Scientific curriculum

DILETTA GOGLIA

1. Personal information

- Born in Civitavecchia (RM), on September 15, 1996
- Nationality: Italian.
- Domiciled and resident in Pisa, Italy.
- Phone: +39 340 320 4546
- Institutional e-mail: d.goglia@studenti.unipi.it
- Personal e-mail: diletta.goglia@gmail.com
- Portfolio: https://dilettagoglia.netlify.app
- Google Scholar profile
- LinkedIn profile
- Medium profile
- GitHub profile

2. Qualifications

- Research fellow (borsista di ricerca) in Machine Learning e Data Science for HumMingBird EU Horizon 2020 project, from September 2021 to March 2022.
  Research topic: "Big data for migration studies". Understand and nowcast human migration by integrating traditional and non-traditional sources (such as social Big Data).

- Enrolled in A.Y. 2019/20 and currently student of the Master of Science course in Computer Science, Artificial Intelligence curriculum, at University of Pisa.

- Bachelor's Degree in Digital Humanities (Informatica Umanistica) with the highest marks, at University of Pisa in A.Y. 2017/18, with final grade 110/110.
  Title of the bachelor thesis: *Linked Open Data per il Cultural Heritage italiano: conversione e pubblicazione nel Semantic Web del Fondo antico del Comune di Novi Ligure*. Supervisor: Andrea Marchetti. Co-supervisors: Angelica Lo Duca, Enrica Salvatori, Vittore Casarosa.

- Scientific High School concluded in 2015 at "Liceo Scientifico Galileo Galilei" in Civitavecchia, with final grade 95/100.

3. Grants and awards

- *March 2022:* selected speaker at *Measuring Migration: How? When? Why?* conference, University of Oxford (abstract proposal winners at the call for submission).

- *March 2022:* selected speaker at *Digitization of Migration Research Methods: Promises and Pitfalls* workshop, University of Warsaw (abstract proposal winners at the call for submission).
  .

- *December 2021:* student on mission abroad attending the Consortium Meeting of HumMingBird EU Horizon 2020 project at University of Salamanca.

- *July 2021*: candidate winner of the call for the allocation of a research grant in "Machine learning and Data Science to study and predict human migration by integrating data from social networks and traditional sources", at HumMingBird EU Horizon 2020 project.

- *February 2021* and *October 2020*: candidate winner of two calls (1st and 2nd semester) for the allocation of part-time collaborations aimed at supporting and mentoring students of the Computer Science courses at the Department of Computer Science of the University of Pisa.

- *March 2020*: candidate winner of the ranking for merit for the assignment of a part-time collaboration at the Computer System of the Medical School, University of Pisa.

- *March 2018*: candidate winner of the ranking for merit for the assignment of a part-time collaboration for the e-learning project (assistance to teachers and technical-administrative staff), at University of Pisa.

4. Scientific activity

Description of research activity

Use of non-traditional data sources to nowcast migration trends through Artificial Intelligence technologies.

In the last years, many researchers have proposed the use of non-traditional data sources to study migration, including so-called social Big Data such as online social networks. Many types of data exist, still very scattered and heterogeneous: in the variety of this context, integration is not straightforward. In general, these works have been performed by Computer Scientists, since they require special skills to meaningfully embed and combine traditional and novel datasets.

Our work describes the use of alternative types of data, a new multi-feature dataset and a new indicator that could significantly contribute to the study of migration and to forecast emerging trends through the use of Artificial Intelligence technologies.

This approach is intended to find an alternative methodology to ultimately answer open questions about the human mobility framework (i.e. nowcasting flows and stocks, studying integration of multiple sources and knowledge, and investigating migration drivers).

For this purpose we provide the **Multi-aspect Integrated Migration Indicators (MIMI) dataset** that we built by integrating official data about bidirectional human mobility (i.e. traditional flow and stock data, gathered from sources like Eurostat and United Nations) with multidisciplinary features and original indicators, including the **Facebook Social Connectedness Index (SCI)**.

This latter measures the relative probability that two individuals across two countries are friends with each other on Facebook: it guarantees an ethical and anonymized collection of information on users and their friendships. The inclusion of this indicator in the dataset enables it to be exploited as a non-traditional way to describe, understand and nowcast international migration.

Then we introduced a new measure, likewise included in the dataset, consisting in a formulation built to mirror the mathematical structure of the SCI: the **Bidirectional Migration Probability (BMP) index**, which takes into account both the inflows and outflows shared by two countries, and measures the relative

probability of a person to be a migrant from country *i* to *j* and vice versa. BMP indicator allows to portray and predict bilateral migration trends relying on the intensity of social networking, since it shows significant correlations with SCI.

We believe SCI and our integrated dataset can be employed to study migration drivers, along with other traditionally used measures (cultural differences indicators, GDP, distance between countries, etc), through **Machine Learning techniques**, so that to link and combine the statistical and computational study of migration phenomenon with **interdisciplinary perspectives** (geographical, demographic, economic, sociological, anthropological). Indeed, the knowledge combined in the dataset is designed to develop a ML model able to extract novel information, analyse patterns and, from the strength of Facebook connectivity between countries, nowcast and forecast both present and future bilateral migration trends.

The long-term perspective of this research is to build trustworthy and reliable predictions for future changing by using new ways of measuring and characterizing international migration, and by using advanced technologies such as Artificial Neural Networks.

Presentations in conferences and workshops (including scheduled)

Conferences and workshops attended as student on official mission abroad on behalf of the University of Pisa, presenting the results of my research activity.

- Conference "Measuring Migration: How? When? Why?", organized by the University of Oxford's Migration and Mobility Network (MMN) and Nuffield College, Oxford (UK), June 9-10, 2022.
- Workshop "Digitization of Migration Research Methods: Promises and Pitfalls", Centre of Migration Research, University of Warsaw, Warsaw (Poland), May 30, 2022.
- "Consortium Meeting - HumMingBird EU Horizon 2020 project", Salamanca (Spain), December 9-10, 2021.

Publications

- Goglia, D. (2022). *Multi-aspect Integrated Migration Indicators (MIMI) dataset*. Zenodo. DOI: doi.org/10.5281/zenodo.6360651

Work in progress

- Goglia, D.; Pollacci, L.; Sîrbu, A. (2022) *Dataset of Multi-aspect Integrated Migration Indicators.* MDPI Data.
- Goglia, D.; Jisu, K.; Pollacci, L.; Spyridon, S.; Iacus, S.; Sîrbu, A. (2022) *International mobility between UK and Europe around Brexit:a data-driven study.*

Autonomous works

- Goglia, D. (2020). *Amplifying bias, automating racism. Ethically framing the issue of algorithmic discrimination.* Academia.edu. https://bit.ly/3uwVu1x

- Goglia, D. (2019). *Linked Open Data per il Cultural Heritage italiano: conversione e pubblicazione nel Semantic Web del Fondo antico del Comune di Novi Ligure*. Bachelor's Thesis. Academia.edu. https://bit.ly/3M1X4OC

- Serie of articles about Computer Science, wrote and published for Analytics Vidhya available on my Medium.com profile: https://medium.com/@d.goglia

Research interests

- Computational Social Science
- Migrations studies and prediction
- Multidisciplinary applications of Computer Science
- AI ethics

5. MSc exams

The Italian unit of measurement for complexity and time required for each exam is expressed in CFU (Crediti Formativi Universitari). The usual range goes from 3 to 12 CFU.

- Intelligent Systems for Pattern Recognition, 6 CFU (prof. D. Bacciu): **30 cum laude**
- Smart Applications, 9 CFU (prof. V. Gervasi): **30/30**
- Data Mining, 9 CFU (prof.ssa A. Monreale): **30/30**
- Robotics, 6 CFU (prof.ssa C. Laschi): **30/30**
- Social and Ethical Issues in IT, 6 CFU (prof. V. Gervasi): **30/30**
- Machine Learning, 9 CFU (prof. A. Micheli): **28/30**
- Semantic web, 6 CFU (prof. V. Bartalesi Lenzi): **28/30**
- Artificial Intelligence Fundamentals, 6 CFU (prof.ssa M. Simi): **26/30**

6. University projects

List of the most recent projects, related to the current and previous academic years.

- "Artificial Neural Network from scratch". *Machine Learning course, Professor Alessio Micheli.*
  Implementation of a neural network capable of solving, through stochastic gradient descent, classification and regression tasks.

- "Among AI's" videogame. *Smart Applications course, Professor Vincenzo Gervasi.*
  Class of software development, following *Agile* method, for creating a video game as a smart application. It includes a recursive neural network, elements of fuzzy logic, and the implementation of a smart behavior obtained through *cellular automata* and rule-based programming for player-behavior.

- "Customer Analysis". *Data Mining course, Professor Anna Monreale.*
  Analysis of customer buying behaviour and trends through Data Mining techniques, Clustering algorithms and Machine Learning.

- "Gated Recurrent Neural Network for timeseries prediction". *Intelligent Systems for Pattern Recognition course, Professor Davide Bacciu.*
  Deep Learning model made with a Long-Short Term Memory neural network to make future predictions on energy consumption.

- "Hidden Markov Model with Gaussian emissions for regime detection". *Intelligent Systems for Pattern Recognition course, Professor Davide Bacciu.*
  Python analysis of timeseries data through a Graphical Model in order to perform a clustering and to extract intervals of consumption.

- "Image processing with SIFT algorithm". *Intelligent Systems for Pattern Recognition course, Professor Davide Bacciu.*
  Python analysis using SIFT algorithm for extraction and description of salient features of a dataset of images.

7. Working activity at the Athenaeum

Related to "Grants and Awards" section above.

- [*Oct 2020 – May 2021*] Teaching assistant of JavaScript & Typescript at the course of "Lab I" to a class of students at the first year of the bachelor's degree, at Department of Computer Science, University of Pisa.

- [*Nov 2020 – Jul 2021*] Tech support, at Computer Science area of Medical School, Santa Chiara Hospital, Pisa. Creation and development of an E-R model for a knowledge base of medicinal plants @Pharmacy Department, University of Pisa, with Microsoft Access database. Creation and development of a website for usability and accessibility of the KB with Microsoft SharePoint.

- *[Nov 2018 – Sep 2019]* Technical assistance to university staff during events and conferences organization, for "MediaLab e-learning project", at University of Pisa.

Reference letters available on request.

Pisa, Italy

April 12, 2022