Analyzing Global Human Migration Patterns: A Novel Application Using Advanced Data Analysis Techniques

Dharani Kumar, Aryan Rohilla, Ekansh, Avinash Barala, Lakshay Bansal

March 10, 2024

Abstract

This proposal outlines a research project to develop a comprehensive application to analyze global human migration patterns, leveraging datasets from the United Nations and similar agencies. The project emphasizes on the importance of understanding migration trends in a rapidly changing world driven by climate change, economic instability, and geopolitical shifts. In the course of this research, the team aims at identifying and using state-of-the-art Information Retrieval and Machine Learning techniques to achieve the goal of a Natural Language application which has a vast knowledge in this field.

Keywords: Information Retrieval, Human Migration, Random Forest Analysis, NLP applications

1 Introduction

The phenomenon of global human migration presents complex patterns that are influenced by a multitude of factors, including environmental changes, socio-political dynamics, and economic opportunities. Accurate analysis and understanding of these patterns are crucial for effective policy making and humanitarian efforts. [1]. This project would be an effort towards utilizing large datasets on international migration stock, influenced by a variety of underlying factors, to build an application enabling users to gain insights for personal growth, in addition to assisting the society as a whole.

2 Problem Statement

Understanding global migration patterns is essential for addressing challenges related to refugee crises, planning for economic development, and ensuring the well-being of millions of migrants worldwide. In addition to solving these problems, the outcome of this research could also benefit individuals who are immigrating and communities as a whole [2]. A tool that provides detailed insights into these patterns can aid in making informed decisions at both the policy and individual levels.

Existing models and applications offer limited insights into migration patterns, often constrained by static datasets and lacking user interactivity. There is a need for a **dynamic tool that not only analyzes but also provides peronalized inputs on migration trends in real-time**, allowing users to **explore scenarios based on various presets and fine-tuning instructions**. *Our dataset on news articles from all the countries in the world would aid us in providing individuals with customized responses to their objectives.*

3 Literature Survey

One of the earliest publications in the field of Human Migration is the work of Walter Adams [2]. It covers various factors leading to the refugee crisis on a global scale and issues surrounding it. One key aspect highlighted in the document is the significance of settlement schemes and international assistance in addressing the refugee problem. It underscores the importance of focusing on the numbers of refugees and the quality of support and integration efforts provided to them. The document points out that the refugee issue is not solely quantitative but rather qualitative, requiring a more holistic approach to ensure successful settlement and reintegration of displaced individuals.

The study by Black et al [6] and the Foresight report on Migration and Global Environmental Change depict migration through a relationship between human capacity, vulnerability to environmental change, and various psychosocial and socioeconomic factors. Marotzke et al [4] and Lilleør and van den Broeck [5] explored the poverty-climate-migration nexus in a laboratory setting, focusing on economic factors. The analysis shows that net-negative migration is concentrated in areas with high environmental stress and medium-low to medium-high adaptive capacity. Income, drought risk, and education are primary factors in explaining net-negative migration in areas with high environmental stress. The study highlights the importance of integrating societal dimensions in the quantitative analysis of the environment-migration nexus.

In a report published by Nature, Venla et al [7] have created a global dataset of annual net migration between 2000 and 2019, followed by a study highlighting the magnitude and impact of net migration, differentiating between rural and urban migration and associating migration patterns with socioeconomic factors and climate conditions. The study reveals global trends in net migration, with certain regions experiencing positive net migration due to urbanization and economic opportunities, while others face negative migration influenced by factors like conflict and economic downturns. The report also covers the importance of sub-national analysis for policy design and international cooperation, emphasizing the complex dynamics of migration that vary significantly across different scales and regions.

One of the pioneering works in the field of Human Migration is the paper titled "Modelling the Age and Sex Profiles of Net International Migration" by James Raymer, Qing Guan, Tianyu Shen, Sara Hertog, and Patrick Gerland [8]. This paper addresses the critical need for accurate estimation and projection of populations by focusing on refining the modeling of age and sex profiles of net international migration. It introduces a methodology to enhance the estimation of age and sex profiles of net international migration, crucial for the United Nations Department of Economic and Social

Affairs' Population Division's World Population Prospects. By distinguishing between immigration and emigration flows, the method improves accuracy, even in the absence of direct migration data. Empirical validation using data from Sweden and the Republic of Korea demonstrates its effectiveness. The model's flexibility allows for application to countries with limited migration data. Overall, the paper highlights the importance of accurate migration data and offers a promising solution to enhance global population estimates.

Another document titled "World Population Prospects 2022: Data Sources" by the United Nations Department of Economic and Social Affairs, Population Division, offers a meticulous account of methodologies and data sources employed in estimating global population statistics. It commences with an overview of the Population Division's pivotal role in analyzing population data to inform UN policies. Geographical and income classifications are elucidated, including distinctions between developed and developing regions, and special groups such as LDCs and SIDS. Each country's section outlines population estimates adjusted for census data, surveys, and official estimates, addressing potential undercounts and age heaping. Methodologies for fertility, child mortality, overall mortality, and migration estimates are delineated, drawing from diverse sources including demographic surveys, birth registrations, and refugee statistics. This comprehensive approach ensures the accuracy and reliability of population projections, supporting global and national planning endeavors.

Various observations have been made in this literature survey, including a **detailed analysis of global migration caused by various environmental, socioeconomic, and political factors, accompanied by appropriate measures to make this work available to the masses**. *The additional gap identified in the currently published work is the use of multivariate data analysis, that we aim to bridge using the dataset of news articles which we have created.*

4 Methodology

We would be dealing with the available datasets, like UN's Global Migration Database, World Population Prospects, etc., using suitable Information Retrieval techniques. Furthermore, we plan to use a combination of machine learning algorithms [3], geospatial analysis, and natural language processing (NLP) techniques to analyze the retrieved migration data. Additionally, we will develop a user-friendly interface capable of translating natural language queries into data queries, facilitating personalized exploration of migration trends for users.

5 Evaluation

The effectiveness of our tool will be evaluated through user studies, comparing the insights generated by our application with existing data and trends. We will also assess the accuracy of our algorithms through back-testing with historical data.

6 Contributions

The workload, over the course of this semester, would be distributed as follows:

Dharani: Literature Review, Language Model, Information Retrieval

Aryan: Literature Review, Database Creation

Avinash: Language Model, Information Retrieval

Ekansh: Language Model, Front-end/API

Lakshay: Language Model, Front-end/API

References

[1] Naval, Naveen and Rizun, Nina and Ciesielska, Magdalena and Alexopoulos, Charalampos and Saxena, Stuti, Identifying the Push and Pull Factors for Migration Patterns Vis-À-Vis Smart Cities' Context (April 27, 2023). Available at SSRN: https://ssrn.com/abstract=4430561 or http://dx.doi.org/10.2139/ssrn.4430561

[2] Adams, W. (1939). Extent and Nature of the World Refugee Problem. The Annals of the American Academy of Political and Social Science, 203, 26–36. http://www.jstor.org/stable/1021882

[3] Venla Niva et al 2021 Environ. Res. Lett. 16 114019

[4] Marotzke, J., Semmann, D. & Milinski, M. The economic interaction between climate change mitigation, climate migration and poverty. Nat. Clim. Chang. 10, 518–525 (2020). https://doi.org/10.1038/s41558-020-0783-3

[5] Helene Bie Lilleør, Katleen Van den Broeck, Economic drivers of migration and climate change in LDCs, Global Environmental Change, Volume 21, Supplement 1, 2011, Pages S70-S81, ISSN 0959-3780

[6] Black RE, Victora CG, Walker SP, Bhutta ZA, Christian P, de Onis M, Ezzati M, Grantham-McGregor S, Katz J, Martorell R, Uauy R; Maternal and Child Nutrition Study Group. Maternal and child undernutrition and overweight in low-income and middle-income countries. Lancet. 2013 Aug 3;382(9890):427-451. doi: 10.1016/S0140-6736(13)60937-X. Epub 2013 Jun 6. Erratum in: Lancet. 2013. 2013 Aug 3;382(9890):396. PMID: 23746772.

[7] Niva, V., Horton, A., Virkki, V. et al. World's human migration patterns in 2000–2019 unveiled by high-resolution data. Nat Hum Behav 7, 2023–2037 (2023). https://doi.org/10.1038/s41562-023-01689-4

[8] Raymer, J., Guan, Q., Shen, T., Hertog, S., Gerland, P. (Year). Modelling the Age and Sex Profiles of Net International Migration. Journal Name, Volume(Issue), Pages. http://bit.ly/3wPS3qo