



Joint ICTP-IAEA Advanced School/Workshop on Machine Learning in Citizen Science | (SMR 3824)

27 Feb 2023 - 03 Mar 2023
ICTP, Trieste, Italy

01 - ALI Mohammed Endris

Data Policy and Governance for Public Service.

02 - ATMAJA Bagus Tris

An acoustic-based emergency condition detection for elderly people living alone using keyword spotting

03 - BHOWMICK Ranjan Amiya

Role of Data Science and Machine Learning in Managing Plant Invasion Under Climate Change: A Case Study with ILORA Database in India

04 - BOURENANE Abdelghani

Natural Language Processing for Environmental Citizen-Generated Data to Study Climate Change Risk and Design Adaptation Techniques in Agricultural Areas, Grape Cultivation as a Case Study

05 - CALDERS Stijn

the Radio Meteor Zoo, a Zooniverse project

06 - MALATIM Zakariya Tariq

Citizen research about the nuclear science research in Libya

07 - OLIVERA MARRERO Orestes José

Machine Learning in progress: Python and Sklearn to improve nuclear training.

08 - OPOKU-NTIM Irene

Home Radon Testing, Environmental Health Literacy, and Efficacy in the Greater Accra Region of Ghana: A Citizen Science Approach

09 - PRAJAPATI Jagdish Jyoti

Projected Distribution of 218 Invasive Plant Species in India under climate change: Identification of Hotspots and Management Implications

10 - RODRIGUES MIRANDA Caetano

Mapping and engaging general public towards a low carbon society: introducing machine learning on citizen science programs

11 - SAAD Mushfiqur Rahman

Leveraging the Power of Citizen Science and Machine Learning to Transform Low-Income Communities

12 - VOSKANYAN Marine

Citizen Science as a Tool to Monitor the Water Pollution. A Case Study of Lake Sevan, Armenia

The Importance of Data Policy and Governance for a Citizens Project using ML in Ethiopia

Endris M. Ali¹, Getahun T. Metaferia²

¹*Adama Science and Technology University, Ethiopia*

²*Adama Science and Technology University, Ethiopia*

Recently, Machine Learning(ML) based service delivery plays a significant role to enhance business decision-making for different sectors. Naturally, data is the backbone of this approach, and its model is data-intensive for training and learning for any business decision. Currently, large data are generated from different sectors in different formats in everyday business activities. However, the absence of a regulatory data collection framework and access policy awkward its impact on delivering smart public service[1]. Different actors such as AU [1], Ethiopian Govt [2], and UN [3] show the urgency to have a formalized legal framework for collecting, organizing, preserving, accessing, and presenting a data set for business problem demonstrated. In support of that, Ethiopian Public Health Institute(EPHI) tries to prepare a data policy guideline for data management and sharing[4]. Plus, data privacy and security is a core aspect of preparing data policy and management with respect to the nature of the application domain. The focus of this study is to present contextual facts and figures on data management policy for the Machine Learning model to enhance public services. The concern is Why an open data policy process and data sharing plan requires in citizens' projects. We present participatory data collection models that need the full involvement of citizens. More importantly, the heterogeneous nature of data is considered in terms of source, format, size, and privacy (personal and public) levels of data from the public. This case study tries to address opportunities, and challenges faced to design data management policy in the Ethiopian context with respect to digital transformation plans.

- [1] Amani Abou-Zeid. AU Data Policy Framework. Addis Ababa. **08**, 84 (2022).
- [2] Abiy Ahmed ALi. Report. Digital Ethiopia 2025 – a Strategy for Ethiopia's Inclusive Prosperity. **2**, 155 (2021).
- [3] Taffere Tesfachew. Report. Ethiopia's Drive to Advance Digital Transformation. **4.2**, 155 (2022).
- [4] EPHI Report. Guideline for Data Management and Sharing August Addis Ababa, Ethiopia. **2**, 37 (2019).

An acoustic-based emergency condition detection for elderly people living alone using keyword spotting

Bagus Tris Atmaja, Akira Sasou

National Institute of Advanced Industrial Science and Technology, Japan

The number of elderly people living alone in such countries is increasing, triggering the need to provide aid in an emergency condition. In such conditions for elderly people, it is difficult and needs time to call an emergency number (e.g., "911" in US or "110" in Japan). Keyword spotting can be utilized to sense this emergency condition. If the elderly people say (or shout) "help... help... ", the system will be triggered and call the emergency number (or dedicated line number). In this work, we simulate the detection of an emergency condition by keyword spotting using machine learning. We trained the keyword (i.e., "help") and other keywords (other words than "help") to the system, so the system be able to accurately detect the emergency condition.

Keyword spotting is an essential feature in modern hands-free voice control devices, where the user speaks a predefined keyword to "wake-up" (hence, also known as a wake or wake-up word) the device before speaking a complete command or query to the device [1]. This technology has been implemented successfully for Google Assistant, Amazon Alexa, Apple Siri, and Microsoft Cortana. Borrowing the success of these keyword spotting implementations, we propose a similar idea to detect the emergency condition of elderly people living alone.

Previous works have been done using a microphone array that tries to map the pattern of movement of elderly people based on footstep sounds [2]. Although this method is passively monitoring the condition of people, the need for a microphone array adds to the cost and maintenance. In this work, we propose to use a single microphone to detect an emergency condition. Although the work focus on elderly people, the proposed method can be applied to other people in emergency conditions.

- [1] S. Myer and V. S. Tomar, "Efficient keyword spotting using time delay neural networks," in Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, 2018, vol. 2018-September, pp. 1264–1268, doi: 10.21437/Interspeech.2018-1979.
- [2] A. Sasou, N. Odontselgel, and S. Matsuoka, "An acoustic-based tracking system for monitoring elderly people living alone," in ICT4AWE 2018 - Proceedings of the 4th International Conference on Information and Communication Technologies for Ageing Well and e-Health, 2018, vol. 2018-March, no. Ict4awe 2018, pp. 89–95, doi: 10.5220/0006664800890095.

Role of Data Science and Machine Learning in Managing Plant Invasion Under Climate Change: A Case Study with ILORA Database in India

**Amiya Ranjan Bhowmick^{1*}, Jyoti Jagdish Prajapati^{1,3}, Achyut Kumar Banerjee²,
Abhishek Mukherjee³**

¹*Department of Mathematics, Institute of Chemical Technology, Mumbai*

²*School of Life Sciences, Sun Yat-sen University, China*

³*Agricultural and Ecological Research Unit, Indian Statistical Institute, Kolkata*

**Presenting Author*

Biological invasions represent one of the major environmental challenges for the conservation of global biodiversity and the continuation of ecosystem services and have a serious impact on human and animal health and economic development (Simberloff et al., 2013). Pant et al. (2021) and Prajapati et al. (2022) have created a comprehensive database, named, Indian Alien Flora Information (ILORA) (<https://ilora2020.wixsite.com/ilora2020>) which contains data for 14 variables related to ecology, biogeography, introduction pathway, socio-economy, and distribution of 1747 alien vascular plant species from 22 national and global sources. The data set is expected to assist a wide range of stakeholders involved in India's scientific research, policy formulation, and decision-making related to Invasive Alien Plant Species (IAPS). In recent years, several modeling tools have been used to predict the consequences of climate change on the spatial distribution of animal species, weeds, and crops in India. However, most of the studies are confined to one or a very limited species. In this talk, we plan to discuss how the ILORA database will be useful in implementing policy regulations with the aid of data science and machine learning applications (Banerjee et al., 2021a). In particular, the identification of the important factors that are responsible for the establishment of the introduced plant species in the new region has been understood (Banerjee et al, 2021b). We shall discuss the potential future scope of this database and how this can aid in the computation of the probability of successful invasion by already established plants.

[1] Simberloff, D., Martin, J.-L., Genovesi, P., Maris, V., Wardle, D.A., Aronson, J., Courchamp, F., Galil, B., García-Berthou, E., Pascal, M., 2013. Impacts of biological invasions: what's what and the way forward. *Trends Ecol. Evol. (Amst.)* 28, 58–66.

[2] Banerjee, A. K., Khuroo, A. A., Dehnen-Schmutz, K., Pant, V., Patwardhan, C., Bhowmick, A. R. and Mukherjee, A. An integrated policy framework and plan of action to prevent and control plant invasions in India. *Environmental Science & Policy*, 124, 64-72, 2021.

[3] Pant, V., Patwardhan, C., Patil, K., Bhowmick, A. R., Mukherjee, A., & Banerjee, A. K. (2021). ILORA: A database of alien vascular flora of India. *Ecological Solutions and Evidence*, 2, e12105.

[4] Prajapati, J., Singh, A., Patil, K., Bhowmick, A.R., Mukherjee, A., Huang, Y. and Banerjee, A.K. (2022), An occurrence data set for invasive and naturalized alien plants in India. *Ecology*. Accepted Author Manuscript e3794.

[5] Banerjee, A. K., Prajapati, J., Bhowmick A. R., Huang, Y., Mukherjee, A. Different factors influence naturalization and invasion processes – A case study of Indian alien flora provides management insights. *Journal of Environmental Management*, 294, 113054, 2021.

Natural Language Processing for Environmental Citizen-Generated Data to Study Climate Change Risk and Design Adaptation Techniques in Agricultural Areas, Grape Cultivation as a Case Study

Abdelghani Bourenane¹ , Nassim Abdeldjallal Otmani¹

¹Institute of Electrical and Electronic Engineering, University of Boumerdes

Climate change is likely to directly impact food production across the globe, and World agriculture faces a serious decline within this century due to global warming[1]. The region of Baghlia, Boumerdes, Algeria most known for grape cultivation, is highly impacted by climate change effects including but not limited to: sudden temperature variations, and groundwater scarcity.

In this work, we engage local farmers in studying the effects of climate change on local grape cultivation as well as possible solutions to integrate a local adaptation policy, this is done through leading a series of focus groups and interviews with farmers, in addition to relevant government records and surveys, by then all data generated in the form of text is subjected to an NLP model, to figure out patterns, analyze belief and sentiment and suggest corresponding adaptation techniques.

The output of this work helps to understand the complex and dynamic interrelationships between water, and energy use in grape agriculture, using Machine Learning techniques and based on qualitative data generated by the community, the ML model suggests approaches to manage the region's limited resources sustainably, as well as balancing different resource user goals and interests, while maintaining the integrity of ecosystems.

This work is considered as a conceptual approach to using ML to better understand and systematically analyze the interactions between the natural environment and human activities. Embodying ML in the process helps to identify and manage trade-offs and to build synergies through responses, and working towards more coordinated management and use of water and energy resources across the region allowing for more integrated and cost-effective planning, decision-making, implementation, monitoring, and evaluation.

We conclude our work by reflecting on the merits and benefits of using NLP in our case study and in general for researching climate change adaptation policy..

[1] Anupama Mahato Junior Research Fellow, Guru GhasidasVishwavidyalaya, Koni, C.G , India, Climate Change and its Impact on Agriculture

the Radio Meteor Zoo, a Zooniverse project

The BRAMS network consists of one beacon in Dourbes and more than 40 receiving stations that observe permanently. Analyzing all observations manually is practically undoable. For the underdense meteors, the signature is simple and the automatic detection algorithms that we use, work quite well. However during meteor showers, more complex shapes appear from overdense meteors. In those cases, the human eye remains the best detector. So in August 2016 we have launched, in collaboration with Zooniverse, a citizen science project called the Radio Meteor Zoo. In this citizen science project, volunteers help us identify meteors in our observations by drawing a rectangle around everything they consider to be a meteor. Each spectrogram is seen by 10 different volunteers, and aggregated.

Citizen research about the nuclear science research in Libya

Tariq Zakariya Malatim¹

¹Libyan Atomic Energy Establishment

In this study, Tajoura Nuclear research reactor has been the objective in a citizen research, the public were participated in this experiment via social media (Facebook).

A post has been published in a general Libyan public group which contains more than one million person, the post is an article which contains general information about the research reactor of Libya, although all the information are exist in the Libyan atomic Energy Establishment site as well as in the IAEA Research Reactor Database. However, the style of post writing was been changed in order to reach the public in easy and understandable way, photos have been used to make the article more interesting, the title of the article was “ Know more about the Libyan Nuclear reactor”.

The main aim of the post is to determine the awareness of the public about the national nuclear research facility, furthermore to give a brief introduction about the facility and its utilizations, as well as to know how the public think about the nuclear reactors in general, and how we can use it effectively in research field.

The results were very surprising and far than I expected, the reaction with post was very strong, more than 1500 person react with post (likes), also there were about 395 comments and 94 post shares from the public and this reaction was very strong especially if you put in mind that it is a scientific article. Comments were from different classes and backgrounds, they were young, old, highly educated, and uneducated, persons in the study. In comments, a very interesting discussion was made, there were positive comments and negative comments as well, so it was a good opportunity to give and take, to share and learn. In addition, something very interesting happened which took the study to another direction, more than 10 very famous public pages with huge number of followers shared the post (without even asked them to do that) which was very exciting ,the reaction of the post in one of the pages had more than 3500 react, about 492 comment, and 130 shares. Furthermore the post was came as trend in Facebook, that even a very public Arab news channel made a promo for the reactor, also a famous Arab Facebook page wanted to have a press interview with me to more about the subject.

Machine Learning in progress: Python and Sklearn to improve nuclear training.

Olivera Marrero, Orestes¹

¹ *National Commission of Atomic Energy, Argentina.*

The usefulness of information for decision-making is a commonplace truth in organizations; however, in practice, building an integrated data center for hosting and using them is a great challenge. This is probably the case because it is not only a technological and security issue, but also because it involves cultural aspects. The state of the art in the National Commission of Atomic Energy, Argentina (CNEA) shows this in the actual context of recession and culture changing that's happening. The acquisition of servers and hardware is a problem that generally affects public organizations, of which the CNEA is a part, and inevitably makes work difficult.

One of the responsibilities of the Intellectual Capital Department is to provide the technologies and services to make visible the necessities of nuclear training that will represent the next researchers inside the institution. Before the year 2020 the state of the technologies and services to accomplish this were below the line of satisfaction in terms of modern data accessibility, compilation and sharing [1]. With the main idea of implementing machine learning services and data analysis to upgrade the scholarship's services that the Department is responsible for, a new approach was taken. The first step was to develop a new system for the postulations of aspirants to the scholarships. This system represented the starting point to the creation of a new data hub and a new attitude because we made a change in technologies using Python as backend. The next step was to create a CRUD application to store the most crucial data of the scholarships and mentors. The database uses Sqlite3 due to its flexibility and low space requirement letting us overtake the hardware support [2]. This application implements an API access point with GraphQL Ariadne. This access point is accessible for all the members of the institution inside the LAN network and protected with Json Web Tokens [3]. It is important to highlight that the data culture of the organization is in the process of change, but the storage of information in unstructured data such as Excel's files, PDF, emails, etc.[4] With this implementation we make our contribution to the culture change and help to accomplishing data government [5]. The core of this application is to serve as data input for a model using Sklearn [6] to classify the scholarship's themes and offer similar themes to the user in order to improve the communication between sectors and to avoid repeated investigation subjects. It's important to highlight that the training topics of the scholarships offered by the CNEA synthesize the core of the training in which it is expected to train/prepare new generations of young technicians and professionals, under the guidance of experienced personnel.

[1] Choi, J. O. (2020). Comparison of Update Performance by File System of Mobile Database SQLite3. *Journal of the Korea Institute of Information and Communication Engineering*, 24(9), 1117-1122.

[2] Ahmed, S., & Mahmood, Q. (2019, November). An authentication based scheme for applications using JSON web token. In *2019 22nd international multitopic conference (INMIC)* (pp. 1-6). IEEE.

[3] Buneman, P., Davidson, S., Fernandez, M., Suciu, D. (1996). Adding structure to unstructured data. In: Afrati, F., Kolaitis, P. (eds) *Database Theory — ICDT '97. ICDT 1997. Lecture Notes in Computer Science*, vol 1186. Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-62222-5_55

[4] Fleckenstein, M., Fellows, L. (2018). Data Governance. In: *Modern Data Strategy*. Springer, Cham. https://doi.org/10.1007/978-3-319-68993-7_8

[5] Razno, M. (2019, April). Machine learning text classification model with NLP approach. In *COLINS 2019. Volume II: Workshop*.

Home Radon Testing, Environmental Health Literacy, and Efficacy in the Greater Accra Region of Ghana: A Citizen Science Approach

Irene Opoku-Ntim¹, Frank Koblah Quashie¹

¹Nuclear Application Centre, Ghana Atomic Energy Commission, Ghana

Lung cancer is primarily brought on by radon exposure globally. In Ghana, few people, however, test their homes for radon. Radon testing should be made more widely available, while radon exposure should be reduced. A convenience sample of 100 non-scientist homeowners from four towns in the Greater Accra area were recruited for this long-term, mixed-methods study using a citizen science approach. They were trained to test their homes for radon using a low-cost continuous radon detector, report results, and take part in a focus group to rate their testing experience. The objective was to assess how environmental health literacy's (EHL) efficacy had changed over time. For the purpose of assessing EHL linked to radon testing and mitigation, participants completed online surveys at baseline, post-testing and 6 months later. Over time, citizen scientists reported a significantly higher level of EHL, health information efficacy, and radon testing self-efficacy. While the perceived ability of citizen scientists to contact a radon mitigation professional significantly increased, neither their opinions that radon mitigation would lessen the risk of radon exposure nor their ability to employ a radon mitigation professional changed over time. To fully grasp the function of citizen science in radon reduction at home, more study is required.

Projected Distribution of 218 Invasive Plant Species in India under climate change: Identification of Hotspots and Management Implications

Jyoti J. Prajapati^{1,2}, Abhijit Singh², Achyut K. Banerjee³, Amiya R. Bhowmick¹, and Abhishek Mukherjee²

¹Department of Mathematics, Institute of Chemical Technology, Mumbai, India

²Agricultural and Ecological Research Unit Biological Sciences Division, Indian Statistical Institute, Giridih, Jharkhand, India

³School of Life Sciences, Sun Yat-sen University, Guangzhou, Province - Guangdong, China

India has several invasive alien plant species (IAPS), which critically impact biodiversity, ecosystem services, and economic development. Understanding the pattern of distribution of IAPS under different climate change scenarios in India will be informative to prioritize its management decisions.

In this study, we investigate the potential distribution of 218 different plant species which are already identified as invasive according to the most recently available comprehensive national repository of invasive plants in India, ILORA (<https://ilora2020.wixsite.com/ilora2020>). To minimize the uncertainty associated with the prediction, we utilize the ensemble modeling approach in which nine different machine learning models have been implemented. Selection of the covariates has been done using the Variance Inflation Factor. We categorized the species into different groups concerning growth habits, duration of species, and climatic niche traits. Projected distributions for the species have been compared between groups by identifying the hotspots, that is, regions of high occurrence probability.

Our study shows that no significant range expansion of the IAPS is observed within each group as a whole under climate change. However, individually, some IAPS have shown significant expansion and the majority of them have shown contraction. This may be attributed to the fact that in the hotspot regions, climate does not change significantly under different scenarios. Positively, hotspots have not reflected the rarity of species where expansion or contraction is happening. Conserved areas where hotspots under current and future climate scenarios coincide need to be prioritized for management actions.

Mapping and engaging general public towards a low carbon society: introducing machine learning on citizen science programs

Caetano R. Miranda¹

¹ *Instituto de Física da Universidade de São Paulo and Research Centre for Greenhouse Gas Innovation, São Paulo, Brazil - cmiranda@if.usp.br*

Most processes involving greenhouse gases, energy transition and climate change issues take place at scales far beyond our daily life, ranging from molecules to planetary scales. This imposes challenges to disseminating the knowledge, regulations and technologies developed by researchers to policymakers and the general public perception. To map the public perception and social pressure on technologies towards a low-carbon society, we develop a platform by combining machine learning methods on science citizen programs within two strategies:

i) mapping the public perception in social networks and ii) engaging the general public through science-art, immersive and interactive experiences, citizen science programs through the molecularium. Citizen science provides the opportunity for the public to contribute to large datasets and engage in scientific research. At first, we apply methods based on social physics and machine learning to the available data on social media. Using data from the Twitter platform, we could identify and rank the key topics and debates that are established between different agents on the energy transition to a low-carbon society. Moreover, we establish the molecularium, which is a non-formal experimentation space of education and diffusion of knowledge, both physical and virtual, where numerous immersive and interactive experiences of the phenomena at the molecular scale using sonification and virtual reality to represent climate change and ideas related to a low-carbon society [1]. The interaction between research and the community is through molecular modelling based on volunteer computing and immersive experiences at the molecular scale. The experiences could be interactive molecules in virtual reality, where it is possible to receive visual and audible stimuli when you pull, squeeze, distort and move these molecules. One challenge is sensory responses and the quick feedback to control these responses during the immersive experiences with realistic physical descriptions. To tackle this issue, machine learning methods have been introduced to model the interatomic interactions coupled to interactive molecular dynamics experiences, which can trigger different sound or visual responses according to the user's actions, improving the immersion experience [2]. In summary, we developed digital approaches based on machine learning methods to map the social perception based on the big data available on social media and engage the public through immersive experiences by introducing and projecting future scenarios on climate change and technologies toward a low-carbon society. In collaboration at RCGI with: Ana Paula Gomes Vasconcelos, Dindara Silva Galvão, Gustavo Chagas de Moraes, Helen Takamitsu, and Miguel Vera.

[1] Miranda, C. R., Applications of sonification on molecular simulation and data mining, Emerging Technologies in Scientific Data Visualisation - Cecam Workshop., 2018

[2] Miranda, C. R., Ambiences at the nanoscale: inclusion experimentations on science-art, extended realities and social entrepreneurship, 2022 Fall Meeting, Brazilian Physics Society. (2022)

Leveraging the Power of Citizen Science and Machine Learning to Transform Low-Income Communities

Mushfiquir Rahman Saad¹,

¹*Wireless Communication@UKM*

Leveraging technology is usually expensive and hard to integrate in low-income regions due to lack of expertise, education and scientific understanding. Additionally it is often expensive to gather scientific tools of research (both hardware and software) in these communities further creating difficulty in community based participation. Technology is also often seen more of a barrier than enabler there, which promotes a negative loop of using analogue methods of intervention in policy and community levels. This case study will discuss and help visualize the impact of citizen science and relevant machine learning within the confines of these communities to show how this scientific approach works for the bottom of the pyramid. Provisions of improved economic activities, poverty alleviation, community involvement and community action will be highlighted with recent examples of development activities which are harnessing the power of accessible science and machine learning.

Citizen Science as a Tool to Monitor the Water Pollution. A Case Study of Lake Sevan, Armenia

Environmental specialist

M. Voskanyan¹

¹University of Traditional Medicine of Armenia

The increasing demand of freshwater is influenced by the dynamic of the climate, the economic development and the growing number of populations. To protect the water resources, a new approach such as citizen science should be considered. This paper describes the monitoring plan and success factors of the citizen science project of water quality in Lake Sevan, one of the contributing lakes in the entire South Caucasus environmentally and politically. Due to some external factors, such as the flow of household, agricultural and industrial wastewater, large waterlogged forested areas and the rising temperature, the water quality of the lake is worsening over time. Based on a 5-grading scale, the water quality of the lake is classified as the 4th class which means it almost loses its recovering ability. Visually, the blooming also starts to appear on the water surface. The problem becomes exacerbating since there is not enough data about water quality due to the lack of human resources. The investments of the government and the international organizations, such as GIZ, UNDP also have not been adequate to solve this issue. Based on the aforementioned urgency, it needs a holistic involvement from stakeholders including the volunteering of the citizens through CSc to monitor water quality of Lake Sevan. Meanwhile, there is still a few existing academic literature of CSc in the water area in the RA. In the parameter measurement, similar to the aforementioned condition, there is still a limited number of resources as well, such as in water levels [1], chemical and physical parameters [2]. This paper will answer the research questions on: how the water quality has been changing and which appropriate Citizen Science monitoring plan design should be applied. The review of water quality data from requested and publicly accessible governmental resources for the last 3 years and a literature review of international citizen science projects for this case study are used as the method. The results show that the water quality of Lake Sevan is classified as 4th grade which means in the critical level. A good water quality monitoring plan which includes success factors should be implemented to observe the dynamic of the lake. Finally, involving trained and knowledgeable citizens in a water quality data collection in a properly designed Citizen Science project will help to fill the existing data gap for decision making, and for prospering scientific research.

[1] Little, K.E., Hayashi, M., Liang, S., **54** (3), 317–324 (2016).

[2] Alender, B., et al 15 (3). **3**, 15 (2016) .