Paakat
Revista de tecnología y sociedad

Estimation of Internet availability at home by a score prediction function

Estimación de la disponibilidad de internet en casa por medio de una función de predicción score

http://dx.doi.org/10.32870/Pk.a12n23.735

Djamel Tourdet*
http://orcid.org/0000-0003-2833-4128
El Colegio de la Frontera Norte, México

Received: January 21, 2022
Accepted: August 03, 2022

ABSTRACT

The availability of Internet at home is one of the most important steps to enjoy the benefits that come from its use. In this sense, knowing the weight implied by the availability or lack of the network becomes strategic in the design of policies and actions to intensify the socio-territorial penetration of the Internet and the empowerment of its users. The aim of this research was to develop a function for predicting Internet availability in households, based on microdata from the 2020 population and housing census conducted by the National Institute of Statistics and Geography (INEGI). This type of prediction, in addition to clarifying the socio-territorial conditions involved in the physical availability of the Internet in homes contribute to the structuring of an intervention context for the public and private factors of digital disengagement. In methodological terms, the prediction function is based on a random sample of 11 000 census interviewees, divided into two groups of equal consistency: the first has Internet at home and the does not. To generate the prediction, a linear discriminant function was used and valued by a score function. The quality of the prediction was estimated through an internal evaluation with Bootsrap (75% of the sample) and an external one (25% of the sample), both of which were complemented with an analysis of the ROC curve and LIFT graph. The resulting function exhibits a sensitivity of 76.9%, a specificity of 20.5% and a discriminant capacity of 86.7% in a 95% confidence interval. With these efficiency parameters, the found function becomes a possible effective prediction tool for both reflection and action against the digital divide in its initial stage.

Keywords
Internet at home;
digital divide;
predictive function.

RESUMEN

Disponer de internet en casa es uno de los pasos más importantes para gozar de los beneficios que se desprenden de su uso. En este sentido, conocer el peso que implica la disponibilidad o la carencia de la red se vuelve estratégico en el diseño de políticas y acciones que permitan intensificar la penetración socioterritorial de la web y el apoderamiento por parte de los usuarios. Esta investigación tuvo por objetivo elaborar una función de predicción de la disponibilidad de internet en los hogares, con base en los microdatos del censo de población y vivienda de 2020 realizado por el Instituto Nacional de Estadística y

Palabras clave
Internet en casa;
brecha digital;
función predictiva.

* PhD in Geography, urban planning and territorial planning from the Instituto de Altos Estudios de América Latina, University of Paris III. Professor-Researcher of the Departamento de Estudios Urbanos y Medio ambiente of El Colegio de la Frontera Norte. Specialist in the issues of the digital divide, territorial and ecological planning and the development of tools for the analysis and diagnosis of socio-environmental contexts.

Geografía (INEGI). Este tipo de predicciones, además de aclarar las condiciones socioterritoriales que intervienen en la disponibilidad física de la red en las viviendas, contribuyen a la estructuración de un contexto de intervención para los factores públicos y privados del desenclave digital. En términos metodológicos, la función de predicción se fundamenta en una muestra aleatoria de 11 000 entrevistados del censo, repartidos en dos grupos de igual consistencia: el primero cuenta con internet en casa y el segundo carece de este servicio. Para la generación de la predicción se utilizó una función lineare discriminante valorada por una función score. La calidad de la función fue estimada a través de una evaluación interna con Bootsrap (75% de la muestra) y otra externa (25% de la muestra), ambas fueron complementadas con un análisis de la curva ROC y una gráfica de LIFT. La función resultante exhibe una sensibilidad de 76.9%, una especificidad de 20.5% y una capacidad discriminante de 86.7% en un intervalo de confianza del 95%. Con estos parámetros de eficiencia, la función encontrada se convierte en un posible instrumento de predicción eficaz tanto para la reflexión como para la acción en el combate a la brecha digital en su etapa inicial.

INTRODUCTION

Similar to what happened with the dissemination of the land line telephone, social and economic actors, in the broadest sense of the word, saw in the Internet a transcendental service that could be rapidly projected within the framework of universal consumption (Graham, 2008; Toudert, 2013; Warf, 2001).

This objective of globalizing access to the network was raised by a legitimacy acquired thanks to the haste of the massification of the internet, a service that conditions access to the benefits of the “digital revolution” and its corollary: the information society (Castells, 2002; Graham, 1998; Eynon *et al*., 2018). Regardless of the relevance of the socio-technological postulates associated with information and communication technologies (ICTs), for the dominant discourse in these contexts, the availability of the web has become an essential step for social integration.

The growing contraction of the analog world makes the social fracture visible between those who have access to the channels of digital supply, those who are waiting to enter and those who will inevitably be left out (Ghobadi & Ghobadi, 2015; Graham & Marvin, 2001). The crossover between the logics of interest –related to the development of the network provision market– and of values –with social inclusion to digital services– strongly intervenes in the first stage of the digital divide: the material availability of the Internet, which comprises the basic foundation of any reflection or action aimed at fostering the appropriation of ICTs and access to their benefits (Ali *et al*., 2020; Toudert, 2019). This last statement was amply validated during the long confinement due to the covid-19 pandemic, a period in which schoolchildren had to have internet service at home to continue their studies (Özge *et al*., 2022).

Internet availability, which often seems to be a situation that has been overcome in the dominant discourse, encounters serious limitations in some individual, social and territorial realities. At present, a significant proportion of the population does not have this service in their homes, which limits their possibilities of connectivity to networks in public, work and school spaces.

In this sense, having internet access in homes still is in national contexts a central objective for public and private policies involved in digital unlocking (Middleton & Chambers, 2010; Pick *et al*., 2014; van Deursen & van Dijk, 2019). When accessibility, privacy and mentoring converge in a residence, it becomes the preferred space to take advantage of all types of network content (Loh & Chib, 2021; Zhao *et al*., 2010).

Thus, having internet at home for an adequate appropriation of ICTs becomes an objective within initiatives and programs in the public and private spheres (Toudert, 2015; Zeqi *et al*., 2019). Knowledge of the socio-territorial contexts that surround the unavailability of the network acquires a strategic value in the design of appropriate policies to address this problem.

The prediction of Internet availability in households through sociodemographic information and housing conditions of the inhabitants is one of the instruments that allow consolidation of the understanding of the different scenarios and can be of great support in the processes of reflection and decision making. Knowing the network access status of individuals and families is a necessary step to undertake in the field of digital unlocking in a successful way.

This study is aimed to develop a discriminating function that by means of a score allows predicting, with a high margin of confidence, the state of internet availability at home, based on microdata from the 2020 population and housing census conducted by the National Institute of Statistics and Geography (INEGI by its acronym in Spanish, 2020).

The resulting prediction function is described in the results section and is developed with the intention of immediately becoming a pilot test, by allowing to detect mainly the unavailability of the web in residences in the framework of the different related socio-territorial contexts.

Theoretical framework

Digital divide and Internet availability at home

Since its appearance in the 1990s in the framework of the four Falling Through the Net studies published by the US Department of Commerce in 1999, the concept of digital divide has been changing around an invariant meaning (Toudert, 2015; van Dijk, 2006). This conceptual constancy stems from the positioning of its meaning in the social and territorial inequalities of access to the opportunities offered by ICTs and to the benefits of the information society (Castells, 2002; Ghobadi & Ghobadi, 2015).

The dynamic perspective of the digital divide alludes to phenomenology, including both the evolution of ICT artifacts and services, and the stages of their individual and social appropriation (Dupuy, 2007; Toudert, 2018; van Deursen & van Dijk, 2015). Thus, the digital divide appears as a complex conceptualization effort involving material, cognitive and social aspects (De Haan, 2004).

From the perspective of internet appropriation, the digital divide intervenes in a multifaceted way: it starts with the barriers of accessibility to the service, continues with the lack of skills for access to digital artifacts and content, and ends with the limitations that intervene in the beneficial use of the network (Toudert, 2018; van Deursen & van Dijk, 2015). Alongside this gradual perspective of the digital divide, and regardless of whether individuals' motivation to use ICTs is important or not (Reisdorf & Grošelj, 2017), service availability is a sine qua non condition for network accessibility (Martínez Domínguez, 2020; Deursen & van Dijk, 2015).

Internet availability suggests a physical situation where there is the possibility of connecting an intermediary device, such as a computer, to the network service, whether it is provided via line, cable, air or satellite. Although this statement may seem obvious today, the confusion between *availability* and *accessibility* led the liberal perspective to consider the digital divide as a product of a lack of motivation on the part of the digitally marginalized population (Graham & Marvin, 2001).

In this sense, availability is the initial stage of the digital divide, as it involves the physical lack of internet service in the place where access to the network is intended (Toudert, 2015; Zeqi *et al*., 2019). This lack is due to several factors that can act individually or together; some of the most important are those related to the market for internet provision, people's socioeconomic conditions and the interest to hire the service when it is available (Barzilai-Nahon, 2006; Reisdorf & Grošelj, 2017; Zeqi *et al*., 2019). Indeed, as Barzilai-Nahon (2006) states, we did not lose sight of the fact that the network and associated technologies, both in their structure and content, are a social and political product.

The inability of the market to provide efficient internet service to all social segments, regardless of their location, constitutes one of the main barriers to its availability at points of greatest stay such as home, school and work (Ali *et al*., 2020; Schleife, 2010; Toudert, 2013), transcendent places for the appropriation of ICTs (Martínez Domínguez, 2020; Toudert, 2018). Network unavailability seems to prevail in contexts of poverty and marginalization, as well as in rural areas of low population density, remote regions and spaces in urban peripheries, confirming a supply still related to large consumption gaps (Graham & Marvin, 2001; Pick *et al*., 2014; Toudert, 2013).

Of course, technological development allows the eruption of alternative platforms to online connectivity in these marginalized spaces, such as low-debit mobile or satellite connection; however, low quality and high cost are still negative factors that come into play when considering contracting this type of service (Graham, 2008; Toudert, 2013; Martínez Domínguez, 2020; Zeqi *et al*., 2019). Under this perspective, it would seem that the lack of internet at home depends on population stratification, which is related to low population density areas where the consistent representation of demographic and socioeconomic communities less likely to use the internet affects the local rate of social penetration and appropriation of ICTs (Pick *et al*., 2014; Toudert, 2013).

In these scenarios, the rates of young people, students, migrants, workers and inhabitants who intensively use technology in their work are often underrepresented (Büchi *et al*., 2016; van Deursen *et al*., 2015). Added to this is the refusal of internet providers to supply the service in this type of areas, the low motivation of some residents to contract the service and even the development of a refractory attitude towards ICT, situations that can be observed as significant barriers when considering the availability of the network in homes (Eynon *et al*., 2018; Lindblom and Räsänen, 2017).

In Mexico, Mora-Rivera and García-Mora (2021) found that the lack of internet in rural spaces translates into difficulties for poverty reduction, compared to urban spaces. In concrete terms, in the Mexican case the set of these limitations translates during 2020 into a relative differential of 27% between urban and rural areas with respect to internet and computer use (INEGI, 2020).

In this national comparison, the gap in internet usage rates between high and low socioeconomic strata reaches 56.6%, 54.4% for computer use and 33.5% for cellular telephony. These figures alone imply a lack of technology use that is still typified by sociodemographic characteristics, which are often confused with poverty and socio-territorial marginalization (Toudert, 2013, 2015; Martínez Domínguez, 2020). The unavailability of the Internet in these contexts is a fact observed since the beginning of public access to this network of digital services (Castells, 2002; Graham & Marvin, 2001).

Although in these scenarios, ways of accessing the Internet have been acquired in accordance with technological advances and the reduction of the costs of alternative connectivity, –which is still not adequate to make available all the content available on the network (Graham, 2008; Toudert, 2013)– the rate of private homes without Internet in Mexico remains high, close to 48% in 2020 (INEGI, 2020). This limitation becomes more discriminating when considering the preponderant role of home connectivity in the access to the different contents hosted on the web.

This is because consulting the Internet at work, school or in public spaces is often accompanied by objective and subjective restrictions that reduce the time available for interaction, the possibility of having some kind of advice and, above all, the freedom to access the entire digital offer of the network (Loh and Chib, 2021; Zhao *et al*., 2010). These shortcomings of internet availability are worsened by service conditions, such as bandwidth and signal regularity, constraints that are taken into account in the accessibility phase of the digital divide (Dupuy, 2007; van Deursen *et al*., 2015).

With all of the above in mind, both the availability and unavailability of Internet service seems to be typified by contextual, sociodemographic, economic and housing condition characteristics with the potential to be predicted in order to generate, among other advances, public policies aimed at improving the mechanisms of socio-territorial appropriation of ICTs.

Internet forecasting: a tool for reflection and action

The prediction of Internet availability at home acquires a strategic value in a general context convinced that the universal coverage of this service is eminent for overcoming the presential limitations of socio-territorial development and integration. Indeed, these visions founded on an assumed digital opportunity that would come to the rescue of communities left on the margins by the "old economy" and the analog world are considerations that often overstress the potential social and territorial impacts of ICTs (Graham, 1998; Toudert, 2015).

This type of perspectives is supported by a discourse imperatively structured around technological transcendence, the determinism of the digital landscape and the substitution of real for virtual space that has been saturating academic reflection and public and private operators involved in the development of ICT appropriation (Graham, 1998; Graham & Marvin, 2001).

It is exactly this type of consensus, which brings together factors coming from different fields and interests, that provides the universality of Internet access with a moral and legitimate sustenance that enthuses the crusades undertaken in the name of the digital divide (Bertrand, 2001; Toudert, 2014). From another perspective, the massification of the service makes it possible to address the iniquities that can arise in times of crisis, such as those experienced during the early stages of the covid-19 pandemic, when part of the schooled population was unable to continue their online activities (Özge *et al*., 2022).

The universality of access to ICTs and mass integration to the benefits of the information society comprise a new paradigm for the convergence of different social agents to promote public action in the direction of digital disengagement (Castells, 2002). Within the framework of these generally voluntary actions, it is clear that the availability of the network is the basic foundation for the adoption and development of technologies.

This is why, in places where the population has the longest stay, the availability of the web is positioned as a transcendent task for all the actors involved in the fight against the digital divide (Ali *et al*., 2020; Toudert, 2013). Contrary to the complexity of implementing collective initiatives to address the other limitations of the digital divide –such as accessibility, digital skills and the beneficial use of the network (Deursen & van Dijk, 2015; Martínez Domínguez, 2020)– these public actions are characterized by an operability that allows the provision of the service to be generalized.

In a recent literature review, Lythreatis *et al.* (2022) mention that most of the latest studies focused on level two of the digital divide, which leaves the issue related to network availability with little research, despite the fact that the still unconnected population remains very important in countries such as Mexico (about 40% of inhabited dwellings, INEGI, 2020).

Digital unleashing projects and programs that are usually undertaken by government or mixed initiatives have been characterized by the fact that they are generally based on predictions that make it possible to forecast their impacts and chances of success. Within the framework of these dynamics, the prediction of ICT appropriation trajectories by means of usage functions of these tools and services is a common practice for both reflection and action.

In this sense, the advent of the unified model for technology adoption and use (UTAUT) is the most widely employed conceptual synthesis tool for predicting ICT adoption and use behaviors (Venkatesh *et al.*, 2012).

In the same group of prediction tools also appear some variants of the UTAUT and, above all, of the staged accessibility to technology (SAT) model that aims to explain the trajectory of ICT use through the four stages of the digital divide: motivation, access, skills and use (Middleton & Chambers, 2010; Toudert, 2019; van Deursen & van Dijk, 2015). In the latter model, van Deursen and van Dijk (2015) include internet availability as an intrinsic dimension of users' material accessibility to ICTs.

Recently, Yu *et al.* (2018) proposed, based on a broad critical reflection, a comprehensive model that allows the analysis of digital inequities that generate the different facets of the digital divide; in this, internet availability is labeled as a micro cause that restricts the social appropriation of ICTs. In fact, the very conceptualization of barriers and limitations to the appropriation of technologies, as stated by Castells (2002), van Deursen and van Dijk (2015) and Toudert (2019), refers to the different phases of the digital divide whose initial stage is characterized by the availability of the internet service, which made ICTs possible in the face of the other artifacts and services.

In addition to user-level modeling through structural equations, as in the cases of Toudert (2019) and van Deursen and van Dijk (2015), regression or factorial models have been developed with a predictive interest in the availability, adoption and use of the Internet. These models are generally developed for territorial contexts, seeking to understand the global and regional appropriation of ICTs by means of explanatory variables from the sociodemographic, economic and political sphere (Mubarak *et al*., 2020; van Deursen & van Dijk, 2019), and in general terms, they highlight the impact on ICT appropriation processes according to age, gender, educational level, income, among others.

However, despite the existence of the aforementioned modeling experiences, there are few studies that are specifically focused on the prediction of Internet availability in households from the variables that are recurrently included in INEGI's population and housing censuses, even when this microdata allows carrying out exploratory processes focused on generating robust predictive functions to determine at the level of individuals the state of Internet availability in homes.

Data and methodology

Generation of the basic prediction function

In general terms, this project applied exploratory statistical and probabilistic prediction methods, widely known and used in the area of social sciences. To carry out the research, microdata from INEGI's 2020 population and housing census were used. From the general universe of those interviewed by the census, a total of 11,000 cases were randomly selected, 5,500 respondents who said they had internet in their homes and another 5,500 who said they did not have the same service. This limitation of the data allows us to operate with a representative and sufficient sample, without having to bear the statistical and operational defects that come with considering a large amount of data.

The selection of variables included in the research was carried out by means of a discriminating analysis on an extensive set of variables that present a thematic and epistemological proximity to the problem of Internet availability (Toudert, 2013). From this, only the discriminating variables that showed an X2 with $P<0.05$ and a test-value largely higher than two, which is a minimum floor for significance, were chosen for the following steps. The retained variables are shown in Table 1, together with their individual quantification: those who have Internet at home in column A and those who do not have this service in column B.

Next, the retained variables were incorporated into a multiple correspondence analysis in order to transform their discrete modalities into continuous factors, an imperative to be able to subsequently operate the calculation of Fisher's linear discriminating function (Droesbeke *et al*., 2005; Steyerberg *et al*., 2010). In this way, about 46 factors were obtained, of which 32 were retained and shown to be statistically significant at $P < 0.05$ and totaling 75.11% of the total variance explained, which characterizes an important level of extraction for the social sciences.

These factors were the basis for the calculation of the linear discriminating function that supported the prediction function and the assignment, by means of a score function of coefficients to the modalities of the variables involved in the study (Steyerberg *et al*., 2010). Indeed, the previously determined factors play only a transitory role in transforming the modes of the variables involved in the study into continuous values, which requires the calculation of Fisher's linear discriminating function for the assignment of a score. For this stage, the basic prediction function was calculated with 75% of the cases of the general sample used in the research and then proceeded to external validation with 25% of the remaining cases.

For the transformation of the prediction function, the transformed coefficients (column D) corresponding to each modality described by the respondent are applied in a pilot test. The resulting coefficients will be summed linearly through an expression of the type: $S = \sum_{k=1}^{n} kxz$, with S characterizing the resulting score, x the variable in question and z the modality of the variable corresponding to each respondent. It should be noted that this same test can be displayed on a digital form to be filled out online or in face-to-face modality.

Evaluation of the predictive performance of the prediction function

The predictive efficiency of the basic function was evaluated through internal and external validation. The internal evaluation consisted of performing an estimation by means of the Bootstrap technique using 2,000 samples that yielded marginal differences with the basic estimate that allowed the prediction function to be accepted (Bleeker *et al*., 2003; Wahl *et al*., 2016). In addition to the above evaluation, it is unavoidable to operate an external validation that considers confronting the basic function developed with 75% of the cases with the result of the same function fed with 25% of the remaining cases of the total sample (Bleeker *et al*., 2003). Both validation procedures are complemented with efficiency indicators derived from the receiver operating characteristic (ROC) curve and the LIFT plot (Vuk & Curk, 2006; Xie & Qiu, 2007).

Table 1. Sociodemographic characteristics and discriminating and transformed coefficients

Variables	A	B	C	D	Variables	A	B	C	D
V5: Where do you receive your health care?					V8: Approved school level				
Not understood	0.83	1.38	3.64	2.23	None	2.33	6.75	-1.41	2.49
IMSS, ISSSTE, state ISSSTE	18.16	7.48	18.53	5.97	Preschool	2.41	3.20	-3.57	1.94
PEMEX, Defense or Navy	0.52	0.15	-3.99	0.31	Elementary	12.74	19.79	0.78	3.04
SSA, IMSS-BIEN-PROSP	14.72	30.29	4.82	2.52	Middle school	12.10	11.74	5.05	4.11
Clinic, private hospital	10.56	6.14	1.05	1.57	High School	9.80	5.78	6.01	4.35
Pharmacy office	4.55	3.92	-2.06	0.79	University	9.43	2.16	-11.3	0.00
Other place	0.66	0.65	-4.03	0.78	Specialty	0.16	0.01	2.97	3.59
V6: Do you consider yourself indigenous?					Postgraduate	1.36	0.23	-3.65	1.92
Yes	11.30	23.26	-4.06	0.00	V9: Can you read and write?				
No	39.03	26.41	2.05	1.54	Yes	48.14	40.96	0.75	1.91
V12: Position at work					No	2.55	8.36	-6.85	0.00
Employee or laborer	33.15	18.48	-0.78	1.03	V13: Do you have medical service?				
Laborer or day laborer	2.08	5.76	-4.88	0.00	Yes	31.87	10.42	1.25	1.20
Paid helper	2.17	3.35	-2.67	0.56	No	25.61	32.10	-3.52	0.00
Employer or employer	2.72	0.86	-1.17	0.93	V17: Size of locality (number of inhabitants)				
Self-employed worker	13.12	10.99	5.48	2.61	Less than 2 500	11.30	28.17	-3.61	0.00
Unpaid worker	2.08	5.23	-3.54	0.34	From 2 500 to 14 999	13.52	13.38	-1.56	0.51
V18: Floor material					From 15 000 to 49 999	10.25	5.19	0.17	0.95
Earth	0.85	5.62	-5.80	29.90	From 50 000 to 99 999	3.37	0.95	8.76	3.11
Cement or pavement	25.13	36.69	-3.34	30.52	100 000 or more	11.55	2.30	10.39	3.52
Wood, mosaic or other	24.02	7.69	7.72	33.30	V19: Number of rooms in housing				
V20: Do you have piped water?					1 to 2	5.92	16.82	-5.98	0.00
Inside the house	39.48	21.86	3.58	2.53	3 to 4	24.84	25.46	-1.95	1.01
Only in the yard or land	9.51	23.45	-6.50	0.00	5 to 6	15.03	6.71	7.92	3.50
No piped water	1.01	4.68	-0.90	1.41	7 to 8	3.21	0.83	12.46	4.64
V21: Does it have drainage or sew					> to 8	1.01	0.18	9.65	3.93
Public network	39.15	23.12	2.40	5.75	V22: Do you have garbage disposal?				
Septic tank or septic tank	8.73	16.42	-2.56	4.51	By truck or cart	42.48	29.58	1.97	3.63
Will flow into a ravine	0.39	1.14	-5.41	3.79	In a container	2.77	1.42	-6.27	1.56
It flows to a river, lake or sea	0.14	0.26	-20.5	0.00	Burned	3.71	16.75	-6.45	1.51
No drainage	1.60	9.06	-6.46	3.52	Buried	0.13	0.55	8.72	5.33
V24: Do you have a television?					At the public dump	0.75	1.00	10.03	5.66
Yes	47.72	39.18	1.23	2.36	Dumped elsewhere	0.16	0.71	-7.57	1.23
No	2.28	10.82	-8.16	0.00	V33: Monthly household income (pesos)				

V25: Do you have a computer, laptop or tablet?					No income	2.05	6.05	-	-
Yes	26.51	3.67	10.00	3.60	De 1 to 5 000	7.61	16.26	-3.32	0.02
No	23.49	46.33	-4.32	0.00	5 001 to 10 000	15.16	15.08	-3.07	0.08
V26: Do you have a land line?					10 001 to 15 000	9.48	5.44	2.24	1.42
Yes	24.68	3.95	9.29	3.27	15 001 to 20 000	7.06	2.33	8.82	3.07
No	25.32	46.05	-3.73	0.00	20 001 to 25 000	3.75	1.16	8.09	2.89
V27: Do you have a cell phone?					More than 25 000	7.37	1.21	10.38	3.47
Yes	48.38	35.07	1.47	2.24	V28: Do you have pay TV service?				
No	1.62	14.93	-7.42	0.00	Yes	28.65	13.67	5.27	2.30
V30: Do you have a video game console?					No	21.35	36.33	-3.87	0.00
Yes	7.25	0.32	18.21	4.96					
No	42.75	49.68	-1.49	0.00					

Notes: (A) % of modality in the total variable for those who have Internet. (B) % of modality in the total variable for those who do not have Internet. (C) Coefficients of the discriminant function. (D) Transformed coefficients.
Source: Prepared by the authors with data from INEGI (2020).

Results

The sociodemographic variables of housing and home possessions, shown in Table 1, were all found to be discriminating in the prediction function (test-value greater than 2, P<0.05), revealing a clear segmentation of the population between those who do and those who do not have Internet access. Indeed, as shown in columns A and B, the population that has internet at home seems to enjoy better socioeconomic integration conditions compared to those who do not have this service. In fact, those who have access to the network, in addition to revealing higher incomes, are also distinguished by better housing conditions and possessions, which tends to characterize more favorable socio-demographic contexts.

High level of education, leadership position at work and residence in larger places are contexts that seem to promote the availability of internet at home. The structure of this availability still seems to be typified in Mexico, regardless of whether age and gender barriers to ICT adoption have been decreasing (Toudert, 2015, 2019).

The development of a prediction function for home internet availability, such as the one described in the previous paragraphs, is based on contextual discrimination and must comply with certain validation parameters that give it predictive capacity and robustness. Thus, the development of the prediction function underwent internal and external validation tests that would allow it to be considered a reliable instrument for estimating Internet availability in homes.

Internal validation of the prediction function

The internal validation of the prediction function consisted of comparing the results of the basic calculation and the resulting estimate using the Bootstrap method of 2,000 samples (Bleeker *et al*., 2003; Wahl *et al*., 2016). The results enlisted in Table 2 show an overall predictive ability of the function produced with the basic estimate of 78.65% for well-classified cases and 21.37% for misclassified cases. In light of these findings, it is perceived that the basic calculation was correct in almost eight cases out of ten, with a higher capacity for those without Internet, with a differential rate of 5.26% of the cases.

The Bootstrap estimation of the prediction function produced marginal differences compared to the basic calculation; these are of the order of 0.09% for the well and poorly classified and a standard deviation of less than half a percentage point. At the level of the two groups analyzed, the estimation differences with the basic calculation are slightly higher for those without internet and these are not in excess of one tenth of a percentage point. From this perspective, the differences found are very marginal and, therefore, everything seems to suggest that the prediction function of the basic calculation is consistent with the Bootstrap estimation, thus meeting the requirements of the internal validation.

External validation of the prediction function

The external validation involving the test calculation with 25% of the overall sample shows few differences compared to the basic calculation of the function (see table 2); these are of the order of 0.05% for the total well classified and 0.06% for the total poorly classified. For the segment of those with internet, the difference with the general calculation strengthens the positively classified with 0.19%, while those without network service strengthen the poorly classified with 0.17%.

Table 2. Internal and external validation of the Fisher's linear discriminating function

Elaboration of the function	Basic calculation		Bootstrap estimation		Function check	Test calculation	
Discriminated groups	Well classified	Misclassified	Well classified	Misclassified	Discriminated groups	Well classified	Misclassified
With internet n %	3136 76.02	989 23.98	3131.44 75.95 [0.57]	992.52 24.06 [0.57]	With internet n %	3144 76.21	981 23.78
No internet n %	3353 81.28	772 18.71	3349.17 81.19 [0.55]	775.83 18.81 [0.55])	No internet n %	3349 81.18	776 18.89
Total n %	6489 78.65	1761 21.35	6481.61 78.56 [0.44]	1768.39 21.44 [0.44]	Total n %	6493 78.70	1757 21.29

Note: Standard deviation is indicated in brackets.
Source: Prepared by the authors with data from INEGI (2020).

Taking all these differences between the test calculation and the basic calculation of the function into account, which vary in all cases between 0.05 and 0.19%, marginal discrepancies are perceived that do not significantly alter the basic results. Thus, the proposed prediction function satisfactorily passes the external validity test and then proceed to review the level of prediction efficiency.

Efficiency of the prediction function and scoring assignment

The ROC curve analysis shown in figure 1 indicates a cut-off score of 61, characterized by a sensitivity of 76.9% and a specificity of 20.5%, corresponding to a discriminating capacity of 0.867. The latter defines the discriminatory power of the prediction function estimated with the area under the curve (AUC) at a confidence interval of 95%, exhibiting a predictive capacity to differentiate those who have internet from those who do not with a probability of 86.7%.

It is believed that this discernment capacity seems acceptable to support public or business policy decisions, this statement is confirmed by the predictive quality of the LIFT graph (Vuk & Curk, 2006), which allows corroborating that the prediction function was able to be right in 77.98% of the predictions with only half of the cases analyzed, managing to show a LIFT ratio of 1.63 at the point of the limit score of 61.

By setting a tolerated error rate of 10% for the prediction function, the result of the discriminating function calculation was used to determine the score function for each of the modalities of the variables involved (see column D of table 1). By means of this assignment, the cases under analysis were assigned to three intervals according to the total score obtained. The highest score (between 64 and 100) pertains to the cases that were predicted to have Internet at home, a low score oscillating between 0 and 56 indicates a prediction of cases without the aforementioned service, while a variable score between 56 and 64 characterizes the cases of undecided assignment that fail to include any of the first two categories mentioned (see table 3).

These same intervals define three prediction zones that yield, in the context of a tolerated error rate of 10%, an agglutination of 37% of cases in the green zone, 33.56% in the red zone and the remaining cases in the orange undecided zone (see table 3). With the same tolerated error rate, the prediction of having Internet at home is almost seven percentage points higher than the prediction of the cases that do not have the service. This difference, when the tolerated error rate is set at 10%, characterizes a prediction with greater efficiency in the forecast of cases that do not have internet compared to those that do have the service.

Figure 1. Quality and efficiency of the prediction function

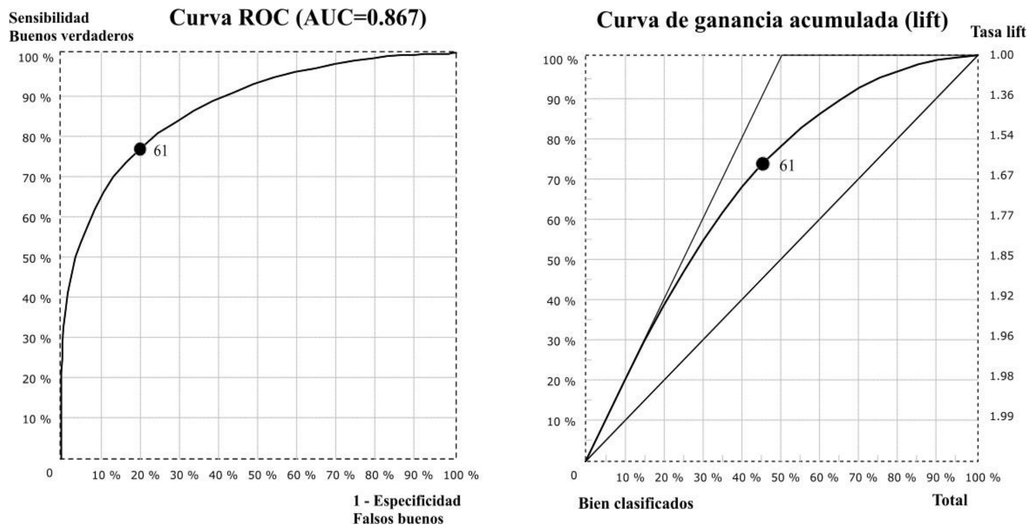

Table 3. Distribution of cases with a tolerated error rate of 10%

Zoning	Red zone [0, 56]	Orange zone [56, 64]	Green zone [64, 100]
With internet n %	551 10.02	1428 25.96	3521 64.02
No internet n %	3143 57.15	1807 32.85	550 10
Total n %	3694 33.56	3235 29.41	4071 37.01

Discussion of results

The findings of this research confirm the rationality of the prediction of Internet in households, which can be transformed into a tool that allows us to determine, at the level of individuals, the needs for network connectivity. This same tool gives rise to a reflection aimed at a planning practice that allows consolidating the conditions of inclusion for new beneficiaries of digital services.

From the perspective of the score function, a higher score given by each of the case studies results in an increasing possibility of having Internet service in homes (see Table 3). This prediction becomes positive from a value of 64 out of a maximum of 100 points, indicating the availability of the service with an effectiveness of a little more than 76% of the cases (see Table 2).

On the other hand, for scores below 64, the prediction indicates unavailability of the service with a success rate of more than 80% of the cases. The difference in the effectiveness of the prediction in favor of the cases that do not have the network service at home increases the utility value of the prediction function, which becomes a possible instrument for public and business policies on issues of inclusion to online digital services (Martínez Domínguez, 2020; Toudert, 2018).

The above becomes strategic at times when accessing the Internet from home becomes an urgent necessity to carry out school and professional (and other) activities, as happened suddenly at the height of the covid-19 crisis (Özge *et al*., 2022). These particular contexts do not diminish the usefulness of predicting which households have Internet service when taking the incidence of other aspects of the digital divide that may affect network users into account (van Deursen *et al*., 2015).

It is true that the lack of proper devices for connection, the lack of skills and conditions for a satisfactory use significantly reduce the network experience and its beneficial use (Loh and Chib, 2021; van Deursen & van Dijk, 2019). However, it should not be lost sight of the fact that about 48% of homes in Mexico still do not have internet, so that connectivity issues are one of the main national problems in the first stage of the digital divide (INEGI, 2020). In these scenarios marked by shortcomings in the appropriation of the digital instrument, the proposal of support actions to overcome these limitations gives the connection to the network a meaning of accessibility to the information society (Castells, 2002).

The potential affordability to increase the levels of social appropriation of the Internet and to promote its beneficial use is conditioned by the availability of physical connectivity to the network. From this perspective, the mere availability of internet service seems to structure an initial condition of digital divide, which tends to fluctuate according to social position and territorial location, defining different contexts for the appropriation of ICTs (Ghobadi & Ghobadi, 2015; Graham & Marvin, 2001).

Once the availability of the Internet is assured, other technical and contextual conditioning factors for accessibility to technologies take place. In the field of technical limitations to accessibility, we can list low bandwidth, inadequacy of query artifacts, which can slow down the interaction with different types of content available on the network, among other factors (Barzilai-Nahon, 2006; Middleton & Chambers, 2010). Regarding contextual constraints, ease of access to connected artifacts, frequency and time of use are recurrently pointed out in the digital divide literature (De Haan, 2004; Zeqi *et al*., 2019). In short, there is a need to overcome limitations to ICT accessibility and to have sufficient knowledge and skills to adequately interact with online digital content (Toudert, 2019; van Deursen & van Dijk, 2015).

Previous research has shown that having internet in different locations of the living space (home, work, school, public space) opens the possibility for greater appropriation of the network (Middleton & Chambers, 2010; Toudert, 2019). In fact, for Zhao *et al*. (2010) the availability of the connection at home allows for greater privacy for the development of personal activities on the web, in addition to the possibility of turning to close relatives to request support for advice on the use thereof. However, the availability of the Internet in the home does not strictly speaking imply accessibility and its beneficial use by each and every member of the household (Toudert, 2019; van Deursen *et al*., 2015).

In addition to the technical and contextual constraints, mentioned in previous paragraphs, other sociodemographic, economic, cultural and territorial variables often have an impact on both the availability and accessibility of technologies (Martinez, 2020; Mubarak *et al*., 2020). Column D of Table 1 shows the transformed coefficients of the discriminating function that make up the score function values given to each of the variable modalities participating in the study. Each time the score obtained for a modality is high, it would indicate that the respondent receives more points that allow him/her to claim the 64 entry points to the green zone, characteristic of those who have internet at home.

The score that characterizes the modalities of the variables involved in the prediction function is not the translation of a direct correlation with the availability of internet at home. In specific terms, the score of these variables is an indicator of a context of social and territorial marginalization that seems to condition the availability of the service.

The modes of the variable that contributes the most points to the function is the floor covering material, with scores per modality fluctuating between 29.90 for earthen floors and 33.30 for those covered with more elaborate materials (see table 1). Taking into account that the difference in points between modalities is not large, the total score is an indicator of the important discriminating value of the variable with a distribution that favors the best processed materials for the floor covering of the houses.

In terms of importance of the score given, the second variable is relatively far from the first one, pointing to the type of drainage and sewage of the house, which ranges from a score of 0 for a drainage leading to a body of water to 5.75 for houses connected to the public network. Here again, the distribution of the highest score seems to follow an increasing qualitative logic. In general terms, for the indicator variables of the forms of public services in housing, an increasing distribution prevails to a large extent according to the qualitative level of service available.

However, this logic does not seem to be confirmed for garbage disposal services, where less appropriate forms are flanked by a higher score, indicating a probable distribution independent of the context of marginalization –for this type of municipal services the provision depends to a large extent on the remoteness and age of neighborhoods and subdivisions–. For the other housing characteristics, an increasing number of rooms plays in favor of a higher score; the same also occurs with home possessions, where having a video game console is tributary to a higher score (4.96).

From the perspective of the social positioning of the case studies, having medical service preferably with the public institutions of IMSS, ISSSTE and state ISSSTE (5.97) seems to give a higher score, perhaps due to the significant density of beneficiaries who belong to the community with internet availability at home. Similarly, self-employed individuals score the highest (2.61), followed by employees and laborers (1.03) who make up the labor landscape in Mexico.

The availability of the network in the case of the self-employed could translate into professional use of the Internet and technologies, which seems to increase with the general population as a function of the level of schooling passed, particularly for secondary, high school and specialty (4.11, 4.35 and 3.59, respectively). Having a university level does not seem to stimulate web availability at home and in the current state of knowledge it is difficult to advance an inconclusive explanation in this regard, although one can speculate on the fact that this social fringe possesses internet in their workplaces and stays connected all day long with mobile data.

Web availability in homes seems to receive a higher score according to the increase in household income, which goes up to 3.47 for incomes above 25,000 pesos per month. Similarly, the score awarded grows linearly with increasing population size of localities, indicating that home internet availability continues to be a primarily urban occurrence, dominated by market density (Graham & Marvin, 2001; Pick *et al*., 2014; Schleife, 2010).

Overall, the variation of the score variation by mode of the variables involved in the study seems to incriminate almost in all cases conditions of socio-territorial marginalization conditioning the low social penetration of the network at home.

In general terms, the study's findings suggest that the network service continues to be socially typified even after more than two decades of being open to public access. This availability seems to be related to higher levels of education, consequent income, optimal conditions and specific housing possessions. These characteristics of social segmentation seem to be expressed more strongly in the context of population densification reflected in the size of towns.

Conclusions

In a context characterized by the growing importance of Internet service in private homes, understanding the logics of its availability and predicting it becomes an opportunity for the promotion of digital integration, from both the public and private spheres. At the same time, it is an advance in empirical knowledge that can nurture reflection on the dynamics of the social appropriation of ICTs. From a practical perspective, the prediction function makes it possible to locate the availability of the web in homes at a personal level, which contributes to the development of policies for the provision of the service, as well as to provide technical and didactic support for better use of the network when the service is available.

To the same extent, the prediction function can enhance coverage expansion projects, public infrastructure investment policies and the design of specific programs for digital unblocking. In this sense, using a business or government database in the prediction function generated allows determining, with high accuracy, whether or not the clients or beneficiaries of social programs have Internet at home.

This prediction was based on a score function that yielded three prognostic zones: negative cases between 0 and 56, undecided cases between 56 and 64, and finally positive cases between 64 and 100 points. The characterization by a score of the variable modalities reveals disparate variations. For example, the variable modes with a higher score that determine the prognosis of Internet availability in homes are related to a better quality of conditions, belongings and services inside the home. In the framework of this logic, the modalities associated with high household income, education levels, professional occupation and access to social security also stand out.

Finally, it should be noted that the vertical qualitative dynamics of these variables are generally accompanied by an increase in the score. From another perspective, both the increase in the size of localities and the assumption of social strata seem to have a positive impact on the prognosis of having a network in households. Taking all these findings together, it can be concluded that the availability of Internet service at home still seems to be influenced by socio-territorial conditions that limit the appropriation of ICTs.

REFERENCES

Ali, M.; Alam, K. & Taylor, B. (2020). Do social exclusion and remoteness explain the digital divide in Australia? Evidence from a panel data estimation approach. *Economics of Innovation and New Technology*, *29*(6), 643-659. https://doi.org/10.1080/10438599.2019.1664708

Barzilai-Nahon, K. (2006). Gaps and Bits: Conceptualizing Measurements for Digital Divide/s. *The Information Society*, *22*(5), 269-278. https://doi.org/10.1080/01972240600903953

Bertrand, N. (2001). Technologies d'information et de communication: quel rôle dans les dynamiques territoriales de développement? *Revue d'Economie Regionale et Urbaine*, (1), 135-152. https://www.cairn.info/revue-d-economie-regionale-et-urbaine-2001-1-page-135.htm

Bleeker, S.; Moll, H.; Steyerberg, E.; Donders A.; Derksen-Lubsen, G.; Grobbee, D. & Moons, K. (2003). External validation is necessary in prediction research: a clinical example. *Journal of Clinical Epidemiology*, *56*(9), 826-32. https://doi.org/10.1016/s0895- 4356(03)00207-5

Büchi, M.; Just, N. & Latzer, M. (2016). Modeling the second-level digital divide: A five- country study of social differences in Internet use. New Media & Society, 18(11). https://doi.org/10.1177/1461444815604154

Castells, M. (2002). *The Internet Galaxy. Reflections on the Internet, business, and society.* Oxford University Press.

De Haan, J. (2004). A multifaceted dynamic model of the digital divide. *Information Technology & Society*, *1*(7), 66-88.

Department of Commerce. (1999). Falling Through the Net: Defining the digital divide. http://www.ntia.doc.gov/legacy/ntiahome/fttn99/contents.html

Droesbeke J.; Lejeune M. y Saporta G. (eds.). (2005). *Modèles statistiques explicative pour données qualitatives*. Editions Technip.

Dupuy, G. (2007). *La fracture numérique*. Ellipses Marketing.

Eynon, R.; Seetjen, U. & Malmberg, L. (2018). Moving on up in the information society? A longitudinal analysis of the relationship between Internet use and social class mobility in Britain. *The Information Society*, *34*(5), 316-327. https://doi.org/10.1080/01972243.2018.1497744

Ghobadi, S. & Ghobadi, Z. (2015). How access gaps interact and shape digital divide: a cognitive investigation. *Behaviour & Information Technology*, *34*(4), 330-340. https://doi.org/10.1080/0144929X.2013.833650

Graham, M. (2008). Warped geographies of development: The Internet and theories of economic development. *Geography Compass*, *2*(3), 771-789. https://doi.org/10.1111/j.1749-8198.2008.00093.x

Graham, S. (1998). The end of geography or the explosion of place? Conceptualizing space, place and information technology. *Progression Human Geography*, *22*(2), 165-185. https://doi.org/10.1191/030913298671334137

Graham, S. & Marvin, S. (2001). *Splintering urbanism: Networked infraestructure tecnological mobilities and the urban conditions*. Routledge.

Instituto nacional de estadística y geografía (INEGI). (2020). Censo de población y vivienda, 2020. INEGI.

Lindblom, T. & Räsänen, P. (2017). Between class and status? Examining the digital divide in Finland, the United Kingdom, and Greece. *The Information Society*, *33*(3), 147- 158. https://doi.org/10.1080/01972243.2017.1294124

Loh, Y. & Chib, A. (2021). Reconsidering the digital divide: an analytical framework from access to appropriation. *Information Technology & People*, *35*(2), 647-676. https://doi.org/10.1108/ITP-09-2019-0505

Lythreatis, S.; Singh, S. & El-Kassar, A. (2022). The digital divide: A review and future research agenda. *Technological Forecasting and Social Change,* 175. https://doi.org/10.1016/j.techfore.2021.121359

Martínez Domínguez, M. (2020). La desigualdad digital en México: un análisis de las razones para el no acceso y el no uso de internet. *Paakat*, *10*(19). http://dx.doi.org/10.32870/Pk.a10n19.519

Middleton, K. & Chambers, V. (2010). Approaching digital equity: is wifi the new leveler? *Information Technology & People,* *23*(1), 4-22. https://doi.org/10.1108/09593841011022528

Mora-Rivera, J. y García-Mora, F. (2021). Internet access and poverty reduction: Evidence from rural and urban Mexico. *Telecommunications Policy*, *45*(2). https://doi.org/10.1016/j.telpol.2020.102076

Mubarak, F.; Suomi, R. y Kantola, S. (2020). Confirming the links between socio-economic variables and digitalization worldwide: the unsettled debate on digital divide. *Journal of Information Communication and Ethics in Society*, *18*(3), 415-430. https://doi.org/10.1108/JICES-02-2019-0021

Özge, K.; Elif, E. & Deniz, E. (2022). Internet access and its role on educational inequality during the COVID-19 pandemic. *Telecommunications Policy*, *46*(5). https://doi.org/10.1016/j.telpol.2022.102353

Pick, J.; Nishida, T. y Sarkar, A. (2014). Broadband utilization in the Indian states: socio-economic correlates and geographic aspects. En J. Choudrie, y C. Middleton (eds.), *Management of broadband technology innovation* (269-296). Routledge.

Reisdorf, B. & Grošelj, D. (2017). Internet (non-)use types and motivational access: Implications for digital inequalities research. *New Media & Society*, *19*(8), 1157-1176. https://doi.org/10.1177/1461444815621539

Schleife, K. (2010). What really matters: Regional versus individual determinants of the digital divide in Germany. *Research Policy*, *39*(1), 173-185. https://doi.org/10.1016/j.respol.2009.11.003

Steyerberg, E.; Vickers, A.; Cook, N.; Gerds, T.; Gonen, M.; Obuchowski, N.; Pencina, M. & Kattan. M. (2010). Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology*, *1*(1), 128-38. https://doi.org/10.1097/EDE.0b013e3181c30fb2

Toudert, D. (2013). La brecha digital en los contextos de marginación socioterritorial de las localidades mexicanas de más de 2500 habitantes: Exploración y discusión. *Comunicación y Sociedad*, (19), 153-180. http://www.scielo.org.mx/pdf/comso/n19/n19a7.pdf

Toudert, D. (2014). Evolución de la polarización territorial de la producción web: ¿para cuándo la esperada descentralización? *Revista Economía, sociedad y territorio*, *XIV*(45), 549-580. http://www.scielo.org.mx/pdf/est/v14n45/v14n45a9.pdf

Toudert, D. (2015). Brecha digital y marginación socioterritorial: el caso de México. En C. Rangel y G. Buzai (eds.), *Geografía Aplicada en Iberoamérica. Avances, retos y perspectivas* (343-370). El Colegio Mexiquense, AC.

Toudert, D. (2018). Connectivité et isolement dans le monde scolaire mexicain. *La Revue Communication, Technologies et Développement*, *5*, 116-126. https://doi.org/10.4000/ctd.374

Toudert, D. (2019). Brecha digital, uso frecuente y aprovechamiento de Internet en México. *Convergencia revista de ciencias sociales*, (79), 1-27. https://doi.org/10.29101/crcs.v0i79.10332

van Deursen, A.; van Dijk, J. y Klooster, P. (2015). Increasing inequalities in what we do online: A longitudinal cross-sectional analysis of Internet activities among the Dutch population (2010 to 2013) over gender, age, education, and income. *TelemaTIC and InformaTIC*, *32*(2). https://doi.org/10.1016/j.tele.2014.09.003

van Deursen, A. & van Dijk, J. (2015). Toward a Multifaceted Model of Internet Access for Understanding Digital Divides: An Empirical Investigation. *The Information Society*, *31*(5). https://doi.org/10.1080/01972243.2015.1069770

van Deursen, A. & van Dijk, J. (2019). The first-level digital divide shifts from inequalities in physical access to inequalities in material access. *New media & society*, *21*(2), 354- 375. https://doi.org/10.1177/1461444818797082

van Dijk, J. (2006). Digital divide research, achievements and shortcomings. *PoeTIC*, *34*(4), 221-235. https://doi.org/10.1016/j.poetic.2006.05.004

Venkatesh, V.; Thong, J. & Xu, X. (2012). Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. *MIS Quarterly*, *36*(1), 157-178. https://doi.org/10.2307/41410412

Vuk, M. & Curk, T. (2006). Lift Chart and Calibration Plot. *Metodoloski zvezki*, *3*(1), 89-108. http://mrvar.fdv.uni-lj.si/pub/mz/mz3.1/vuk.pdf

Wahl, S.; Boulesteix, A.; Zierer, A.; Thorand, B. & van de Wiel, M. (2016). Assessment of predictive performance in incomplete data by combining internal validation and multiple imputation. *BMC Medical Research Methodology*, *16*(1). https://doi.org/10.1186/s12874-016-0239-7

Warf, B. (2001). Segueways into cyberspace: Multiple geographies of the digital divide. Environment and Planning B: *Planning and Design*, *28*(3), 3-19. https://doi.org/10.1068/b2691

Xie. J. & Qiu Z. 2007. Bootstrap technique for analysis: A stable evaluation of Fisher classifier performance. *Journal of Electronics*, *24*(4), 523-527. https://doi.org/10.1007/s11767-005-0261-0

Yu, B.; Ndumu, A.; Mon, L. & Fan, Z. (2018). E-inclusion or digital divide: an integrated model of digital inequality. *Journal of Documentation*, *74*(3), 552-574. https://doi.org/10.1108/JD-10-2017-0148

Zeqi, Q., Shuqin, Z. y Shiding, L. (2019). From the Digital Divide to the Connectivity Dividend Difference: A Connectivity Capital Perspective. *Social Sciences in China*, *40*(1), 63-81. https://doi.org/10.1080/02529203.2019.1556475

Zhao, L.; Lu, Y.; Huang, W. y Wanga, Q. (2010). Internet inequality: The relationship between high school students' Internet use in different locations and their Internet self-efficacy. *Computers & Education*, *55*(4), 1405-1423. https://doi.org/10.1016/j.compedu.2010.05.010

HOW TO CITE

Toudert, D. (2022). Estimation of Internet availability at home by means of a score prediction function. *Paakat: Revista de Tecnología y Sociedad, 12*(23). http://dx.doi.org/10.32870/Pk.a12n23.735