



EUMIGRATOOL REPORT #3

**Daniel Morente González
Nikos Gkevrekis**

August 2023

ITFLOWS Project
EUMigraTool Report #3
August 2023



"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 882986".

Table of Contents

- I. Executive Summary
- II. Introduction
- III. Current architecture of the EUMigraTool
- IV. Types of migration in the EU context: terminology and dataset limitations
- V. Case study one: Nigeria to Germany
 - a. Analysing the Nigeria conflict through the EMT
 - b. Asylum seekers from Nigeria to Germany
- VI. Case study two: attitudes towards migration in Greece
- VII. Last developments of the Twitter model
- VIII. Future considerations

I. Executive Summary

The EUMigraTool (EMT) is a solution-oriented tool developed by ITFLOWS for civil society and local authorities with two main functions. First, predicting arrivals of asylum seekers, and, second, detecting attitudes and potential tensions related to migration and asylum in EU Member States. The EMT works by predicting migration flows for humanitarian purposes, delivering data for the objective of the integration and relocation of refugees and always complying with a robust legal and ethical framework. Further, the tool detects and identifies needs among individuals prior to their arrival in Europe. End users of the EMT can employ it to obtain predictions on the number, gender and age range of asylum seekers and non-recognised refugees entering their respective territories, whilst also obtaining real-time information regarding refugee camps, some major cities and the conflict zones in non-EU countries, which can be also beneficial for their mission. The main objective of the tool is to assist relevant stakeholders manage migration for humanitarian reasons, for the benefit of host societies and for migrants themselves. This is the third report that analyses the most recent advances regarding the EUMigraTool.

II. Introduction

The project ITFLOWS is funded by EU Horizon 2020 and was born as a response to the challenge of employing computational predictive tools to anticipate and develop responses to arriving migrants at the European Union. Specifically, the project addresses how said management should involve equipping the actors who act as first and second respondents to the arrival and reception of migrants and asylum seekers with the necessary tools. And has closely worked with those who promote the successful integration of newly arrived migrants, asylum seekers and refugees who decide to remain in any given EU country.

On the third and final year of the project, the consortium partners have finalized their research work around the complex issue of migration flows, putting at the centre of the equation their multi-stage nature and non-linearity. The scope has remained until the end on humanitarian issues, dealing with the whole process from reception to relocation and settlement. Furthermore, and as this report documents, the project has concluded its development process of the EUMigraTool. This is an evidence-based ICT enabled solution, employing simulation and artificial intelligence models previously approved and under constant review by the members of the Users Board of ITFLOWS. This group of organizations and institutions comprise local governments, policy actors and relevant civil society organisations. Sustained contact has been maintained with the Users Board through the conduction of workshops and special meetings, through which the project developers of the EMT have been able to gather the formers' insights, recommendations and requests for the tool, based on their own experience as field practitioners and the needs that they have when conducting their mission. The conduction of tests in real time environments by this group of users have led to the improvement of the EMT.

The ITFLOWS project is divided into five objectives. First, there is the aim to provide accurate predictions and forecasts of migration arrivals towards the EU. This is fulfilled by analysing the drivers of migration and their intersection with structural elements of expulsion and attraction in origin and destination countries, respectively, taking into account EU policy on migration and the steady input of the members of the Users Board along the development process of the EMT. Second, building the EMT so that it can be employed by first and second respondents to migration, mainly, civil society organisations and municipalities, by making sure that its practical use fits their requirements and needs. In this regard, the active participation of the Users Board in the validation process has been critical.

Third, offer policy recommendations for the proper management of arriving asylum seekers and irregular migrants at EU Member States. This has been conducted at all times within an ethics framework based on the respect of human rights. Fourth, there is the objective of employing public sentiments of local citizens of EU countries to analyse attitudes towards migration and reduce the potential risk of tensions between them and newly arrived individuals. Fifth and final, to develop policy solutions to foster

and enhance the level of integration of migrants and refugees at host countries within the EU.

The way the project sees it, the EMT has the potential of adding simulation and machine learning techniques to the field of migration management and in the daily work and decision-making processes of relevant stakeholders. Humanitarian actors working on the reception and integration of migrants can benefit from such equipment, together with migrants themselves. Precisely because of the wide range of outputs that the EMT could provide its users with, ITFLOWS has dedicated a great deal of time and effort into identifying and contacting potential stakeholders working on the field of migration. Now that the EMT has been released as a functioning application, this report reflects on some of the findings made by the project throughout the last few months and on the responses provided by the Users Board.

The report begins with a summary of the architecture of the EUMigraTool, pointing out some of its technical details in regard of data processing and databases. Then the report details the terminology that has been used both in the project and within the very EMT, according to a pre-established glossary agreed upon by ITFLOWS members. In this section some limitations encountered throughout the development process of the EMT in regard to lexicon and databases are pointed out. The report follows with two case studies, that of the migration from Nigeria to the EU and, particularly, Germany, and Greece in relation to the attitudes towards migration. Later, some of the latest developments of the Twitter model are detailed, to then conclude with some future considerations regarding the EUMigraTool.

III. Current Architecture of the EUMigraTool

As it stands, the EMT has mainly two functionalities.

The first functionality is the one on Predictions, that consists of three models. Firstly, the small-scale model (SSM), offering predictions in the origin countries. It aims at predicting the distribution of arriving asylum seekers and unrecognised refugees at the neighbouring country of the origin state. To perform this task, the EMT employs a generalised and automated simulation development approach and the Flee agent-based simulation code. The SSM works through data obtained by a heterogenous set of sources, mainly, the United Nations High Commissioner for Refugees (UNHCR), the Armed Conflict Location and Event Data Project (ACLED), OpenStreetMap and population data from the City Population database, among others. Secondly, there is the large-scale model (LSM), providing forecast of international protection requests for as long as one month in EU Member States. To perform this monthly predictions, the EMT uses a state of the art machine learning approach that includes neural network architecture and time series analysis. This synthesised methodology allows for correlation analysis using raw data sources and simulations. And, thirdly, the Google Trends analytics model, offering forecasts of asylum applications also in EU countries for the following three months.

Then, the second functionality is the one on Attitudes. To conduct this task, the EMT uses the LSM to collect attitudes towards migration among the local population in the EU Member States acting as destination countries. This is accomplished employing the Twitter Sentiment Analysis model data as input, together with indicators deemed relevant to understand attitudes towards migration. Migration-related keywords are used to process relevant tweets.

These functionalities work by harnessing data from open-source, freely accessible statistic, such as Eurostat, together with information on factors of attraction and expulsion in origin and destination countries, respectively, and public sentiments towards migration.

For further information on the models that articulate the EMT, there are specific Deliverables of the ITFLOWS project that deepen on the subject. For instance, D6.2 on the preliminary release of the tool, or D6.4 on its final release.

This report aims at providing relevant information to policy makers, researchers and actors from the civil society. In particular, it offers the final analysis of the EMT and the case study of Nigeria as an origin country of migration and Greece in regard to the attitudes towards the arrival of migrants and asylum seekers. The document also reflects on the Twitter data and provides some future considerations regarding the employment of the EMT.

IV. Types of migration in the EU context: terminology and dataset limitations

Having arrived at the final stages of development of the EMT, it is necessary to clarify how ITFLOWS has defined certain terms across its deliverables and in the very tool. All the terminology employed by the project is grounded in its Ethical Framework, which is couched within wider European and international human rights frameworks.

First, a migrant is understood as anyone who moves from their country to another, whatever the reasons. There is no universally accepted definition of a “migrant” in international law, but according to the International Convention on the Protection of the Rights of Migrant Workers and Members of their Families (Convention on Migrants), a migrant worker is a person who is to be engaged, is engaged or has been engaged in a remunerated activity in a State of which he or she is not a national. In the absence of a definition in international and EU law, the term “migrant” is a generic socio-legal concept.

In its widest scope and depending on the source, the word “migrant” is also used to refer to asylum seekers and refugees, but ITFLOWS stresses out that these are separated categories regulated by other legal instruments that receive specific coverage in international law. Under the current and accepted interpretations of such law (for example, the Refugee Convention), somebody is considered a refugee as soon as he or she meet the pre-established criteria, regardless of whether the person has been formally recognised as a refugee or not. The idea is that a person does not become a refugee because of recognition, but rather is recognised because he or she is in fact a refugee. Following this logic, ITFLOWS understands “recognised refugees” as individuals who have been formally granted refugee status, and “non-recognised refugees” as anyone who meets the eligibility criteria but have not applied or applied and have not yet been granted asylum by a State.

Finally, the project also made it clear to avoid use the term “illegal migrant”. Regular migrants are individuals who have legal permission, normally through a visa or a residence permit, to stay in the EU. Despite the fact that the term “irregular” to refer to migrants who fall out of this category is also contested, within the context of ITFLOWS it is used to refer to undocumented migrants.

Taking this glossary into account and the importance to respect it, as a result of the very nature of the EMT and its models, some limitations were encountered on the selected datasets during the development process. First, the LSM has not been able to provide accurate predictions for all migration flows arriving in the EU due to the absence of precise and regularly updated datasets for migration. And the problem has been particularly acute regarding irregular migration. Because of this, predictions are only on international protection applications using historic data from Eurostat, and the EMT offers predictions on asylum seekers and unrecognised refugees that would apply for international protection along the subsequent month.

Second, the LSM is based on a Topic Modelling (LDA) algorithm by monitoring the national press. Precisely as a result of this, there were certain limitations due to the asymmetries in different countries regarding the level of freedom of speech and press, and the censorship imposed by national governments. On top of that, available datasets are not always regularly updated or sometimes provide updates for some countries whilst not doing so with others.

And, third, the LSM works on predetermined datasets and is limited to working based on their categories. For instance, and as members of the Users Board have pointed out more than once to ITFLOWS, the model cannot work with the category of “unaccompanied minors”, and this is because it does not exist in the original dataset that the model is trained with.

V. Case study one: Nigeria to Germany

a. Analysing the Nigeria conflict through the EMT

Nigeria is the most populous country in Africa, and the sixth most populous country in the World with a population over 230 million. Moreover, Nigeria is the largest economy in Africa, with a similar size to Austria or Israel. These data together also depict a serious problem: GDP per capita (PPP) is relatively low, lower than in Palestine or Bangladesh. Nigeria's external trade income comes mostly (over 90%) from petroleum related products which are distributed and reinvested in the country rather unevenly.

Nigeria's population is divided in many ways. For a long time in history, three administrative protectorates were established under British rule. The northern region is mostly populated by the Hausa, the West by the Yoruba, and the East by the Igbo. There are more than 300 additional ethnic groups (e.g., Fulani, Tiv, Kanuri) living in Nigeria and they speak around 500 different languages. The official language of Nigeria is English, but there is also use of Hausa, Yoruba, and Igbo languages. Moreover, there is a wide variety of religious practices. Nigeria gives home to the fifth largest community of Muslims and sixth largest community of Christians in the World. Economic division is also present and apparent. Historically the southern region was relying and thrived on international trade while the North had no direct access to markets. This imbalance is still affecting the population: people who live near the southern coasts are wealthier than those who live on the northern, or north-eastern regions.

The agricultural landscape is also changing, and the climate change is affecting Nigeria severely. Northern regions are arid where the desert is expanding. The middle section of Nigeria is a savannah, and the most southern regions are covered with rainforests. On the North-East, at the borders with Chad and Niger, the largest lake located there, called Lake Chad, is drying out. In the 1500s it was as large as Great Britain, but by 1970s it had the size of a third of Ireland, and today it is not larger than the urban area near Barcelona.

Amid these circumstances, there have been several armed conflicts and coups since Nigeria gained full independence in 1960. The last of them was the foundation and aggressive expansion of Boko Haram, and recently the rise of Islamic State's West Africa Province. The conflicts, that sometimes involve these non-state violent actors, are not just driven simply by economic or religious grounds, or a resistance to the ever-increasing corruption, but there is also tension between herders and farmers for food and water.

There is an estimated total number of 3,228,000 displacement movements from Nigeria according to IDPS, 200-400 thousand just in 2022. The most affected areas are Maiduguri, Kano, Katsina, and Sokoto. Maiduguri is the heart of the Boko Haram insurgency, and the most attacked city by them. The city is host to refugee camps and is a town of close to 800,000 people, according to the most recent population estimation.

b. Asylum seekers from Nigeria to Germany

In Germany, the vast majority of asylum applications, 201 out of 246, concern children.

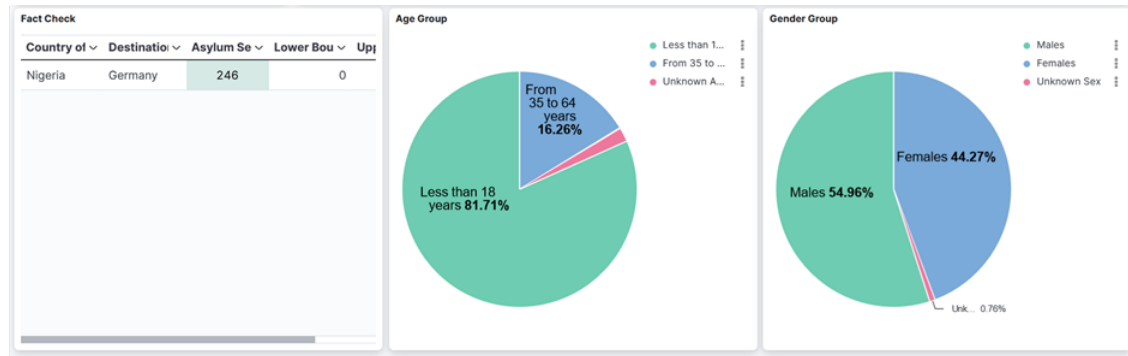


Figure 1: LSM forecasts from Nigeria to Germany

It is possible to make estimations and predictions for migration from Nigeria as the country of origin to Germany as the destination. The case of asylum seekers from Nigeria to Germany is quite accurate in terms of forecasting. As shown in Figure 2 and Figure 3 below, the point forecasts are accurate and effectively reflect the various fluctuations of the dependent variable under study. Even though the intervals are accurate, they cover a wide range. Nearing the end of the project and as of the time this report is written the case of Germany is one of the best performing for predictions of asylum seekers to the EU. As shown in Figure 3, both the data point forecasts (dark blue) as well as the confidence intervals (light blue) explain the variance of asylum applications recorded by Eurostat (orange) excellently.

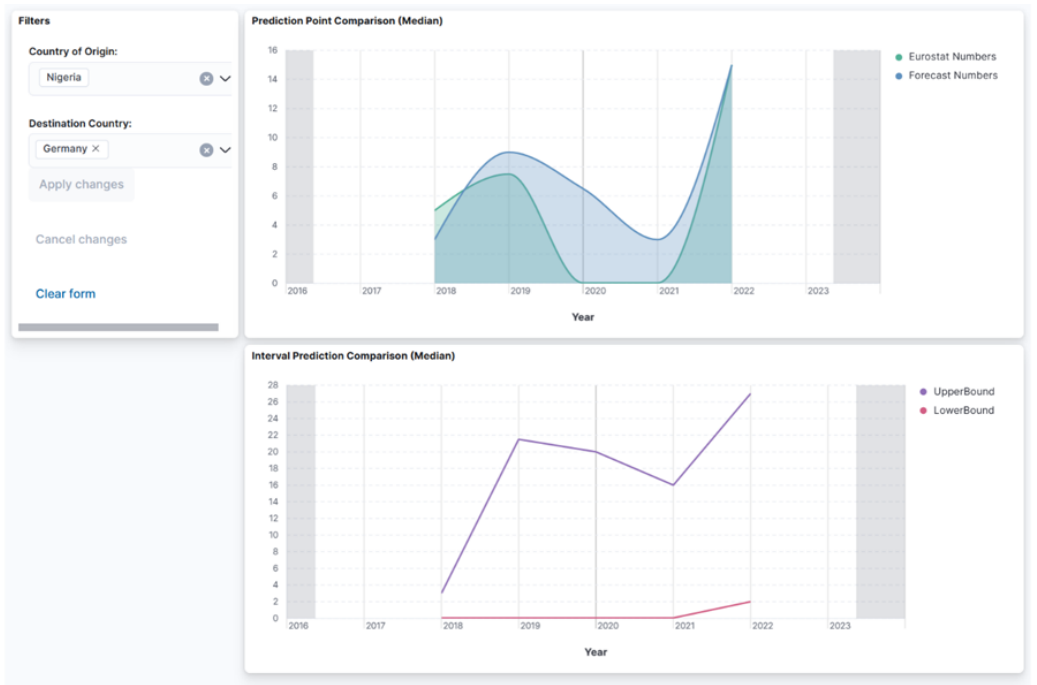


Figure 2: Comparison chart and upper-lower bounds chart of Nigeria to Germany forecasts

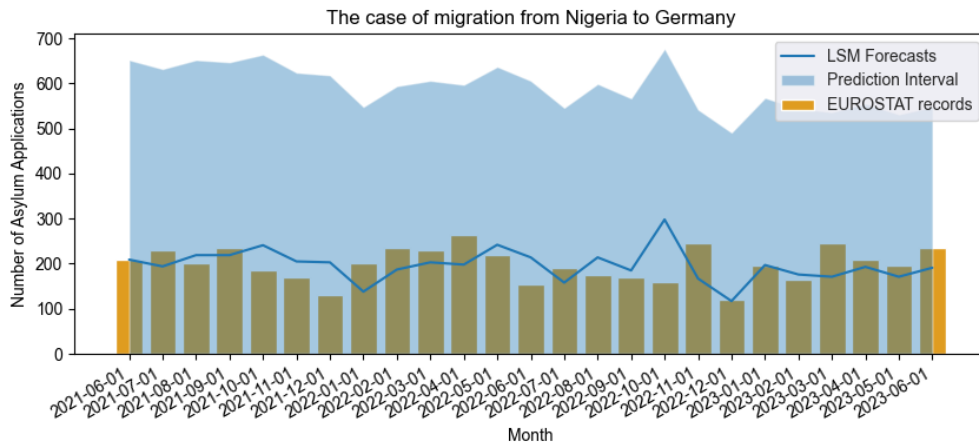


Figure 3: Forecasts and comparison bar chart for Nigeria to Germany

VI. Case study two: attitudes towards migration in Greece

The Twitter analysis shows a map of EU Member States and the sentiment in relation to migration.

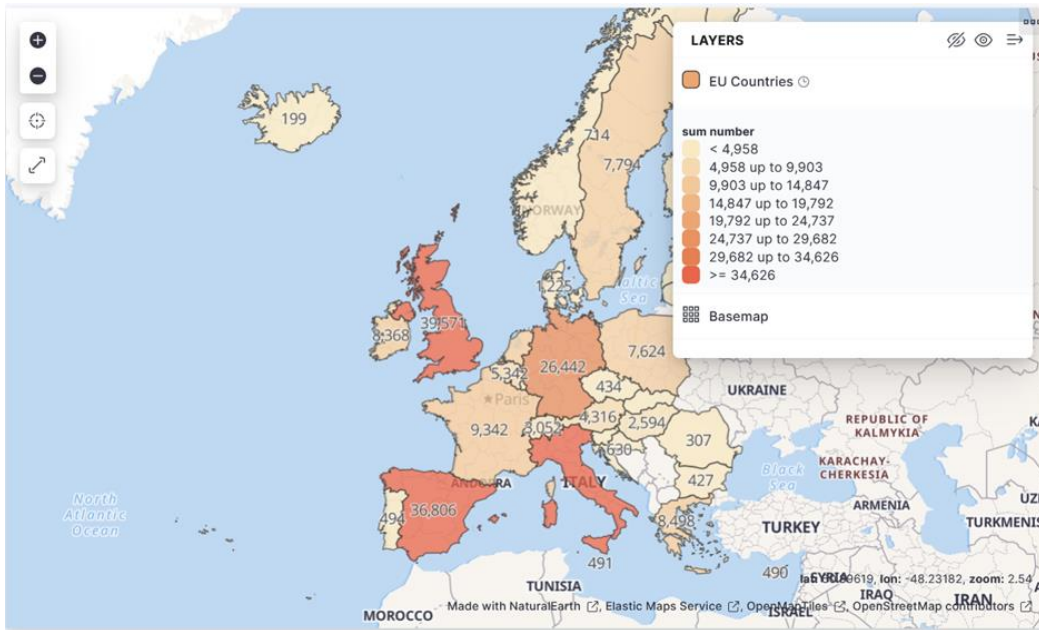


Figure 4: Percentage of tweets related to migration in EU countries

Using Greece as country of interest, it is possible to analyse the distribution of negative, positive, or neutral attitudes in migration-related Twitter messages.

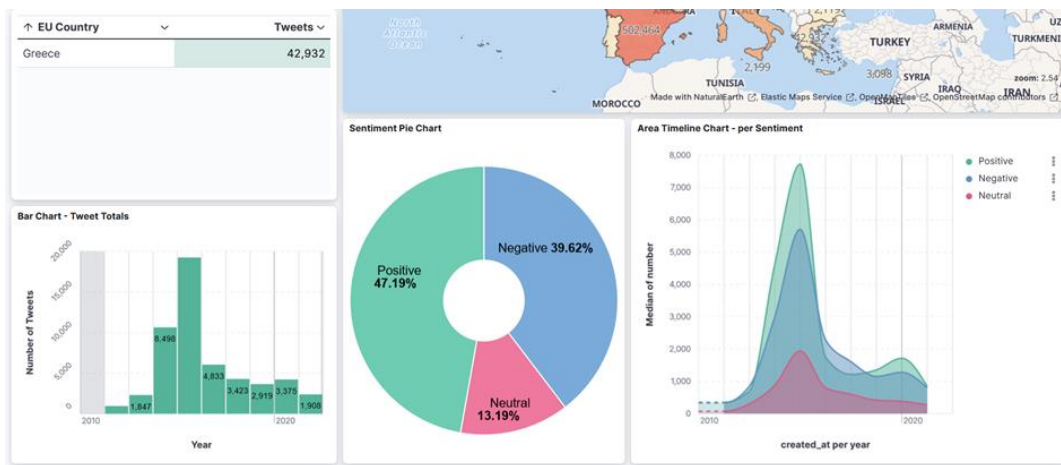


Figure 5: Charts for migration related tweets in Greece

There are two peaks in the number of tweets about migration over time, in 2015 and 2018. It can be observed that the negative sentiments have an all-time-low ratio within the total communications in 2015, and they are culminating in 2017. A gradual improvement is observed after that.

And, over time, there is a fluctuation in the distribution of negative and positive sentiments.

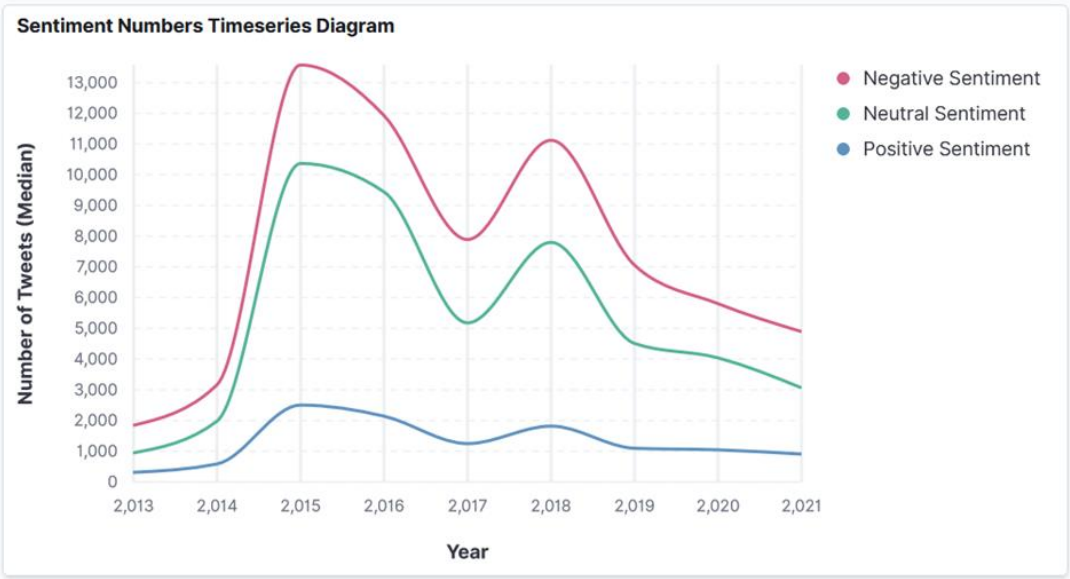


Figure 6: Sentiments trend over the years for tweets in Greece

VII. Last developments of the Twitter model

The Twitter model has experienced numerous modifications that make it incapable to work with new data right now. This part of the report details the substantial alterations that have taken place in the Twitter API, and the subsequent effect on the sentiment analysis model of the EMT. Unfortunately, due to these broad adjustments and the associated price changes to employ the new API service, the model can no longer perform correctly and be updated.

One noteworthy change is how the Twitter API now presents crawled tweets. The API has gone through significant corrections, resulting in changes to, mainly, the content of the data collected. Twitter's translation and language identification processes have shifted considerably. The developers responsible for updating and debugging the model have detected that Twitter has altered the way it groups and recognizes languages. Unfamiliar values in the pipeline's necessary fields cause a cascade of issues that affect the pipeline's diverse phases, including topic modelling, and eventually the final functionality of sentiment analysis and attitudes towards migration. The unfamiliar values in the fields lead to the pipeline's efficient data flow to break down. The presence of these unknown values can thus lead to partial or inconsistent data representations when performing topic modelling. This makes it more challenging to correctly detect and extract significant topics from the corpus of tweets. Consequently, the actual tweets that pass the cleaning and topic modelling process are far too few in relation to before. This shift has translated into considerable difficulties to extract tweets accurately and efficiently to perform the analysis of sentiments and attitudes of citizens of destination countries. By the end of the sentiment analysis, the results are about one third of the original model results previous to the Twitter API update.

As previously stated, another alteration is the financial cost of the Twitter API access. New Twitter regulations have turned the formerly free service to a paid one. All the platform's accounts have been prompted to pay for the API services and the charge has risen notably. The increased price is not feasible financially since the model would entail multiple calls per month to the Twitter API and the pricing model of the new Twitter API augments the cost accordingly to the calls made on the API. Academic accounts stay on free as of now, but the pricing model specifies these accounts as TBD (To Be Decided), which means that the plan is to price for academic accounts and usage as well in the immediate future.

As recent news detail, Twitter has inactivated all academic research accounts including the accounts the EMT developers were employing. After plenty of debate within the Twitter developer community forums and in related online articles, everything seems to indicate that Twitter has completely disabled all the academic accounts. Also, the official developer portal web page for academic accounts suggests new users to subscribe to the Twitter API, which has significant monthly costs.

Considering these circumstances, the model can no longer be updated to the newest data in the EMT dashboards and, as a consequence, the developers have decided to keep the past data that range from May of 2013 to December of 2021 while they are still working on the model.

VIII. Future Considerations

The Users Board provided some feedback regarding what type of additional data could benefit the EMT and potential users. Mainly, three suggestions were made. First, the number of unaccompanied and separated minors. This is a particularly vulnerable collective and special mechanisms and procedures should be put in place prior to their reception to address their needs. Second, family composition, to prepare spaces and resources for specific groups of people that should remain together and that work as emotional support for each other. And, third, it was argued that it would be helpful also to have information on the skills and level of education of migrants. As difficult to obtain as this sounds, it was argued that it would help design and promote mobility programmes for migrants and asylum seekers, and thus benefit or facilitate their labour inclusion. This would be also a great opportunity to match locations within Member States where there is a shrinking workforce with the potential skills that migrants and asylum seekers bring with them. And fourth, some more information on the situation and conditions in the countries of transit along the different migratory routes.

Other members of the Users Board addressed the issue of how the overall EMT design and appearance could be improved to make it more valuable to their organisations. Most end users shared that the design and appearance is friendly to the user and that it is not complicated to work on it, that it is well designed and presents a user-friendly interface that allows to read the results in a visual and graphical way without previous technical knowledge on these types of tools. As a result, the widespread answer was not having additional suggestions to put forward at that moment. Those who mentioned a couple of possible improvements signalled introducing more elements on the visual impact, such as small icons that could reduce even more the times of reading and interpreting the data shown, first, or more detailed information on the countries of origin and transit, second.

Finally, there were a few members of the Users Board who felt that additional functionalities could be added to either the EMT or another tool of similar nature. For example, it could be beneficial to be able to access data on the number of referrals for forced returns in different Member States, with disaggregation in age, gender or country of origin. It was argued that it could be significantly useful to have, since now more than ever there is an effort by civil society organizations and certain media outlets to give visibility and cover the topic of the violation of human rights of migrants and asylum seekers in the external and internal borders of the European Union. And another suggestion was to have a more refined insight into the conditions in which migrants and asylum seekers arrive at reception countries would be to consider the effect of transit countries both in the migratory routes and in the personal experiences of individuals throughout what often is a multi-staged journey with multiple border-crossings.