CONFINEMENT, AN E-LEARNING BOOSTER

Changes study in content e-learning consumption and causality analysis through a Bayesian Network

OLIVER CARRERO MÁRQUEZ, CRISTINA MARÍN PALACIOS, MARÍA BOTEY FULLAT
ESIC University and ESIC Business & Marketing School, Spain

KEYWORDS
Confinement
YouTube
Education
E-learning
Bayesian Networks
COVID-19

ABSTRACT
The confinement has been one of the measures adopted to mitigate the effects of Covid-19 with the greatest effect on society, such as changes in the consumption of content e-learning due to their need to acquire new knowledge. The objective is to establish causal links between periods of confinement and the growth of consumption of e-learning resources. Data on the use of an e-learning channel have been taken and Bayesian Network have been used as a methodological instrument. The results confirm this causal relationship. It shows lines of cause-effect research that could improve e-learning content attending on the user behavior.

PALABRAS CLAVE
Confinamiento
YouTube
Educación
E-Learning
Redes Bayesianas
COVID-19

RESUMEN
El confinamiento ha sido una de las medidas adoptadas para mitigar los efectos del Covid-19 con mayor impacto en la sociedad, como los cambios en el consumo de contenidos e-learning debido a su necesidad de adquirir nuevos conocimientos. El objetivo es establecer vínculos causales entre los periodos de confinamiento y el crecimiento del consumo de recursos de e-learning. Se han tomado datos sobre el uso de un canal de e-learning y se ha utilizado la Red Bayesiana como instrumento metodológico. Los resultados confirman esta relación causal. Muestra líneas de investigación causa-efecto que podrían mejorar el contenido de e-learning atendiendo al comportamiento del usuario.

Received: 16/ 06 / 2022
Accepted: 20/ 08 / 2022
1. Introduction

This article shows the strength of human beings, proactively solving their learning problems, caused by Covid-19 confinement, by accessing e-learning platforms in their autonomous learning process. Streaming platforms have modified the forms of consumption of an increasingly active and empowered audience when it comes to choosing how, when and what content to watch (Benavides & García-Béjar, 2021).

Aligned with Jenkins (2009), we consider that the growth of participatory culture on the Web has also been incorporated into the didactic field. In a context of “transmedia literacy” (Šcolari, 2016) new relationships between subjects, ICT and learning are developing, a reality, in which the objective of this study is developed, which seeks to establish causal links between periods of confinement and the consumption of educational resources broadcasted through YouTube. Based on the hypothesis that the new situation has accelerated the proactivity of the population to take responsibility for their own education (Pereira et al., 2019), through YouTube, we seek future prospective lines focused on the responsible optimization of the educational process provided by this social network.

The pandemic caused by Covid-19 has led to an acceleration in remote learning processes. Chang and Yano (2020) provide an overview of the measures taken by countries affected by confinement, including the expansion of existing channels for distance learning, such as television. Cornock (2020) sees an evolution in the use of ICT, motivated by a greater proactivity of students, more familiar with new technologies, to be responsible for their own learning and to assume the role of prosumer, where it becomes difficult to distinguish between the creator and the consumer of content (Arriaga Azcárate et al., 2016). However, as stated by Orduña-Malea et al. (2020), there is not a large number of studies that particularize the use of educational YouTube channels during the pandemic, the researchers have focused more on the use that universities have made of the platform to communicate and raise awareness among the population (Quijano-Escate et al., 2020). Regarding the pedagogical and cultural possibilities of YouTube, through its e-learning channels the figure of the booktuber appears, represented by Internet users who recommend books in vlog (‘video blog’) format (Vizcaíno-Verdú et al., 2019).

Until the last quarter of 2018, the field that achieves greater development and, to some extent, greater demand is that of video games and entertainment (Herranz de la Casa, 2019). However, the use of audiovisual formats has settled in the educational field, both traditional (Walsh et al., 2019) and informal (Vizcaíno-Verdú et al., 2019). For his part, Álvarez (2020) states that educational videos on YouTube come to meet the new demands of education, not only caused by the Covid-19 situation, but by a new learning model, which picks up the baton of research showing that the use of videos, not only improves comprehensive (Bohloko et al., 2019) and motivational competencies of students (Palazón-Herrera, 2018; Tiernan and O’Kelly, 2019), but also contributes to the construction of their self-learning capacity (Ranga, 2017).

We believe that the growth trend proposed by Šcolari (2016) on collective learning pedagogies, as well as the situation derived from Covid-19, act as accelerators of the aforementioned autonomous learning processes. Moreover, the studies on multiscreen use in video consumption proposed by Chang et al. (2019), could be extrapolated to the content consumption models studied in this article.

Based on Hidalgo-Pérez (2020), who states that the greatest activity on the Internet occurs in the 16-24 age group, these article tries to demonstrate that the age range in the consumption of e-learning is shifting towards a more mature population, perhaps motivated by a greater proactivity demanded by the new employment situation.

Finally, there seems to be a tendency to viralize content. Thus, users who access an e-learning portal for the first time do so on the recommendation of a colleague who provides the url through other channels external to YouTube.

In this context, this paper aims to study the possible causality between the different periods related to confinement, and the growth or decrease in the number of accesses on different devices and operating systems, as well as the profile of users, in order to offer a prospective view that allows creators to optimize the quality and appropriateness of their content. For this purpose, weekly data from two years of accesses to the Zalathun Learning channel and daily data of accesses during the confinement period of the year 2020 have been analyzed, comparing them with those of the same period of time of the two previous years. The Social Network included in Zalathun is located within the communicative needs detected by its authors, to adapt its educational content to a more mature population, perhaps motivated by a greater proactivity demanded by the new employment situation.

According to data provided by the platform itself, academic videos are a perfect complement to face-to-face classes, as they allow learning, motivating to deepen at different paces and asynchronously (Tapia-Jara et al., 2020). In short, the channel aims to act as a knowledge enabling, through the audiovisual materials that enable users to be responsible for their own training cycle. The instrument used to capture the sample was YouTube Analytics.

The Bayesian Networks (BN) methodological tool has been selected to study causality. Compared to other multivariate analysis methods such as logistic regression or linear regression or structural equations, it has numerous advantages such as not requiring any type of distribution in the data or assuming any functional relationship between the variables (López Puga & García García, 2007).
The Bayesian network has been applied to the sample of accesses to the Zalathun Learning channel during the 109 days of confinement in Spain, which has allowed us to deduce that there is a strong causal relationship between the different stages associated with the confinement and the variation in the number of accesses to the portal. As for the changes in the profile of the user who accesses in the confinement stage, there is an increase in age and no changes in the proportions of accesses according to gender are detected. It is also detected that this effect is not immediate and there is a period of adaptation to the new situation, at least 10 days after confinement, in which a slight increase in accesses is observed, possibly at an exploratory level.

The article is structured in five sections. Section 2 presents the conceptual framework necessary to understand the study, where two key concepts are discussed: the increased need for self-training during confinement and the use of YouTube as an educational platform. Section 3 presents the working methodology, the population and the sample used, as well as the description and justification of the methodological tool applied for the modelling of the problem (the BRs). Section 4 offers a detailed analysis of the results of the study. Finally, section 5 presents the final conclusions of the study, as well as the limitations and proposals for future work.

2. Theoretical framework

The irruption of Covid 19 convulsed the educational institution, generating a need for transformation in the teaching-learning processes, which had to be adapted to the new situation of confinement (García-Peñalvo et al., 2020), by telematic and online means (Ricardo-Barreto et al., 2020). This new health reality has meant the need to overcome a methodological change, more focused on proactive attitudes and a strong technological dependence, caused by the lack of physical presence in the classroom.

Thus, e-learning should be framed within what Scolari (2012) defines as new media to refer to those whose main feature is to provide new forms of access to education. Campos-Freire (2015) includes social networks within what he calls digital metamedia, characterized by the development of optimized methods for the economy of attention and where, in our understanding, the educational objectives of educational videos hosted on YouTube fit. In this line, Suárez Rodríguez (2017) warns about the need for new education professionals capable of adapting the message, not only to the new digital formats, but also to the new roles of prosumer audiences, already described by Toffler (1980).

However, in order to meet the new educational needs of these increasingly proactive users, it is necessary to measure their consumption habits, especially in a situation such as the one experienced during the confinement. And it is, in this need, where this work is developed, which aims at a prospective analysis of consumption habits during the pandemic, in order to bring educators closer to the new needs and ways of consuming educational content of users.

2.1. Increased consumption of digital content during the pandemic

On March 14, 2020, a state of alarm was decreed in Spain (Real Decreto 463/2020) due to the spread of Covid-19, which led to a period of confinement, which was adopted by other countries in a staggered manner over time. Isolation led to an increase in the consumption of content over the Internet. An exceptional situation, which accelerated a trend that authors such as Scolari (2016) had already detected prior to the pandemic, where the socio-cultural changes caused by cyberspace were already conditioning the emergence of new collective pedagogies, based on reciprocity and the exchange of roles, under the nomenclature of the prosumer.

Thus, Telefónica (2020) reported an increase in Internet traffic in Spain of 35%. The age range of those individuals who until the confinement spent more time surfing the Net was between 16 and 24 years old (Hidalgo-Pérez, 2020). According to data from Infinia (2020), on a sample of 30 million users, 78.2% of the content consumption habits of Spaniards were linked to the online environment. This study reveals that, among the categories that have been affected by this growth, are the downloading of applications related to online learning. As for the family composition, the study highlights that 21% of the total are households with children, where the female audience predominates (51.11%). And where a greater digital consumption of information was observed was from 18 to 25 years old, with 25.67%.

Such has been the volume of data demanded by users, with Netflix and YouTube as the two channels with the sharpest increase, that the European Commissioner for the Internal Market Thierry Breton asked Amazon Prime, Facebook, Netflix and YouTube, among others, to temporarily reduce the quality of their videos, in order not to overload Internet infrastructures (Nokia, 2020).

In order to monitor this upward trend in the consumption of streaming content, the European Regulatory Body for Electrical Communications agreed, together with the European Commission, to carry out a monitoring of the Internet traffic situation in the member states of the European Union (Berec, 2020).

2.2. YouTube, education and the Edutuber concept

The pandemic caused education professionals to redirect towards synchronous and asynchronous online channels, where YouTube stood out as one of the most used tools worldwide, putting in open, audiovisual resources linked
to both formal and informal multidisciplinary learning (Rangarajan et al., 2019). For example, in Spain, to alleviate these mobility problems, the platform 'Aprendo en casa' (Encinas-Martín, 2020) was created.

In this same line, where learning is democratized (Yammine et al., 2018), authors such as Cheng and Dong (2018) see YouTube as a new way to bring the scientific community closer together. They consider the possibilities of underpinning the open science concept, where disseminators, communicators and educators can share the advances achieved in their respective fields of research (Bautista-Puig et al., 2019). For Hargittai et al. (2018) the success of social networks as tools for knowledge acquisition, is based on the ability to interact and measure such activity, as this allows to extract guidelines towards a more effective e-learning model that favors personal learning environments, as a differential element for knowledge management, through education (Tur et al., 2016). Educators who use YouTube as a platform for the transmission of their teachings are called edutubers and enjoy great recognition within the educational community (Pattier, 2021).

Likewise, the manager of a channel as these characteristics can keep his subscribers informed by posting information on the wall of his community. Communities are also organized on the basis of interests, which leads to infinite forms of social relations, among which, within our field of study, the exchange of skills stands out. For example, among these communities, Vizcaíno-Verdú et al. (2019) speak of booktubers to refer to a set of literary channels where tastes and preferences on the subject are shared.

Everything seems to revolve around the Internet as a dissemination medium that channels and segments its contents, through online portals, online magazines, forums, podcasts and vlogs, among others (Amarasekara & Grant, 2018). This new context of searching for contrasted sources of knowledge, beyond formal training, is considered by Sulaimanu, et al. (2019) a gateway to informal learning and the cultivation of soft skills, where the user takes responsibility for their own learning by turning to different sources of information (Becker et al., 2017; Sharples et al., 2016).

2.3. Previous studies and hypotheses

Studies by Scolari (2016) indicate that, prior to the pandemic, there was already a tendency for internet consumption to grow, generating the emergence of new collective learning pedagogies. In addition, according to the Telefónica report (2020), internet traffic has increased by 30%. Therefore, we propose as a first hypothesis H1: Covid-19 confinement measures are the cause of the growth in accesses to e-learning portals and that the user accesses the contents of the e-learning channel after a period of adaptation.

Some authors (Chang, et al., 2019) consider that working in a multi-screen video streaming environment is growing due to the technological facilities and that it captures their attention better, improving the cognitive load of the learner in their self-learning. Taking into account that some brands, such as Apple, provide applications for the management of this environment (between Macintosh and Iphone) we put forward the second hypothesis. H2: There is a relationship between desktop devices and mobile devices or between their operating systems, which could be related to the consumption of content under the multiscreeen modality.

According to Hidalgo-Pérez (2020), the age range with the highest number of Internet activity is from 16 to 24 years old. In this new scenario, where teleworking has grown, we propose as a third hypothesis H3: The age of consumers of educational content increases during confinement, due to the need for autonomous learning.

It is intuited that users who access for the first time to an e-learning channel do it through their mobile device. Either by recommendation of another person, who provides them with the web address through social networks, or because they have first made a Google search of the topics that are of interest to them. This leads us to think that it is possible that the origin of the accesses is mainly from external sources to the portal studied and to consider a possible growth of external accesses due to new users who connect to the portal for the first time (Mojarro Alíaño et al., 2019).

3. Methodology

A quantitative analysis of the data has been carried out, initially the descriptive characteristics of the data have been studied and then a Bayesian Network has been modelled for the study of causality.

The stages are mainly based on a data analysis process and are as follows:

- Data collection: The sample to perform this work, was obtained from the data generated by the Zalathun channel using Google Analytics tool.
- Data preparation: First, an exploratory analysis of the data was performed, in which the basic statistics are obtained. Next, we proceeded to the selection of the attributes or variables needed for the causality analysis, as well as the transformation of the data to a required domain type.
- Tool selection: Different data analysis and intelligence tools were reviewed and tested (Structural Equations, Clustering, …) and finally, BN was selected as the most suitable for the study of causality relationships. The tool used for the analysis was OpenMarkov 0.4.0, with the Hill-Climbing structure learning algorithm and the K2 score evaluation metric.
- Model generation and evaluation: Preliminary BN models are obtained and finally, through various
graphics and the conditioned probabilistic analysis to the different periods, the models with the best results are established.

Analysis of the results: We proceed to the analysis of the results obtained, for which we resort to the observation and comparison of the metrics provided by the techniques. Subsequently, the interpretation of the BRs and the causal relationships obtained from the data sample is described.

3.1. Population and sample

This study analyses the effect of confinement on the number of accesses to the Zalathun portal in Spain in order to analytically justify the causality between the period of confinement (third period) and the strong growth of accesses to the portal. It also analyses which devices and operating systems have the greatest effect.

We have used as variables the ratios of accesses (of one day with respect to the same day of the previous year) with the devices and operating systems, and as a temporal space we have treated a reduced period (109 days) close to 13/03/2020, date on which the strict confinement in Spain (DC) began, that is, from March 1 to June 17, 2020. This study has also been carried out in the same period of 2019, allowing to compare the behaviour in the two years.

The variables that participate in the study (Table 1) corresponding to devices and operating systems were discretized, associating to each day the quartile in which the growth ratio per accesses in the day is found, therefore, the variables take values between 1 and 4.

Table 1. Variables involved in the study

<table>
<thead>
<tr>
<th>Devices</th>
<th>Operating Systems</th>
<th>Other Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>DESKTOP</td>
<td>WINDOWS</td>
<td>DiaRel</td>
</tr>
<tr>
<td>MOBILE</td>
<td>MACINTOSH</td>
<td></td>
</tr>
<tr>
<td>TABLET</td>
<td>ANDROID</td>
<td></td>
</tr>
<tr>
<td></td>
<td>IOS</td>
<td></td>
</tr>
</tbody>
</table>

Source: Authors, 2022.

For the temporary review we performed to the transformation of the data to a required domain type by defining the variable DiaRel as follows (DC = Confinement Start Date = 13/03/2020):

- DiaRel = 1 if date is before DC, from 01-03 to 12-03
- DiaRel = 2 if the date $\in [DC, DC +10 \text{ days}]$, from 13-03 to 22-03.
- DiaRel = 3 if the date $\in [DC +11 \text{ days}, DC +61 \text{ days}]$, from 23-03 to 12-05.
- DiaRel = 4 if the date is after the date DC+61 days, from 13-05 to 7-06

The periods have been determined using the following criteria:

- A pre-confinement period is taken to identify ‘normal’ ratio levels prior to the event (DiaRel=1).
- Ten days after the DC are considered, because it was a situation that had never been experienced before, which produced certain fears and confusion and required some time to assimilate. In addition, both the companies and the educational institutions took a few days to prepare their resources and define how to act after the CF. It was also necessary for individuals to adapt, as they had to provide themselves with the means to continue in some way their essential activities: work, studies, contact with family and friends, logistics at home. All of this has meant that in many homes the computer is already a device for individual use, even for children, and is one of the main tools for work or study (DiaRel = 2).
- The third period is formed by the days of strict confinement (except the first 10 days), so that it is possible to identify the access rates during those days (DiaRel = 3).
- The fourth period is constituted by the dates after the confinement, this allows to evaluate if some digitization needs have been maintained after the confinement, being the ratios in this period higher than the ratios of the previous year (DiaRel = 4).

3.2. Instrument used for sample collection

The access data for each device and operating system have been obtained from the statistical results provided by the Google Analytics tool on the Zalathun portal.

Google Analytics is a web analytics service offered by Google that tracks and reports website traffic, as of 2019 it is the most widely used web analytics service on the web. Google Analytics is used to track website activity such as session duration, pages per session, bounce rate, number of people using the site, along with information about the source of that traffic.

Google Analytics is implemented with “page tags”, in this case, called Tracking Code, this is a piece of JavaScript code that the website owner adds to each page. The tracking code runs in the client’s browser when the client...
browses the page (if JavaScript is enabled in the browser) and collects visitor data and sends it to a Google data collection server as part of a web beacon request.

A potentially negative impact on data accuracy comes from users deleting or blocking Google Analytics cookies. Any individual web user can block or delete cookies, resulting in a loss of data from those visits. Another limitation of Google Analytics for large websites is the use of sampling in generating many of its reports. To reduce the load on its servers and provide users with a relatively quick response to their query, Google Analytics limits reports to 500,000 randomly sampled sessions at the profile level for its calculations.

3.3. Choice of the methodological tool used. Adequacy and functionality

In the last decades, a multitude of methods have been developed that create Artificial Intelligence for different types of problems and applications. One of these methods are the so-called expert systems. One type of expert system widely used nowadays for describing causal dependencies is the one based on BNs. In essence, it consists of a graph that represents a set of known variables and the dependence relationships between them, in order to estimate the probability of unknown variables.

Among the existing techniques of intelligent data analysis we have used BN because we consider it to be the most appropriate for the study of causal relationships, as compared to other methods (Structural Equations, Clustering, Regression,...) it has the following advantages: in the case of having discrete variables (as in our study) it does not require distribution conditions to the variables, it shows any type of causal relationships (linear or not), it can include the knowledge of experts and provides the conditional probabilities (López Puga & García García, 2007).

3.4. Study development stages

A sample of 109 days close to the COVID-19 confinement in Spain (from 01/03/2020 to 17/06/2020) was analysed, and for the same days in 2019 and 2018.

Based on the data obtained from the channel, they are transformed into the required domain type and the DiaRel variable is defined with values from 1 to 4 depending on the period to which the access date belongs. To study the effect of the DiaRel variable on the number of accesses in the different devices and operating systems, the “Daily growth ratio” indicator of one year with respect to the previous one was used, in order to avoid periodic effects of the series, that is, to distinguish between the effect due to a period of confinement or the effect due to a period of repetitive growth in the series (for example, in each year: the schooling period or the Christmas or summer holiday period).

The “Daily growth ratio” has been defined as the number of accesses on a given day in year $X$ divided by the number of accesses on the same date in year $X-1$. Therefore, the calculation of the2020 ratio involves data from 2020 and 2019 and for the2019 ratio it involves data from 2019 and 2018. These two ratios allow a comparison to be made between the relative growth ratios of 2019 and 2020.

3.5. Study development stages

BNs were first introduced in the 1920s. Since then, they have been “reinvented” by a large number of researchers under different names, such as causal networks, or probability networks, among others. More recently they have been applied as alternatives to classical expert systems oriented to decision making and prediction under uncertainty in probabilistic terms. BNs belong to the set of techniques aimed at modelling (López Puga, et al., 2007)

BN have been applied to very different fields (medicine, law, environment, psychology, communication, ...), among them, decision making processes in automated learning models oriented to automatic tutoring, creating a human-machine communication that facilitates the development of the CSL learning model.

A BN is a probabilistic graphical model that makes it possible to represent relationships of probabilistic dependence and independence between a collection of data. It is defined by means of an acyclic directed graph, in which each node represents a variable and each arc a probabilistic dependence (Pearl, 2001). The nodes that are not connected represent variables that are conditionally independent of each other. Each node has a probability function associated with it that takes as input a particular set of values of the node's parent variables and returns the probability of the represented variable.

The probability functions of each variable or node in the network is characterized by a conditional probability table where the values that the variable can take depending on the values taken by the set of variables on which it depends are represented. In this sense, following Cowell et al. (1999), and considering, that we are working with an acyclic directed graph $G$ with a set of nodes $V$, for each $v \in V$ we must specify the conditional distributions of $X_v$ given its “parents” $\mathbf{X}_{\text{pa}(v)}$. If we assume that this density is $p \left( X_v | \mathbf{X}_{\text{pa}(v)} \right)$, the joint density function would be:

BRs provide a compact way of representing knowledge (López et al., 2005). They facilitate the communication of results as they are graph-based and allow explicit documentation of assumptions and uncertainties, which
makes them easier to understand and use than most modelling frameworks. In addition, they allow combining knowledge given by a human expert with data collected in a database and show dependency relationships between variables. They are based on Bayes’ Theorem of probability theory. They also allow uncovering the underlying causal structure in a dataset (Neapolitan and Morris, 2004). Furthermore, in the context of inference, BRs allow bidirectional inferences to be made; that is, from effects to causes and from causes to effects (Gámez, 1998).

However, despite their advantages, it is important to be aware of several limitations. BNs generally represent processes with continuous probability with difficulty. The distributions require conversion to an equivalent discrete space to facilitate computation. In addition, their ability to incorporate qualitative (and possibly subjective) data is often considered an advantage, however, the use of expert opinion is a potential source of bias.

In a BN, the information provided by one or more observed variables (evidence) propagates through the network and updates our belief about the unobserved variables. This process is called inference. In addition, the system “learns” the conditional probabilities that describe the relationships between the variables from the data.

A BN can be built manually, with the help of experts, or automatically, by applying learning algorithms. There are several algorithms to identify the network that best fits the data, such as the PC algorithm or the Hill Climbing algorithm, which use evaluation functions using metrics such as K2, BIC and BDeu (Carlson et al., 2007).

The K2 algorithm attempts to find an optimal network in terms of the likelihood of the database for each candidate network. In contrast, the PC algorithm attempts to determine the network structure through statistical tests of independence. Neither method is absolutely superior to the other and both operate with discrete variables. Finally, it is possible to combine expert knowledge with machine learning, thus allowing the obtaining of BNs well-adjusted to reality and to the data.

This analysis has been performed with OpenMarkov 0.4.0. using the Hill-Climbing structure learning algorithm and the K2 score evaluation metric. In this study we are not interested in fitting pairs (input, output), but in increasing the structural knowledge of the available data, so it is an unsupervised learning model and the tests that evaluate the accuracy of the prediction (confusion matrix, ROC curve, ...) will not be applied.

4. Results

4.1. Sample Statistic Description and comparisons in different periods

As indicated above, the Zalathun portal is accessed from different countries, mainly Spain and Latin America, with Spain having the highest number of accesses (11% in Argentina, 14% in Colombia, 41% in Spain, 22% in Mexico and 12% in Peru).

In all the countries indicated, the percentage of accesses by men is higher than that of women. The country where the difference is lower is Spain (61% men) and the country where the difference is greater is Peru (79% men).

Regarding the age distribution (Table 2), an increase in age in Spain is observed in the period of confinement with respect to the percentage of the last two years. This may respond, on the one hand, to the need to consult the portal for work in companies without having the support of the usual working group as available as in the face-to-face case. On the other hand, it is possible that some accesses by school-age minors are made through the computer of another adult nearby who provides the device and the user during the time of confinement.

Table 2. Percentage of accesses in Spain in different periods by age

<table>
<thead>
<tr>
<th>Age brackets</th>
<th>Period 09-2018/09-2020</th>
<th>Period 03-2020/06-2020</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age 13-17</td>
<td>0.00 %</td>
<td>0.10 %</td>
</tr>
<tr>
<td>Age 18-24</td>
<td>33.20 %</td>
<td>29.60 %</td>
</tr>
<tr>
<td>Age 25-34</td>
<td>32.70 %</td>
<td>30.00 %</td>
</tr>
<tr>
<td>Age 35-44</td>
<td>19.40 %</td>
<td>20.80 %</td>
</tr>
<tr>
<td>Age 45-54</td>
<td>11.10 %</td>
<td>14.00 %</td>
</tr>
<tr>
<td>Age 55-64</td>
<td>3.10 %</td>
<td>5.10 %</td>
</tr>
<tr>
<td>Age +65</td>
<td>0.50 %</td>
<td>0.50 %</td>
</tr>
</tbody>
</table>

Source: Authors, 2022.

Table 3. Percentage of accesses in Spain in different periods by origin of access


<table>
<thead>
<tr>
<th>Access origin</th>
<th>Period 09-2018/09-2020</th>
<th>Period 03-2020/06-2020</th>
</tr>
</thead>
<tbody>
<tr>
<td>External</td>
<td>43.60 %</td>
<td>42.40 %</td>
</tr>
<tr>
<td>YouTube Searches</td>
<td>16.00 %</td>
<td>12.50 %</td>
</tr>
<tr>
<td>Playlist</td>
<td>11.10 %</td>
<td>13.30 %</td>
</tr>
<tr>
<td>Suggested Videos</td>
<td>5.70 %</td>
<td>3.60 %</td>
</tr>
<tr>
<td>Other</td>
<td>23.60 %</td>
<td>28.20 %</td>
</tr>
</tbody>
</table>

Source: Authors, 2022.

In terms of accesses, the majority are made with a desktop device (82%). The next device with the highest frequency of accesses is the mobile phone (10%), while the number of tablets is much lower (1%). These proportions are very similar in the period 01/03/2020 - 17/06/2020, with mobile accesses increasing by 2 points (12%).

The most used operating systems are those associated with a Desktop (Windows 58% and Macintosh 20%). As for the operating systems associated with mobile (Android 7% and IOS 3%) increase their percentages in the period of the confinement environment, being the new values 9% and 3.5% respectively.

After an analysis of correlations, in the period 20/03/01 - 20/06/14, it is observed that there is a correlation between MACINTOSH and IOS (Pearson correlation coefficient=0.602 and significance at 0.05). H2 This could be due to the possibility provided by Apple to synchronize content between IOS and MACINTOSH, allowing the mobile phone to be used as a second screen.

As for the growth ratios in Spain for 2019 and 2020 in the period 01-March to 17-June, it can be seen in Figure 1 that from the tenth day of confinement (13-March) the speed of growth of accesses (ratio) is much higher than that observed in 2019 and that after confinement (fourth period) it remains higher than the year 2019.

**Figure 1. Accesses ratios, for 2019 and 2020 (growth rate)**

4.2. Causality study through a BR in Spain, comparing growth ratios in 2020 and 2019 during the period 1-March to 17-June

For the causality analysis between confinement period and increase of accesses, a discrete BN has been defined with a machine learning structure, with the Hill Climbing algorithm, using the variable DiaRel and the access ratios to the variables associated with the device and the operating system with which it has been accessed. This network has been compared with 2019 ratio data and the same network with 2020 ratio data.

To initially incorporate the expert knowledge, a BN pattern has been used (Figure 2, see legend) allowing the addition of new learned connections. The initial pattern has been created with the expert knowledge, which reflects the functional dependencies existing between the devices (DESKTOP and MOBIL) and the operating systems that support them (WINDOWS and MACINTOSH; ANDROID and IOS), connecting in turn the devices to the DiaRel variable. This structure allows us to study the effect of the different periods defined in the confinement environment on the growth rate of accesses through the different devices and operating systems. The TABLET device has not been included in the pattern because it has very few accesses.
To construct the BN, the growth ratios for each of the variables have been calculated, understanding the growth ratio as the quotient between the number of accesses per day in 2020 and the number of accesses per day in 2019 (in each of the variables considered). Subsequently, each variable has been discretised into 4 values, assigning values 1, 2, 3 and 4 corresponding to the number of the interquartile range to which it belongs. The values of the variable DiaRel are also 1, 2, 3 and 4 and correspond to the different periods within the confinement environment (described in section 3.1 population and sample). Subsequently, from the “learned network” and with the indicated pattern, using the OpenMarkov tool, we proceeded to obtain the associated BN and the probabilities of the conditional BN.

The results obtained for the year 2020 are also shown in Figure 2, showing the general structure of the network obtained. The conditional probabilities of each of the variables studied (devices and operating systems) for each value of the DiaRel variable (periods 1, 2, 3, 4) are shown in Table 3.

Table 4. Conditional probabilities of each variable to the different states of DiaRel for 2019 and 2020
The results indicate that in addition to the relationships predicted in the pattern, there is one more relationship DiaRel $\rightarrow$ MACINTOSH for 2019 and 2020 and two other relationships, the weak relationship between WINDOWS $\rightarrow$ MACINTOSH for 2019 and MOBILE $\rightarrow$ DESKTOP for 2020.

The DiaRel $\rightarrow$ MACINTOSH ratio was already contemplated in the initial pattern, but indirectly, through the DESKTOP variable. However, Table 3 shows that it is the variable with the greatest sensitivity to changes in the DiaRel variable, presenting a probability of 0.7 that the values of the MACINTOSH ratio are in quartile 4 for the period of confinement (DiaRel=3), and less than 0.15 for the rest of the values. That is, within the accesses with computer, those with an Apple have presented higher activity in the confinement period in 2020, however the opposite happens in 2019. (H2)

The relationship MOBILE $\rightarrow$ DESKTOP of 2020, presents a Pearson correlation coefficient of 0.534, significant at 0.05. When studying the conditional probabilities of the DESKTOP variable to the MOBILE variable, it is obtained that, when the MOBILE values are high, the DESKTOP variable presents a very high probability (0.83) of having a very high growth ratio. That is to say, in the confinement period in 2020, many accesses from MOBILE produced accesses in DESKTOP, perhaps as a consequence of a first search from MOBILE that later materializes from DESKTOP. (H2)

As for the results of probability conditional on values of the variable DiaRel, in 2020 all the variables present high values (quartile 4) in the period of confinement (DiaRel=3) with a probability between 0.42 and 0.70, much higher than the probability of value 4 in other periods of the same year and almost doubling the probability of that period in 2019 for all the variables. This indicates that there is causality between the increase in the number of accesses and confinement (H1).

At the beginning of the confinement (DiaRel=2), there is a slight increase in the ratio of the number of accesses in all the variables with respect to period 1 of the same year. It could be interpreted as a period of ‘adaptation’ in which the user begins to sense the need to ‘be autonomous’ in their training and begins to access the portal to study the suitability to their needs. (H1)

In the study, equivalent to the previous one, carried out for 2019 data, the behaviour is very different. Small variations in the conditional probability of each of the variables are observed when changing the values of DiaRel (Table 3), i.e., there is no dependence on the different periods considered in this variable, so it can be confirmed that the strong variation in 2020 is due to the different periods related to confinement, since the rest of the variations (trend, periodicity, ...) have been eliminated when dealing with the ratio of one year with respect to another and not with the number of total accesses.

Source: Authors, 2022.
5. Discussion and conclusions

Although, as Scolari (2016) indicates, the trend towards greater consumption of online content is widespread, our study confirms that this increase has been significantly higher during the confinement caused by Covid-19. Telefónica’s report (2020) on Internet traffic reveals that this has increased by 30%, most of it linked to the online environment (Infinia, 2020). This work not only corroborates the aforementioned data, but also focuses on the field of education, more specifically on the e-learning portal Zalathun Creatividad, providing evidence of causality between the different stages associated with confinement and the number of accesses to the channel. We have been able to establish two clearly differentiated periods. A first stage that we have defined as “awareness” or adaptation, which covers the 10 days before the confinement, where a small advance in the number of entries is already observed. And a second period, in the middle of the confinement, where, for reasons related to teleworking and online training, an upturn in the number of queries to the channel was observed.

Derived from the increase in visits caused by teleworking, it is also observed that the age range of the target audience of the channel has increased by 5%, moving from ages between 18 and 24 years, coinciding with the trend stated by Hidalgo-Pérez (2020), to ages between 35 and 64, with the range of 45 to 54 being the fastest growing during the period studied (H3).

This growth may be due to two aspects. Some minors with on-line training needs access with users of parents or adults who in other circumstances would not enter e-learning portals. On the other hand, professionals (over 35 years old) with the need to make reports or other studies without the possibility of having a face-to-face support from a colleague.

The BRs also provide information about the consumption habits of educational content. The results show a relationship between desktop and mobile devices or between their operating systems (H2), which confirms the multiscreen viewing of explanatory videos. In other words, everything seems to point in two directions. The first would have to do with the use of the mobile phone as an exploratory prelude to its subsequent consumption and replication of the proposed activity via e-learning, through a desktop or laptop, either PC or MAC. The second causal relationship would have to do with a synchronous use of both devices. In this way, the user would reproduce the contents from his smartphone and execute, at the same time, the tutorials from his computer. This variant completes the statements of Chang et al. (2019) by establishing synergies between the simultaneous use of screens.

Although throughout the history of the canal most of the accesses to it have come from external sources, a growth in these sources has been observed throughout the period of the confinement under study. The study has been carried out on a first period of confinement, however, we cannot consider the health effects of Covid-19 as finished, so we will continue to experience changes in the freedoms of mobility, and therefore in the consumer’s behaviour in different aspects.

As a limitation of the study, the small number of subscribers to the channel studied can be considered and it would be recommended as a future line of research to replicate the study to YouTube e-learning channels with a larger number of subscribers. It would be convenient, for its extrapolation to the consumption of e-learning content in general, to include in the study the information corresponding to other e-learning portals and to extend the period of study.

We are currently at the beginning of a new wave of Covid-19, although it is possible that we will not suffer a confinement as strict as that of March-May 2020, it is likely that we will suffer individual quarantines. It is expected that some of the social changes experienced in this period will be maintained over time, generating new forms of communication, training, and work performance. All this leads us to think about the possibility of carrying out new studies to describe these changes and to indicate in a predictive way the necessary actions to adapt education to the new protagonist age groups, as well as to the consolidation of multitasking and multiscreen consumption habits, pointed out in this work.
References


Hidalgo-Pérez, M. (2020, April 18). Dime quién eres y te diré cómo matas el tiempo en la red [Tell me who you are and I’ll tell you how you kill time online]. Diario El País. https://tinyurl.com/ycjn5dlw


