**Texto

Descripción generada automáticamente**

**Maestría Economía Aplicada**

**MECA 4107: Big Data and Machine Learning for Applied Economics**

**Presentado por el grupo 5: Carlos Avilán y Francisco Ortiz**

**Abstract**

**Introducción**

En nuestro trabajo de grado pretendemos evaluar el impacto del Proyecto Nacional Conectividad de Alta Velocidad -PNCAV el cual nace en el 2013 en el marco del Plan Vive Digital con el objetivo de impulsar la masificación del Internet de alta velocidad, y alcanzar una cobertura del 100% para el 2023 en algunos municipios de las regiones de la Amazonía, Orinoquía y Chocó. Este despliegue de infraestructura pretende generar un desarrollo regional aumentando las oportunidades de transformación socioeconómica.

En el mundo existen estudios del impacto positivo en la economía a causa de la masificación del Internet banda ancha. Sin embargo, en Colombia no existen estudios recientes que puedan determinar el efecto que tienen esta clase de políticas y en particular del PNCAV. Ante esto, nuestro trabajo de grado tiene como fin responder la siguiente pregunta: ¿Cuál es el efecto del aumento de la penetración de Internet banda ancha en el Producto Interno Bruto (PIB) en estos municipios?

Para poder determinar este impacto de la conectividad de internet en la economía es necesario conocer el PIB de los municipios de Colombia. Sin embargo, esta información solamente existe del año 2005 al 2009 (datos CEDE). Para ello en este trabajo final establecimos modelo que fuera capaz de predecir este PIB municipal del año 2010 al 2020 para cada municipio de Colombia en base a diferentes predictores que explican de una u otra forma el PIB.

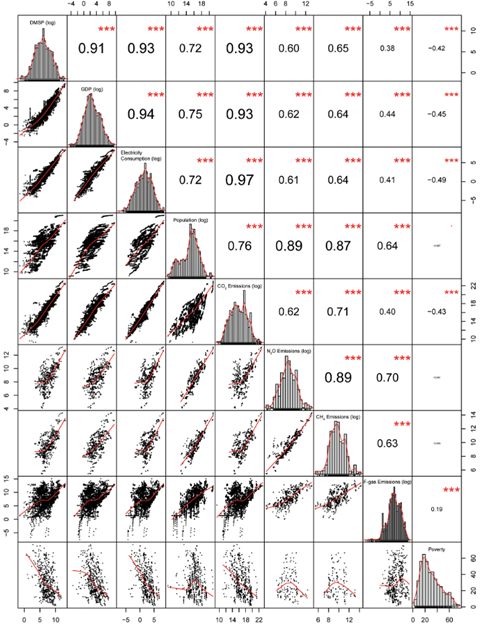
**Datos**

En la tabla 1 se resumen los datos utilizados para cada una de las variables utilizadas. Los datos utilizados son a nivel municipal por cada año y fueron obtenidos del Panel Municipal del CEDE a excepción de los datos de luminosidad que fueron obtenidos a partir del desarrollo realizado por el profesor Eduard Martinez para los años 2005 a 2018. Los datos de luminosidad del 2019 y 2020 fueron estimados en base a este desarrollo previo.

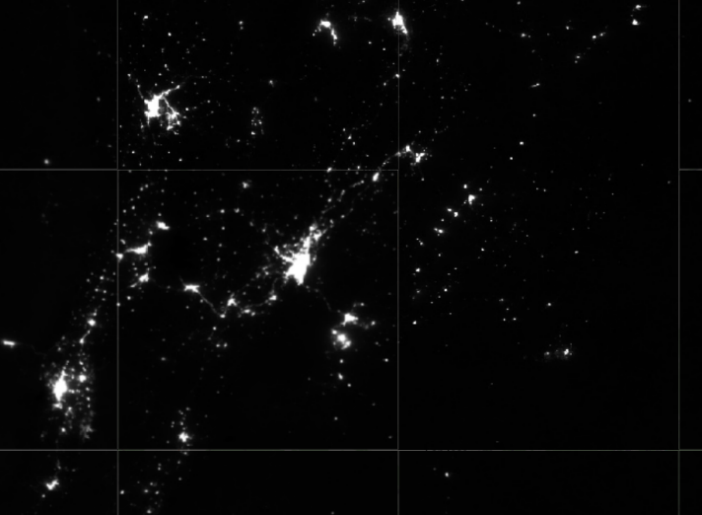
Igualmente, en la tabla 1, contiene la sección de bibliografía en donde se referencian los estudios que dan validez a cada una de las variables como explicativas del PIB. En particular los datos de luminosidad son útiles para este trabajo ya que existe una fuerte correlación entre las luces nocturnas y las medidas del Producto Interior Bruto (PIB), a nivel nacional, estatal y regional.

Por ejemplo, algunos estudios como el de Proville et al. (2017) encontraron fuertes correlaciones entre los datos de luminosidad nocturna y el consumo de electricidad, las emisiones de CO2 y el PIB. En la figura 1 se aprecia esta correlación.

En la figura 2 se aprecia la luminosidad nocturna de Colombia para el año 2020. En particular en el centro del país donde se encuentra Bogotá se observa un área de alta intensidad lumínica. Esto acorde a la información del DANE (2020), Bogotá D.C. es la región que más aporta al PIB del país.



Gráfica 1. Correlaciones entre los datos de luminosidad nocturna y el consumo de electricidad, las emisiones de CO2 y el PIB

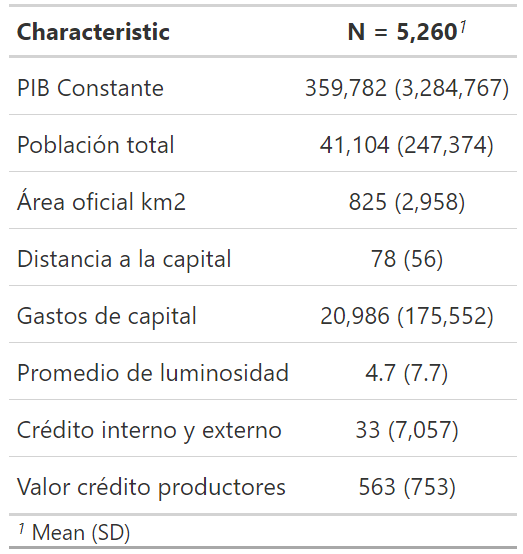


Gráfica 2. Intensidad lumínica nocturna en Colombia. Año 2021.

|  |  |  |
| --- | --- | --- |
| **Variable** | **Fuente de datos** | **Bibliografía** |
| PIB municipal (2005 - 2009) | Sánchez, F. y España, I. DANE, Censo 2005. CEDE | No aplica. (Variable de interés) |
| Datos de luminosidad (satélite VIIRS) 2005 - 2020 | A harmonized global nighttime  light dataset. Eduard Martinez. | [Using luminosity data as a proxy for economic statistics](https://www.pnas.org/doi/10.1073/pnas.1017031108) |
| Número de habitantes (2005 - 2020) | (pobl\_tot,  DANE. Proyecciones de población) | [The Role of Population in Economic Growth](https://journals.sagepub.com/doi/full/10.1177/2158244017736094) |
| Gastos de capital (Inversión) relacionados con la formación bruta de capital fijo y otros  (2005 - 2020) | (g\_cap, Buen gobierno CEDE) | [ShowProperty (urf.gov.co)](https://www.urf.gov.co/webcenter/ShowProperty?nodeId=%2FConexionContent%2FWCC_CLUSTER-070589%2F%2FidcPrimaryFile&revision=latestreleased) |
| Crédito interno y externo  (2005 - 2020) | (finan\_credito, Buen Gobierno CEDE) | [El efecto del microcrédito en el PIB de Colombia, 2005-2018 (redalyc.org)](https://www.redalyc.org/journal/118/11866602001/html/) |
| Distancia lineal a la capital del departamento  (2005 - 2020) | (discapital, Características generales CEDE) | [The Contribution of Economic Geography to GDP per Capita (oecd.org)](https://www.oecd.org/norway/42506177.pdf) |
| Valor crédito pequeños productores  (2005 - 2020) | (vrf\_peq, Agricultura CEDE) | [Effects of credit on national and agricultural GDP, and poverty: a developing country perspective | SpringerLink](https://link.springer.com/article/10.1007/s43546-021-00146-6) |
| Área oficial municipio en km²  (2005 - 2020) | (areaoficialkm2, Características generales del CEDE) | [Competitive Cities Economic Growth - TCdata360 (worldbank.org)](https://tcdata360.worldbank.org/stories/competitive-cities/) |

Tabla 1. Variables utilizadas para predecir el PIB

Para el entrenamiento de nuestro modelo se utilizaron datos del 2005 al 2009 que es el periodo de tiempo donde existe la información del PIB municipal (base de datos *train*). Para las predicciones se utilizaron los datos del 2010 al 2020 (base de datos *test*). En las siguientes tablas se encuentran las estadísticas descriptivas, tanto para la base de datos de *entrenamiento* como la de *prueba*:

 Interfaz de usuario gráfica

Descripción generada automáticamente con confianza media

Tabla 2. Estadísticas descriptivas base de datos train. Tabla . Estadísticas descriptivas base de datos test

Como se aprecia en las tablas 2 y 3, los modelos fueron entrenados con 5260 observaciones. La base de datos para predecir tiene 12064 observaciones. Se aprecian datos parecidos en población con una media de 41104 habitantes para *train* y 43556 habitantes para *test*. La distancia a la capital en km se encuentra en 78 y 79, respectivamente. Se encuentran diferencias en los datos de luminosidad, en la de *train* el promedio es de 4.7 *Day-Night Band radiance* y en *test* es de 7. Esto tiene sentido ya que se espera que el pasar de los años la intensidad lumínica aumente, así mismo ocurre con variables como los “gastos de capital”, el “crédito interno y externo” y el “valor de crédito de los productores”.

**Modelo**

**Resultados**

**Conclusiones**

**Datos y código**

**Referencias**

Proville J, Zavala-Araiza D, Wagner G (2017) Night-time lights: A global, long term look at links to socio-economic trends. PLoS ONE 12(3): e0174610. <https://doi.org/10.1371/journal.pone.0174610>

DANE (2021). Cuentas nacionales departamentales. https://www.dane.gov.co/index.php/estadisticas-por-tema/cuentas-nacionales/cuentas-nacionales-departamentales