

## The digital divide: An approach through machine learning classifiers

Andrés Aleán<sup>1</sup>, Manuel Vicente Nieto Mengotti<sup>2</sup>

<sup>1</sup>IDEEAS, Universidad Tecnológica de Bolívar, Colombia, <sup>2</sup>Department of Economics, Universidade da Coruña, Spain

---

### Abstract

*In 2022, 2.9 billion people worldwide lacked access to the internet, thus being unable to benefit from the digital economy (WEF, 2022). Moreover, lacking internet access at home can further exacerbate existing educational and economic inequalities. Thus, it is crucial not only to identify the sociodemographic profile of households that lack internet access, but of those most vulnerable to lacking internet access in the future (Hidalgo et al., 2020). This study applies several widely used machine learning classifiers (logit regression, naïve Bayes, linear discriminant analysis, k-nearest neighbors and random forest; James et al., 2021) to analyze the main socioeconomic internet access drivers for the Mexican population, using household surveys for the period between 2016 and 2020 (INEGI, 2020). Our principal result is that income, education level, and rurality are the main factors determining lack of internet access, both present and future; and that gender and occupation only play a secondary role in explaining the digital divide. These results can inform the formulation of public policies with the aim to secure universal access to the internet, and thus prevent the widening of existing inequalities in development.*

**Keywords:** *Internet access; Machine learning; Forecasting and nowcasting.*

---

This document was prepared within the framework of Manuel Vicente Nieto Mengotti's postdoctoral research stay at IDEEAS in 2022-I.

Hidalgo, A., Gabaly, S., Morales-Alonso, Uruña, A. (2020). The digital divide in light of sustainable development: An approach through advanced machine learning techniques. *Technological Forecasting and Social Change*, 150.

Instituto Nacional de Estadística y Geografía (INEGI) (2020). *Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH): Nota técnica*. INEGI.

James, G., Witten, D., Hastie, T., Tibshirani, R. (2021). *An introduction to statistical learning*. Springer.

World Economic Forum (WEF) (2022). *World Economic Forum Annual Meeting Report*. World Economic Forum.