ieee.es
Instituto Español de Estudios Estratégicos
1970 — 2020

Documento
Opinión

46/2021 22/04/021

Fernando Toboso Marqués *

Demystifying algorithms. Appreciating data, the past and proxies

Visitar la WEB Recibir BOLETÍN ELECTRÓNICO

Demystifying algorithms. Appreciating data, the past and proxies

Abstract:

This article, although it mentions the drawbacks of algorithms or the dangers of social networks, aims to highlight some of the aspects of algorithms with an intelligence analysis approach, related to decision making, examining how they function and what they are based on.

These aspects lead us to understand the importance of algorithms' raw material: data. Those that are available, proxies, are used an infinite number of times and they are often not optimal for representing the concept studied and may therefore distort the result.

In spite of this, references, values and sequences similar to proxies are used in other probabilistic approaches, also to know the present and predict the future. This is what happens with *superforecasters*.

In both instances we are looking at a search for quantities and figures; in short, this is about numbers and their management.

Keywords:

Algorithm, data, proxies, superforecaster

.

Desmitificando los algoritmos. Apreciando los datos, el pasado y los valores sustitutivos

Resumen:

El presente artículo, aunque mencione los inconvenientes de los algoritmos, o los peligros de las redes sociales, pretende destacar algunos de los aspectos de los algoritmos con un enfoque desde el análisis de inteligencia, relacionado con la toma de decisiones, detallando cómo operan y en qué se basan.

De esos aspectos se desprende la importancia de su materia prima: los datos y cómo infinidad de veces se recurre a los que hay disponibles: los valores sustitutivos, que muchas veces no son los óptimos para representar el concepto estudiado y por tanto pueden distorsionar el resultado final.

A pesar de ello, referencias, valores y secuencias similares a los valores sustitutivos son utilizados en otros enfoques de carácter probabilístico, también para conocer el presente y predecir el futuro, como es el caso de los Superpronosticadores.

En ambas ocasiones se recurre a la búsqueda de cantidades y cifras, en definitiva, a números y su gestión.

Palabras clave:

Algoritmo, datos, valor sustitutivo, superpronosticador

Cómo citar este documento:

TOBOSO, Fernando. *Demystifying algorithms. Appreciating data, the past and proxies*. Documento de Opinión IEEE 46/2021. http://www.ieee.es/Galerias/fichero/docs_opinion/2021/DIEEEO46_2021_FERTOB_Algoritmos_ENG.pdf y/o enlace bie[3] (consultado día/mes/año)

***"Anything that is not measured does not interest me." Ivan Redondo*[1]**

Data: enormously abundant resources, easy to obtain, but extremely valuable

Precious metals, energy resources, water, food, the most fertile valleys, the most sheltered natural harbors… these have always been disputed property. In recent decades, new elements have been added to this list: coltan, rare earths, new maritime routes, space… but an immaterial element, "revolutionary" and located in cyberspace, stands out in its own right: data. Hundreds of billions of data. They are extremely easy to obtain, they can be processed thanks to advances in computer science, and they enable those who collect and process them to become super-millionaires and super-powerful.

We are all aware of the immense power of technology multinationals and technology in general. However, many experts affirm that data is the most valuable resource today, more valuable than gold and oil. If we recognize that data is one of the most valuable strategic elements, we should also prioritize its analysis in strategic think tanks, transcending the fields specializing in Big Data.

Data is generated when we sign into a website, or in any interaction we engage in on the internet, whether in a browser, with our credit card or supermarket points program, on our social networks (including from our contacts), etc. Everything is recorded with data associated with us, difficult or impossible to delete. Data are the essential components of algorithms. Many companies sell their users' data and there are whole companies that are dedicated to buying and selling data banks. For example, in November 2018 the German NGO Tactical Tech published a list of 329 companies specializing in electoral campaigns[2].

Referring to the best minds of a generation, the great expert Marta Peirano stated: “Those who did not work at Facebook, Google, Twitter, LinkedIn, Amazon or Groupon looking for ways to improve their ads were on Wall Street, writing algorithms capable of digesting

[1] VALLS, Fernando, **"**Un día de trabajo con Iván Redondo: así funciona la sala de máquinas de Moncloa”. *La Información.com,* July 14, 2019. Available at: https://www.lainformacion.com/espana/ivan-redondo-sanchez-moncloa-psoe/6506465/
All web links in this document are active as of 1 March 2021

[2] Tactical Tech. "Who's Working for Your Vote?" *November 29, 2018.* Available at: https://ourdataourselves.tacticaltech.org/posts/whos-working-for-vote/

industrial quantities of market data to make buying and selling decisions in microseconds."[3]

Data arouses interest, and not only because of its commercial exploitation in sectors related to advertising, marketing, or high-frequency trading on the stock market. It has long since made the leap to other apparently far-removed sectors, such as social sciences or politics (as the Tactical Tech listing confirms) and this is so because its usefulness in those other fields has been proven. The data of millions of individuals, collected, interpreted and segmented according to various factors such as the psychological-personality factors in the Cambridge Analytica scandal, has helped those who use it win elections and referendums.[4]

Expert opinions have lost much of their prestige. For decades, competent and authoritative experts have proposed diametrically opposed diagnoses and treatments, based on the same facts. Faced with this situation, the temptation to seek an objective assessment, aseptic and technical, removed from the buzz and based purely on numbers, is gaining adepts. The solution would lie in pure and flawless technology. This explains the blunt statement (*"anything that is not measured does not interest me"*) that opens this article and it also has to do with the ease of evaluating results and acting accordingly.

Decisions that are rational and… economical

One of the main applications of intelligence is that it makes it possible to project, enabling decision-makers to make better decisions. Consequently, Heuer and Pherson established the "Decision support" function, which includes methods such as SWOT analysis, Decision Matrix, etc.[5] Likewise, in some studies on intelligence analysis, techniques such as game theory or criteria weighting (AHP, Analytic Hierarchy Process; Electre Method, etc.) are included in the field of "decision aid". They do not strictly

[3] PEIRANO, Marta, "El enemigo conoce el sistema". Editorial Debate, 2019

[4] Netflix, **"**The Great Hack**"**, a documentary about Cambridge Analytica. 2019

[5] HEUER Richards J. Jr. and PHERSON Randolph H., "Técnicas analíticas estructuradas para el análisis de inteligencia". Editorial Plaza y Valdés, 2015. Spanish version of the 1st edition of this book, 2010, pp. 279 to 301. Review available at: http://www.plazayvaldes.es/libro/tecnicas-analiticas-estructuradas-para-el-analisis-de-inteligencia/1493/.

correspond to intelligence, as the factors are known and there is not so much uncertainty. However, they help to make rational decisions.

We might think that the main reason for creating algorithms is because they fulfill this function: making objective, data-supported decisions, avoiding subjective decisions by individuals with different criteria. However, this is only partially the case, as very often an algorithm is created because previously there was a workforce reduction and it was necessary to optimize the resources that still remained and reduce costs. This reduction is achieved through suitable planning and resource allocation.

In fact, algorithms' closest antecedent is "operations research" (also called "operational analysis"), a branch of mathematics that underwent a boom during World War II, where the allies used it to achieve maximum efficiency in the use of their resources. Given its success, after the war it was widely used in logistics and planning.

Many algorithms constitute an initial decision phase, one that is mechanical and relatively easy, often a broader filter, with a subsequent phase involving people who are specialists in the matter in question and who fine-tune the final solution.[6] If the cut-backs have been so significant that there are not two phases and everything is done by the algorithm, the risk of error increases.

Also because of resource optimization, an algorithm will have an initial version and continuous updates. Many companies risk their future on the whole process being successful and enabling them to be better than their competition, so feedback and improvement must be constant.

Algorithms. A brief review. What are they, what they do and how they do it, and how long it takes

According to the Royal Spanish Academy (RAE), an algorithm is an "ordered and finite set of operations that makes it possible to find the solution to a problem". In practice they

[6] OLÍAS, Laura, "Cuando el inspector de trabajo es un algoritmo: avances y temores sobre las sanciones laborales automatizadas". *El diario.es,* January 29, 2021. Available at:

https://www.eldiario.es/economia/pasa-si-inspector-trabajo-robot-avances-temores-sanciones-automatizadas-aprobado-gobierno_1_7177360.html

are executed by computer, so they have become synonymous with software or computer programs. But the creator of the "like" on Facebook goes further. Complex algorithms can already be considered artificial intelligence[7] and of course they learn automatically (*machine learning*). Without a doubt they are in a different category and "compete" with our Homo sapiens brain and with our very human weaknesses. They do not need to overcome our intelligence and our strengths; it is enough for them to take advantage of our weaknesses.

There are many types of algorithm. Some of their most common applications are: arranging or evaluating; making a ranking of universities, or of cities that are ecological, "friendly" or offering a better quality of life;[8] selecting, recommending or predicting success (news, music, series); analyzing risks (of recidivism, payment default, health, accidents, etc.); planning supplies or schedules; optimizing delivery or patrol routes to prevent robberies and crimes, or reflecting hot spots (of crime) in an area.[9]

There are also algorithms for sports performance; in baseball and basketball, they are called PECOTA and CARMELO, respectively.[10] They are used to predict the performance of today's players, comparing them with the statistics of other, similar players from the past and consequently signing promising talents before rival teams can do so. And of course this happens in soccer as well,[11] although with uneven success.

Algorithms have been a reality for years, used very extensively, including in Spain. Their scope includes things we might not suspect, from offering different results to the same Google search (tailored to each user on their computer, "optimized" according to their

[7] ROSENTEIN, Justin, interview at minute 47 in *The Social Dilemma*, a documentary available at Netflix, 2020.

[8] PLAZA, Analía, "Qué hay detrás de los rankings de 'mejores ciudades para teletrabajar' en los que España sale tan bien parada". *Eldiario.es,* December 8, 2020. Available at: https://www.eldiario.es/economia/hay-detras-rankings-mejores-ciudades-teletrabajar-espana-sale-parada_1_6481578.html

[9] PREDPOL. Predictive Policing Company. Available at: https://www.predpol.com

[10] TOBOSO, Fernando, "Grado de acierto de estimaciones y pronósticos. criterios de evaluación de la metodología y de la calidad de los análisis". *Instituto Español de Estudios Estratégicos*. July 19, 2016 Available at: http://www.ieee.es/Galerias/fichero/docs_marco/2016/DIEEEM12-2016_Metodologia_CalidadAnalisis_FernandoToboso.pdf

[11] DRIBLAB. Big Data consulting football professionals. Available at: https://www.driblab.com/es/

search history), to determining the waiting time on a phone call to customer service (depending on how "good" a customer we are) or uncovering tax fraud.[12]

In the words of an award-winning data scientist: “First, the 'real world’ problem or challenge is presented, the key features are identified, and finally it is transformed into a statistical problem using algorithms. The solution may be obtained in a week, a month or a year depending on the complexity of the problem, although sometimes there is no solution.”[13]

If we have the data and we know what we want, the software will find an initial solution in margins measured in hours or days. It will require verification and refining in new versions (weeks, months), in an improvement process that is endless, as it will translate directly into economic benefit. Once the data is collected, the incredible shortening of the timelines the algorithm requires to provide the solution is another advantage that facilitates their implementation and expansion, in addition to their low cost and accuracy.

Another factor that explains their success, especially with regard to algorithms used in social networks, is the incorporation of psychological techniques to maintain the user's attention and even their addiction. Mechanisms theorized since 1948 by B.F. Skinner (demand, action, reward, variable interval reinforcement) and since 1998 by B.J. Fogg (motivation, ability and a prompt), which have already shown their effectiveness in slot machines and their levers.[14]

In summary, the algorithm is a scientific tool of the first order, as it describes a phenomenon, discovers regularities and patterns and predicts future behaviors.

The importance of proxies in algorithms

It can be very difficult to capture a qualitative concept of "real" life (safety, quality, prestige, excellence, friendly, healthy, solvent, etc.) and translate it into measurable characteristics

[12] GONZÁLEZ, Ignacio and MATEOS, Alfonso, “¿Se puede destapar el fraude fiscal con algoritmos?”. *Observatorio Social de “La Caixa”,* October 2020. Available at: https://observatoriaociallacaixa.org/es/-/se-puede-destapar-el-fraude-fiscal-con-algoritmos

[13] CORDERO, Sara, “Fátima El Baghdadi, ganadora de los Global Awards 2020”, *El plural*, January 1, 2021. Available at: https://www.elplural.com/economia/empresas/fatima-baghdadi-ganadora-globant-awards-2020-no-genio-persona-disciplinada-esfuerza-objetivos_256142102

[14] Peirano, Ibid, Chapter 1

(statistical variables). If this is achieved, we will have captured its essence and a successful model can be replicated and reproduced, also obtaining the desired success.

It is often sufficiently demonstrated that the abstract notion (risk of credit default, for example) depends on certain variables, which in the words of Dr. Cathy O'Neil[15], PhD in mathematics from Harvard University, would be called “relevant values” (in the example above: income, debt amount and bill payment history), but there may be legal limitations on their use (individual consent, data protection laws, prohibition of use in marketing). Nor can data be obtained that imply discrimination based on sex, race, religion, etc. Likewise, it may happen that they are new or unique phenomena and there is little information from which to start.

In these cases, algorithm designers look for other substitute variables (proxies), even though the influence of these new variables on the definition of the qualitative concept has not been sufficiently demonstrated. However, they are no limitations on obtaining or processing these data; a paradise for new technology companies, which do not suffer from the regulations affecting other large companies such as the telecommunications giants. Thus, in the United States it is illegal to use intelligence tests to hire a person, so they are replaced by personality tests (online tests), although these may be drawn from many different sources, some more rigorous than others, and do not measure the same thing.

One of many examples and representative of the use of an algorithm with proxies instead of relevant values, included by Dr. O'Neil in her book,[16] is the one used by North American insurance companies to calculate the price of automobile insurance. Here, instead of setting the rate based on the driver's history (fines, accidents, etc.) it does so based on their socioeconomic factors and consumption patterns (and those of their friends and neighbors, as they are supposedly similar). This translates into the price of insurance, which hurts many drivers, even if they are not any worse at the wheel.

Likewise, for "big brother", the once important *WHAT are we doing?* has been replaced by the much more affordable proxy *WHERE are we?*, available to phone companies and

[15] O'NEIL, Cathy, “Armas de Destrucción Matemática”. Spanish edition Captain Swing, 2017
[16] O’NEIL, Ibid, p. 206 and “The truth about car insurance”. *Consumer Reports*. July 30, 2015. Available at: Car Insurance & Auto Insurance Special Report - Consumer Reports

Google (and through it, other third parties). Likewise, the content of our emails has been replaced by their metadata.

Since we underestimate our privacy and the importance of our own data, we surrender them carelessly or resignedly to gain access to or use of attractive programs. The result is that there is no shortage of data; it appears in the most unusual places and for the most unusual purposes. Tens, hundreds or thousands of data can be selected for the same individual or object, under the assumption that this large quantity enables a better representation of the quality sought. An approximation will be made and with the new versions of the algorithm it will be perfected.

Proxies offer an approximation to the abstraction that they define, so their key features are similarity and easy observation. Logically, the more correlation there is between them and the qualitative concept, the better they will embody it. Occasionally there may not be a clear correlation and only time will determine if we are dealing with relevant or proxy values. However, most of the time there is no such doubt and a conscious choice is made for one or another based on their availability or the effects pursued. Thus, the proxies that most favor the desired outcome will be chosen.

These approximations constitute an initial reference that avoids major errors; they offer a starting point. Their success in the substitution will be uneven, and aside from considerations about the aim pursued, this will distinguish a good analyst, as the relationship between these values and the concept they should represent may not be obvious.

The management of all this data is simple due to computing advances, unthinkable a few decades ago. Big Data has been with us for years and materializes in these algorithms that encode the past and offer a probable vision of the future.

Finally, this process is validated. The media or social response; the academic, economic or athletic success; or the response in the field to which the concept belongs confirms that we have captured it.

If we were to make a comparison between relevant values, proxies and their use in algorithms, its summary would be as follows:

	Relevant values	Proxies
Correlation with concept to be studied	Strong and proven	"passing grade"
Purpose	Provide solution/s to the problem posed	Provide solution/s to the problem posed Frequently favor a certain option
Quantity of values	Very limited, around ten	Very numerous, can be hundreds or thousands
Value access and management	Regulated and protected	Unregulated; sold or ceded endlessly
Extraction timeframe	Data from the past	Data from the past
Transparency	Those taken into account are usually disseminated	They are not disseminated completely

Table1: Prepared by the author

Although proxies do not reflect reality, they do dazzle us with their brilliant technological packaging.

These proxies, used in quantitative processes, would be comparable to analogies in qualitative processes. Historians, for example, frequently resort to analogies, contrasting present situations with past situations, comparing and contextualizing their similarities and differences, to facilitate the explanation of a current event and its possible evolution. There is, therefore, a search for known values or situations from the past that may be useful to us in the present for the near future.

From abstract quality to quantitative variables

Taking our cue from the magnificent book *Weapons of Math Destruction* by Dr. O'Neil,[17] a great specialist in the field, we will use one of the oldest and most representative algorithms as an example. This algorithm establishes the ranking of the best North American universities. It is assumed that it will evaluate the notion of "university quality and prestige", with the desired model to be imitated (the requirements to be met) being that of the most renowned universities, such as Yale, Harvard, Stanford or MIT, which already have great international prestige. They have come up with the "formula for success" that must now be captured and replicated.

This task was tackled by a magazine going through hard times (*US News & World Report*) that wanted to increase its audience, by evaluating no fewer than 1,800 colleges and universities. Its list was released in 1988. 75% was prepared with an algorithm and the rest was based on the opinions of university employees. The results seemed to be sensible and did not incite controversy, but in subsequent years the list became a sacrosanct reference throughout the country, producing a feedback loop. It rewarded the well-ranked schools and penalized others. Nobody could have imagined the reception and success they obtained; the magazine disappeared in 2010 but rankings of all kinds expanded, due to the great acceptance and credibility with which they were received.

The usual way to proceed is to break down a complex topic into many simpler topics (or questions/variables) that can be observed (and translated into objective quantities). This quantity would not be problematic to manage and would theoretically bring us closer to understanding the essence of the concept studied.

Thus, of the tens or hundreds of possible variables, some will be mentioned as an example:

Variable 1: Tuition price

Variable 2: Credit price

Variable 3: Location in rural or urban area (small, medium or large city, pre-defined)

[17] O'NEIL, Ibid, p. 65-86

Variable 4: Number of admission applications received for each available spot

Variable 5: Percentage of students who pass the first year of studies

Variable 6: Percentage of students who graduate

Variable 7: Number of students per class in science subjects

Variable 8: Number of students per class in humanities subjects

Variable 9: Number of lab hours per subject

Variable 10: Number of hours dedicated to lectures or seminars per subject

Variable 11: Number of lectures given by Nobel Laureates

Variable 12: Number of university professors who have obtained a Nobel Prize

Variable 13: Number of articles by professors in prestigious journals (previously identified)

Variable 14: Number of citations of these articles

Variable 15: Ease of access for people with reduced mobility

Variable 16: Number of rooms and accommodations on the campus itself

Variable 17: Number of rooms and accommodations within 10 kilometers of the campus

Variable 18: Number of rooms and accommodations between 10 and 30 kilometers from the campus

Variable 19: Amounts donated to the university by former students

Variable 20: Amounts donated to the university by companies and private foundations

Variable 21: Percentage of former students who have found a job nine months after graduating

Examining different facets regarding professors, students, facilities (sports, classes, laboratories, libraries and study rooms, accommodation, dining rooms, cafeterias, etc.), we could effortlessly reach a hundred variables to capture the excellence of those universities in the North American "major league" (there are hundreds of others that would not even be shown in the ranking).

One important point is how the answers to these variables are obtained, taking the large number of universities evaluated into account; a huge job for the editors of a magazine in trouble. Some of them are a matter of public record and could be obtained by said editors (such as the price of tuition, though in 1988 this was not available on the internet). Others can only be answered by the university itself, such as the *percentage of former students who have found a job nine months after graduating*. How did the universities investigate this? In some cases, they wrote a letter to the former student and, if no reply was received, they took it to mean that the former student was very busy because they were working. In most cases we don't know how they reached their conclusions.

As this ranking has been prepared for more than 30 years, we can be benevolent and assume that some variables and some universities must have been audited, with a more demanding investigation into their responses. But when the only possible option is to obtain certain data for the evaluation through the evaluated party itself, the objectivity of said data is at risk.

Whether the writers obtain the answers to the variables or it is the universities that answer a questionnaire, it is the journalists themselves, a group of experts or the algorithm programmers who decide which variables to include and which not, as well as the weight they give to each of them. In this case, the decision fell to the journalists, who did not include variables such as the cost of tuition and university fees. If those variables were included, the universities that theoretically served as a model (Yale, etc.), would not appear among the top spots. In other words, the design of the algorithm, by choosing and prioritizing some data and discarding others, makes it possible to point its results in a certain direction.

An algorithm is a closed, opaque package that is not subject to appeal. It does not admit explanations or claims; IT IS DATA. Data and everything that surrounds them must be perfect by definition, alien to subjective opinions and clothed with SCIENCE's aura of authority.

With this ranking, the students and their families would apparently know what to expect. They could make a rational decision and choose the best university.

A poorly ranked university, however, could only imagine a percentage of the factors evaluated, not the rest. So, to improve its position in the ranking, at best it would make a blind attempt to improve certain aspects, to see their influence on the following year's listing. Thus, rather than academic improvement, its objective would be to satisfy the algorithm, since its ability to attract students and donations, and ultimately its survival, could depend on its position on the list.

Faced with the problems of this ranking or other, similar ones, what could be a possible solution? Dr. O'Neil tells us: "*Instead of creating a new ranking, the U.S. Department of Education published tons of data on a website. And now students can ask their own questions about what interests them, such as the student ratio, the graduation rate and the average debt of students at the end of their studies. They do not need to have knowledge of statistics or know the weighting of the different variables. The software itself, like the websites of online travel agencies, creates individual models for each person.*"[18] Students can maybe filter the search according to ten criteria, those they choose as being most important, since they know which relevant values reflect the excellence they seek in the University. They will gladly exchange dozens of criteria for a very limited number of them. Fortunately, more and more websites offer this alternative adapted to the user, which allows them to be the ones who choose and decide.

Proxies and Superforecasters

In his book *Superforecasting*,[19] North American psychologist Philip Tetlock also explains and highlights the importance of certain values, comparable to the proxies seen above, which help provide specific answers.

The book follows the continuation of his investigative work about expert political judgment, begun in 1984. In this Good Judgment Project, he continues to ask questions about future events (in margins measured in months), which only admit quantitative answers; figures or percentages that facilitate their evaluation and reveal experts in making probabilistic forecasts: the superforecasters.

[18] O'NEIL, Ibid, p. 86. Quote translated into English from the Spanish version of the book.
[19] TETLOCK, Philip. "Superforecasting". Crown Publishing Group 2015. "Superpronosticadores", Spanish edition, Katz editores, 2017.

At the beginning of 2015 he asked one of his characteristic questions: What probability of jihadist attacks is there in eight European countries (detailed), from January 21 to March 31, 2015?[20] Subsequently, he analyzed the response of the superforecaster who hit closest to the mark. This expert first looked for an initial reference in similar events from which to start, to avoid major errors, then modified it according to the particular and concrete circumstances of those first months of 2015. He looked for that first reference in the past, in a solid, objective and known data, the number of jihadist attacks in those eight countries in the last 5 years, which constitutes the average number of attacks per year. In statistics, this would be the base rate. This initial success or error in finding a similar value from the past is essential, and distinguishes the superforecaster, as it will persist and notably condition the final result.

Before making his final estimate, the superforecaster published it on an internal forum of the project, for review by his forecaster peers. He could correct his forecast until the last date of the forecast period, although everything was on the record.

This procedure and this general-to-particular order (a reference value taken from the past, correction with specific circumstances of the present), is repeated in other examples in the book. Tetlock defines it in accordance with Daniel Kahneman (psychologist and Nobel Prize winner in Economics for his contributions regarding human judgment and decision-making under uncertainty), with his “outside view” and subsequent “inside view”[21] and also in line with Heuer and Pherson's "Outside-Inside Thinking."[22]

Tetlock also draws attention to our natural tendency to replace, even inadvertently, a complicated question with a much simpler one, calling it "attribute substitution.”[23]

All of the above ties in with what was previously seen in the search for proxies and also with Enrico Fermi's way of thinking, which Tetlock and his team admire, detailing how Fermi estimated the number of piano tuners in Chicago.[24] This great physicist solved a very complicated question for which he had barely any data, by means of simpler questions that were easier for him to assess, in a logical sequence that reduced the

[20] TETLOCK, Ibid p. 126
[21] Kahneman, Daniel, “Thinking, Fast and Slow”, New York, Farrar, Straus and Giroux, 2011. Spanish edition "Pensar rápido, pensar despacio", Editorial Debate, 2012.
[22] HEUER, Ibid, p. 213
[23] TETLOCK, Ibid, p. 47
[24] MONTEJANO, Javier, “Preguntar con éxito: el problema de Fermi”. *Club mba,* May 22, 2015. Available at: https://www.club-mba.com/2015/05/22/preguntar-con-exito-el-problema-de-fermi/

margin of error. He undoubtedly wasn't right with the exact figure, but he was right with its magnitude.

Conclusions

The idea is not to convert anyone into a data scientist, an impossible task for those who do not have adequate training, but rather to:

Understand the transcendental value of data, and how the processes of obtaining and selecting it condition and introduce bias into the final solutions offered by algorithms, which are ubiquitous and highly influential mechanisms in everyday life.

Highlight the importance of proxies and other similar widely-used techniques, which are often the only affordable alternative despite their drawbacks, representing a whole way of thinking to solve complex and abstract questions through simpler and more concrete processes.

Serve as a reminder that the results or estimates offered by both algorithms and superforecasters are not indisputable and unchangeable. To be successful, they must be continually reviewed and refined.

*Fernando Toboso Marqués**
Lieutenant colonel (reserve) of the Spanish Army
and professor of analysis methodology at CISDE