



No Village Left Behind: A Moroccan Data-driven Platform for Effective Aid Coordination



North Africans in ML

Abdelaziz Bounhar¹, Adnan Anouzla², Ahmed Lekssays³, Asmaa Zizaan⁴, Basma Chourane⁵, Fatima Zahra Qachfar⁶, Hafsa Ouifak⁴, Imane Momayiz⁷, Loubna Ben Allal⁸, Maroua Razzouqi⁹, Meryem Jebrane, Mustapha Ajeghrir¹⁰, Nissrine Hatibi¹¹, Nouamane Tazi⁸, Soundouss Messoudi¹², Yassir Bendou¹³ and Youssef Adarrab

¹Télécom Paris, ²ENSMR, ³Qatar Computing Research Institute, ⁴UM6P, ⁵Seven-app, ⁶University of Houston, ⁷Equancy, ⁸Hugging Face, ⁹FST Mohammedia, ¹⁰Hrflow.ai, ¹¹ENSIAS Rabat, ¹²Heudiasyc, ¹³IMT Atlantique

nt3awnoumorocco@gmail.com

Introduction

- In the aftermath of the devastating 6.8 magnitude earthquake that struck Morocco's High Atlas Mountains on September 8, 2023, Nt3awnou's platform swiftly emerged to optimize relief coordination. This involved efficiently orchestrating resources to aid those affected.
- Consolidating diverse and heterogeneous data sources is complex, particularly in low-resource dialects like the Moroccan one.
- In this work, we employed NLP techniques to convert collected data into a refined and usable dataset. Key challenges included authentic data collection in crises, accurately identifying similar-named rural villages ('douars') and obtaining precise geolocation despite inaccuracies in mapping APIs (e.g. Google Maps and OpenStreetMap).

Pipeline in a Nutshell

- Curating reference 'douars' data by aligning two datasets with geo-ontology categories, utilizing phonetic distance measures.
- Gathering needs/interventions, involving structured information from NGOs and individual requests in Moroccan Dialect via a chatbot.
- Data processing: transcribing audio messages, cleaning, parsing locations, and mapping them to reference douars' data using matching techniques.
- Categorizing requests (shelter, food, medicine) via a multilingual model.

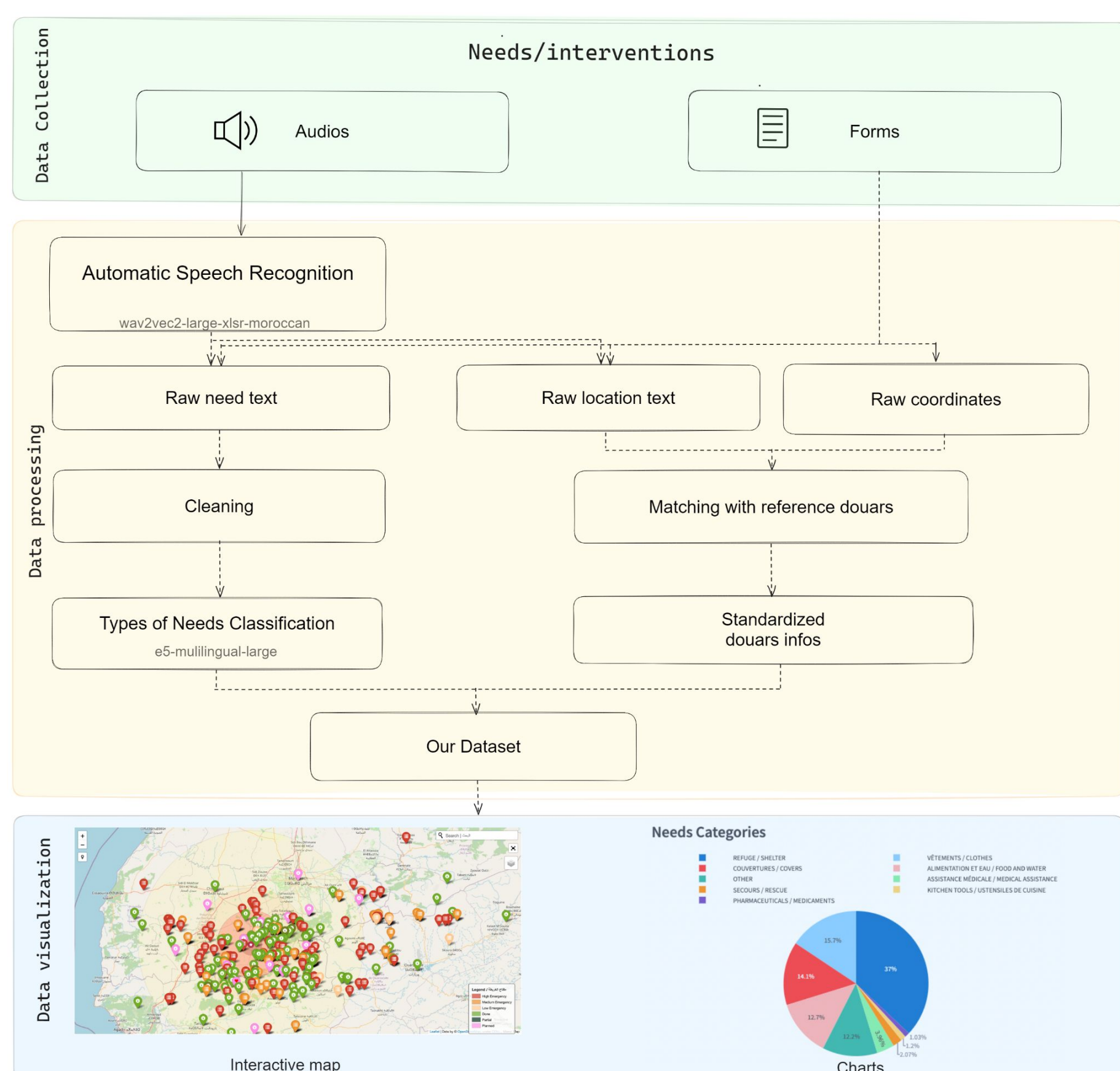


Figure 1 : The overall pipeline of the data-driven platform.

We aggregate incoming requests into pre-defined population needs using the Fuzzy-Matching algorithm [1]:

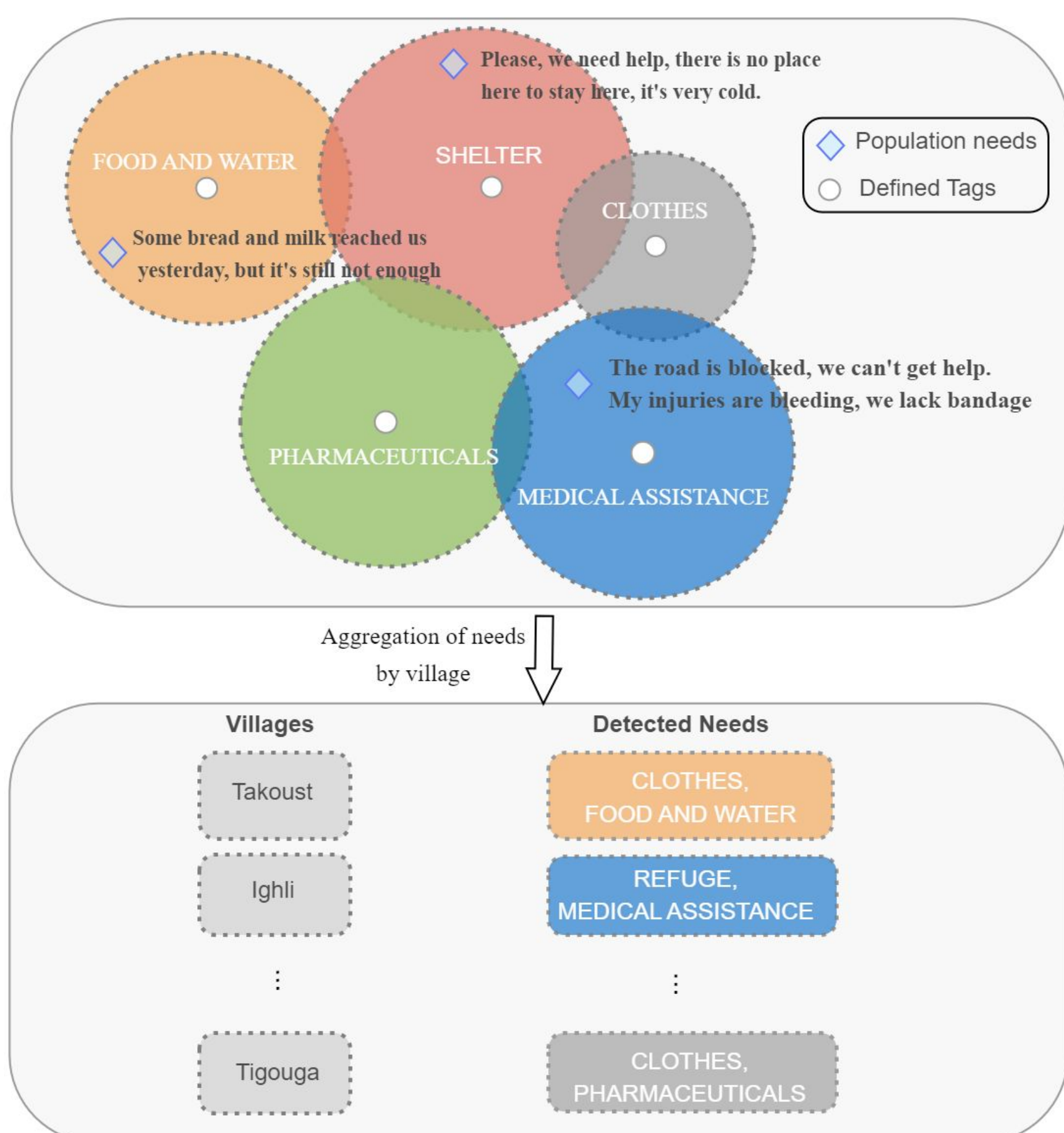


Figure 2 : Requests automatic classification.

Results

The collected and processed dataset is presented on an interactive map:

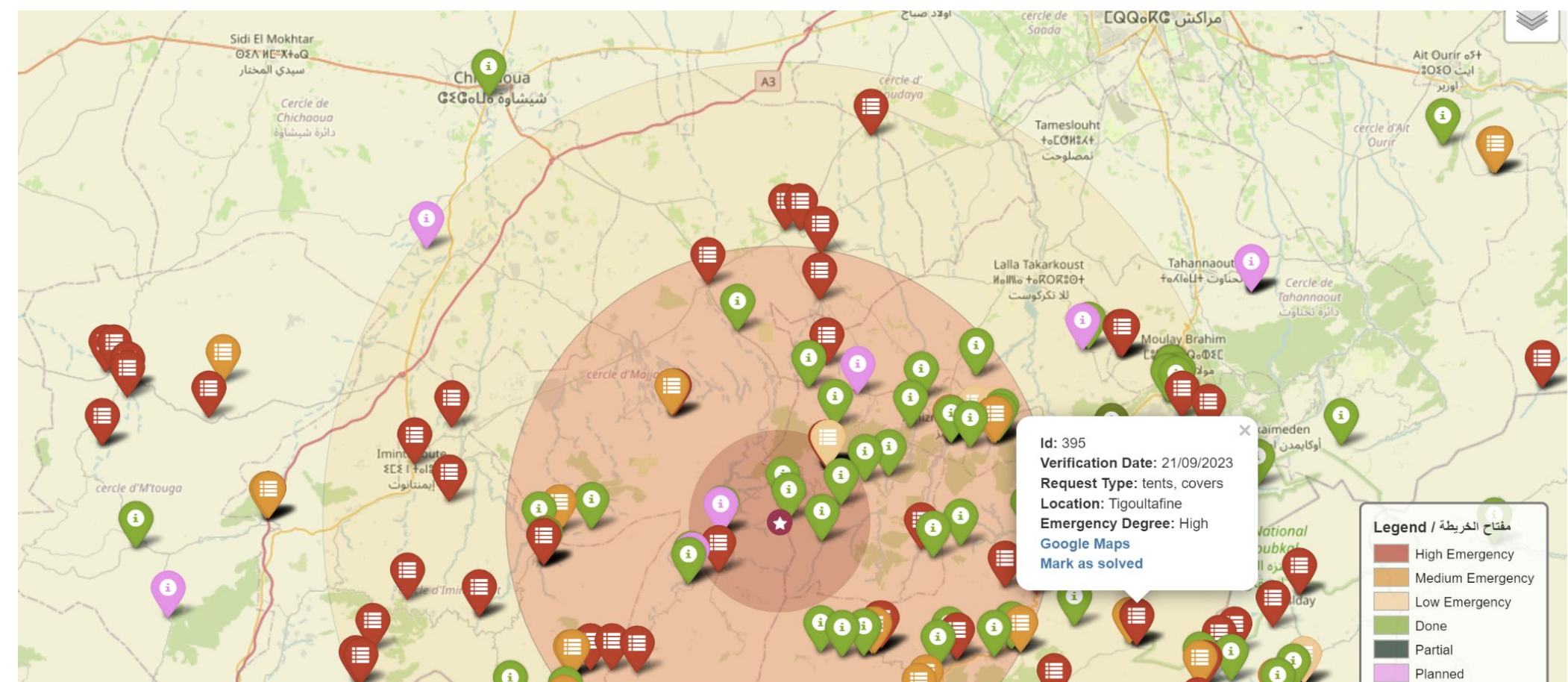


Figure 3 : Interactive map of the aid requests and interventions.

Using manually labeled data, we evaluate the performance of the different processing algorithms:

Tasks	#Data	Methods	Metrics	Categories	Results
Automatic speech recognition	306	wav2vec2-large-xlsr-moroccan [2]	Average Word Error Rate (WER)	—	53.54%
Matching textual locations with reference douars	66	Fuzzy-Matching [1]	Percentage of correct matches	Province	22%
Needs and supplies classification	40	Text embeddings by weakly-supervised contrastive pre-training [3]	Accuracy	Cercle	25%
				Commune	52%
				Douar (fr)	60%
				Douar (ar)	70%
				Clothes	87.5%
				Shelter	76.9%
				Food/Water	76.9%
				Pharmaceuticals	70%
				Covers	50%
				Rescue	50%
Medical Assistance	50%				
Other	53.8%				

Table 1: Results of the preprocessing tasks.

We provide summary metrics on the processed data, and the results show a high demand for sheltering and food.

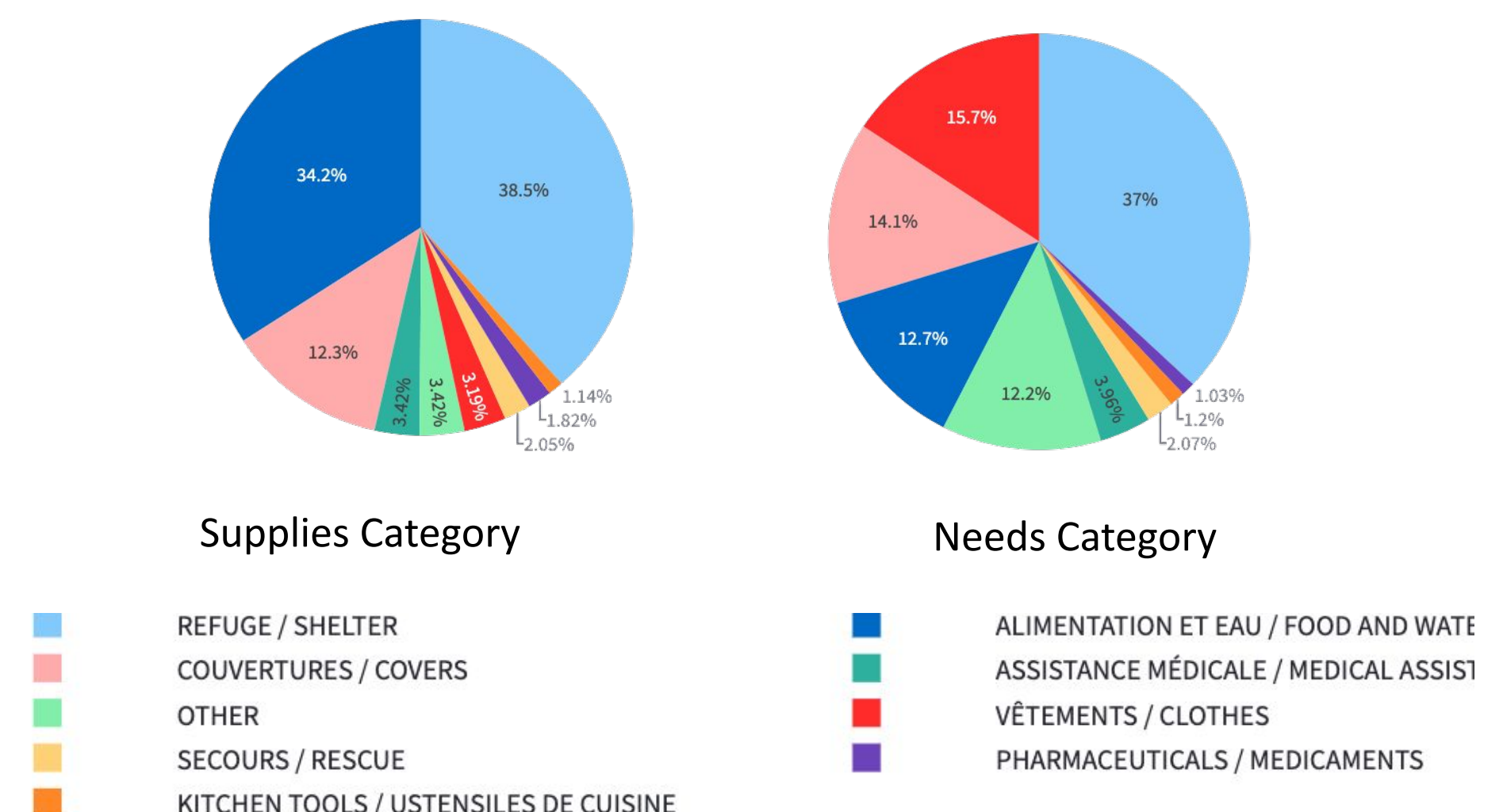


Figure 4 : Statistics of the needs and supplies.

Conclusion

- The presented Moroccan data-driven platform is designed to facilitate the coordination of aid for areas in need following the earthquake by collecting data, preprocessing it to get a clean dataset, and visualizing results and statistics on our online platform.
- Future work aims to involve social media analytics to gather additional updates on the situation of the affected areas and to develop methods for assessing damages through satellite imagery data.

Acknowledgements

We want to express our sincere gratitude to other contributors in this project: Ahmed Bendaouia, Ali El Filali, Amal El Amrani, Anass Amejjoud, Asmae Khalid, Ayoub Bami, Aziz Tamma, Faïçal Naciri, Halab Khidhr, Hind El Ajouti, Imane Briki, Imane Houmir, Issam Sahraoui, Khalid Baba, Louis Ulmer, Miloud Belarbia, Mohammed Chahboun, Mohammed Saika, Rachid Serraj, Walid Bouibes, and others for their invaluable assistance with data collection and cleaning. Thank you for your vital roles in our success.

References

- [1] Chaudhuri, S., Ganjam, K., Ganti, V., & Motwani, R. (2003, June). Robust and efficient fuzzy match for online data cleaning. In Proceedings of the 2003 ACM SIGMOD international conference on Management of data (pp. 313-324).
- [2] Baevski, A., Zhou, Y., Mohamed, A., & Auli, M. (2020). wav2vec 2.0: A framework for self-supervised learning of speech representations. Advances in neural information processing systems, 33, 12449-12460.
- [3] Wang, L., Yang, N., Huang, X., Jiao, B., Yang, L., Jiang, D., ... & Wei, F. (2022). Text embeddings by weakly-supervised contrastive pre-training. arXiv preprint arXiv:2212.03533.

