

People of Data

Meta-learning for fake news detection surrounding the Syrian war: An interview with co-author Roaa Al Feel

Roaa Al Feel^{1,*}

¹American University of Beirut, Beirut, Lebanon

*Correspondence: rba15@mail.aub.edu

<https://doi.org/10.1016/j.patter.2021.100394>

Roaa Al Feel, an early-career researcher, discusses her passion for using data science for social good. She uses data to reflect living conditions of society, and in the paper published with *Patterns* in November, the team explores machine learning techniques for the detection of fake news around the Syrian war, demonstrating the efficacy of meta-learning techniques when tackling datasets of a modest size.

Tell us about your background (personal and/or professional)

I hold a master's degree in computer science, primarily in machine learning, data science, and big data, from the American University of Beirut (AUB). I currently work as a software developer and data scientist at the United Nations Economic and Social Commission of Western Asia. Before this, I worked as a data scientist at Data-Pop Alliance, a nonprofit organization that aims to change the world with data with the works of social and data experts. Additionally, I have worked for years as a research assistant at the American University of Beirut on several data science and machine learning projects, one of which is this work. Throughout my masters and years of work as a data scientist, my focus has been to utilize traditional and non-traditional data sources to understand, analyze, and help improve the living conditions of the multiple sectors of society (Figure 1).

What motivated you to become a data researcher? Is there anyone/anything in particular that helped guide you on your path?

During my master's studies, my research focused on the different aspects of detection of fake news in the Syrian War, which later became my master's thesis. From this research, we published FA-KES¹ and recently our journal submission to *Patterns*.² This work was my initial inspiration on how much impact data science and machine learning can have on society and social good. From this project on-

ward, I decided on the career path of researching the many ways data can be leveraged to understand and improve the living conditions in a society. Seeing how my work reflects what my country and the region are going through increases my motivation every day; especially when I see results that reflect the situation in Lebanon in the middle of the many crises we are facing. The many results that can be extracted from data can be used by policymakers to influence better decisions that target the most vulnerable. All of this makes me feel that, through the type of work that I do, I am somehow helping my country get through the challenges. My biggest inspiration to pursue this field has been Dr. Fatima Abu Salem, the first author of our *Patterns* article, who was my thesis advisor and with whom I worked on multiple research projects during my time at AUB. She has always been a great role model and an inspiration to me and other students on the role of women in challenging gender stereotypes and in the field of data science. She also encouraged me a lot personally to pursue this field for its broad, interesting, and challenging opportunities.

What is the definition of data science in your opinion? What is a data scientist? Do you self-identify as one?

To me, data science is the integration of mathematics, programming, and analysis skills with any other skill asset (health, policy making, business, etc.) to derive useful results to better improve both fields. In

addition to the previously mentioned impact in social studies, data science skills can also be used for medical, business, economic, and many other purposes. A data scientist's role varies from collecting, cleaning, understanding, and analyzing data to produce useful derivations and results. I self-identify as a data scientist and hope to continuously learn more about the field as it grows.

Why did you decide to publish in *Patterns*?

We decided to publish this article in *Patterns* due to its high relevance to our project: sharing knowledge and expertise about data science, for us originally in the field of media analysis and social good, in addition to the reusability of machine learning findings across the different fields of data science. Whether our reader is interested in media analysis or not, they can benefit as a data scientist from reading our work.

What is the role of data science in the domain/field that you work in? What advancements do you expect in data science in this field over the next 2–3 years?

As a current UN organization member and a former NGO member, data science in my field is mostly used to analyze traditional (survey data) and non-traditional (varying from mobile telecom data to social media data) to help policymakers understand the past and current living conditions, the needs of the society, followed by deriving conclusions





Figure 1. Roaa Al Feel

and policy recommendations for the policymakers to benefit from. I can already see the continuous progress that data science is making and the increasing appreciation and understanding of the value of data science in this field. This varies from the many works that get published every day on the utilization of data science and machine learning for social good to the increased job openings in NGOs and UN organizations requesting data scientists. I believe the next 2–3 years will witness great breakthroughs and higher cooperation from the governments to leverage the benefits of data science in this field.

How can data science help domain science/manufacture/industry/policy/society, and how can collaborations between academia and industry/policy/society be started?

From my experience, the key to a successful collaboration is the understanding of the importance and the added benefit of data science. This can be achieved when data scientists share related work with the collaborators and explain to them the role that data science played in achieving the results. Additionally, the collaborators must be aware of the importance of sharing meaningful and useful data to achieve optimal results. Without data sharing, it can be difficult to get the entire benefit of data science.

Did you encounter any particular difficulties? Were there any specific challenges about data/data management/FAIR data sharing that you dealt with? How did you overcome them? Can others use the solutions you used to overcome these challenges?

Just like any other data science project, we encountered some difficulties during data collection. The difficulty in our project was the lack of labeled data. Collecting news articles from different sources can be achieved through the many scraping methodologies. However, labeling articles as “true” and “fake” was challenging. Since our data were centered around the Syrian war, our solution was to compare articles’ casualties’ content against the Violations Documentation Center (<https://www.vdc-sy.info/index.php/en/about/>) by calculating the Gower distance between the article content and the ground truth, followed by creating two clusters: true and fake. The details of this solution were published in our paper, FA-KES.¹ With this labeled dataset, we were then able to extract media and text features that are not specific to Syria or war, meaning that while the data are centered around the Syrian war, it is not limited to only the Syrian war. This was proven in our *Patterns* article, where we showed that the trained model performed well on other known fake news datasets.²

How did you come to collaborate?

This work was co-authored by professors and graduate students from the computer science department and a professor from the media department at the American University of Beirut. It was important for the computer science team to collaborate with the media team to inspire features and verify the logic behind the theories suggested, in addition to having an expert review a sample of the results for verification. Additionally, the collaboration was also within the computer science team between the computer science professors themselves, who are all experts of data science, machine learning, and big data, in addition to the graduate students. I was one of the graduate students working

on this project, which was, in addition to our FA-KES publication, part of my master’s thesis work.

How important was the collaboration to the success of the paper? How important do you think collaboration is in general to research?

The interdepartmental collaboration was crucial to the project due to the project’s targeting of the media and the news fields rather than being purely technical. The media expert’s opinion was particularly useful in reviewing the features suggested, in addition to a sample of the results to compare what an expert would consider fake or true as opposed to what the machine learning models would consider fake or true. The collaboration between the experts of machine learning, data science, and big data of the computer science department was just as important to ensure using state-of-the-art technical approaches and to verify the technical side of the theories, features, and methodology. I believe collaboration is extremely important in research to ensure the quality and variety of experiments, ideas, inputs, and results.

REFERENCES

1. Abu Salem, F.K., Al Feel, R., Elbassuoni, S., Jaber, M., and Farah, M. (2019). FA-KES: A Fake News Dataset around the Syrian War. Proceedings of the International AAAI Conference on Web and Social Media 13, 573–582. <https://ojs.aaai.org/index.php/ICWSM/article/view/3254>.
2. Abu Salem, F.K., Al Feel, R., Elbassuoni, S., Ghannam, H., Jaber, M., and Farah, M. (2021). Meta-learning for fake news detection surrounding the Syrian war. *Patterns* 2, 100369.

About the author

Roaa Al Feel holds an MSc in computer science from the American University of Beirut and a BSc in computer science from the Lebanese American University. Her research specializes in the topics of machine learning, data sciences, and big data. She worked as a data scientist at Data-Pop Alliance, a collaborative laboratory created by the Harvard Humanitarian Initiative and other partners. She currently works as a data scientist and software developer at the United Nations Economic and Social Commission for Western Asia (UN-ESCWA). Her work focuses on leveraging traditional and non-traditional data sources to assist policy makers in targeted decision making for social good.