**Propuesta trabajo final – Big Data**

**Grupo 5: Carlos Avilán y Francisco Ortiz**

*Esta propuesta de trabajo final nace de la necesidad de realizar una predicción para el desarrollo de nuestra tesis.*

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 es necesario conocer el PIB de los municipios de Colombia. Sin embargo, esta información existe solamente del 2005 al 2009 (datos CEDE). Para ello en este trabajo final pretendemos establecer un modelo que sea capaz de predecir este PIB municipal en base a distintos predictores que explican de una u otra forma el PIB.

En donde la matriz X, incluye algunos de los posibles predictores. La siguiente información está desagregada a nivel municipal:

|  |  |  |
| --- | --- | --- |
| **Variable** | **Fuente de datos** | **Bibliografía** |
| Datos de luminosidad (satélite VIIRS) | 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 | (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 | (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 | (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 | (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 | (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 | (areaoficialkm2, Características generales del CEDE) | [Competitive Cities Economic Growth - TCdata360 (worldbank.org)](https://tcdata360.worldbank.org/stories/competitive-cities/) |

Para lograr esto pretende realizar lo siguiente:

1. Consolidar datos (especialmente los datos de luminosidad). Unir base de datos y realizar limpieza de estas.
2. Estimar varios modelos con diferentes combinaciones de variables independientes.
3. Utilizar todas las herramientas vistas en la clase de Big Data, con el fin de estimar el mejor modelo posible.
4. Comparar cada modelo por medio de diferentes parámetros de eficiencia y escoger el mejor.
5. Describir los resultados obtenidos y concluir.