

Materia: Taller de programación

Alumnos: Jimena Terán, Juan Lynch, Renzo Falciglia

Fecha: 23/12/2024

Propuesta Final: "Ecosistemas Productivos y su Impacto en Mercados Laborales Locales"

Introducción

Esta propuesta de investigación aborda la siguiente pregunta: ¿cómo la presencia de ecosistemas productivos afecta la probabilidad de éxito de nuevos proyectos industriales? El análisis se enfoca específicamente en evaluar si los proyectos tienen mayor probabilidad de completarse cuando se ubican en áreas con alta concentración de industrias del mismo sector o relacionadas.

La relevancia de esta pregunta surge de dos aspectos fundamentales. Primero, en un contexto de recursos limitados, queremos identificar los determinantes que maximizan la probabilidad de éxito de las inversiones industriales. Segundo, la literatura sugiere que la concentración geográfica de industrias relacionadas genera externalidades positivas a través de mercados laborales especializados, proveedores compartidos y derrames de conocimiento, pero existe evidencia limitada para países en desarrollo.

Como análisis complementario, examinaremos el impacto de estos ecosistemas productivos en los mercados laborales locales. Específicamente, investigaremos cómo la consolidación de clusters industriales especializados afecta variables como niveles salariales y calidad del empleo formal en los aglomerados urbanos cercanos,

poniendo particular atención a efectos heterogéneos por nivel educativo y experiencia laboral.

Revisión de la literatura

La literatura sobre clusters industriales y aglomeración económica tiene sus bases modernas en Porter (1998), que argumenta que las empresas obtienen ventajas competitivas al ubicarse cerca de otras firmas e instituciones del mismo sector o sectores relacionados. Porter enfatiza que estas ventajas surgen no solo de la proximidad física, sino de las complejas redes de interacción entre empresas, proveedores especializados, instituciones de investigación y organismos gubernamentales.

Greenstone et al. (2010) demuestran que la llegada de grandes plantas manufactureras genera incrementos significativos en la productividad de las empresas locales existentes, un efecto que atribuyen tanto a vínculos en la cadena de suministro como a derrames tecnológicos. Ellison et al. (2010) cuantifican la importancia relativa de diferentes fuerzas de aglomeración, encontrando que los vínculos insumo-producto y los mercados laborales compartidos explican gran parte de los patrones de co-localización industrial.

Incorporando técnicas modernas de machine learning, Delgado, Porter y Stern (2016) desarrollan una metodología para identificar clusters industriales y sus interrelaciones, utilizando análisis de redes y clustering que superan las limitaciones de las clasificaciones industriales tradicionales. Los autores construyen medidas de similitud entre industrias basadas en la co-localización, flujos de empleo y

patrones de ocupación compartidos, lo que les permite identificar grupos de industrias que tienden a beneficiarse de ubicarse cerca unas de otras.

Datos

Los datos que podrían utilizarse para dar respuesta a la pregunta de investigación propuesta son los de la base de datos "*Industrial Info Resources (IIR)*", para el período 2015-2024. Dicha base contiene información detallada de más de 3.000 proyectos industriales de gran escala (superiores a USD 1 millón). Se suele utilizar en industrias como la energía, minería, manufactura, etc. La información incluye proyectos de todas partes del mundo, ya que la IIR se especializa en rastrear y recopilar la información de los proyectos industriales a lo largo del mundo.

Esta base de datos incluye información exhaustiva sobre la ubicación exacta (coordenadas, provincia, ciudad) de los proyectos, el sector industrial al que pertenecen y el subsector (minería, energía, manufactura, etc.), el tipo de proyecto (greenfield, expansión, mantenimiento), los montos de la inversión en USD, el estado del proyecto, las fechas de inicio y finalización planificada, la probabilidad de ejecución, el empleo estimado (construcción y operación), la capacidad productiva, la tecnología utilizada, y una amplia base de contactos empresariales y detalles técnicos de cada proyecto.

Toda esta información se presenta de manera estructurada en diferentes tablas de manera relacional. Es decir, las tablas se encuentran relacionadas entre sí por diferentes variables (keys), lo que permite hacer diferentes análisis con la

información de distintas tablas. A su vez, puede consultarse la información en distintos formatos como csv, excel e incluso puede consultarse a través de APIs.

Cabe destacar que la base de datos IIR es de carácter privado y que será necesario verificar los términos de uso para la investigación académica o identificar fuentes alternativas de datos similares.

Por otro lado, para un análisis sobre el impacto de los proyectos industriales en las condiciones del mercado laboral de las zonas donde se ubican, se podrían utilizar datos de la EPH para los 32 principales aglomerados urbanos. Esto permitiría examinar cómo la consolidación de clusters industriales especializados afecta los niveles salariales o la calidad del empleo formal en los aglomerados cercanos, controlando por características individuales, efectos fijos por tiempo y región, y tendencias específicas por aglomerado. Se podría hacer especial énfasis en los efectos heterogéneos por nivel educativo y experiencia laboral.

Finalmente, también podría utilizarse la API de Mercado Libre trabajada durante el curso, para evaluar el volumen de los envíos por ciudad, como un *proxy* del poder adquisitivo local. Esto sería útil para validar “efectos derrame” de los proyectos industriales en el consumo local. Consideramos que estas tres fuentes de información serían suficientes para poder investigar sobre la pregunta de investigación propuesta más arriba e incluso sobre el análisis secundario que se planteó.

Metodología

En cuanto a la metodología a utilizar, una de las técnicas principales sería el análisis de clusters, para identificar ecosistemas productivos naturales, combinado con el método de *k-nearest neighbors* (*k-NN*) para examinar patrones de similitud y complementariedad entre proyectos. Esta técnica se refiere a una herramienta estadística de aprendizaje no supervisado que permite agrupar observaciones en subconjuntos homogéneos denominados *clusters* o grupos.

Esta técnica es particularmente útil para identificar ecosistemas productivos, ya que los *clusters* resultantes pueden representar regiones o sectores en los que los proyectos comparten características similares, como por ejemplo su ubicación, el tipo de industria, los recursos utilizados, las capacidades tecnológicas, entre otros.

Para esta propuesta de investigación particular, el método será utilizado para analizar patrones de similitud y complementariedad entre los proyectos, evaluando cómo las características de un proyecto se relacionan con las de sus vecinos más cercanos en términos de ubicación, industria o recursos.

El algoritmo de *k-NN* asigna una clase o un valor a una observación en función de las etiquetas de sus *k* vecinos más cercanos, lo que lo convierte en una herramienta flexible tanto para tareas de clasificación como de regresión.

Para validar los resultados, se podría implementar una estrategia de *cross-validation*, comparando el poder predictivo del modelo con predictores basados únicamente en factores tradicionales de localización.

Por último, para el análisis secundario sobre el impacto de los *clusters* industriales en las condiciones laborales de la población que vive en la misma zona, se podría utilizar la técnica de regresiones múltiples, vinculando los datos de los proyectos industriales con la información de la EPH y de la API de Mercado Libre. La utilización de estas dos fuentes secundarias de información permitiría controlar los resultados por las características individuales de los individuos y aislarlas de los efectos directos de los *clusters* industriales. Se podrían implementar también modelos como los de Diferencia en Diferencias (DiD), para evaluar los cambios en las condiciones socioeconómicas de la zona antes y después del establecimiento de un *cluster*.

Los resultados a los que se podría llegar con esta propuesta serían particularmente relevantes a la hora de evaluar las distintas políticas públicas orientadas al sector industrial. Más específicamente, podría servir para destinar de una manera más eficiente los recursos escasos con los que cuenta el país para llevar adelante su política industrial.

Podrían destinarse recursos a incentivar el establecimiento de clusters que permitan por un lado incrementar las posibilidades de éxito de los proyectos industriales y, por otro lado, permitiría mejorar las condiciones socioeconómicas de los lugares donde se radiquen los proyectos.

Resultados esperados

Se espera que los resultados de esta investigación nos permitan entender cómo los ecosistemas productivos influyen en el éxito de nuevos proyectos industriales, expliquen si los proyectos tienen mayor probabilidad de completarse según su

ubicación y cuál es el impacto de los mismos en el mercado laboral y el consumo local.

Se espera principalmente confirmar si los proyectos industriales que se ubican en áreas con alta concentración de industrias relacionadas tienen una mayor probabilidad de completarse exitosamente, o en su defecto, cuáles son los patrones de aglomeración específicos que maximizan las probabilidades de éxito.

Respecto al mercado laboral, esperamos encontrar que el hecho de generar aglomerados de proyectos industriales mejora significativamente los niveles salariales y la calidad del empleo formal de la zona.

En relación al consumo local, se espera validar si los proyectos industriales exitosos también tienen un impacto positivo en el consumo local, utilizando como proxy el volumen de envíos registrados en la API de Mercado Libre.

Finalmente, respondiendo a cada uno de estos interrogantes, se espera poder elaborar evidencia útil para formular políticas públicas que permitan generar estrategias industriales más eficientes.

Limitaciones

A pesar de que se plantea una metodología sólida, existen varias limitaciones que podrían afectar los resultados esperados:

En primer lugar, tenemos la limitante de la disponibilidad de datos. Como se mencionó previamente, la base de datos IIR es de acceso privado, lo que podría limitar el alcance del análisis si no se logra obtener acceso total a la información

necesaria. Alternativamente, podría ser necesario recurrir a bases de datos públicas o menos detalladas, lo que podría afectar la calidad y profundidad del análisis.

Si bien tenemos pensado acudir a la EPH, esta podría no capturar completamente los efectos de los clusters industriales, especialmente en zonas no urbanas o menos representadas en la muestra. Este sesgo podría subestimar o sobrestimar los impactos observados.

Todo esto podría llevar a que los resultados no se puedan generalizar a todas las regiones, ya que quizá no se pueda capturar completamente los efectos de todos los aglomerados industriales.

Por otro lado, aunque se planea utilizar el modelo de Diferencias en Diferencias (DiD), la identificación de relaciones causales puede verse limitada por factores no observados que también influyen en los resultados, como decisiones políticas locales, variaciones climáticas o inversiones públicas previas que no están adecuadamente controladas.

Con relación al método de análisis general, el éxito del mismo dependerá de la calidad y especificidad de las variables disponibles. Variables faltantes o mal definidas podrían reducir la precisión del modelo, afectando la interpretación de los resultados.

Con respecto al análisis de impacto de estos aglomerados en el consumo local, utilizar el volumen de envíos de la API de Mercado Libre como un *proxy* del consumo local podría no captar completamente otros factores que influyen en el poder

adquisitivo y patrones de consumo en las zonas analizadas. Este *proxy* podría ser más representativo en zonas urbanas que en rurales.

Finalmente, eventos externos como crisis económicas, epidemias como la que ya hemos tenido en 2020 o conflictos internacionales podrían distorsionar los resultados y dificultar la identificación de patrones consistentes en los datos.

Bibliografía

Porter, Michael E. "[Los clusters y la nueva economía de la competencia](#)". Harvard Business Review 76, núm. 6 (noviembre-diciembre de 1998): 77-90.

Greenstone, M., Hornbeck, R., & Moretti, E. (2010). Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings. Journal of Political Economy, 118(3), 536–598. <https://doi.org/10.1086/653714>

Ellison, Glenn, Edward L. Glaeser, and William R. Kerr. 2010. "What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns." American Economic Review, 100 (3): 1195–1213. DOI: 10.1257/aer.100.3.1195

Mercedes Delgado & Michael E. Porter & Scott Stern, 2016. "[Defining clusters of related industries](#)." Journal of Economic Geography, vol 16(1), pages 1-38.