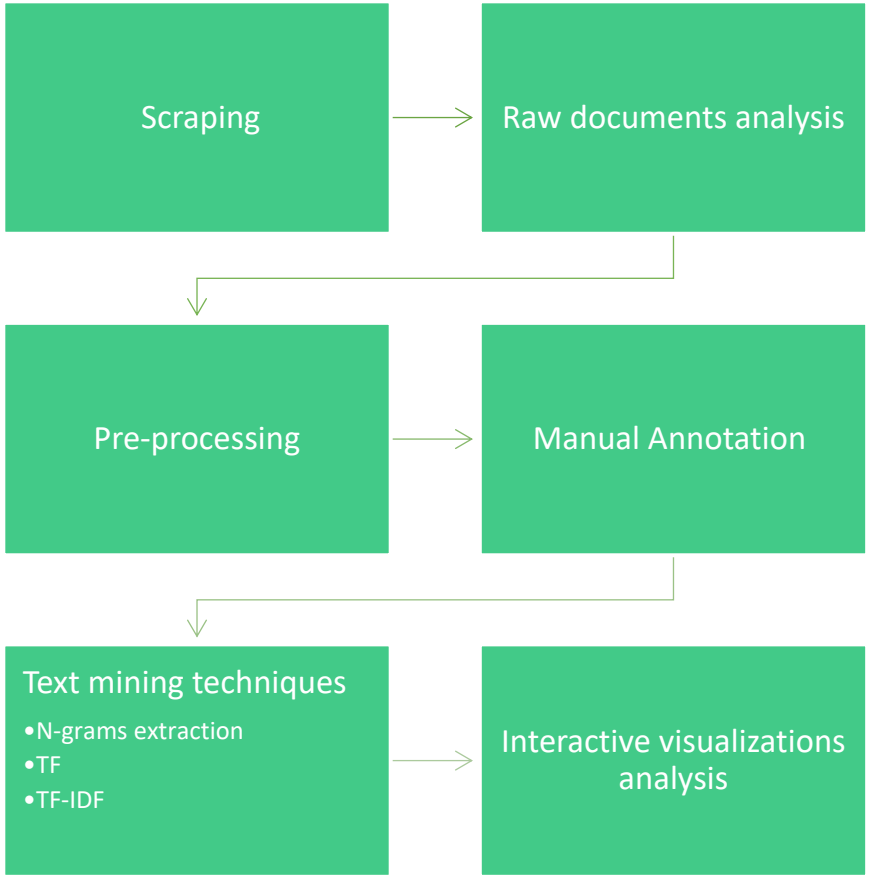
Argument Mining - I

Source:

European Commission’s consultations collection (European Commission 2023b)

Data Preprocessing:

1. Reduced the number of total topics from 91 to 40
2. Anonymized the data
3. Remove unnecessary words
4. Random sapling (500 comments each lang)



FASTLISA

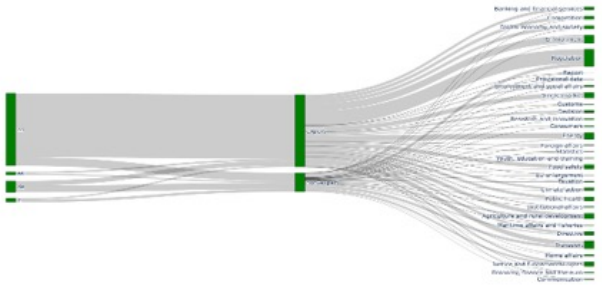
Preliminary studies

- DATA ELABORATION

Argument Mining - III

Considerations on Data Exploration

- Non expert users are mainly Italian and German
- Regulation and Transport have the highest number of paticipants
- Spain is the country with the lowest participation rate



COnsiderations on TF and TF-IDF results

•Automatically detect and extrac t frequently occurring phrases from the responses
•Discourse indicators like “because” and “however” + mod al verbs emerge from all corpora
•European citizens from differen t countries share similar worries and preferences towards the activities

Final Remarks

Arguments analysis was a valuable tool for understanding how to classify and manage the commentsts, independently from their language to gather important considerations:

- We base our work on a theory that studies argumentation schemes, where only premises and conclusions are classified (Mochales and Moens 2011)
- Explicit if a unit is in support or against the main thematic
- Explicit their relations using similarities

Software developement

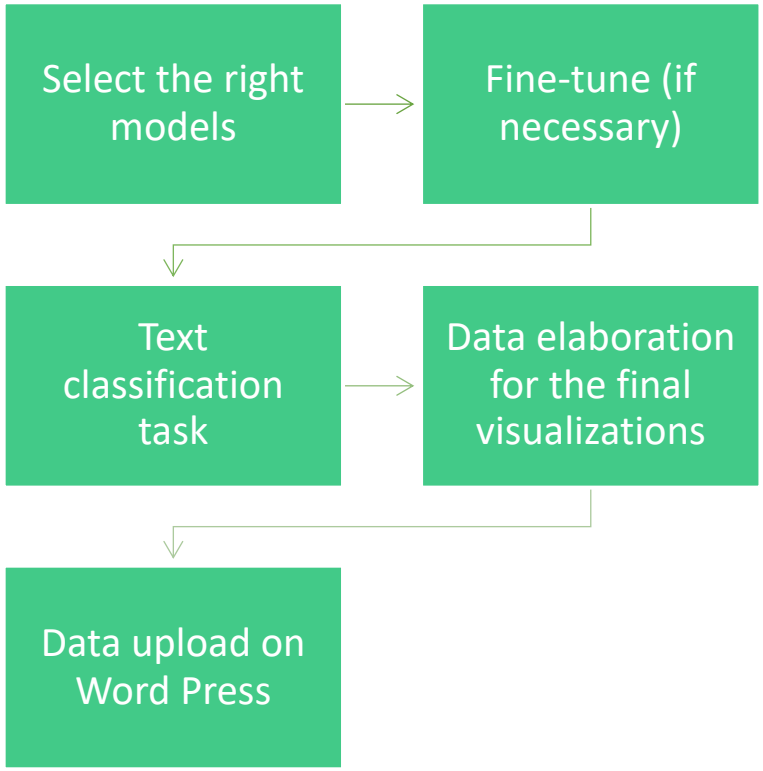
- DATA ELABORATION

General workflow

Do not start building models froms scratch, reuse (models and training data) when possible.

Five different classification tasks :

1. Hate speech presence
2. Target identification
3. Bias motivation identification
4. Claims and premise s identification
5. Support/contro relationships identifification



FASTLISA

Software development

- DATA ELABORATION

Classificators Selection

Testing of available hate speech detectors in each language (it, de, es).

Selected models for HS classification:

The model we relieve on for argument mining task:

[mDeBERTa-v3-base-xnli-multilingual-nli-2mil7]

Further training is needed for...

- Target
- Bias motivation
- Claims and premises

Models used for finetuning

Software development

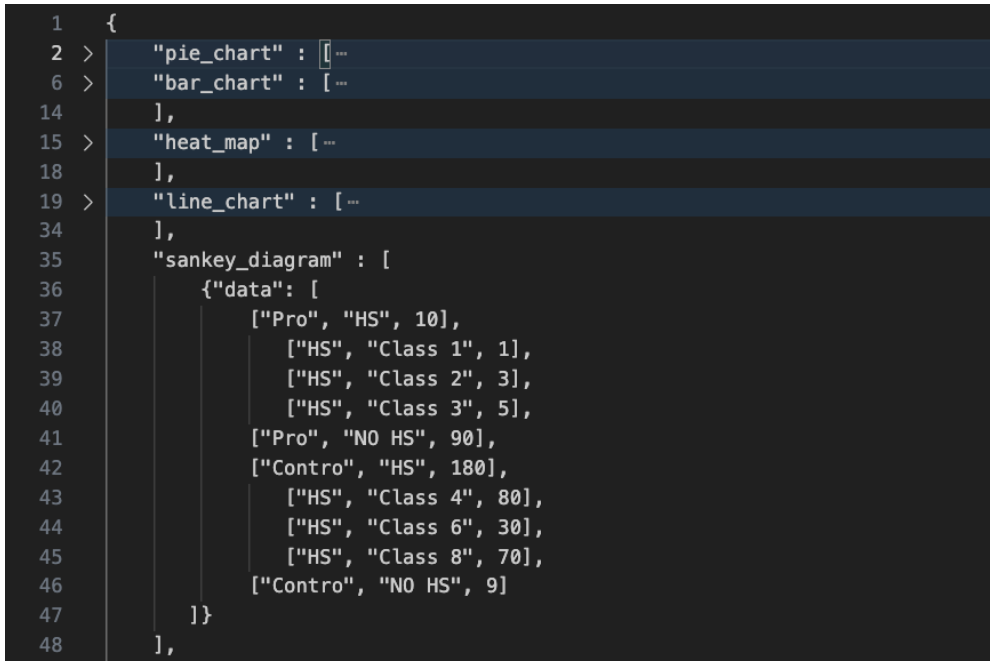
- DATA ELABORATION

Steps:

- Detect hate speech presence
  (pie-chart)
- Detect hate speech targets
  (bar-chart, tree-map)
- Detect hate speech bias motivations
  (heat-map, treemap, sankey diagram)
- Detect argument units (claims and premises)
  (arguments map)
- Predict support or cotraddiction relationships between argument units
  (sankey diagram, arguments map)
- Compute duration time of the discussion, log data of the interactions.
  (line-chart and counters)

Results elaboration

- DATA ELABORATION



```
{
    "pie_chart" : [...
    "bar_chart" : [...
    ],
    "heat_map" : [...
    ],
    "line_chart" : [...
    ],
    "sankey_diagram" : [
        {"data": [
            ["Pro", "HS", 10],
               ["HS", "Class 1", 1],
               ["HS", "Class 2", 3],
               ["HS", "Class 3", 5],
            ["Pro", "NO HS", 90],
            ["Contro", "HS", 180],
               ["HS", "Class 4", 80],
               ["HS", "Class 6", 30],
               ["HS", "Class 8", 70],
            ["Contro", "NO HS", 9]
        ]}
    ],
```

Dissemination

- Save the results in JSON format on local machine
- Upload the JSON file to FAST LISA's server

Limitations:

A future fine-tuning iteration might be needed for improving of the system.
W deal with low reasource language, i.e., lack of ready-to-consume NLP tool.

FASTLISA

RESULTS VISUALIZATION

FASTLISA

Interactive Dashboard

- RESULTS VISUALIZATIONS

Context: Displaying relevant information about the discussion, including the title, proponent, date, and a brief summary of the central argument.

Transparency: Info buttons providing detailed information about the purpose and data represented in each visualization.

Help Section: Offering support and information about the dashboard's features and functions.

Privacy: Presenting data in an aggregated and anonymized manner.

FASTLISA

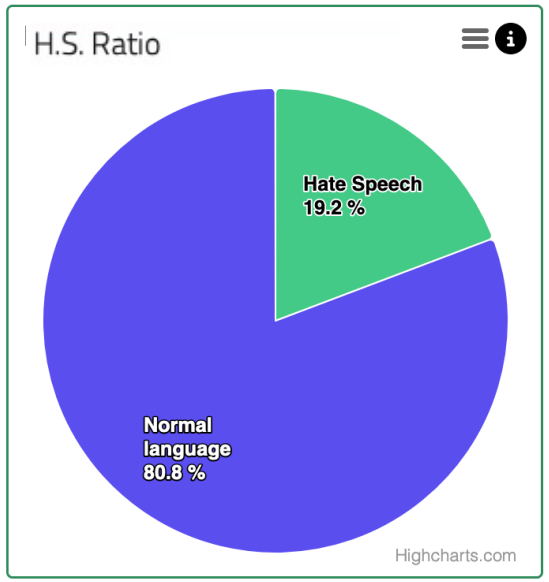
Left Column

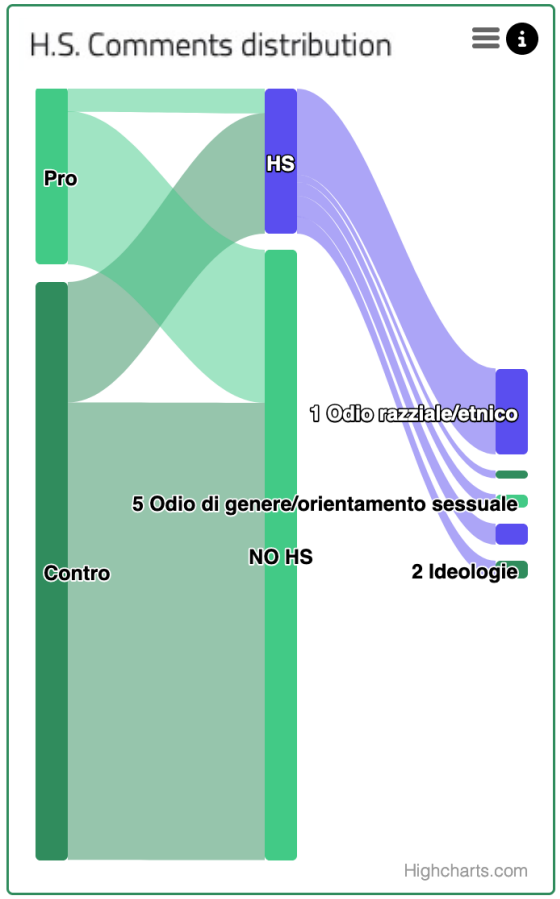
“level of severity”

- RESULTS VISUALIZATIONS

A **pie chart** provides a clear ratio between Hate Speech and neutral comments.

Quickly grasping the **conversation's tone** and **hateful language presence**.





A **sankey diagram** displays the composition of hate speech within pro and counter comments.

Effectively illustrating the flow and **distribution of comments**. Highlights **hate speech presence** and **different classifications**.

Simplifies hate speech analysis, aiding in comprehension and discussion argumentations' dynamics.

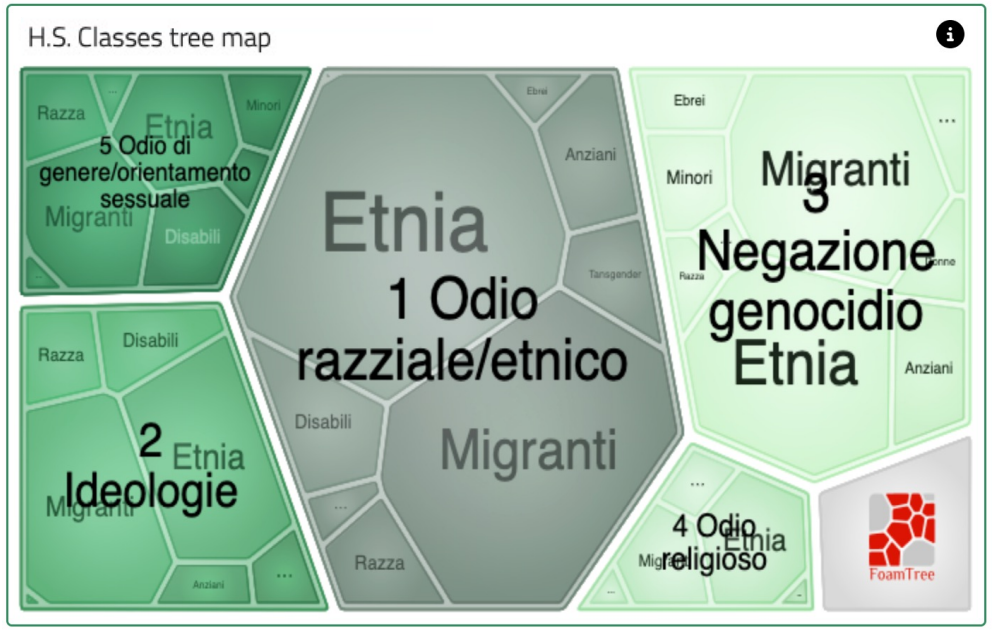
FASTLISA

Central Area

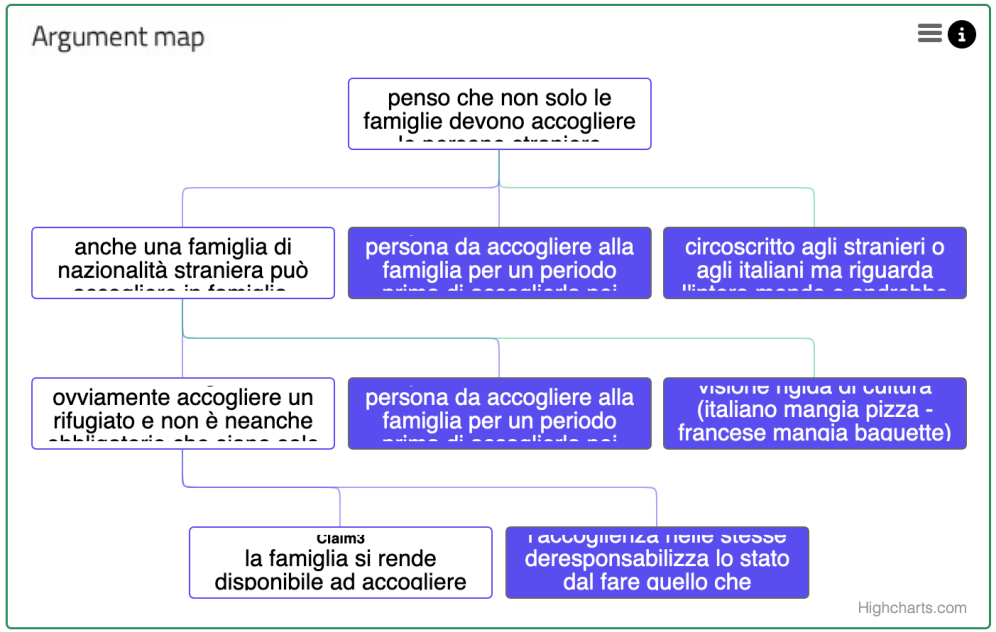
“lexicon” & “hate speech classification”

- RESULTS VISUALIZATIONS

A tree map displays hierarchical relationships between Hate Speech classes and frequently used words in classified Hate Speech comments. Helps users understand the content associated with each class (e.g., racial, religious).



An argument map highlights primary arguments and their relationships within the discussion. Offers a critical overview of the discussion's progression. Enables users to form well-informed opinions by visualizing key debating points.



FASTLISA

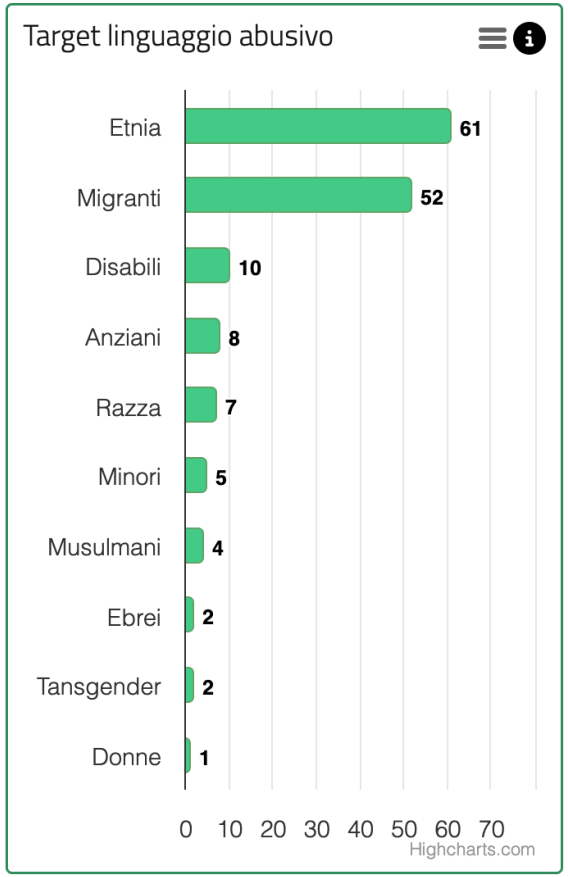
Right Column

"purpose"

- RESULTS VISUALIZATIONS

A **heat map** illustrates the relationships between bias motivations within the discussion.
Reveals **correlations among various biases**, aiding in understanding their presence and overlap.

Utilizes a classification system categorizing biases, including **ethnic origin**, **race**, **religion**, **nationality**, **gender**, and **migrant status**.



The target group victim of the offense is examined through a **bar chart**. It showcases the distribution of Hate Speech comments based on their respective targets.

Distinguishes between **single** and **group targets**, including **specific characteristics** like gender, race, and age.

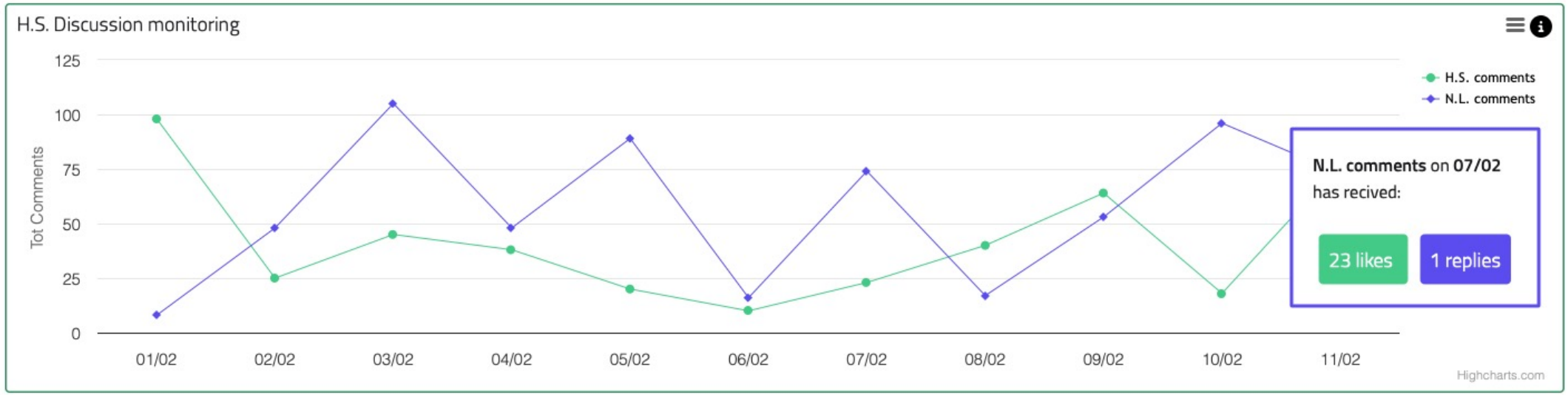
FASTLISA

Bottom Area

“discussion monitoring”

A **line chart** tracks trends in both neutral and Hate Speech comments throughout the entire discussion. Offers a comprehensive view of the conversation's progression, identifying points of increased Hate Speech prevalence or vice versa. Enhances understanding of discussion dynamics and patterns, aiding in analysis and critical evaluation.

A representation of **comments engagement** showcases the number of likes and replies received by classified Hate Speech comments. Enables users to gauge the influence of Hate Speech on the discussion and the attention it receives from other users. Valuable for understanding Hate Speech's potential impact on the community and promoting awareness and critical evaluation.



FASTLISA

Report

- RESULTS VISUALIZATION

Purpose

Support the understanding of data through brief descriptions of the visualized data

Structure

1. Introduction
   - Discussion name
   - Consultation period (start and end-date)
   - Proposing office/comunality
2. Dashboard purposes
3. Introduction to classes
4. Description (with picture) of each viz
5. Conclusion/duiscussion

Thank you.

FASTLISA