

# Dirección de Desarrollo Digital

Unidad de Científicos  
de Datos



**El futuro  
es de todos**

**DNP**  
Departamento  
Nacional de Planeación

## ESTIMACIÓN DEL ÍNDICE DE DESARROLLO DE LAS TIC (IDI) POR DEPARTAMENTO Y MUNICIPIOS MEDIANTE MODELOS DE ESTIMACIÓN EN ÁREAS PEQUEÑAS

### Entidad

Dirección de Desarrollo Digital

### Sector

TIC

### Lenguaje

R.

### Fuente de datos

1. Terridata
2. SISBEN III
3. ECV – DANE
4. Panel Municipal del CEDE – Uniandes
5. Registros administrativos

### Presentación

El Índice de Desarrollo de las TIC (IDI), es un índice compuesto que combina 11 indicadores simples en una medida de referencia. Se utiliza para monitorear y comparar los desarrollos en las tecnologías de información y las comunicaciones (TIC) entre países a lo largo del tiempo. El índice está diseñado para ser global y reflejar los cambios que tienen lugar en países con diferentes niveles de desarrollo de las TIC. El proceso de desarrollo de las TIC y la evolución de un país para convertirse en una sociedad de la información se puede describir utilizando el modelo de tres etapas ilustrado en la preparación, intensidad e Impacto de las TIC que son las dimensiones que conforman el mencionado índice. En el presente trabajo se utilizan los modelos de estimación en áreas pequeñas con el fin de pronosticar el valor del IDI a nivel departamental y municipal empleando la información presente en la base de datos de la Encuesta de Calidad de Vida (ECV) del DANE, SISBEN III del DNP y datos de diferentes fuentes administrativas.

*The ICT Development Index (IDI), published annually since 2009, is a composite index that, until 2017, combined 11 indicators into a benchmark measure. It is used to monitor and compare developments in information technology and communications (ICT) between countries over time. The Index is designed to be global and reflect the changes that take place in countries with different levels of ICT development. The process of ICT development and the evolution of a country to become an information society can be described using the three-stage model illustrated in the preparation, intensity and impact of the ICTs that are the dimensions that make up the aforementioned index. In the present work, estimation models are used in small areas in order to forecast the value of the IDI at the departmental and municipal levels using the information present in the database of the Quality of Life Survey (QLS) of DANE, SISBEN III of the DNP and data from different administrative sources.*

### Objetivo general

Identificar un modelo de estimación en áreas pequeñas que a partir de diferentes registros administrativos genere pronósticos para el IDI a nivel departamental y municipal.

### Objetivos específicos

1. Identificar un conjunto de variables (a nivel departamental y municipal) que estén correlacionadas al IDI y que sean interpretables para el indicador.
2. Identificar los modelos de estimación en áreas pequeñas que permitan realizar los pronósticos del IDI a nivel departamental y municipal.

3. Desarrollar un algoritmo que permita al modelo de estimación en áreas pequeñas aprender de la información disponible en la ECV 2018 y así, generar estimadores consistentes de los parámetros de interés en el estudio, pronósticos adecuados (a nivel departamental y regional) que permitan medir la precisión de las estimaciones a nivel municipal.

### Metodología

El Índice de Desarrollo de las TIC (IDI), que se publica anualmente desde 2009, es un índice compuesto que, hasta 2017, combinó 11 indicadores en una medida de referencia. Se utiliza para monitorear y comparar los desarrollos en las tecnologías de información y las comunicaciones (TIC) entre países a lo largo del tiempo. Los principales objetivos del IDI son medir:

- El nivel y la evolución a lo largo del tiempo de los desarrollos de las TIC en los países y la experiencia de esos países en relación con otros.
- Progreso en el desarrollo de las TIC tanto en los países desarrollados como en los países en desarrollo.
- La brecha digital, es decir, las diferencias entre países en términos de sus niveles de desarrollo de las TIC.
- El potencial de desarrollo de las TIC y la medida en que los países pueden utilizarlas para mejorar el crecimiento y el desarrollo en el contexto de las capacidades y habilidades disponibles.

El Índice está diseñado para ser global y reflejar los cambios que tienen lugar en países con diferentes niveles de desarrollo de las TIC. Por lo tanto, se basa en un conjunto limitado de datos que se pueden establecer con una confianza razonable en los países en todos los niveles de desarrollo. El IDI reconoce las TIC como facilitadoras del desarrollo. El proceso de desarrollo de las TIC y la evolución de un país para convertirse en una sociedad de la información se pueden describir utilizando el modelo de tres etapas ilustrado en la Figura 1:

- **Etapas 1 - Preparación para las TIC:** refleja el nivel de infraestructura en red y el acceso a las TIC.
- **Etapas 2 - Intensidad de las TIC:** refleja el nivel de uso de las TIC en la sociedad.
- **Etapas 3 - Impacto de las TIC:** refleja los resultados de un uso más eficiente y efectivo de las TIC.

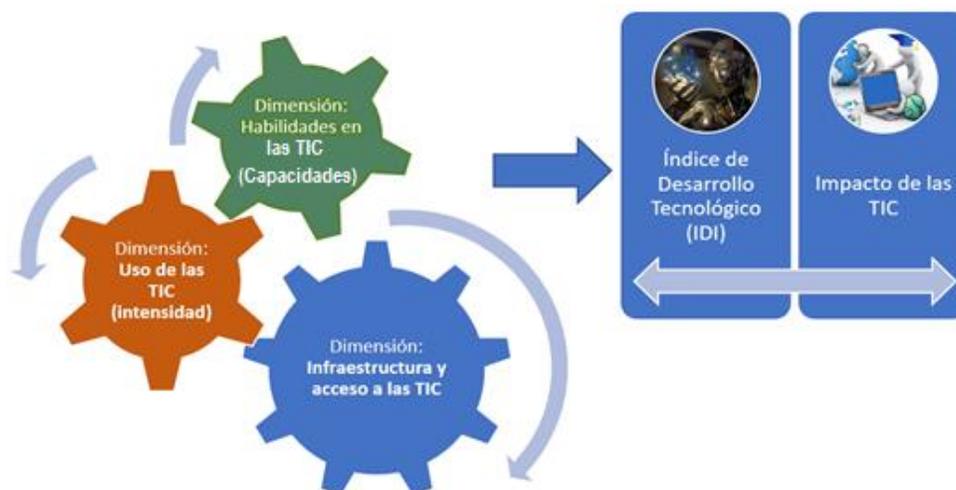


Figura 1: Tres etapas en la evolución hacia una sociedad de la información.

El IDI se divide en tres subíndices y un total de 11 indicadores (Figura 2).

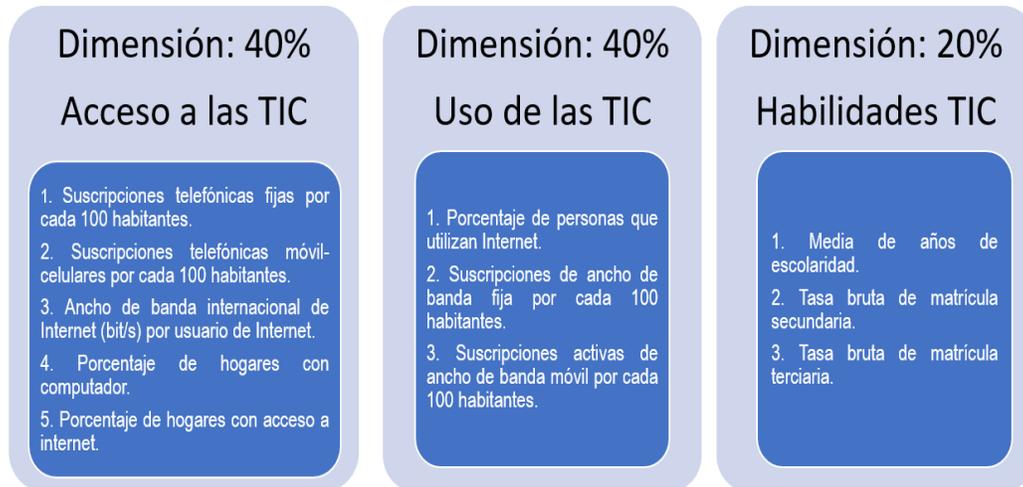


Figura 2: Composición del IDI y participación de las dimensiones.

- **Subíndice de acceso:** captura la preparación de las TIC e incluye cinco indicadores de acceso e infraestructura (suscripciones de teléfonos fijos, suscripciones de teléfonos móviles y celulares, ancho de banda internacional de Internet por usuario, hogares con computador y hogares con acceso a Internet).
- **Subíndice de uso:** captura la intensidad de las TIC e incluye tres indicadores de intensidad y uso (personas que utilizan Internet, suscripciones de banda ancha fija y suscripciones de banda ancha móvil).
- **Subíndice de habilidades:** busca capturar capacidades o habilidades que son importantes para las TIC. Incluye tres indicadores proxy (años promedio de escolaridad, inscripción a educación secundaria bruta e inscripción y educación terciaria bruta). Como se trata de indicadores proxy en lugar de indicadores que miden directamente las habilidades relacionadas con las TIC, este subíndice tiene menos peso que los otros dos subíndices en el cálculo del IDI.

A continuación, se proporcionarán más detalles sobre los pasos involucrados en el cálculo del IDI.

#### **Indicadores incluidos en el IDI**

La selección de indicadores se basó en ciertos criterios, incluida la relevancia para los objetivos del Índice, la disponibilidad de datos y los resultados de diversos análisis estadísticos, como el análisis de componentes principales (PCA). Los siguientes 11 indicadores están incluidos en el IDI (agrupados por Tres subíndices: acceso, uso y habilidades).

#### *Infraestructura TIC e indicadores de acceso*

Los indicadores incluidos en este grupo proporcionan una indicación de la infraestructura TIC disponible y el acceso de los individuos a las TIC básicas. Los datos para todos estos indicadores son recopilados por la International Telecommunication Union (ITU).

- *Suscripciones telefónicas fijas por cada 100 habitantes.*

El término "*suscripciones de teléfono fijo*" se refiere a la suma de las líneas analógicas activas de teléfono fijo, las suscripciones de Protocolo de voz sobre Internet (VoIP), las suscripciones de bucle local inalámbrico fijo, los equivalentes de canales de voz de la Red Digital de Servicios Integrados y los teléfonos públicos fijos.

- *Suscripciones telefónicas móvil-celulares por cada 100 habitantes.*

El término "*suscripciones de teléfonos móviles y celulares*" se refiere al número de suscripciones a un servicio público de telefonía móvil que proporciona acceso a la red telefónica pública conmutada que utiliza tecnología celular. Incluye tanto el número de suscripciones de postpago como el número de cuentas prepagas activas (es decir, las cuentas que han estado activas durante los tres meses anteriores).

- *Ancho de banda internacional de Internet (bit/s) por usuario de Internet*

El término "*ancho de banda internacional de Internet*" se refiere a la capacidad utilizada del total de ancho de banda internacional de Internet en megabits por segundo (Mbit/s). El ancho de banda de Internet internacional utilizado se refiere al uso promedio de todos los enlaces internacionales, incluidos los cables de fibra óptica, los enlaces de radio y el tráfico procesado por las estaciones terrestres de satélite y los telepuertos a los satélites orbitales (expresados en Mbit/s). Se tienen en cuenta todos los enlaces internacionales utilizados por todo tipo de operadores, es decir, operadores fijos, móviles y satelitales.

- *Porcentaje de hogares con computador*

"Computadora" se refiere a una computadora de escritorio, computadora portátil (portátil), tableta o computadora de mano similar. No incluye equipos con algunas capacidades informáticas integradas, como televisores inteligentes o dispositivos con telefonía como función principal, como teléfonos móviles o teléfonos inteligentes.

Hogar con una computadora significa que la computadora está disponible para que la utilicen todos los miembros del hogar en cualquier momento.

- *Porcentaje de hogares con acceso a internet*

Hogar con acceso a Internet significa que Internet está disponible para que lo utilicen todos los miembros del hogar en cualquier momento. Los datos se obtienen por países a través de encuestas nacionales de hogares y se proporcionan directamente a la ITU por las Organizaciones de Servicios Nacionales (OSN) o se obtienen por la UIT a través de su propia investigación, por ejemplo, de los sitios web de las OSN.

#### *Indicadores de uso de las TIC*

Los indicadores incluidos en este grupo captan la intensidad y el uso de las TIC. Los datos para todos estos indicadores son recopilados por la ITU.

- *Porcentaje de personas que utilizan Internet*

El término "*personas que usan Internet*" se refiere a las personas que usaron Internet desde cualquier ubicación y para cualquier propósito, independientemente del dispositivo y la red utilizados, en los tres meses anteriores. El uso puede ser a través de una computadora (es decir, computadora de escritorio,

computadora portátil, tableta o computadora de mano similar), teléfono móvil, máquina de juegos, televisión digital, etc.). El acceso puede ser a través de una red fija o móvil.

- *Suscripciones de ancho de banda fija por cada 100 habitantes.*

El término "suscripciones de banda ancha fija" se refiere a las suscripciones fijas para el acceso de alta velocidad a la Internet pública (un Protocolo de Control de Transmisión (TCP) / conexión IP) a velocidades descendentes iguales o superiores a 256 kbit / s. Esto incluye cable módem, DSL, fibra-al-hogar / construcción, otras suscripciones de banda ancha fija (cableada), banda ancha por satélite y banda ancha inalámbrica fija terrestre.

- *Suscripciones activas de ancho de banda móvil por cada 100 habitantes.*

El término "suscripciones activas de banda ancha móvil" se refiere a la suma de las suscripciones de banda ancha móvil de datos y voz y las suscripciones de banda ancha móvil solo de datos a la Internet pública. Cubre las suscripciones que realmente se utilizan para acceder a Internet a velocidades de banda ancha, no las suscripciones con acceso potencial, aunque este último puede tener teléfonos habilitados para banda ancha.

#### *Indicadores de habilidades TIC*

- *Media de años de escolaridad*

El término "años promedio de escolaridad" es el número promedio de años de educación de la población de un país, excluyendo los años dedicados a la repetición de años escolares. Se estima utilizando la distribución de la población por grupo de edad y el nivel de educación más alto alcanzado en un año determinado, y datos de series de tiempo sobre la duración oficial de cada nivel de educación.

- *Tasa bruta de matrícula (secundaria y terciaria)*

Según el Instituto de Estadísticas de la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura, la tasa bruta de matrícula es "la matrícula total en un nivel específico de educación, independientemente de la edad, expresada como un porcentaje de la población oficial elegible en edad escolar correspondiente al mismo nivel de educación en un año escolar dado".

A partir de la ECV se construirán indicadores aproximados a los definidos por la ITU y que se resumen en la siguiente Tabla 1.

Subíndice	Servicio	Indicador ITU	Indicador Propuesto
<b>Acceso a las TIC</b>	Telefonía	Suscripciones de telefonía fija por cada 100 mil habitantes	Porcentaje de hogares con telefonía fija
	Telefonía	Suscripciones de telefonía móvil por cada 100 mil habitantes	Suscriptores de celular
	Banda Ancha	Ancho de banda de Internet Internacional (bits/s) por usuario de internet	Ancho de banda de internet internacional
	Dispositivos	Porcentaje de hogares con computador	Porcentaje de hogares con computador
	Internet	Porcentaje de hogares con acceso a internet	Porcentaje de hogares con acceso a internet

<b>Uso de las TIC</b>	Internet	Porcentaje de personas que usan internet	Porcentaje de personas que usan internet
	Internet	Suscripciones de banda ancha fija por cada 100 mil habitantes	Porcentaje de hogares con internet fijo
	Internet	Suscripciones de banda ancha móvil por cada 100 mil habitantes	Porcentaje de hogares con internet móvil
	Internet	N/A	Frecuencia de uso de internet
	Dispositivos	N/A	Porcentaje de personas que usan computador
<b>Habilidades TIC</b>	Dispositivos	N/A	% de personas con teléfono celular inteligente
	Educación	Años promedio de escolaridad	Años de escolaridad
	Educación	Años promedio de escolaridad	Años de escolaridad
	Educación	Crecimiento de la tasa de matrícula secundaria	Cobertura de educación secundaria
	Educación	Crecimiento de la tasa de matrícula terciaria	Cobertura de educación terciaria
	Conocimiento	N/A	Actividades que sabe hacer en el computador
	Conocimiento	N/A	Para cuales servicios utiliza internet

*Tabla 1: Variables identificadas en la ECV que son empleadas para la estimación del IDI.*

Teniendo en cuenta los indicadores trazadores y las preguntas que permiten su conformación, se realizó un análisis de sensibilidad descriptivo del IDI con el cual se buscó determinar los cambios y estabilidad que podría sufrir el proceso de estimación en función de la variabilidad observada a lo largo del tiempo.

### **Modelos de estimación en áreas pequeñas**

Los modelos de áreas pequeñas (SAE, por sus siglas en inglés) corresponden a aquellos modelos de enlace explícito y efectos aleatorios específicos por dominio o área de interés, lo que permite cuantificar la variabilidad entre los dominios que no es capaz de explicar las variables auxiliares consideradas en el estudio.

En particular, los modelos de estimación en áreas pequeña involucran efectos aleatorios específicos por dominio de interés que explican la variación entre estos más allá de lo explicado por las variables auxiliares incluidas en el modelo. El uso de modelos SAE ofrece varias ventajas:

- (i) Los diagnósticos del modelo se pueden usar para encontrar modelos adecuados que se ajusten bien a los datos. Dichos diagnósticos del modelo incluyen análisis de residuales para detectar desviaciones del modelo asumido, selección de variables auxiliares y medidas diagnóstico para detectar observaciones influyentes y/o atípicas.
- (ii) Las medidas de precisión de cada dominio pueden asociarse con cada estimación de “área pequeña”, a diferencia de las medidas globales (promedios por dominio de interés) empleados por los estimadores sintéticos.

- (iii) Los modelos lineales mixtos, así como los modelos no lineales, como los modelos de regresión logística y los modelos lineales generalizados con efectos de área aleatorios, pueden ser empleados como parte del predictor lineal. Los modelos SAE permiten, adicionalmente, manejar estructuras de datos complejas, como la dependencia espacial y las series de tiempo.
- (iv) Los desarrollos metodológicos recientes para los modelos de efectos aleatorios se pueden utilizar para lograr inferencias precisas en modelos SAE.

Los modelos de estimación en áreas pequeñas se pueden clasificar en 2 tipos:

- Modelos de nivel agregado (área):  
Son aquellos que relacionan los estimadores directos en los dominios de interés con covariables que son específicas del área.
- Modelos a nivel de unidad:  
Son aquellos modelos que relacionan los valores unitarios de una variable de estudio con las covariables específicas de la unidad.

Resulta importante señalar que los modelos a nivel de unidad asumen que los valores dentro de la muestra en los dominios de interés obedecen al modelo poblacional asumido. Para los modelos de área se asume la ausencia de muestra informativa en los dominios de interés en situaciones donde solo algunos de estos son seleccionados en la muestra, esto implica que los valores obtenidos en las áreas muestreadas (estimadores directos) obedecen al modelo poblacional asumido.

En el presente proyecto se están empleando los modelos SAE a nivel área para realizar una estimación inicial del parámetro de interés (IDI a nivel municipal). Los mencionados modelos se definen mediante la siguiente ecuación:

$$\hat{\theta}_i = z_i^T \beta + b_i v_i + e_i$$

Donde

- $i = 1, 2, \dots, m$  denota el subíndice asociado a cada dominio o área pequeña,
- $\hat{\theta}_i$  denota el estimador del parámetro de interés  $\theta$  (IDI) en el  $i$ -ésimo dominio (municipio),
- $z_i^T$  denota el conjunto de variables auxiliares observadas para cada dominio específico. Para el presente proyecto denota la información auxiliar disponible para cada municipio de interés.
- $\beta = (\beta_1, \dots, \beta_p)^T$  denota el vector de coeficientes de regresión ( $m > p$ ) que representan el efecto (fijo) que tiene cada variable auxiliar sobre la estimación del IDI para cada municipio.
- $b_i$  denota un conjunto de constantes positivas.
- Los  $v_i$  son los efectos aleatorios específicos de cada dominio o área de interés. Se supone que los  $v_i$  son independientes e idénticamente distribuidos (iid) con

$$E_m(v_i) = 0, \quad V_m(v_i) = \sigma_v (\geq 0),$$

Siendo  $E_m$  y  $V_m$ , el valor esperado y la varianza del modelo lineal considerado. En resumen, para el proyecto en cuestión se supone que cada municipio tiene un efecto aleatorio independiente que sigue una distribución de probabilidad con media 0 y varianza  $\sigma_v$ , es decir,  $v_i \sim^{iid} (0, \sigma_v^2)$ .

El parámetro  $\sigma_v^2$  es una medida de homogeneidad de los municipios (áreas) después de tener en cuenta la información disponible en las covariables  $z_i^T$ .

- $e_i$  denota el error del diseño de muestreo empleado que está asociado al  $i$ -ésimo municipio. Se supone que los  $e_i$  son independientes e idénticamente distribuidos (iid) con

$$E_p(e_i|\theta_i) = 0, \quad V_p(e_i|\theta_i) = \psi_i (\geq 0),$$

También es habitual suponer que las variaciones asociadas al diseño de muestreo,  $\psi_i$ , son conocidas.

**Nota:** El modelo asume que  $v_i$  y  $e_i$  son independientes.

La metodología SAE requiere utilizar información auxiliar, al nivel de desagregación geográfica deseado, para estimar el modelo mediante el cual se busca inferir el valor cierto del IDI. Esta metodología impone dos condiciones para esta información auxiliar, en este caso a nivel municipal: i) que explique de forma general el desarrollo de las TIC y ii) que se exógena a las variables incluidas en el índice. Un ejemplo de lo anterior, podría ser estadísticas demográficas, que explican el desarrollo general de programas de las TIC a nivel municipal, pero no explican a nivel de hogar o persona, sus resultados sobre acceso, uso o habilidades.

Siguiendo este lineamiento, se elaboró una base de datos con cifras de Terridata<sup>4</sup> del DNP y del panel municipal<sup>5</sup> de la Universidad de los Andes. En la Tabla 2 se presentan las temáticas de las variables auxiliares utilizadas para el cálculo del IDI a nivel municipal:

Fuente	Temáticas
DNP – Terridata	Vivienda y acceso a servicios públicos
	Educación y Salud
	Conflicto armado y seguridad ciudadana
	Finanzas públicas y desempeño municipal
	Percepción ciudadana
	Ambiente y Ordenamiento Territorial
	Economía y Pobreza
	Mercado laboral
Uniandes – Panel Municipal	Características generales
	Fiscal
	Conflicto y violencia
	Agrícola y tierra
	Educación

Tabla 2: Temáticas de las variables auxiliares para SAE.

<sup>4</sup> Terridata es un repositorio, buscador y herramienta de visualización de datos a nivel municipal, departamental y regional desarrollado por el DNP. El objetivo de este instrumento es contar con indicadores estandarizados y comparables que den cuenta de los resultados en diversas dimensiones socioeconómicas de todas las entidades territoriales del país. (DNP, 2019).

<sup>5</sup> El Panel Municipal del CEDE es una iniciativa del Centro de Estudios sobre Desarrollo Económico de la Universidad de los Andes para consolidar, en una sola base de datos, información a nivel municipal. (Uniandes, 2019).

## Resultados

### IDI por departamento

Se debe señalar el valor a nivel nacional del IDI para 2018, con un resultado de 5,36 es igual al reportado por la ITU en su publicación para el año 2017, por lo que se espera que se encuentre muy cerca del valor ITU 2018 (el cual no se publica debido a que el IDI se encuentra en un cambio metodológico).

En la Figura 3 se presentan los resultados obtenidos de la estimación de IDI a nivel departamental, destacando el desarrollo de las TIC de Bogotá, D.C, con un valor de 7,12 seguido de Valle del Cauca (6,15), el promedio nacional fue de 5,36. En contraste, los resultados más bajos se encuentran en los departamentos de Amazonas (1,90), Vichada (1,65), Vaupés (1,57) y Guainía (1,54), todos correspondientes a la región de la Amazonía-Orinoquía colombiana.

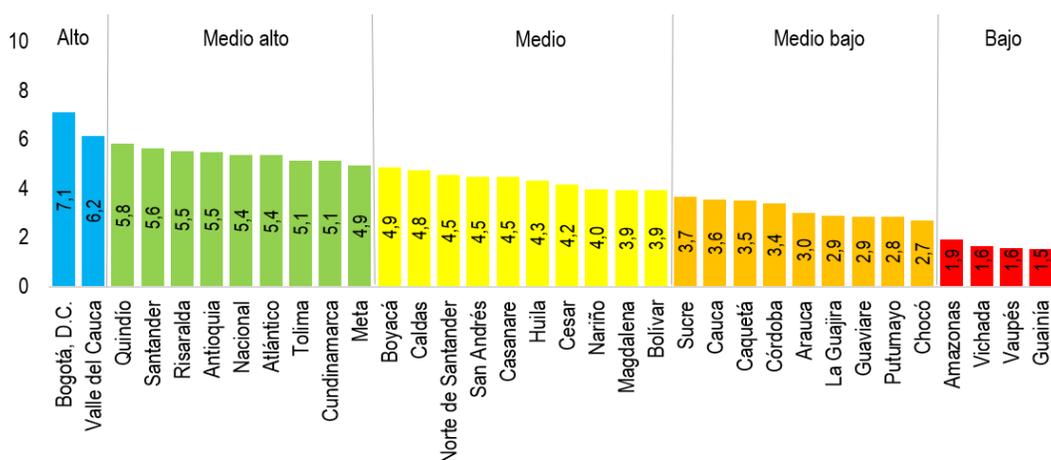


Figura 3: Resultados departamentales IDI a nivel departamental 2018.

En las Figuras 4, 5 y 6 se presentan los resultados IDI a nivel departamental 2018 desagregados por los subíndices Acceso a las TIC, Uso a las TIC y Habilidades TIC respectivamente.

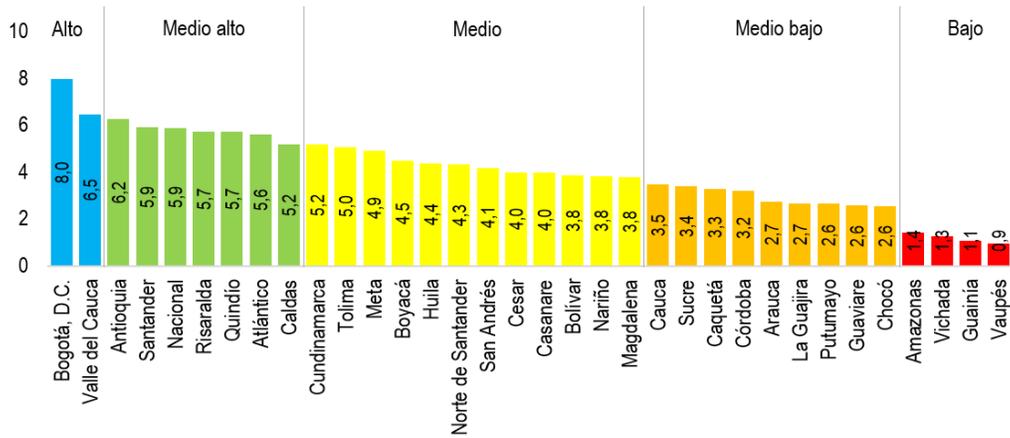


Figura 4: Resultados departamentales IDI 2018 – Por subíndice Acceso a las TIC.

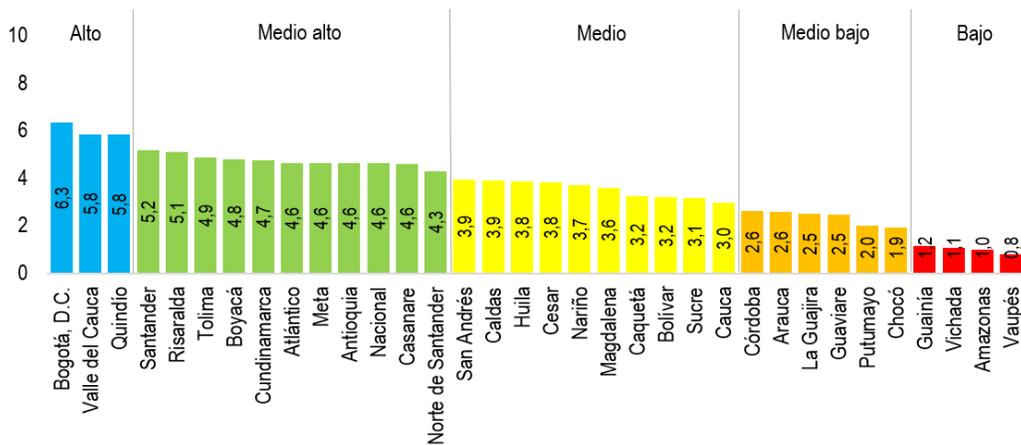


Figura 5: Resultados departamentales IDI 2018 – Por subíndice Uso de las TIC.

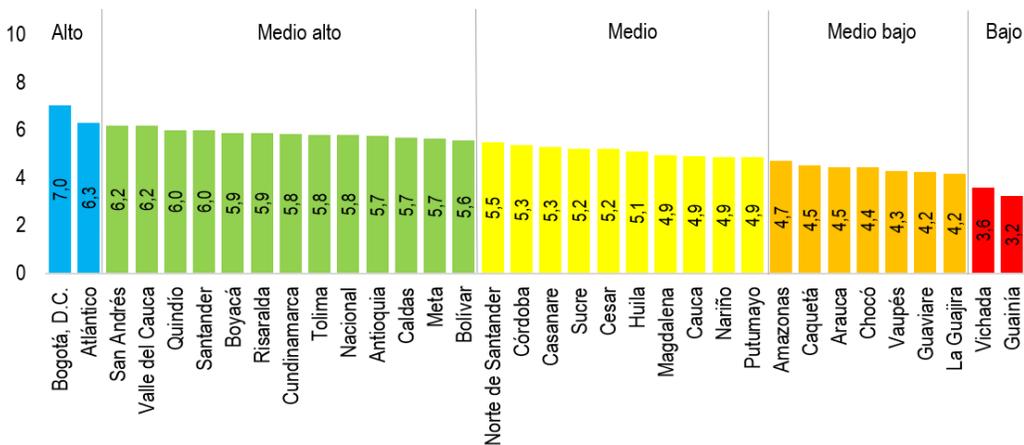


Figura 6: Resultados departamentales IDI 2018 – Por subíndice Habilidades para las TIC.

### **IDI por municipio**

En la Tabla 3 y Tabla 4 se ilustran los 10 municipios con resultados de mayor y menor valor en el IDI 2018 respectivamente. Se aprecia que de los resultados de mayor valor destacan ciudades capitales y ciudades cercanas, lo cual implica que puede haber un efecto de cercanía espacial en cuanto el desarrollo de este tipo de tecnologías.

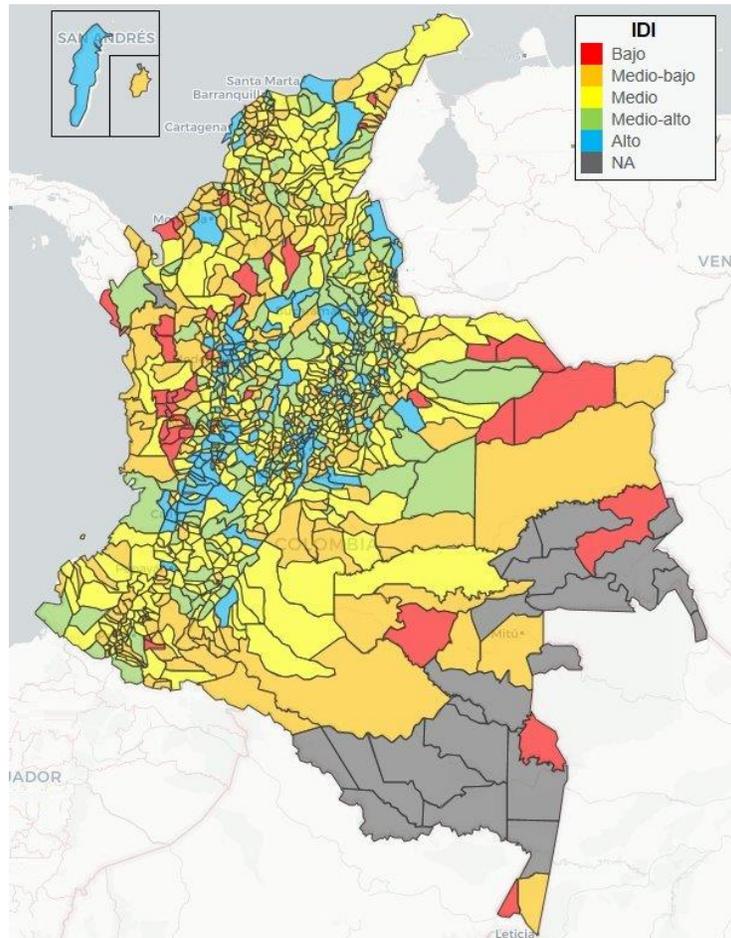
Departamento	Municipio	IDI
Cundinamarca	Cajicá	7,41
Cundinamarca	Fusagasugá	7,28
Valle del Cauca	Candelaria	7,25
Quindío	Armenia	7,23
Cundinamarca	Soacha	7,19
Caldas	Manizales	7,16
Antioquia	Medellín	7,15
Bogotá, D.C.	Bogotá, D.C.	7,12
Antioquia	Envigado	7,11
Cundinamarca	La Calera	7,08

Tabla 3: Municipios con resultados de mayor valor en el IDI Regional 2018.

Departamento	Municipio	IDI
Antioquia	Valdivia	1,14
Chocó	Cértegui	1,12
Antioquia	Nariño	1,08
Chocó	Sipí	0,84
La Guajira	La Jagua del Pilar	0,82
Antioquia	Murindó	0,73
Vaupés	Taraira	0,64
Vichada	Santa Rosalía	0,57
Amazonas	Puerto Nariño	0,26
Guaviare	Miraflores	0,18

Tabla 4: Municipios con resultados de menor valor en el IDI Regional 2018.

Para tener una mejor perspectiva de los resultados obtenidos, en la Figura 7 se presenta un mapa de calor de los resultados del IDI 2018 para los municipios de Colombia.



*Figura 7: Mapa de resultados IDI 2018 a nivel municipal.*

### **Conclusiones y recomendaciones**

1. Los modelos SAE pueden ser empleados para realizar la estimación de IDI en aquellos municipios en los cuales hay ausencia de información.
2. El modelo SAE puede ser ajustado según las covariables que la Dirección de Desarrollo Digital consideren deban ser analizadas dentro del estudio.

### **Socialización**

Los resultados fueron socializados con la Dirección de Desarrollo Digital.