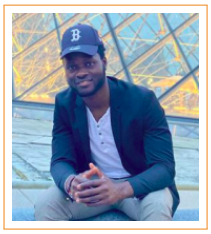

Esso-Ridah BLEZA

Teaching and Research Statement

Previous Research Experience

My research helps answer the following question. Exploit real-time temporal data using machine learning methods for the recognition of phenomena related to public health. My previous research has helped answer this question by studying environmental data from sensor networks for the recognition of human activity on the one hand and for the detection of aerobiological emissions in ambient air on the other hand.

Thus, I have experience in exploiting data types constituting large spatio-temporal flows characterized by variables that are often redundant, and whose distributions are asymmetric (non-Gaussian) and inflation of heavy-tailed zeros. To effectively exploit this data, I was led to propose innovative methodological and algorithmic developments.

At first, I was interested in reducing the volume of data by proposing two types of approaches: (a) segmentation of a data stream from a dynamic propagation algorithm based on a piecewise linear approximation and a complexity criterion (minimum description length) (b) unsupervised classification based on kernel clustering methods. This made it possible to see that classical distances or similarity measures are not very effective in segmenting zero-inflation data and that there are not enough studies in the literature on this subject.

In a second time, I was able to propose original detection and supervised classification algorithms, by proposing a combination of logistic regressions by random forest. These methods have been implemented in two applications. The first focuses on the (non-intrusive) monitoring of elderly people for better care in the event of loss of autonomy or accident (fall, illness, lack of activity, etc.). The second study focused on the prevention of allergy risks related to the diffusion of pollen in the outdoor air. The algorithms developed are one of the major bricks put into production to produce and broadcast preventive pollen risk alerts on mobile applications of a French startup (LIFY AIR) since March 2022.

Research Goals

45000 Orléans – Frnace
(0033)651 675 195 • *esso-ridah@bleza.fr*

My previous research is more detailed on My Page[1], but I would like to focus on my interests and the consequent and valuable work of exploring, understanding, and preparing datasets from real-life applications. More specifically for the pollen risk prediction application, I conducted work ranging from understanding the operation of sensors to deploying an alert solution.

The processing of temporal data, which is particularly zero-inflation and heavy-tailed, deserves to be approached by a special approach to make the best use of the information. I will build on my thesis and my experiences to better understand all the issues on this type of data. Moreover, one of my perspectives is to devote work to the search for a measure of similarity adapted to multivariate time data with zero inflation and heavy tails based on the number of common zeros between individuals.

Indeed, sensor data containing many zeros seem to pose problems at usual distances (based on L_p type standards) for similarity measurement. The number and position of zeros on each vector is a characteristic signature form of a sensor data signal. For example, the total absence of particles at the moment of measurement characterizes individuals with the presence of zeros in almost the entire line while the presence of a given particle marks fewer zeros on the characteristic line of the individual. I could easily show that vectors with the same numbers of zeros in the same positions are more likely to represent the same particle. This expertise in experimental measurements provides me with a solid foundation on which I can build new theoretical explanations.

Big data mining and machine learning have become essential to automate the information processing necessary to solve application problems in all sectors of activity, whether they are economic, health, industrial production, energy distribution, transport, space, defense, environmental management or the societal fabric, etc. So, I have a huge range of contributions that I can make.

The transversality of the projects I worked on allowed me to judiciously mobilize different data analysis techniques and to conduct a good methodological approach in my work. My expertise in exploiting different types of multivariate multi-source temporal data provides me with a strong ability to contribute and deepen methods for any subject. This is regardless of the new or complex nature of the data to be processed that may constitute the data of future applications.

Teaching

My teaching interests cover mathematics, statistics that I practiced in high school but especially the fundamentals of Machine Learning and its practice in python and R programming language. Indeed, despite the environment of my thesis in business, I was able to accompany the master's students in statistics of the University of Rennes 1 as a practical work assistant and mentor for their academic projects on Machine Learning.

[1] https://bleza.fr

I look forward to inviting undergraduate students from the classroom into the research lab. While I can teach a wider range of courses at undergraduate and graduate level if needed, I would feel more comfortable starting in the areas of data analytics, Machine/Deep Learning. I think I could teach the following courses with a high level of expertise:

• Undergraduate: Introduction to Python/R, Introduction to Statistics and Probability
• Second cycle: Data Mining (ACP, ACM, etc.) Machine Learning (supervised and unsupervised learning), Deep Learning (Perceptron, NLP, CNN, etc.)

Motivated by the practices of the techniques used in Data Science, I am also interested in the development of innovative courses applicable to artificial intelligence for environmental topics.

I will help students read current journal articles during their academic project, practice writing article reviews and presenting synthesis lectures (revision) on these advanced topics during their projects. The class could also be seen as a development of graduate research programs as well as the practice of professionalization for students.

I will cultivate the following attitudes in students who complete courses or research with me:

• demonstrate understanding of rigorous mathematical tools for design/analysis
• outline public-oriented communication strategies, (Justify and explain their work simply and easily, i.e. make the complex topics they work on more accessible)
• establish sound ethical/social reasoning. (Think long-term, not just short-term)