

EVALUACIÓN DE NUEVAS HERRAMIENTAS Y TÉCNICAS DEL BIG DATA PARA PROYECTOS DEL BANCO INTERAMERICANO DE DESARROLLO

Emmanuel Letouzé (Desarrollo Partes I, II & III, Edición y aprobación final)
Nicolás de Ligny Tandefelt (Traducción Partes I, II, III, Desarrollo de metodología, realización
de entrevistas para casos, Producción de Introducción y conclusión)

Índice

Resumen Ejecutivo

Introducción

I. Marco conceptual

1. Desarrollo y programas complejos
2. Visión general de Big Data y metodologías de ciencias de datos

II. Casos de estudio

1. Metodología
2. Casos de estudio por país

A. Bolivia: "Programa de Expansión de Infraestructura Eléctrica" (BO-L1190)

B. Colombia: "Sistemas Estratégicos de Transporte Público (SETP)" (CO-L1091)

C. Perú: "Proyecto para la Mejora de la Eficiencia en la Gestión de la Inversión y las Contrataciones Públicas" (PE-L1231)

D. Venezuela: "Programa Apoyo Centro Acción Social por la Música Fase II" (VE-1017)

E. Ecuador: "Movistar Ecuador Financiamiento de Equipos" (IDB Invest-12261-01)

III. Conclusión y Recomendaciones Generales

Bibliografía

Resumen Ejecutivo

El presente estudio identifica herramientas innovadoras de Big Data a partir de casos de estudios basados en cinco proyectos en vía de implementación por el Banco Interamericano de Desarrollo (BID) en los países de la Comunidad Andina (CAN) con el fin de que nuevas tecnologías puedan ser aprovechadas e implementadas en estos u otros proyectos de inversión de la región. De esta manera, se analizan proyectos de diversas áreas estratégicas, con objetivos y contextos locales diferentes. Los equipos operacionales y representantes de cada país estudiado fueron consultados sobre los avances del proyecto, así como sobre las necesidades específicas para proponer herramientas que puedan ser adoptadas de manera realista y generen un impacto concreto. Asimismo, el estudio aborda las implicaciones de las nuevas tecnologías en cuanto a sus limitaciones y desafíos. El público objetivo son todos los actores relevantes a la toma de decisiones, tanto los equipos operacionales del Banco que impulsan, planifican y ejecutan los proyectos de inversión, así como los altos funcionarios del servicio público y autoridades de gobierno.

Palabras clave: Ciencia de Datos, Datos Masivos, Analítica de Datos, Energía, Infraestructura, Transporte, Inversión Social, Tecnologías de la Información y la Comunicación, Reforma del Estado.

Introducción

Los organismos multilaterales de préstamos tales como el Banco Interamericano de Desarrollo (BID) son capaces de reunir actores de diversos sectores y con intereses propios para llevar a cabo iniciativas ambiciosas para la creación y desarrollo de bienes públicos. En efecto, entre los roles que asume el BID está el de ser una fuerza motora en la adopción a nivel regional de buenas prácticas por gobiernos e instituciones. Tanto los países miembros, como el BID se benefician al usar tecnologías digitales en la toma de decisiones basadas en evidencia. Sin embargo, adoptar estas herramientas por lo general requiere de un soporte en infraestructura y recursos humanos capacitados. Por lo que este documento argumenta que el BID juega un rol importante en generar visibilidad de los beneficios que aportan las nuevas tecnologías al desarrollo.

La cantidad de información que es capturada a diario por sensores que miden comportamientos adquisitivos, acciones en redes sociales, videos y fotografías digitales, y señales de telefonía celular o GPS ha crecido de forma exponencial. IBM estima que se generan aproximadamente 2.5 quintillones de bytes de datos diarios y que 90% de los datos almacenados en 2017 fueron producidos en los últimos dos años¹, redefiniendo sociedades anteriormente caracterizadas por su escasez de datos, a sociedades en donde la información producida por sus ciudadanos sobreabunda. Como el resto del mundo, los países andinos han observado una adopción acelerada de hardware y sensores de las tecnologías de la información y la comunicación (TICs), las cuales han permitido revelar de manera precisa y detallada patrones de comportamiento individuales y de grupo.

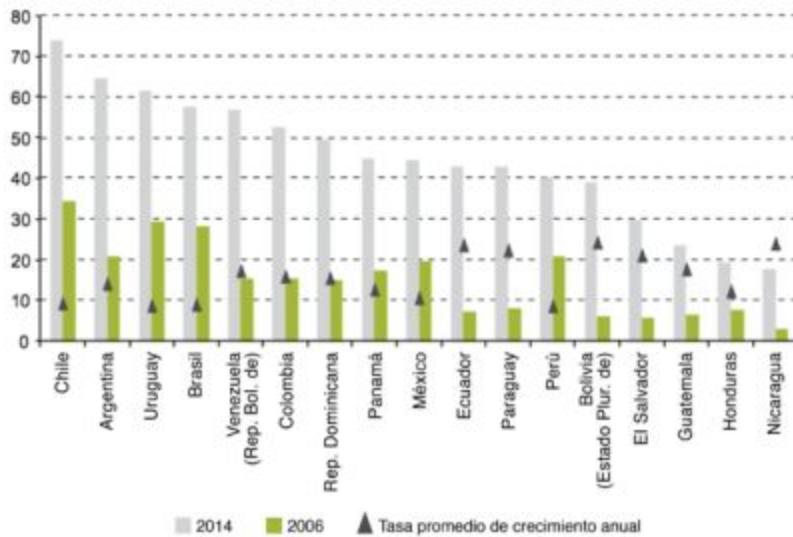
Un mayor conocimiento sobre estas innovaciones, así como la toma de conciencia alrededor de su uso sistematizado harán que las poblaciones latinoamericanas y sus gobiernos sean partícipes de la “revolución de los datos”, una nueva revolución económica y social comparable al impacto transformador de lo que constituyó la electrificación masiva anteriormente. Los medios de comunicación han contribuido indudablemente a llamar la atención alrededor del uso de tecnologías y herramientas para el análisis y el aprovechamiento de datos. Sin embargo, la utilización de estas nuevas tecnologías no garantiza un impacto transformador si no se toman en cuenta otros factores y elementos clave, tales como el ecosistema local y consideraciones relacionadas al contexto sociopolítico. La incapacidad de coordinación entre expertos y actores institucionales para aprovechar los datos celulares durante la crisis del Ébola en 2014 en África Occidental, descrito también como “Ébola, un desastre de los datos masivos”² constituye un ejemplo de fracaso sobre el carácter inherentemente prometedor de los datos masivos.

¹ IBM, “10 Key Marketing Trends for 2017 and Ideas for Exceeding Customer Expectations”, 2017: <https://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=WRL12345USEN>

² McDonald, Sean. “Ebola: A Big Data Disaster – Privacy, Property, and the Law of Disaster Experimentation”, The Centre for Internet and Society, Jan. 2016: <https://cis-india.org/papers/ebola-a-big-data-disaster>

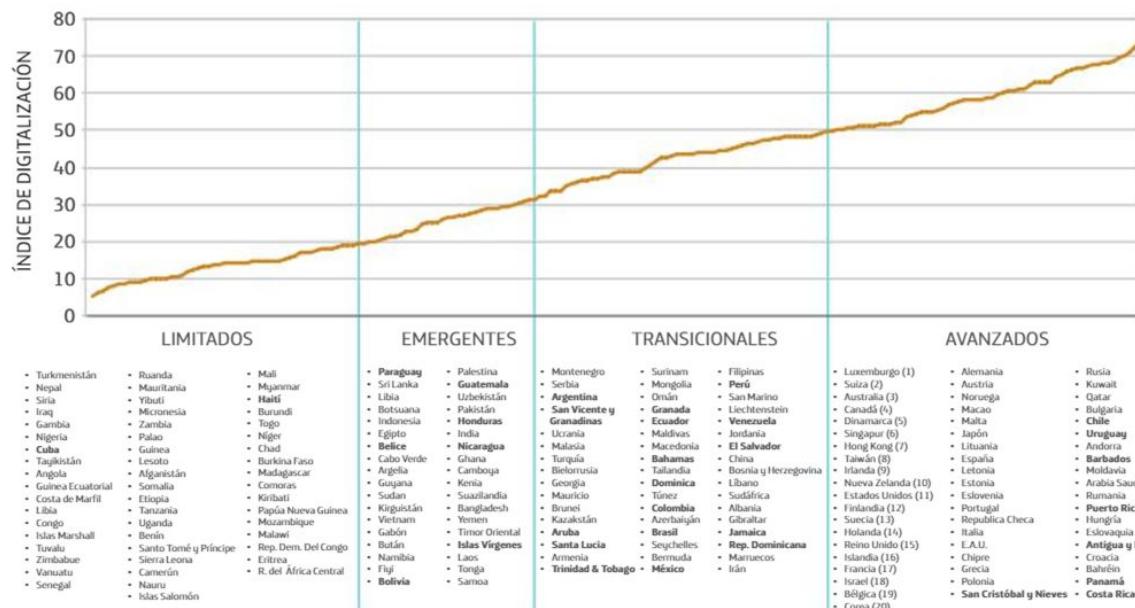
El proceso de digitalización de países latinoamericanos ha sido heterogéneo. En 2013, países como Chile, Panamá y Uruguay ya se encontraban en una etapa avanzada mientras que Bolivia o Paraguay en una etapa emergente. No obstante, Bolivia ha experimentado desde 2006 un aumento significativo del uso poblacional de Internet.

Uso de Internet por porcentaje de población en 2006 y 2014



Fuente: CEPAL, 2015

Desarrollo de la digitalización, 2013



Fuente: Raúl Katz. El ecosistema y la economía digital en América Latina. 2015. URL: <http://cet.la/bloq/course/libro-el-ecosistema-y-la-economia-digital-en-americalatina/>

Este estudio tiene como objetivo identificar nuevas técnicas y herramientas del Big Data para mejorar la ejecución y/o resultados esperados de cinco proyectos en vía de implementación financiados por el Banco Interamericano de Desarrollo (BID) en países de la Comunidad Andina (Bolivia, Colombia, Ecuador, Perú y Venezuela). El informe se estructura de la siguiente manera:

En una primera instancia, se introducirá el marco conceptual enfocado en contextualizar el uso de tecnologías y metodologías relacionadas al Big Data en programas de desarrollo, particularmente en el área de monitoreo y evaluación (I). En una segunda instancia, se abordarán los cinco casos de estudios que abarcan diversas áreas estratégicas y se propondrán herramientas innovadoras que podrían ser consideradas para estos u otros proyectos similares (II). Finalmente, se realizará una reflexión en torno al análisis previo, abordando las implicaciones como los desafíos y límites del Big Data para el desarrollo, y a partir de ello se sugieren recomendaciones generales (III).

El público objetivo son todos los actores relevantes a la toma de decisiones, tanto los equipos del Banco que impulsan, planifican y ejecutan los proyectos de inversión, así como los altos funcionarios del servicio público y autoridades de gobierno.

I. Marco conceptual

1. Desarrollo y programas complejos

Las perspectivas sobre el desarrollo han cambiado significativamente desde 2005, cuando altos funcionarios de más de cien países receptores de ayuda y agencias donantes se reunieron en París para definir los pasos para mejorar la eficacia de la ayuda al desarrollo. Los pasos concretos acordados en la Declaración de París sobre la Eficacia de la Ayuda formularon una hoja de ruta para mejorar la calidad de la ayuda a través de una serie de medidas específicas para la implementación, indicadores de desempeño para evaluar el progreso y un marco de responsabilidad mutua entre los países donantes y países receptores. Los cinco principios de la eficacia de la ayuda son los siguientes: (1) la apropiación (los países en desarrollo establecen sus propias estrategias para el desarrollo, mejoran sus instituciones y enfrentan la corrupción), (2) la alineación (los países donantes alinean su apoyo con estos objetivos), (3) la armonización (los países coordinan sus acciones, simplifican los procedimientos y comparten información para evitar la duplicación), (4) la gestión de los resultados (los países receptores y los países donantes se enfocan en generar y medir los resultados), y (5) la responsabilidad mutua (los países donantes y los países receptores son responsables de los resultados del desarrollo y su impacto). Estos principios también han servido de base para otros compromisos, adaptados a contextos específicos: la Declaración de Bogotá (que se concentra en los principios de ayuda efectiva en la cooperación Sur-Sur), los Principios de Estambul (sobre el papel de la sociedad civil) y la Declaración de Dili (sobre la efectividad de la ayuda en Estados frágiles afectados por conflictos). A partir de ese momento, la efectividad de la ayuda se reemplazaría por el concepto de "efectividad del desarrollo".

Como resultado del compromiso global a la Agenda 2030, los modelos tradicionales de intervención están mostrando sus límites. Los modelos burocráticos nacidos después de la Segunda Guerra Mundial respondieron contra problemas técnicos que podrían estandarizarse, respuestas reducidas a "mejores prácticas", pero mostraron limitaciones importantes para enfrentar problemas complejos y ambiguos en entornos específicos.³ De hecho, la Agenda 2030 enfatiza la necesidad de enfoques que tengan en cuenta la integralidad de las dimensiones de desarrollo sostenible (económica, social y ambiental) donde las decisiones en un dominio afectan los resultados en otros.

³Andrews Matt, Pritchett Lant and Woolcock Michael, "Escaping Capability Traps through Problem-Driven Iterative Adaptation (PDIA)", Center for Global Development, Working Paper 299, June 22, 2012: <https://www.cgdev.org/publication/escaping-capability-traps-through-problem-driven-iterative-adaptation-pdia-working-paper>

Si bien los países donantes y la comunidad internacional abogan por la efectividad, la reforma institucional, la participación, la apropiación local y el empoderamiento, la mayor parte de la investigación y el trabajo sobre el desarrollo todavía funciona en un paradigma de causalidad lineal predecible y se caracteriza por el enfoque en la responsabilidad y la estructura vertical comando/control. Mientras tanto, ha surgido un mundo de entendimiento diferente que ayuda a explicar fenómenos dinámicos complejos mediante el uso de ideas y conceptos como la no linealidad, el borde del caos, la autoorganización, la emergencia y la co-evolución.

Los programas de desarrollo han sido reconocidos cada vez más por su carácter multifacético, desafiando seriamente la validez de los diseños de evaluación convencionales que asumen una relación lineal entre los insumos del programa y los resultados. En consecuencia, a menudo se requieren diseños de evaluación "de respuesta a la complejidad".⁴ En este contexto, es probable que en un futuro próximo muchas de las fuentes de datos utilizadas para el monitoreo y evaluación (M&E) del programa se generen de forma pasiva a través del uso de nuevas tecnologías, en lugar de recopilarse en los estudios de M&E que se usan comúnmente en la actualidad.⁵

Los Objetivos de Desarrollo Sostenible deben identificar nuevos caminos hacia la sustentabilidad para navegar por la intrincada interacción de la investigación y el desarrollo, las inversiones públicas y privadas, la política, la planificación y la compleja toma de decisiones por parte de múltiples actores. También se requiere una movilización sin precedentes de conocimiento global e información precisa y oportuna. En la era de los teléfonos móviles, la banda ancha inalámbrica, la detección remota y la inteligencia artificial, la recopilación de datos podría ser considerablemente más rápida. Para apoyar programas de desarrollo, los gobiernos y las organizaciones internacionales de desarrollo deben invertir en un sistema de información casi en tiempo real con el fin de producir datos confiables con una base anual o trimestral.

2. Visión general de Big Data y metodologías de ciencia de datos

Los enfoques tradicionales de M&E presentan los datos después de que se finaliza la implementación, cuando solo es útil para diseñar un nuevo programa en lugar de mejorar uno existente. Con una duración de proyectos que abarcan hasta cinco años o más, las pérdidas de eficacia son importantes. Los implementadores de desarrollo rara vez saben dónde se ubican en relación con sus objetivos.⁶

⁴ UN Global Pulse, "Integrating Big Data into the Monitoring and Evaluation of Development Programmes", 2016, page 54: http://unglobalpulse.org/sites/default/files/IntegratingBigData_intoMEDP_web_UNGP.pdf

⁵ Ibid.

⁶ Pritchett, L., Samji, S., and Hammer, J., "It's All About MeE: Using Structured Experiential Learning ("e") to Crawl the Design Space", Center for Global Development, Working Paper 322, 2013 <https://www.cgdev.org/publication/its-all-about-mee-using-structured-experiential-learning-e-crawl-design-space>

Sin embargo, en las últimas décadas ha habido algunas mejoras en el campo de M&E. En el aspecto científico y tecnológico, el aumento de los ensayos de control aleatorios (ECA), así como la adopción de nuevos dispositivos y servicios digitales han generado cambios en este campo. Aunque los ECA han sido durante mucho tiempo la herramienta más utilizada para la evaluación sumativa, este enfoque también ha sido el más discutido entre los analistas de Big Data. Un enfoque exclusivo en los ECA es sujeto a críticas dentro de la comunidad de M&E en gran medida debido, entre otras razones, a un enfoque limitado en uno o un pequeño número de resultados generalmente cuantitativos, una falta de atención al proceso de implementación del proyecto (la evidencia a menudo está disponible después de años), los altos costos y el contexto dentro del cual se diseñan, implementan y evalúan los programas. Algunos evaluadores también enfatizan la necesidad de contemplar otras perspectivas y argumentan que no hay una sola manera de identificar o evaluar los resultados del programa.

Se han realizado varios intentos para responder a estos y otros desafíos. Por ejemplo, hay un llamado a nuevos enfoques, como a una "evaluación ágil", que se centra en un acceso más barato a los datos y en la etapa de implementación en lugar de los resultados, o el concepto de "ECA+", que combina diseños de evaluación experimentales con enfoques cualitativos.⁷ La mayoría de los diseños de evaluación tienen como objetivo determinar si existe evidencia creíble (estadística, basada en la teoría o narrativa) de que los programas han alcanzado sus objetivos previstos y la lógica de muchos procesos de evaluación, incluso aquellos que se consideran los más rigurosos, no están identificando debidamente resultados que no fueron especificados en el diseño del programa. La combinación de enfoques ECA con métodos mixtos (el llamado diseño "ECA+") puede fortalecer la credibilidad de estas metodologías y capturar importantes consecuencias no deseadas.⁸

A pesar de los grandes avances, como la fuerte caída de la pobreza extrema, el despliegue de la "revolución de datos" presenta grandes oportunidades para procesos de toma de decisiones más ágiles, transparentes y responsables. Aprovechar las nuevas fuentes de datos, los enfoques analíticos y las tecnologías para hacer que los programas de desarrollo sean más efectivos y, en general, contribuir a que las políticas estén más basadas en la evidencia (o informadas), contribuirá en gran medida a inclinar la escala hacia lo positivo.

Aunque la utilización de este tipo de herramientas parezca incremental a lo que se ha usado tradicionalmente, al pasar de un enfoque tradicional a uno basado en tecnologías digitales, se pueden obtener los siguientes resultados:

⁷ UN Global Pulse, "Integrating Big Data into the Monitoring and Evaluation of Development Programmes", 2016, page 27: http://unglobalpulse.org/sites/default/files/IntegratingBigData_intoMEDP_web_UNGP.pdf

⁸ Bamberger, M., Tarsilla, M., Hesse-Biber, S., "Why so many "rigorous" evaluations fail to identify unintended consequences of development programs: How mixed methods can contribute", Evaluation and Program Planning, Vol. 55, April 2016, pages 155-162: <https://www.ncbi.nlm.nih.gov/pubmed/26874234>

- *Mejores medidas:* La incorporación de nuevas metodologías generarán cambios significativos en los procesos de monitoreo del desarrollo. El análisis de metadatos de llamadas telefónicas (en inglés, *Call Detail Records* - CDR) para medir la movilidad de la población, o el estudio de registros de transacciones electrónicas para análisis sobre la pobreza son solo algunos ejemplos de aplicaciones basadas en datos que ofrecen claras ventajas sobre los métodos tradicionales, como la recopilación de datos más rápida, la evaluación en tiempo real y una mayor granularidad de la información. Implementar nuevas tecnologías también podría contribuir a medir el avance hacia los indicadores de Nivel III de los ODS para los cuales no se ha formalizado ninguna metodología.
- *Mejores procesos:* el Big Data ofrece a los profesionales del desarrollo la capacidad de obtener información en tiempo real sobre el bienestar de la población y orientar las intervenciones de ayuda para grupos vulnerables de una manera que se adapte más a sus necesidades específicas. Las imágenes satelitales, los registros de teléfonos móviles, el rastreo por GPS y el análisis de las redes sociales son solo algunos ejemplos de las múltiples fuentes que se pueden utilizar para estructurar bases de datos integradas que sintetizan información sobre las diversas problemáticas de desarrollo (como el cambio climático, las epidemias, los conflictos, los recursos hídricos, y la seguridad alimentaria). La intersección de metodologías abrirá posibilidades para métricas e indicadores más dinámicos, lo que podría fomentar el aprendizaje de todos los actores relevantes.

Dada la rápida adopción del Big Data en el desarrollo internacional, es probable que se adopten sistemas de información integrados, que comenzarán a generar, analizar y sintetizar datos para la selección, el diseño, la gestión, el monitoreo, la evaluación y la difusión de programas. Sin embargo, aún quedan muchos desafíos por delante, como por ejemplo las restricciones legales, y el progreso dependerá claramente de la disponibilidad de datos y de la capacidad para medirlos de manera segura.

A continuación, se detallan los pasos sugeridos por Global Pulse para implementar una evaluación que aprovecha el uso del Big Data y las nuevas tecnologías⁹:

1. Identificar la información requerida para la evaluación.
2. Datos clasificados en fuentes convencionales (encuestas, publicaciones gubernamentales, grupos focales) y fuentes de Big Data (redes sociales, teléfonos móviles, registros financieros, etc.) generadas por las TIC.
3. Fuentes potenciales de Big Data categorizadas según disponibilidad y exhaustividad:

⁹ UN Global Pulse, "Integrating Big Data into the Monitoring and Evaluation of Development Programmes", 2016, page 85:

http://unglobalpulse.org/sites/default/files/IntegratingBigData_intoMEDP_web_UNGP.pdf

- Datos ubicuos se generan en todas partes al mismo tiempo (por ejemplo, datos meteorológicos sobre la lluvia, la temperatura);
- Datos no ubicuos que no se generan en todas partes y pueden ser clasificados en dos tipos: Específicos del sistema: generados dentro de un sistema específico (como los datos que solo se refieren a una organización particular o contexto geográfico); Fuera del sistema (por ejemplo, factores contextuales externos, como el clima económico en la región o la presencia y calidad de la infraestructura en todo el estado).

Dada la importancia de enfoques de métodos mixtos, también puede ser útil, en algunos casos, clasificar los datos según sean cualitativos o cuantitativos.

4. Datos categorizados en seis tipos según el tipo de pregunta y evidencia proporcionada:
 - Descriptivo
 - Exploratorio
 - Inferencial
 - Predictivo
 - Análisis causal (atribución)
 - Mecanicista
5. Cuatro fases del ciclo del proyecto donde se puede utilizar la evidencia:
 - Diseño
 - Implementación y monitoreo
 - Evaluación y aprendizaje
 - Sostenibilidad
6. Análisis de los indicadores potenciales para evaluar la calidad y la adecuación de los indicadores generados por Big Data y las TIC

Lista de verificación para evaluar la calidad y la adecuación de Big Data y los indicadores generados por las TIC

		Rating 1–5*
Relevant	Directly relates to the topic being measured	
Comprehensive	Covers all important aspects of the topic	
Clear and unambiguous	Easy to interpret and use without any confusion or misinterpretation.	
Cost effective	The data is inexpensive to collect	
Accessible	The data is easy to access (physically and in terms of the format) and does not require difficult administrative procedures to access.	
Monitorable	The data can easily be collected over time	
User friendly	Uses concepts that are widely accepted and easy to use	
Stable over time	The definitions or collection methods do not change over time	
Comparable across agencies and locations	All agencies define the indicator in the same way. Does not vary in different regions	
Ethical	Collecting or using the data does not pose ethical concerns	

* Rating: 1 = This dimension is ignored or inadequately addressed. 5 = fully addressed

Fuente: Global Pulse Integrando el Big Data en el M&E de los programas de desarrollo

El trabajo de desarrollo ha producido algunos "beneficios rápidos", áreas en las que la Inteligencia Artificial también puede facilitar nuestras vidas al etiquetar y clasificar los datos cuantitativos, realizando una diferenciación simple entre imágenes y objetos, y la identificación amplia de personas y grupos. A medida que la IA siga evolucionando, su relevancia para el M&E probablemente también crecerá.

3. ¿Cómo acceder a datos en manos de actores privados?

II. Casos de estudio

1. Metodología

Esta sección abordará cinco proyectos en vía de implementación financiados por el BID en Bolivia, Colombia, Perú y Venezuela, así como un proyecto de *IDB Invest* en Ecuador. Antes de abarcar cada país por separado, es necesario explicar la metodología general utilizada para la construcción y análisis de estos casos de estudio.

Factores de selección: ¿Por qué se escogieron estos proyectos?

País	Proyecto	Sector
Bolivia	"Programa de Expansión de Infraestructura Eléctrica" - (BO-L1190)	Energía
Colombia	"Sistemas Estratégicos de Transporte Público (SETP)" - CO-L1091	Transporte
Perú	"Proyecto para la Mejora de la Eficiencia en la Gestión de la Inversión y las Contrataciones Públicas" - PE-L1231	Reforma y Modernización del Estado
Venezuela	"Programa Apoyo Centro Acción Social por la Música Fase II" - VE-1017	Inversión Social
Ecuador	"Movistar Ecuador Financiamiento de Equipos" - <i>IDB Invest</i> - 12261-01	Telecomunicación, Medios y Tecnología

Se seleccionaron proyectos que cubrieran diversos sectores estratégicos con el fin de explorar y proponer una mayor cantidad de técnicas y herramientas innovadoras. Los proyectos estudiados pertenecen a los sectores de Energía, Transporte, Reforma y Modernización del Estado, Inversión Social, Telecomunicación, Medios y Tecnología, áreas en las cuales se ha desarrollado literatura relacionada a la explotación de datos y casos exitosos de aplicación de estas herramientas en otros países del mundo.

Otro criterio de selección fue limitarse a aquellos proyectos que todavía se encuentran en vía de ejecución para poder interactuar con los equipos del BID que todavía están asociados a estos proyectos. De esta manera, se analizan proyectos en los cuales eventualmente se podrían considerar la adopción de las herramientas propuestas.

Marco analítico: ¿Cómo se construyeron los casos de uso?

En una primera fase, se realizan entrevistas con los diferentes jefes de operaciones y de proyecto para recabar información que permita conocer las etapas de avance del proyecto y que determine las lagunas y necesidades actuales. El objetivo preliminar consiste en evaluar las herramientas o técnicas que ofrezcan mejoras observables en un proyecto de desarrollo, ya sea en la implementación y/o resultados esperados. A través de estas entrevistas, se busca también capturar información relativa a los tipos de datos a los cuales se podría tener acceso y a las capacidades del equipo (ej. conocimientos de herramientas y métodos, hardware y software disponible, *know-how* y habilidades necesarias). Esta etapa incluye identificar obstáculos que podrían presentarse durante la adopción de tales herramientas, de nuevo entendiendo que el éxito de lo propuesto dependerá también de la capacidad de los equipos locales en implementarlas.

En una segunda fase, y con esta información inicial, se consultan expertos que identifican técnicas y herramientas concretas con el fin de construir casos de uso que contemplen las ventajas, desventajas y/o limitaciones de las técnicas y datos disponibles, incluyendo posibles riesgos de privacidad o seguridad.

Cada caso de estudio contiene la siguiente estructura analítica:

- Breve descripción del proyecto que incluye resumen de antecedentes, objetivos, y monto del presupuesto.
- Avances y lagunas del proyecto que describe la fase de implementación en la cual se encuentra, resultados logrados hasta la fecha, entre otro tipo de información relevante, así como las dificultades asociadas a la gestión de los proyectos y/o logros de los objetivos.
- Técnicas y herramientas digitales que pueden ser aplicadas para atender necesidades concretas del proyecto, de manera realista y práctica, así como discutir sus desventajas y/o limitaciones.

Presentación de objetos para consideración: ¿Qué tipo de herramientas se proponen?

En base a las discusiones con los equipos del Banco, se lleva a cabo una investigación extensiva utilizando material público y privado, la cual es complementada por recomendaciones de académicos especializados en distintas áreas del *Big Data*. Se consideran técnicas para sistemas de almacenamiento y manipulación de los datos, sistemas de procesamiento de datos, metodologías de modelación y formas de análisis basados en estos datos. Por ejemplo, se contemplan herramientas para capturar y revisar las “migajas digitales” generadas por

dispositivos de navegación por satélites (ej.: GPS), redes de sensores inalámbricos, y metadatos de llamadas telefónicas.

2. Casos de estudio por país

A. Bolivia: "Programa de Expansión de Infraestructura Eléctrica" (BO-L1190)

Estado y avance del proyecto

El proyecto "Programa de Expansión de Infraestructura Eléctrica" (BO-L1190) perteneciente al sector de Energía fue aprobado en 2018 y consiste en un proyecto con un costo total de USD 78,000,000. El proyecto BO-L1190 se alinea a los objetivos propuestos por el "Programa de Fortalecimiento del Sector Eléctrico" (BO-L1189) que apoyó reformas institucionales orientadas a desarrollar una política de eficiencia energética y a poner en operación proyectos de inversión, entre ellos el proyecto BO-L1190.

El proyecto BO-L1190 consiste en el financiamiento de dos inversiones específicas: (1) desarrollar proyecto de alumbrado público eficiente, y (2) desarrollar infraestructura de líneas de transmisión. El objetivo final es apoyar la diversificación de la matriz energética de Bolivia y promover el uso eficiente de la electricidad con el fin de reducir emisiones de dióxido de carbono y contribuir al desarrollo de las localidades en las áreas de intervención.

Con respecto a la inversión específica para un alumbrado público eficiente, el proyecto busca sustituir luminarias convencionales de alto consumo, como por ejemplo sistemas de vapor de sodio de alta presión, y reemplazarlas por luminarias LED que posean alto flujo luminoso y alta eficiencia lumínica. Para ello, se utilizarán drones para georreferenciar todos los puntos de luz en las calles con el fin de identificar la potencia de las luminarias y eventualmente reemplazar la tecnología convencional por luminarias LED. Entre los resultados esperados por la adopción de estas tecnologías, se encuentran la reducción de costos de operación y reemplazo, reducción de consumo energético, aumento del tiempo de vida útil de acuerdo a las necesidades específicas de las ciudades, e indirectamente contribuir a la seguridad y mejora de calidad de vida de la población beneficiada.

Recomendación: Evaluación del impacto del proyecto en los niveles de pobreza mediante datos satelitales y teléfonos móviles

Dado que uno de los objetivos del proyecto es mejorar la calidad de vida de la población beneficiada, se recomienda utilizar big data para evaluar una posible mejora en los niveles de ingreso en las zonas beneficiarias. El objetivo sería entender de qué manera la expansión de infraestructura eléctrica incide en el nivel de desarrollo económico de las poblaciones.

Para esto, se recomienda medir la pobreza utilizando datos agregados de datos geoespaciales y operadores móviles. Los enfoques tradicionales para medir y abordar la pobreza dependen en gran medida de los datos de censos y encuestas, que en la mayoría de los países de ingresos bajos y medios no están disponibles o están desactualizados. Utilizando nuevos tipos de datos, sería posible medir la evolución de niveles de pobreza, incluso en zonas rurales de difícil acceso que sufren de una ausencia de información.

Existen diferentes enfoques para calcular los indicadores del nivel de vida, incluyendo la creación de índices unidimensionales y multidimensionales, así como el uso de métricas monetarias o no monetarias. Las métricas monetarias identifican la pobreza como un déficit en el consumo (o ingreso) y miden si los hogares o los individuos se encuentran por encima o por debajo de una línea de pobreza definida. Por el contrario, los indicadores basados en activos definen el bienestar de los hogares según la propiedad de estos activos (por ejemplo, refrigeradores, radios o bicicletas), las características de vivienda y el acceso a servicios básicos como agua limpia y electricidad.

La utilización de Sensores Remotos y de Metadatos de Llamadas telefónicas permite también abordar la evaluación del nivel de pobreza:

- Sensores Remotos (en inglés, *Remote Sensing - RS*) y Datos del Sistema de Información Geográfica (en inglés, *Geographic Information Systems - GIS*): se pueden usar para medir la distancia a carreteras y ciudades, i.e acceso a mercados e información, o para medir la intensidad de la luminosidad nocturna, i.e infraestructura de alumbrado público e índices de propiedad de activos en hogares.
- Metadatos de Llamadas Telefónicas (en inglés, *Call Detail Records - CDRs*): se pueden utilizar para rastrear movimientos de teléfonos móviles a un nivel agregado. La estructura y el alcance geográfico de las redes de llamadas individuales pueden correlacionarse con los flujos de remesas y oportunidades económicas.

Los datos de *RS* y *CDR* capturan distintas características complementarias que describen aspectos relevantes de las condiciones de vida y comportamiento humano. Además, dado que estos se recopilan continuamente, la capacidad de producir mapas precisos utilizando estos tipos de datos ofrece la promesa de un monitoreo subnacional continuo requerido por los Objetivos de Desarrollo Sostenible.

Caso de estudio

Steele et al., 2017 trabajaron con fuentes superpuestas de *RS*, *CDR* y datos basados en encuestas tradicionales de Bangladesh para estimar con precisión tres medidas diferentes de pobreza. El método de evaluación propuesto por Steele et al. proporciona un enfoque flexible para modelar la pobreza y representa el primer intento de construir mapas predictivos de pobreza utilizando una combinación de datos de *CDR* y *RS*.

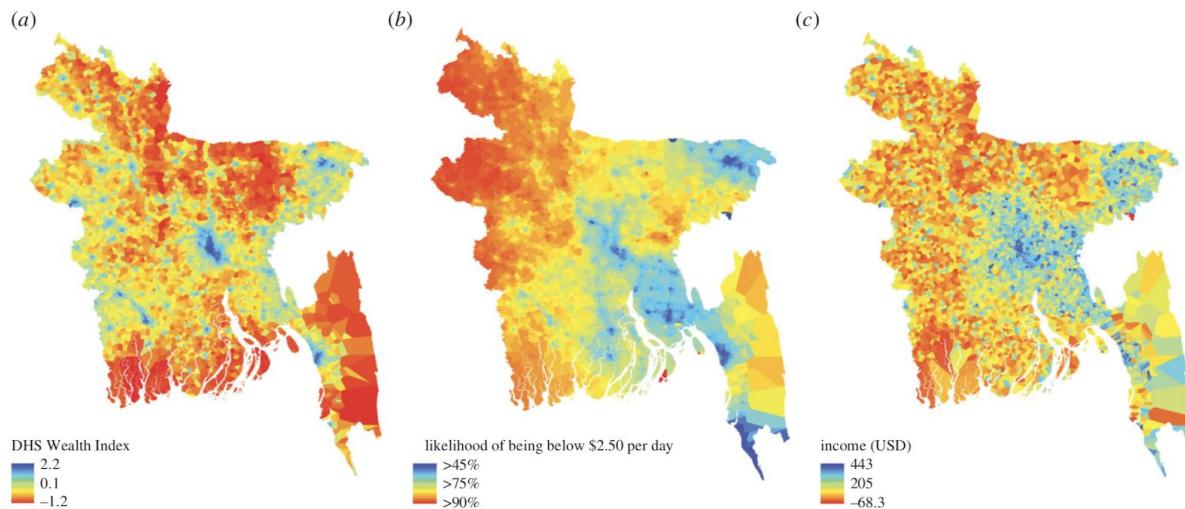


Figura: Mapas de predicción a nivel nacional de Bangladesh para el ingreso de riqueza promedio. Los mapas se generaron utilizando las funciones de metadatos de llamadas telefónicas, datos de RS y modelos geo estadísticos Bayesianos. El color rojo indica áreas más pobres en los mapas de predicción (Steele et al., 2017)

La metodología propuesta representa una alternativa robusta y económica a los índices de pobreza basados en encuestas. Para el caso de estudio BO-L1190, la investigación de Steele et al demuestra que una combinación de fuentes alternativas de datos logra medir de forma continua y con alta resolución espacial niveles socioeconómicos.

Algunos aspectos que podrían evaluarse en relación con este proyecto son los siguientes:

- Evolución del grado de luminosidad en las zonas con nueva conexión eléctrica: esto permitiría evaluar la pobreza utilizando un indicador basado en activos, i.e en el número de electrodomésticos o las características de infraestructura de las viviendas y barrios (alumbrado público)
- Evolución de los patrones de residencia en las zonas recientemente conectadas para medir efectos en términos de atractividad de población
- Evolución de los patrones de movilidad de la población residente en zonas recientemente conectadas para medir efectos en términos de actividad económica y redes sociales

Plan de trabajo sugerido

Posibles fuentes de datos (big data)	De libre acceso: <ul style="list-style-type: none"> - Imágenes satelitales generadas por NOAA (National Oceanic and Atmospheric Administration)
---	--

	<p>En venta:</p> <ul style="list-style-type: none"> - Digital Globe es una empresa especializada en imágenes satelitales y mapas, ofreciendo diferentes servicios de análisis y visualización. <p>Otras:</p> <p>Para acceder a datos de telefonía móvil (CDR) es necesario entablar acuerdos con los operadores privados que los generan. El BID cuenta con un acuerdo marco con Telefónica cuyo alcance podría ser estudiado para integrar este tipo de análisis.</p> <p>Una plataforma como OPAL (Open Algorithms) también espera poner a disposición de actores del desarrollo este tipo de datos bajo una modalidad freemium.</p>
Capacidades Requeridas	<p><i>Capacidades humanas:</i> se requiere un jefe de proyecto que pueda coordinar la interacción entre la ciencia de datos y las necesidades en términos de M&E que existan en este proyecto.</p> <p>El equipo científico debe contar con capacidades técnicas en analítica, programación, bases de datos, y procesamiento de datos masivos: debe incluir un data scientist con experiencia en GIS. También se sugiere contar con un experto legal que evaluará los aspectos éticos y de privacidad.</p> <p><i>Infraestructura Tecnológica:</i> es posible que se necesiten servidores físicos o en la nube para almacenar y procesar la cantidad de datos necesarios para llevar a cabo la evaluación. Esto dependerá del modelo de colaboración que se defina con el controlador de los datos. Si se utiliza una plataforma como OPAL, no se necesitaría mayor infraestructura tecnológica pues se podría acceder directamente a los indicadores.</p> <p>Programa para el análisis:</p> <p>La herramienta considerada en Steele et al. fue desarrollada con R. R es un entorno y lenguaje de programación con un enfoque al análisis estadístico. Para este tipo de análisis el entorno es capaz de cargar programas llamados paquetes como glmuti para la selección de modelo, o las herramientas estadísticas (integrated nested Laplace approximations) de R-INLA</p> <ul style="list-style-type: none"> • Sistemas de información geográfica: ArcGIS, QGIS, Google Earth Engine
Riesgos y Limitaciones -	<p>Los mapas de luminosidad obtenidos por la NOAA sólo están disponibles a partir de 2011, lo cual reduce la posibilidad de contar con datos históricos sobre las zonas estudiadas.</p>

	<p>La utilización de CDR para medir comportamientos individuales como patrones de migración, movilidad y redes sociales debe tomar en cuenta posibles riesgos éticos relacionados con invasión de la privacidad o identificación de grupos vulnerables. Por lo tanto, se recomienda utilizar datos agregados y anónimizados y mantener en los estudios un nivel de granularidad cuya precisión no sea tan elevada.</p> <p>Es igualmente importante reflexionar sobre cómo incluir a las poblaciones que serán estudiadas en las decisiones y los resultados que se obtendrán con sus datos.</p>
Próximos pasos	<p>Para llevar a cabo esta evaluación se sugiere:</p> <ul style="list-style-type: none"> - Identificar los impactos que se quieren evaluar, los datos actuales con los cuales se trabajaría, y los posibles datos alternativos que podrían llenar lagunas existentes. - Identificar la modalidad de acceso a estos nuevos datos (acceso libre, acuerdo, venta, etc.) y entablar discusiones con los controladores o facilitadores de datos - Identificar los riesgos éticos y de privacidad y definir un plan de acción para contrarrestarlos - Definir una propuesta de piloto para probar la metodología de la evaluación en una zona / tiempo restringidos - Reunir un equipo con las capacidades necesarias

B. COLOMBIA: "Sistemas Estratégicos de Transporte Público (SETP)" (CO-L1091)

Estado y avance del proyecto

El proyecto "Sistemas Estratégicos de Transporte Público (SETP)" (CO-L1091) perteneciente al sector de Transporte fue aprobado en 2011 y consiste en una operación de préstamo de un costo total de USD 340,000,000 que comenzó a ejecutarse en 2013. Aunque el financiamiento del proyecto acababa en 2018, se trató la extensión por dos años más hasta 2020 debido a ciertos retrasos y dificultades.

El objetivo principal del proyecto CO-L1091 consiste en apoyar al gobierno de Colombia en la implementación de sistemas de transporte público en ciudades de porte medio, particularmente en Pasto, Popayán, Armenia y Santa Marta, con el fin de desarrollar una movilidad urbana

sostenible.

Estos componentes se refieren principalmente a (i) construcción de infraestructura de transporte, como por ejemplo rutas estratégicas, puentes peatonales, paraderos, y (ii) sistemas tecnológicos para la puesta en operación, en particular sistemas de gestión y control de flota, sistemas de control de tráfico, y sistemas de recaudo, entre otros componentes relacionados a la gestión del proyecto y adquisición de predios, planes de compensación y reasentamiento de poblaciones afectadas.

El proyecto ha avanzado de manera general en el área de infraestructura donde el diseño y la construcción se realizan de acuerdo a las particularidades de cada ciudad. Sin embargo, no ha habido la maduración necesaria para garantizar una puesta en operación exitosa que se alcance no sólo a través de la incorporación de herramientas tecnológicas sino también desde el punto de vista de la capacidad organizacional de los entes gestores locales. La ciudad de Pasto cuenta con un centro de control muy avanzado y en el cual se realizó una inversión muy importante en sistemas similares a los utilizados en grandes ciudades como París o Nueva York. Sin embargo, este tipo de sistemas, además de ser costosos, no están necesariamente adaptados al tamaño medio de la ciudad, sobre todo teniendo en cuenta que, desde el inicio de la ejecución del proyecto en 2013, se han diseñado técnicas y herramientas más actualizadas. Por otro lado, tanto Pasto, como Armenia, Popayán y Santa Marta, no cuentan con personal técnico especializado capacitado para operar estos sistemas, por lo cual la capacidad de análisis de datos es igualmente débil. Esto se debe en parte a la dificultad de retener las personas capacitadas dentro de los entes gestores locales.

Por lo tanto, las mayores dificultades del proyecto se refieren a la puesta en operación de los SETP, no sólo por la inexistencia de sistemas tecnológicos en las ciudades de Armenia, Popayán y Santa Marta, sino también por sistemas demasiado costosos y no actualizados que están poco adaptados a la realidad local, como en el caso de Pasto, además de la insuficiencia de capacidades técnicas para operar estos sistemas tecnológicos dentro de los entes gestores de las cuatro ciudades del proyecto.

Recomendación: Sistema basado en Metadatos de Llamadas Telefónicas (*Call Detail Records - CDRs*) para la Gestión del Tráfico

El sector del transporte y la movilidad ha sido una de las principales áreas de utilización de Big Data en los últimos años. En lugar de utilizar cronómetros o encuestas a hogares para calcular tiempos de viaje y matrices de origen-destino, es posible basarse en sensores GPS de los buses o en metadatos de llamadas telefónicas para obtener esas informaciones de forma mucho más precisa, frecuente y práctica.

También se están desarrollando alianzas público-privadas entre gobiernos locales y empresas del sector (Waze, Mobike, etc.), permitiéndole a las autoridades locales acceder a los datos

recolectados por estas aplicaciones sobre el perfil de sus usuarios, los trayectos recorridos e incluso los reportes de accidentes¹⁰. Todo esto representa una oportunidad de mejorar la infraestructura y la gestión de los sistemas de transporte públicos movilizando datos existentes, lo cual puede reducir drásticamente los costos y los tiempos necesarios para obtener este tipo de información.

En el caso del proyecto COL-1091, la utilización de fuentes de Big Data podría ayudar a mejorar la planificación del sistema de semaforización, adaptar el diseño de las rutas a la demanda para reducir la sobreoferta y atender la demanda existente, así como proporcionar recomendaciones a usuarios para mejorar la eficiencia colectiva del sistema.

Para ello se propone realizar un análisis de registros de teléfonos móviles (en inglés, Call Detail Records - CDRs), los cuales no solo pueden revelar patrones de movilidad humanos, sino que también permiten gestionar la demanda de movilización para una mayor eficiencia del sistema.

Los análisis de los CDRs pueden complementarse con métricas disponibles en sistemas actuales (o futuros) de gestión de tráfico urbano que incluyan datos de cámaras en ubicaciones clave, y archivos de tipo de sistemas de información geográfica (en inglés, *Geographic Information Systems - GIS*). Otros datos relevantes incluyen datos de redes sociales geolocalizados, el número de carriles, capacidades de carreteras, y la estimación de tiempo de viaje sin tráfico (disponible a través de la API de Google Maps).

Datos de telefonía móvil y patrones de movilización

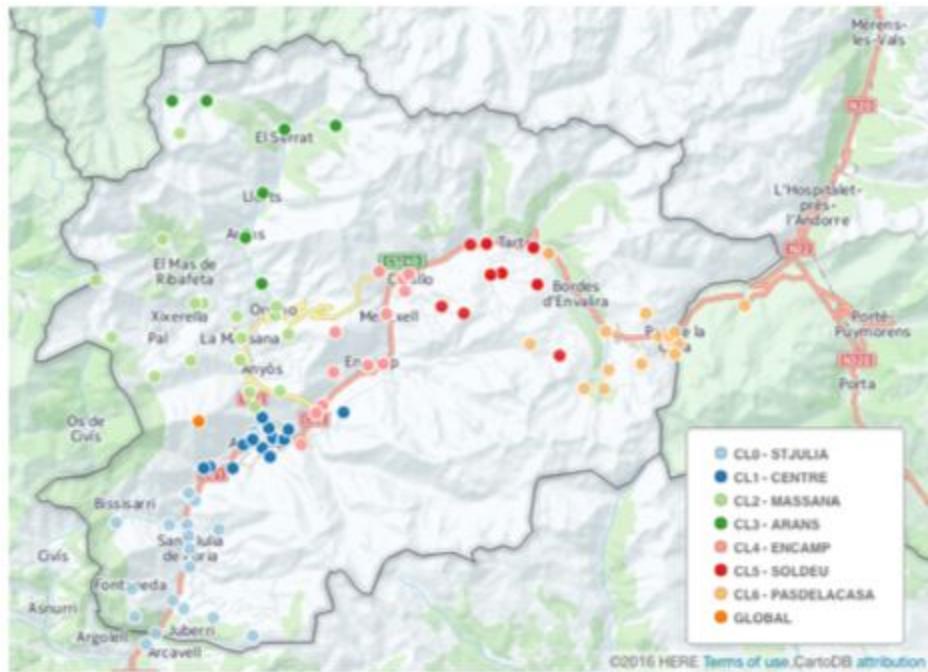
Los Metadatos de Llamadas Telefónicas (CDR) son registros generados por operadores de telecomunicaciones que reflejan el uso y los atributos de los servicios como el tipo de interacción (llamadas de voz o mensajes de texto), la fecha y duración de la llamada. La información de geolocalización, como la ubicación de la antena, suele estar disponible en los CDR. Con esta información es posible calcular patrones de movilidad individuales y colectivos, y estimar flujos de tráfico en las ciudades.

interaction	direction	correspondent_id	datetime	call_duration	antenna_id
call	in	8f8ad28de134	2012-05-20 20:30:37	137	13084
call	out	fe01d67aecccd	2012-05-20 20:31:42	542	13084
text	in	c8f538f1ccb2	2012-05-20 21:10:31		13087

Figura: Campos de muestra de CDR¹¹

¹⁰ International Consortium for Data collaboratives, World Bank, IADB, IMF www.datapartnership.org

¹¹ <http://bandicoot.mit.edu/>



Distribución de torres celulares en Andorra coloreada por ciudad (Leng et al, 2016)

Por otro lado, los CDRs también han sido utilizados para mejorar la eficiencia de los sistemas de transporte y la congestión del tráfico, generando estrategias de gestión en la demanda de viajes. La siguiente recomendación está basada en el estudio de Leng et al, 2016¹² que destaca la efectividad de la sinergia entre los comportamientos colectivos para aumentar la eficiencia del sistema.

En las ciudades colombianas, esta metodología podría ser utilizada para adaptar el diseño de las rutas y la cantidad de buses asignados a cada ruta. También podría ser utilizado para hacer recomendaciones a viajeros sobre rutas, destinos y horarios para movilizarse.

La metodología se divide en tres pasos principales:

1. Inferencia de la demanda de viajes en vehículos a lo largo de enlaces de carretera basados en CDR.
2. Preferencias personales de ubicación (necesidades) basadas en trazas de ubicación.
3. Predicción de la siguiente ubicación basada en trazas históricas y comparación de la ubicación de la recomendación.

¹² Leng, Y., Rudolph, L., Pentland, A., Zhao, J., Koutsopoulos, H., "Managing travel demand: Location recommendation for system efficiency based on mobile phone data", Proceedings of Data for Good Exchange (D4GX), 2016: [arXiv:1610.06825](https://arxiv.org/abs/1610.06825)

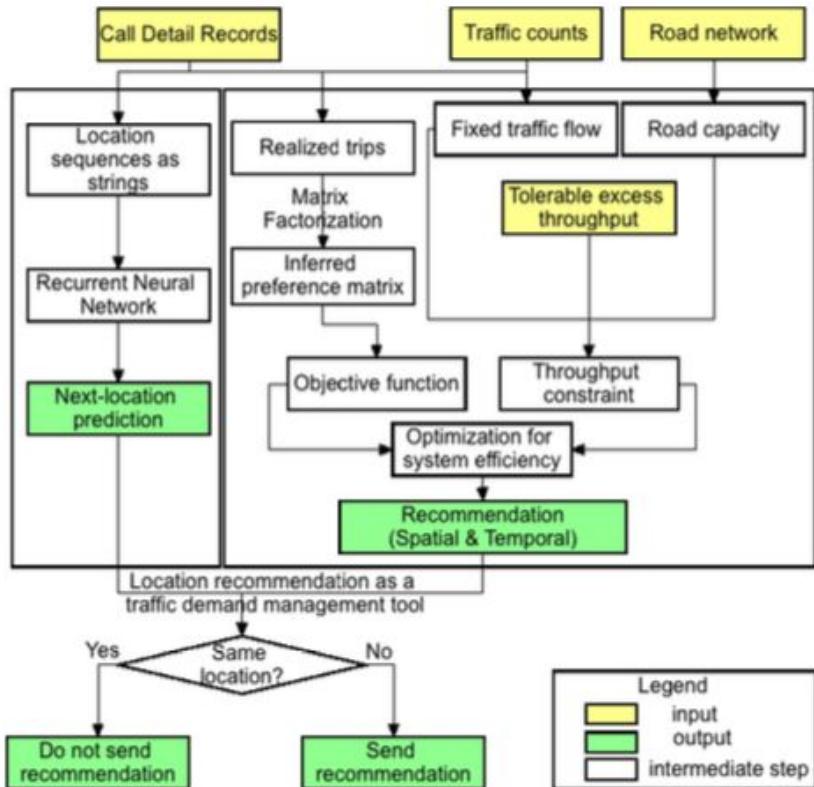


Figura: Marco metodológico¹³

El objetivo de las recomendaciones es optimizar las preferencias de ubicación y de horario de los viajeros con la restricción de una congestión aceptable. Para la aplicación en situaciones reales, sería también necesario desarrollar un sistema de comunicación eficiente con los usuarios y con los centros de controles del sistema de transporte.

Posibles fuentes	Para acceder a los CDR, el Banco puede optar por elaborar un acuerdo con operadores de telefonía móvil quienes pueden poner a disposición o vender datos e indicadores basados en CDR. La plataforma OPAL en Colombia también permite a los usuarios acceder a indicadores basados en CDR, y podría ser un medio para autoridades locales y para el Banco de utilizar estas fuentes de datos privadas sin costo o por un costo moderado.
-------------------------	---

¹³ Leng, Y., Rudolph, L., Pentland, A., Zhao, J., Koutsopoulos, H., "Managing travel demand: Location recommendation for system efficiency based on mobile phone data", Proceedings of Data for Good Exchange (D4GX), 2016: [arXiv:1610.06825](https://arxiv.org/abs/1610.06825)

	<p>Los indicadores obtenidos con CDR pueden ser completados y mejorados con datos obtenidos de aplicaciones de movilidad (Waze, Google Maps, etc.), con análisis de actividad en redes sociales</p> <p>Si las cuentan con sistemas de cámaras o conteos de pasajeros también es importante incluir estos datos en los análisis de flujos de movilidad.</p>
Capacidades Requeridas	<p>Capacidades en infraestructura: Para realizar los análisis de movilidad a través de los CDR puede optarse por una infraestructura ligera que sólo reciba los resultados y no la totalidad de los datos. Por ejemplo, al utilizar la plataforma OPAL, los cálculos se realizan en los servidores de las mismas empresas y los usuarios no necesitan comprar los servicios de almacenamiento y procesamiento de datos.</p> <p>Capacidades humanas: Si bien la ciudad puede optar por externalizar los servicios de análisis computacional de los datos, se debe contar con personal capaz de interpretar correctamente los resultados para utilizarlos de forma eficiente.</p>
Riesgos y Limitaciones	<p>Puesto que los CDR están en posesión de compañías privadas de telefonía móvil, es necesario desarrollar un convenio de colaboración para poder acceder a ellos. Se deben tomar en cuenta los siguientes riesgos jurídicos:</p> <ul style="list-style-type: none"> - posible incumplimiento del convenio por una parte - retrasos en el proceso de firma del acuerdo - incompatibilidad de los intereses entre las partes <p>Dado que una parte de la recomendación incluye mantener un flujo de comunicación con los usuarios, es importante evaluar las probabilidades de adopción de las recomendaciones por los mismos y los mejores canales de comunicación disponibles.</p> <p>Se ha identificado la falta de personal capacitado en las administraciones de las ciudades, lo cual representa una limitación importante a este tipo de proyectos, pues se trata de los usuarios finales. Para hacer frente a este obstáculo, se sugiere desarrollar un componente de formación y construcción de capacidades para incentivar una mayor apropiación de estas herramientas. También se podrían explorar alianzas con universidades o centros de estudio que</p>

	tengan las capacidades instaladas y quieran contribuir a desarrollar y gestionar el sistema.
Próximos pasos	<ul style="list-style-type: none"> - Seleccionar una ciudad del proyecto para llevar a cabo una prueba piloto y un objetivo inicial entre, por ejemplo: mejorar el sistema de semaforización, adaptar el diseño de las rutas o la cantidad de buses, estudiar patrones de movilidad problemáticos, etc. - Identificar las fuentes de datos y las diferentes formas posibles de acceder a ellas (convenio con compañías, OPAL, fuentes de libre acceso, etc.) - Identificar los diferentes actores involucrados (empresas, funcionarios locales, usuarios finales, socios académicos, etc.) - Desarrollar una propuesta detallada del piloto con indicadores claros de éxito y retroalimentación

C. PERÚ: "Proyecto para la Mejora de la Eficiencia en la Gestión de la Inversión y las Contrataciones Públicas" (PE-L1231)

Estado y avance del proyecto

El "Proyecto para la Mejora de la Eficiencia en la Gestión de la Inversión y las Contrataciones Públicas" (PE-L1231) perteneciente al sector de Reforma y Modernización del Estado fue aprobado en 2018 y consiste en una operación de préstamo de un costo total de USD 73,750,000. Las dos contrapartes principales son la Dirección General de Inversión Pública y el Organismo Supervisor de las Contrataciones del Estado (OSCE)

El objetivo principal del proyecto PE-L1231 es mejorar la calidad de la gestión de la inversión y las contrataciones públicas de Perú a través de la creación de un nuevo sistema que fortalezca una mejor coordinación interinstitucional, de planificación y transparencia de las inversiones, generando así una gestión más eficiente. El programa contiene los siguientes componentes: (i) mejora de la capacidad del marco institucional, incluyendo el desarrollo de metodologías para el ciclo de proyectos e instrumentos de apoyo, (ii) desarrollo e implementación de un sistema informático de gestión de la inversión pública interoperable con otros sistemas de gestión de las finanzas públicas, (iii) mejora de la capacidad del capital humano.

Respecto al segundo componente, el sistema actualmente utilizado por OSCE está concebido como un sistema digitalizado que archiva documentos relativos a las inversiones y contrataciones públicas anteriores, resultando en una sumatoria de archivos PDF de los cuales

no se logra extraer información específica de forma ordenada y rápida. Dentro de este contexto, el proyecto busca impulsar la adopción de un nuevo sistema que podría generar más beneficios que la sola digitalización de archivos. La tecnología que el proyecto considera desarrollar seguiría con las funciones de archivador, pero también incorporaría varios módulos que extraerán información precisa para analizar datos y realizar estudios macroeconómicos aplicados. Asimismo, se buscaría generar una plataforma que permita la interacción entre los distintos actores institucionales relevantes con el fin de complementar datos e información. Por lo tanto, este proyecto implica una transformación tecnológica no sólo desde el punto de vista de la creación de un nuevo sistema, sino también del desarrollo de toda una infraestructura que lo soporte.

Mientras que el proyecto se encuentra actualmente en la fase de pre-arranque, la dificultad principal está en concebir un nuevo sistema que sea integral y adaptado a la realidad peruana, y que pueda a su vez generar valor de los datos contenidos en los archivos digitalizados del sistema actual.

Recomendación: Extracción de información de documentos PDF utilizando Text Analytics

Pasar de un corpus de información textual (no estructurado) almacenado en formato digital (por ejemplo, PDF) a una salida estructurada puede ser una tarea exigente, denominada genéricamente como el "cuello de botella en la adquisición de conocimiento". Esta tarea es muy exigente en cuanto a recursos (especialmente recursos humanos con suficiente experiencia para analizar los documentos de origen) y también requiere una gran cantidad de conocimientos, ya que quien esté haciendo la anotación debe saber cómo leer el documento en un dominio determinado.

El análisis de texto puede ayudar a descubrir patrones con un significado oculto en los repositorios de datos. De este análisis se podría obtener informaciones valiosas para mejorar la eficiencia en la gestión de inversiones y la contratación pública. Por lo tanto, se propone el uso de análisis de texto para descubrir información nueva, previamente desconocida, mediante la extracción automática de información de documentos PDF.

Una de las propuestas de valor sería identificar los sectores o proyectos donde consistentemente se observan errores presupuestarios que conllevan gastos inesperados en el futuro, con el fin de mejorar las asignaciones presupuestarias o modificar las políticas de aseguramiento en esos sectores. También se podría establecer una lista de sectores en donde se observan pocos proveedores, para impulsar políticas de competencia y diversificación de proveedores. Finalmente, se podrían identificar patrones de corrupción o de sobrecostos.

Metodología

Text Analytics se considera a menudo como el siguiente paso en el análisis del Big Data, y se refiere al análisis del contenido de texto en contenido digital almacenado como documentos, correos electrónicos, blogs, tweets y otras formas de comunicación textual.

Una aplicación típica de análisis de texto consta de los siguientes pasos y tareas: a partir de una colección de documentos, la herramienta de explotación de texto recupera un documento en particular y lo procesa previamente verificando el formato y los juegos de caracteres. Luego pasa por una fase de análisis de texto, a veces repitiendo técnicas hasta que se extrae la información. La estrategia subyacente en todos los componentes es encontrar un patrón (de una lista o de un proceso anterior) que coincida con una regla, y luego aplicar la regla que anota el texto. El proceso resultante proporciona información "estructurada" o semiestructurada que se puede utilizar para la creación de la Base de Conocimientos ("Knowledge Base Building").

Una de las técnicas que se han desarrollado en el área de Text Analytics es la extracción de información. Como describe Moreno, Redondo 2016, el software de extracción de información (en inglés, *Information Extraction - IE*) tiene la capacidad de identificar frases y relaciones clave dentro del texto. Lo hace buscando secuencias predefinidas en el texto, un proceso que generalmente se denomina coincidencia de patrones y es basado en expresiones regulares. La forma más popular de *IE* se denomina reconocimiento de entidad (en inglés, *Named-Entity Recognition, NER*) y es diseñado para ubicar y clasificar elementos del texto en categorías predefinidas (usualmente en ontologías preestablecidas). Las técnicas de *NER* extraen características como los nombres de personas, organizaciones, ubicaciones, expresiones temporales o espaciales, cantidades, valores monetarios, valores de acciones, porcentajes, entre otra información. Algunas de las herramientas relevantes para esta tarea son: Apache OpenNLP, Stanford Entidad y LingPipe (compatible en idioma castellano).

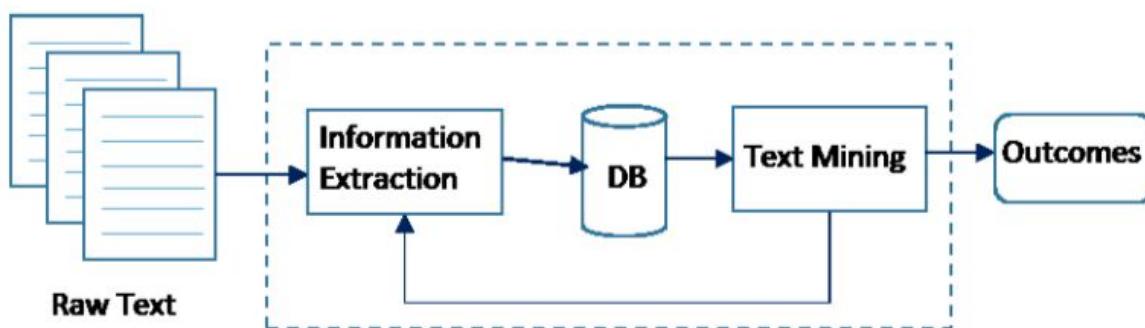


Figura: Marco de explotación de texto (Moreno, Redondo 2016)

La extracción de información puede desempeñar un papel importante en el área de gestión de inversiones y contratación pública. Este tipo de sistema permitiría anotar elementos de interés, con el fin de identificar un rango de elementos de información específico que serían relevantes para los actores institucionales, tales como:

- Proyecto de caso
- Nombres de las organizaciones involucradas
- Roles de las partes interesadas
- Nombres de grupos de interés, directivos, clientes
- Naturaleza del proyecto, es decir, usar palabras clave para clasificar el proyecto por área/temática
- Cifras
- Cualquier otra información relevante del proyecto.

El interés de este tipo de herramienta es que permitiría extraer la mayor cantidad de información contenida en un documento de una forma eficiente y escalable. Adicional a generar y brindar acceso a una librería de archivos completos de licitaciones públicas, esta herramienta sirve como cimiento para un análisis mucho más detallado que incorpore elementos cualitativos y metadatos (determinar cuáles empresas están consistentemente ganando contratos o si existe un patrón de coordinación entre diversos actores, entre otros). La herramienta entonces puede ser comparada con la cantidad de horas hombre que se tendrían que dedicar para poder procesar y extraer la información relevante. Aunque este proceso sería extenso y propenso al error humano de cierta forma brindaría un mayor nivel de confianza en la información capturada, sobre todo si se tiene procesos de verificación para auditar posibles inconsistencias en la información. Una forma de hacer escalable un proceso de captura manual sería a través de *Crowdsourcing* (del inglés crowd –multitud– y outsourcing –recursos externos–) en español como *colaboración abierta distribuida* o *externalización abierta de tareas*, se podría considerar entonces herramientas como **Amazon Mechanical Turk (AMT)**. Esta plataforma sirve para trabajos simples y de bajo precio unitario que requieren un cierto nivel de inteligencia que una máquina no puede hacer. Sin embargo, esta opción no es probablemente sostenible en el largo plazo considerando la periodicidad y el volumen de los archivos de compras públicas.

En una fase posterior, el proceso de contratación puede enriquecerse integrando índices de mercado que pueden ser desarrollados a través de big data para establecer presupuestos mejor informados. Otra área de estudio puede ser la utilización de datos de redes sociales (Twitter, Facebook) para analizar la percepción de los ciudadanos frente a la calidad de los productos comprados.

Posibles fuentes	Esta propuesta se basa casi exclusivamente en los documentos PDF que ya están en posesión de la OSCE. Una forma de incrementar el valor de los análisis puede ser cruzando los
-------------------------	---

	datos con otro tipo de información disponible como el presupuesto histórico de los sectores estudiados para compararlo con los gastos incurridos.
Capacidades Requeridas	<p>Infraestructura: Se debe encontrar una herramienta que pueda traducir las imágenes en elementos para ser categorizados en la base de datos. Una herramienta comúnmente utilizada para cumplir con esta función es llamada OCR por sus siglas en inglés para Optical Character Recognition. Los programas que pueden ser utilizados para el análisis del texto son: Apache OpenNLP, Stanford Entidad y LingPipe (compatible en idioma castellano).</p> <p>Humanas: Será necesario contar con un científico de datos para diseñar los patrones de análisis iniciales y desarrollar la herramienta. Esta debería estar contar con una interfaz simple que permita a cualquier funcionario de la OSCE utilizarla sin problema. Se recomienda también contar con un especialista en comunicaciones para subrayar el impacto el proyecto y generar adhesión entorno a la nueva herramienta.</p>
Riesgos y Limitaciones	<p>El proyecto no presenta mayores riesgos de privacidad. Tiene un componente importante de transparencia y eficiencia del sector público que debe ser bien comunicado y explotado para sacar el mayor provecho.</p> <p>Puede que se encuentren dificultades iniciales en el reconocimiento de texto, dando lugar a inconsistencias en el análisis. Para ello, se sugiere realizar primero el análisis de una muestra de documentos que permita identificar y corregir problemas.</p>
Próximos pasos	

D. VENEZUELA: "Programa Apoyo Centro Acción Social por la Música Fase II" (VE-1017)

Estado y avance del proyecto

El proyecto "Programa Apoyo Centro Acción Social por la Música Fase II" (VE-1017) perteneciente al sector de Inversión Pública fue aprobado en el 2007 y consiste en una operación de préstamo de un costo total de USD 211,000,000.

El objetivo del proyecto VE-1017 es contribuir a la inclusión y protección social de Venezuela, y en particular mejorar las condiciones de vida de los niños, niñas y jóvenes, a través del fortalecimiento y expansión del Sistema Nacional de Orquestas Juveniles e Infantiles de Venezuela ("Sistema"). Para lograr este objetivo, el programa contiene tres componentes: (i) desarrollo estratégico del Sistema, en particular en las áreas académica-musical, gerencial, financiero-presupuestaria y de captación de recursos, (ii) inversiones físicas, en particular el financiamiento en el diseño y construcción de siete centros regionales, y (iii) desarrollo de un sistema de supervisión y evaluación que mida el impacto social y a nivel individual del Sistema y de la implementación del programa.

El proyecto VE-1017 se alinea dentro de los objetivos de una operación anterior (VE-105) también apoyada por el Banco y que permitió una expansión substancial del Sistema manteniendo la calidad del servicio. En su primera fase, el programa también fortaleció el prestigioso método académico-musical único en el mundo y los procedimientos administrativos-financieros del Sistema, sobre todo ante el incremento del número de beneficiarios y orquestas. En su segunda fase, el proyecto busca una expansión mayor del Sistema, incluyendo la compra de instrumentos musicales para responder a la nueva demanda, consolidando así los logros del proyecto VE-105. Además, el proyecto VE-107 busca promover una presencia del Sistema más importante a nivel subnacional y menos centralizada en Caracas, a través del financiamiento en infraestructura de siete Centros Regionales.

Sin duda, la característica distintiva del Sistema con respecto a otras iniciativas de orquestas juveniles en el mundo es la concepción de la música como vehículo para la inclusión y protección social más que una vía de carrera profesional. Es así que alrededor de la música, se han establecido programas sociales, en donde por ejemplo madres de bebés y niños pequeños asisten a conciertos, pero también a cursos sobre alimentación, lactancia, prevención de embarazo, entre otros temas que van más allá de lo recreativo. La estrategia comunicacional del Sistema consiste principalmente en conciertos que se realizan a todos los niveles, desde lo internacional hasta lo comunitario. Siguiendo el objetivo de inclusión social, busca integrar tanto población vulnerable como no vulnerable - promoviendo así la interacción entre personas pertenecientes a distintas categorías sociales.

El Sistema ha desarrollado, con la ayuda del Banco, un plan de negocios para expandirse a nivel mundial, desarrollando un sello de calidad que compruebe la debida implementación del Sistema en otros países. La puesta en operación de este nuevo plan de negocios puede ser un área para incorporar herramientas tecnológicas que apoyen una puesta en operación exitosa.

Recomendación: Evaluación del impacto del proyecto en redes e interacciones sociales

Los estudios demuestran que existe una fuerte correlación entre relaciones individuales

diversas socialmente y el desarrollo económico de las comunidades¹⁴. El Sistema, al estimular intercambios entre individuos de diversas clases sociales, puede contribuir a este objetivo. Por lo tanto, se sugiere trabajar con fuentes de big data para evaluar el impacto del programa en las redes sociales y los comportamientos de los estudiantes que hacen parte del programa. Esto podría proporcionar métricas objetivas sobre el impacto de los programas de música actuales y futuros en el país.

Los enfoques tradicionales adoptados en la psicología social para comprender el comportamiento humano incluyen encuestas. Sin embargo, en teoría, las encuestas tienden a ser costosas y, a menudo, inadecuadas para el análisis a largo plazo. Además, tienden a ser subjetivas y pueden ser sujetas a sesgos de percepción¹⁵.

Para el análisis de las interacciones sociales, sería posible utilizar registros de las interacciones de los usuarios (registros de llamadas, SMS y Bluetooth). Otra opción sería movilizar redes sociales como Facebook, Twitter o Whatsapp.

El comportamiento interpersonal a largo plazo se puede utilizar para comprender y medir aspectos relacionados con la comunidad como la cercanía, la confianza, la fuerza de los lazos y la amistad. Los registros de WiFi y GPS se utilizan para un análisis más sólido (pero a menudo más amplio) de los patrones de movilidad y ubicación conjunta de las personas. Un estudio de Shmueli et al. 2014¹⁶ muestra cómo se pueden usar las señales de interacción social basadas en llamadas, mensajes de texto y Bluetooth para predecir la confianza en la relación y cuantificar su papel en la mediación de la persuasión.

Shmueli et al. El 2014 demostró que, a través del uso de una muestra rica y densa de las vidas de más de 100 participantes que viven en una sola comunidad durante un año¹⁷, es posible (1) predecir relaciones de confianza mediante métodos computacionales basados en la detección pasiva (de llamadas, mensajes de texto e interacciones de Bluetooth) y el análisis de red, y (2) la confianza puede tener un impacto significativo en la persuasión social: es significativamente más eficaz que la cercanía de los lazos para determinar la cantidad de cambio de comportamiento.

¹⁴ Nishikata, T., Hardjono, T., & Pentland, S., "Social Capital Accounting", MIT Media Lab, October 2018: <https://www.media.mit.edu/publications/social-capital-accounting/>

¹⁵ Fogg, BJ., "Persuasive technology: Using computers to change what we think and do", Morgan Kaufmann, 2003

¹⁶ Shmueli, E., Singh, V. K., Lepri, B., & Pentland, A., "Sensing, understanding, and shaping social behavior", IEEE Transactions on Computational Social Systems, April 2014, Vol. 1, Issue 1, pages 22–34. <http://doi.org/10.1109/TCSS.2014.2307438>

¹⁷ El conjunto de datos Amigos y Familia (en inglés, *Friends and Family dataset*), introducido por primera vez por Aharony et al. utiliza la plataforma de telefonía móvil, encuestas, recibos, extracto de tarjeta de crédito y una aplicación de Facebook para recopilar un conjunto de datos inmensamente rico y denso sobre las vidas de los 130 participantes (aproximadamente 64 familias).

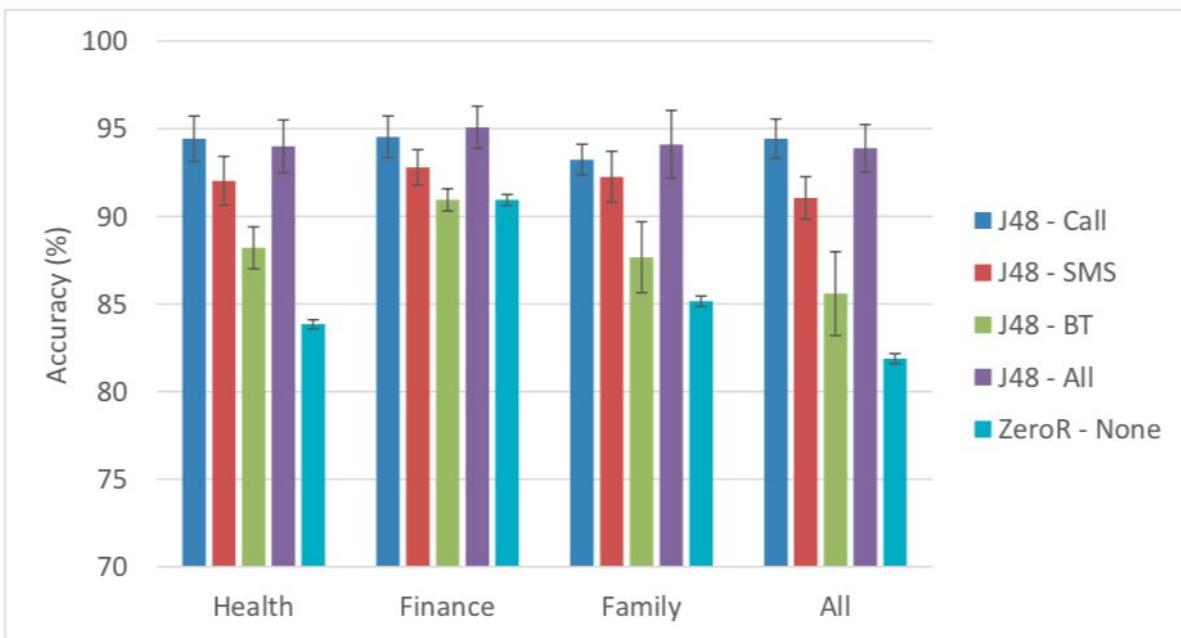


Figura: Clasificación del nivel de confianza: resultados de precisión. Las barras de error representan los intervalos de confianza del 95% (en las ejecuciones de 10 veces) (Shmueli et al. 2014).

Se requeriría mucho trabajo en terreno para obtener datos similares de interacción de personas para evaluar movilidad económica.

Se han desarrollado metodologías alternativas para evaluar e identificar cómo los individuos pasan sus horas de ocio. Una evaluación a la escala propuesta requiere que registros manuales sean realizados según cómo las personas recuerdan usaron su tiempo en el transcurso del día. Alternativamente se organizan con expertos actividades de monitoreo diario “shadowing” de las actividades del hogar. Sin embargo, ambas metodologías resultan ser costosas, requieren de mucho entrenamiento a las personas realizando la evaluación y propensas al error humano en la recolección. Por eso proponemos considerar la utilización de llamadas celulares, texto y posición GPS ya que permite cerciorarse con registros físicos como fue utilizado el tiempo y los costos asociados a la utilización de estas herramientas hacen que sean escalables abarcando un mayor número de personas.

Posibles fuentes	De libre acceso: Análisis de las interacciones registradas en redes sociales abiertas como Facebook o Twitter. Otros: CDRs, accesibles a través de convenios con operadores telefónicos o de una plataforma como OPAL.
Capacidades Requeridas	Infraestructura: En función de la modalidad de análisis y almacenamiento de datos

	<p>acordada con los controladores de datos, puede que sea necesario adquirir servidores para almacenar los datos.</p> <p>Programa para el análisis:</p> <p>En el estudio considerado se procesó y analizó la información utilizando la herramienta Weka, que es una recolección algoritmos de machine learning para tareas de minería de datos. Permite preparar y clasificar los datos hacer regresiones análisis de grupos, minería por reglas de asociación y visualización de los resultados.</p> <p>Humanas:</p> <p>Para desarrollar la evaluación se necesita un coordinador para la interacción entre la ciencia de datos y las necesidades en términos de M&E que existan en este proyecto.</p> <p>El equipo científico debe contar con capacidades técnicas en analítica, programación, bases de datos, y procesamiento de datos masivos: debe incluir un data scientist y puede incluir un experto en visualización de datos. También se recomienda incluir un experto legal que evaluará los aspectos éticos y de privacidad y un experto en comunicaciones para manejar las interacciones con el público evaluado.</p>
Riesgos y Limitaciones	<p>Existe un riesgo al recolectar esta información en un contexto político y social tenso en el país. Por un lado, los ciudadanos pueden sentir que su privacidad está siendo vulnerada al realizar este tipo de estudios. Por otro lado, existiría el riesgo de que el gobierno nacional o los gobiernos locales requieran acceso a la información recolectada.</p> <p>Para llevar a cabo el análisis, se requeriría obtener el consentimiento de los estudiantes y padres de los centros de música, y obtener datos también de las diferentes compañías con las cuales operan sus celulares. El proceso de negociación puede ser complejo, por lo que se sugiere limitar el estudio a tendencias globales y desarrollar una metodología robusta de protección a la privacidad.</p>
Próximos pasos	<ul style="list-style-type: none"> - Desarrollar una propuesta preliminar de evaluación que detalle las ventajas de utilizar CDR. - Hacer consultas en los diferentes Centros de Música para identificar comunidades que estén interesadas en participar en una evaluación piloto. - Evaluar los riesgos de seguridad y privacidad y diseñar una hoja de ruta para contrarrestarlos. - Llevar a cabo dos evaluaciones piloto utilizando las dos metodologías diferentes para probar el grado de eficacia del análisis de los CDR para este tipo de estudio.

E. ECUADOR: "Movistar Ecuador Financiamiento de Equipos" (IDB Invest-12261-01)

1. Estado y avance del proyecto

El proyecto "Movistar Ecuador Financiamiento de Equipos" (IDB Invest-12261-01) perteneciente al sector de Telecomunicaciones, Medios y Tecnología fue aprobado en el 2018 y consiste en un proyecto de IDB Invest que otorga una línea de crédito no comprometida de USD 50,000,000,000.

El objetivo principal del proyecto 12261-01 es masificar la compra y el uso de dispositivos móviles, en particular Smartphone, para habilitar un mayor acceso a la banda ancha de clientes de Movistar Ecuador, en particular personas naturales de estratos más vulnerables y pequeñas y medianas empresas (PYMES). A través de la extensión de una línea de crédito, la cancelación de las facturas podrá hacerse dentro de un periodo de 24 meses. Este programa ha sido desarrollado anteriormente en Chile y busca extenderse en otros países de América Latina, como Costa Rica, Guatemala y Panamá.

El préstamo del Banco no se otorga directamente a los clientes, sino que se hace a través de Movistar Ecuador quien entregan una cartera diversificada con cuotas más accesibles de acuerdo al promedio del ingreso. Dentro de la cartera compartida, se encuentran tanto clientes existentes de la compañía que no tenían telefonía de datos inteligentes y clientes nuevos. Por lo tanto, el Banco no realizó una demografía de poblaciones para apuntar a un grupo en particular, sino que se basó en la cartera compartida por Movistar Ecuador. Mientras que el objetivo principal consiste en masificar el acceso a tecnologías de 3G, 4G y LTE, una fase futura podría apuntar a ciertas poblaciones vulnerables.

Con respecto a fuentes de datos disponibles para la implementación del proyecto, el equipo del Banco se basa en información contenida dentro de la cartera específica al proyecto 12261-01. Esta información incluye nombres, apellidos, número de identificación fiscal, número de cuotas que se han utilizado para la compra del celular, monto por el que se compró el celular, dirección del individuo, entre otros datos.

Recomendación: Uso del comportamiento móvil en el pago del préstamo

Puede ser costoso proporcionar préstamos pequeños, especialmente a poblaciones remotas. Como consecuencia, alrededor de 2 mil millones de personas carecen de cuentas bancarias. Muchos hogares en los países en desarrollo carecen de antecedentes financieros formales, lo que dificulta que los bancos otorguen préstamos y que los prestatarios potenciales los reciban.

También puede ser difícil replicar los enfoques tradicionales del mundo en desarrollo para mejorar el acceso a la financiación a través de sucursales bancarias y oficinas de crédito.

Los enfoques tradicionales agregan información sobre la gestión histórica del crédito del individuo. Sin embargo, pocos hogares en los países en desarrollo interactúan con las instituciones formales que generan los datos necesarios. Como resultado, los prestamistas tienen muy poca información formal sobre los posibles prestatarios. Curiosamente, muchos han mantenido una rica historia de interacción con una institución formal durante un período prolongado - su actividad de teléfono móvil, registrada por su operador.

Las firmas de comportamiento en los datos del teléfono móvil pueden predecir el incumplimiento del préstamo (pago del préstamo), utilizando registros de llamadas.¹⁸

Así pues, se propone utilizar datos de empresas de telecomunicaciones para evaluar el riesgo crediticio de potenciales beneficiarios. Esto podría representar una nueva forma de préstamo para llegar a la población sin servicios bancarios y otros servicios, y evaluar su efectividad en un análisis retrospectivo. Esta metodología puede ser particularmente útil en América Latina, en donde se estima que en un país de ingreso medio solo el 34% de los adultos tienen cuentas bancarias, mientras que el 89% de los hogares poseen teléfonos móviles.

Como lo muestran Bjorkgren y Grissen (2018), es posible utilizar métodos de bajo costo para identificar inversiones rentables, utilizando información sobre prestatarios potenciales que ya están siendo recopilados por las redes de telefonía móvil. Los autores demuestran que los indicadores de comportamiento derivados de los registros de transacciones de teléfonos móviles son predictivos del pago del préstamo.

El método revela el potencial de alcanzar una precisión predictiva útil, garantizar la estabilidad en el tiempo y que la información tradicional recopilada por bancos es sólo ligeramente complementaria a los indicadores calculados.¹⁹

Los hallazgos de Bjorkgren, Grissen (2018) sugieren que los matices capturados en el uso de los teléfonos móviles pueden reducir las asimetrías de información y, por lo tanto, pueden formar la base de nuevas formas de préstamos de bajo costo. Junto con el dinero móvil, estas herramientas están habilitando un nuevo ecosistema de servicios financieros digitales. Este ecosistema está conduciendo a lo que parece ser una revolución en el acceso a la financiación en el mundo en desarrollo.

Los datos para este tipo de método incluyen metadatos de llamadas y mensajes de texto, pero no dinero móvil o información de recarga. Se espera que el rendimiento aumente con datos más ricos y muestras más grandes, observadas durante períodos de tiempo más largos.

¹⁸ Bjorkgren D, Grissen D (2018) Behavior revealed in mobile phone usage predicts loan repayment. SSRN Electronic Journal, December 2017: <https://arxiv.org/pdf/1712.05840.pdf>

¹⁹ Ibid.

	Correlation with repayment	t-stat	Number of Features
Demographics and loan characteristics			2
Age	0.073	2.35	
Female	-0.039	-1.26	
Loan term	-		
Loan size	-		
Credit Bureau			36
Has a credit bureau record	-0.022	-1.89	
Fraction of debt lost	-0.046	-3.86	
Phone usage			5,541
<i>Categories</i>	<i>High performing example feature:</i>		
Periodicity	-0.163	-5.27	796
	SMS by day, ratio of magnitudes of first fundamental frequency to all others		
Slope	0.126	4.06	44
	Slope of daily calls out		
Correlation	0.111	3.57	224
	Correlation in SMS two months ago and duration today		
Variance	-0.104	-3.34	4,005
	Difference between 80 th and 50 th quantile of SMS use on days SMS is used		
Other	0.100	3.07	542
	Number of important geographical location clusters		

Tabla: características individuales que incluyen 5,541 por uso del teléfono (Bjorkgren, Grissen 2018)

Este enfoque ya se ha utilizado desde su propuesta inicial en el 2010 por Bjorkgren & Grissen. Hoy en día, ya existen más de 68 productos de crédito digital con 11 millones de prestatarios y en Kenia, más personas obtienen préstamos a través de estas nuevas plataformas digitales que a través de la banca tradicional o las microfinanzas.

Posibles fuentes	Para acceder a datos de telefonía móvil (CDR) es necesario entablar acuerdos con los operadores privados que los generan. El BID cuenta con un acuerdo marco con Telefónica cuyo alcance podría ser estudiado para integrar este tipo de análisis. Una plataforma como OPAL también espera poner a disposición de actores del desarrollo este tipo de datos bajo una modalidad freemium.
Capacidades Requeridas	Infraestructura: Soluciones locales o en la nube para almacenamiento, procesamiento, análisis y visualización de los datos relevantes. Es necesario tomar en cuenta regulaciones locales sobre el almacenamiento y tratamiento de información sensible de usuarios. Humanas:

Riesgos y Limitaciones	<p>El proyecto implica utilizar datos personales de una gran cantidad de usuarios, con un impacto directo en la vida de los mismos (obtención o no de créditos). Por lo tanto, se sugiere tomar medidas para mitigar el riesgo de invasión a la privacidad, adoptando buenas prácticas internacionales (respeto de la legislación europea GPDR) y/o llevando a cabo auditorías externas que permitan certificar que el proyecto respeta la privacidad de las personas. También sería recomendable desarrollar un plan sólido de comunicación y participación para que los potenciales beneficiarios del crédito y las personas cuyos datos son examinados entiendan el propósito del estudio, así como la posibilidad de opt-out para las personas que no quieran participar.</p> <p>Es posible que en áreas rurales el comportamiento que esté siendo analizado de un celular genere un análisis erróneo de riesgo al considerar que se trata del comportamiento de un individuo pero que en realidad sea un bien público dentro de una célula familiar.</p> <p>Aunque se considere que facilitar el acceso a instrumentos financieros generaría crecimiento económico en el área. Sin embargo si estos instrumentos son considerados de alto riesgo y presentan altas tasas de intereses, junto con una baja tasa de alfabetización financiera de sus usuarios podría resultar vulnerando a estas comunidades con cargas usureras.</p> <p>El riesgo de abuso, si los usuarios son capaces de aprender que mediante un cierto tipo de patrones en sus comportamiento pueden acceder a una buena clasificación de riesgo. Si suficientes personas siendo consideradas para el análisis alteran su comportamiento eventualmente la herramienta pierde su relevancia.</p>
Próximos pasos	

Conclusión

Reconociendo los aportes con potencial impacto transformador de las nuevas herramientas digitales, su integración deberá ser sistemáticamente considerada como elemento en el diseño de proyectos. Idealmente su implementación también será considerada para mejorar los procesos y resultados de desarrollo esperados en sus proyectos. Dependiendo del enfoque que se les den, estas herramientas - incluso combinadas con técnicas tradicionales y otras no convencionales - pueden ser aplicadas a prácticamente todos los sectores estratégicos. Muchos de los instrumentos innovadores resultan ser costo-efectivos y aportan beneficios a largo plazo que van más allá de un proyecto en específico. Al ser estas implementadas se deben considerar instalarlas dentro de estructuras sostenibles apoyadas por las unidades ejecutoras.

No obstante, lo anterior, para una introducción exitosa de una o algunas herramientas tecnológicas en particular se deben tener en cuenta los siguientes elementos:

- En una primera fase, para una aplicación realista y efectiva, las herramientas digitales deben estar adaptadas al ecosistema local. Por ejemplo, el centro de control de la ciudad de Pasto del proyecto CO-L1091 resultó ser una inversión importante que no consideró todas las características de esta ciudad, empezando por su tamaño mediano. Se estableció un centro de control moderno con sistemas comparables a los utilizados en ciudades como París y Nueva York, pero inoperable por los recursos humanos locales.
- Dependiendo del contexto, estas herramientas adaptadas a la realidad local podrán paulatinamente generar cambios positivos que vayan de la mano de los marcos regulatorios o idealmente promuevan su modernización. La política de reforma energética boliviana apoyada por el BID bajo el proyecto BO-L1189 generó un cambio institucional necesario para una debida implementación de proyectos de inversión que se alinean a los nuevos objetivos del gobierno boliviano y se apoyan en los nuevos marcos regulatorios, tal como el proyecto analizado BO-L1190.

Aspectos específicos de consideración y análisis incluyen:

- Estructuras institucionales y de gobierno, por ejemplo, en cuanto a si son centralizadas o descentralizadas.
- Marco regulatorio moderno e integral que considere por ejemplo aspectos relacionados a la seguridad en la privacidad de datos.
- Nivel de maduración tecnológico en cuanto a la existencia de infraestructura (hardware y software) que soporte la introducción y operación de nuevas herramientas digitales, así como de un capital humano capacitado y estable para operar estas herramientas.
- Otras implicaciones dentro de las estructuras organizativas, por ejemplo, en cuanto al perfil de funcionarios públicos reacios o interesados en procesos de modernización

disruptivos.

Por un lado, el BID tiene la capacidad de movilizar expertos para apoyar una debida introducción de herramientas tecnológicas en proyectos de inversión que tomen en cuenta los puntos de consideración mencionados. Por otro lado, el BID puede jugar un papel adicional importante en establecer acuerdos con entidades tanto públicas como privadas que recolectan y procesan datos, con el fin de facilitar el acceso de estos datos para su utilización en proyectos de desarrollo (ej. Establecer un acuerdo marco para acceder a CDRs con Telefónica Movistar a nivel regional).

Recomendaciones

Anexo: Caja de herramientas

Caja de herramientas - Data-pop

1. Recolectar: reunir datos de personas o dispositivos

Herramientas para recolectar datos de la gente:

- Crowdsourcing: uReport, OpenStreetMap, Ushahidi
- Herramientas de encuestas: Kobo Toolbox

Herramientas para recolectar datos de dispositivos:

- Sensores en smartphones: Bandicoot, Funf library
- Sensores Remotos: Google Earth

2. Almacenar: guardar datos para un uso futuro

Herramientas para almacenar volúmenes grandes de datos / información:

- Base de datos: Hadoop, MySQL, MongoDB
- Soluciones de almacenamiento en la nube: Amazon S3, Dropbox, Google Drive, OneDrive
- Sistemas de control de versión: Git, SVN

Herramientas para almacenar datos sensibles de forma segura:

- Almacenamiento de datos seguro: OpenPDS
- Almacenamiento descentralizado: blockchain (e.g., Bitcoin)

3. Procesar: convertir datos brutos en datos intermediarios

Herramientas para que las máquinas ejecuten instrucciones:

- Computación en la nube: Amazon Web Services, Microsoft Azure, Google Cloud
- Computación descentralizada: MIT Enigma
- Automatización por lotes: IFTTT

Herramientas para que las personas manipulen datos brutos:

- Data wrangling tools: OpenRefine, DataBasic.io

4. Analizar: pasar de los datos a la comprensión

Herramientas para programación y análisis estadísticos:

- Lenguajes de programación: MatLab, SPSS, Stata, Python, R, Julia
- Librerías open source: Scipy, Numpy, Pandas, MatPlotLib, Pylr
- Software y utilidades: Rstudio, Jupyter, Anaconda scientific stack

Herramientas para aplicar modelos y metodologías:

- Herramientas de machine learning: Scikit-learn, Spark, Accord.NET
- Procesamiento de lenguaje natural: Apache OpenNLP, Stanford CoreNLP, LingPipe
- Sistemas de información geográfica: ArcGIS, QGIS, Google Earth Engine
- Inferencia causal: Tetrad
- Grado de propensión: Twang R package

5. Compartir / acceder: hacer que otros puedan acceder a los datos almacenados

Herramientas para compartir / acceder a datos y códigos:

- Plataformas para albergar códigos: GitHub, BitBucket, CDNJS
- Plataformas con agrupamientos de datos: United Nations' HDX
- Interfaces de programación de aplicaciones: Twitter Streaming API, Bit.ly Social Data API

Herramientas para compartir / acceder a datos sensibles:

- Arquitectura de privacidad: OpenPDS
- Estructuras de gobernanza: OPAL project

6. Transmitir / comunicar: compartir información almacenada o recibir información compartida por otros

Herramientas para transferir datos:

- Formatos estándar: csv, json, geojson, xml, shp, xls
- Herramientas de cifrado: AES, RSA, OpenPGP

Herramientas para comunicar información:

- Tecnología de información y comunicación: World Wide Web
- Herramientas de visualización: D3.js, R Shiny, Tableau

Bibliografía

IBM, "10 Key Marketing Trends for 2017 and Ideas for Exceeding Customer Expectations", 2017: <https://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=WRL12345USEN>

McDonald, Sean. "Ebola: A Big Data Disaster – Privacy, Property, and the Law of Disaster Experimentation", The Centre for Internet and Society, Jan. 2016: <https://cis-india.org/papers/ebola-a-big-data-disaster>

Andrews Matt, Pritchett Lant and Woolcock Michael, "Escaping Capability Traps through Problem-Driven Iterative Adaptation (PDIA)", Center for Global Development, Working Paper 299, June 22, 2012: <https://www.cgdev.org/publication/escaping-capability-traps-through-problem-driven-iterative-adaptation-pdia-working-paper>

UN Global Pulse, "Integrating Big Data into the Monitoring and Evaluation of Development Programmes", 2016: http://unglobalpulse.org/sites/default/files/IntegratingBigData_intoMEDP_web_UNGP.pdf

Pritchett, L., Samji, S., and Hammer, J., "It's All About MeE: Using Structured Experiential Learning ("e") to Crawl the Design Space", Center for Global Development, Working Paper 322, 2013
<https://www.cgdev.org/publication/its-all-about-mee-using-structured-experiential-learning-e-crawl-design-space>

Bamberger, M., Tarsilla, M., Hesse-Biber, S., "Why so many "rigorous" evaluations fail to identify unintended consequences of development programs: How mixed methods can contribute", Evaluation and Program Planning, Vol. 55, April 2016, pages 155-162: <https://www.ncbi.nlm.nih.gov/pubmed/26874234>

Steele JE et al., "Mapping poverty using mobile phone and satellite data", Journal of the Royal Society Interface, Vol. 14, Issue 127, February 2017: <https://royalsocietypublishing.org/doi/10.1098/rsif.2016.0690>

Noor, AM., Alegana, VA., Gething, PW., Tatem, AJ., Snow, RW., "Using remotely sensed night-time light as a proxy for poverty in Africa", Population Health Metrics, 2008: <https://pophealthmetrics.biomedcentral.com/articles/10.1186/1478-7954-6-5>

Ghosh, T., Anderson, SJ., Elvidge, CD., Sutton, PC., "Using nighttime satellite imagery as a proxy measure of human well-being", Sustainability, 2013, Vol 5. Issue 12: <https://www.mdpi.com/2071-1050/5/12/4988>

Leng, Y., Rudolph, L., Pentland, A., Zhao, J., Koutsopoulos, H., "Managing travel demand: Location recommendation for system efficiency based on mobile phone data", Proceedings of Data for Good Exchange (D4GX), 2016: [arXiv:1610.06825](https://arxiv.org/abs/1610.06825)

Nishikata, T., Hardjono, T., & Pentland, S., "Social Capital Accounting", MIT Media Lab, October 2018: <https://www.media.mit.edu/publications/social-capital-accounting/>

Shmueli, E., Singh, V. K., Lepri, B., & Pentland, A., "Sensing, understanding, and shaping social behavior", IEEE Transactions on Computational Social Systems, April 2014, Vol. 1, Issue 1, pages 22–34. <http://doi.org/10.1109/TCSS.2014.2307438>

Fogg, B.J., "Persuasive technology: Using computers to change what we think and do", Morgan Kaufmann, 2003

Bjorkgren, D., Grissen, D., "Behavior revealed in mobile phone usage predicts loan repayment", SSRN Electronic Journal, December 2017: <https://arxiv.org/pdf/1712.05840.pdf>