

Universidad Externado de Colombia

Facultad de Economía

Curso de Inteligencia Artificial con Aplicaciones en Economía I

La paradoja de Chapinero. Educación alta, empleo informal

David Leonardo Martínez Pinzón

2 de octubre de 2025

Resumen

Chapinero es el corazón universitario de Bogotá. Aquí se concentra buena parte del capital educativo de la ciudad, miles de estudiantes, decenas de campus y la juventud movilizanre, la voz de quienes con la esperanza de acobjirase de la formalidad invierten millones cada semestre para titular sus habilidades. Bajo la promesa de la profesionalización, se espera que esa formación se traduzca en trabajos estables, seguros y formales. Sin embargo, los datos muestran que esa detras de esas “palabras bonitas” se encuentran tendencias que revelan que el problema excede el marco legal y se encarece a apartir de ramas estrucutrales del mercado. La formalidad laboral no llega como debería y la brecha entre educación e inserción se convierte en una paradoja alarmante. Este proyecto combina dos fuentes que permiten mirar la problemática desde ambos lados del mercado laboral. La estrategia metodológica utiliza modelos supervisados y no supervisados. Regresión logística y Random Forest para analizar la oferta juvenil. K-Means y árboles de decisión para segmentar y explicar las prácticas empresariales. La hipótesis se impone sobre el marco del discurso político de la informalidad, si en Chapinero, con toda su concentración universitaria, la formalidad no está garantizada, el problema no es educativo. El problema es estructural.

Palabras clave: Informalidad Laboral, Juventud profesionalizada, Economía Popular, Mercado laboral, Chapinero.

1. Introducción

Chapinero es el corazón académico de Bogotá. Las universidades llenan sus calles, los estudiantes marcan el ritmo de su vida diaria. La narrativa con la que muchos de ellos invierten

en estas instituciones siempre ha sido que la educación es la llave de la formalidad y la movilidad social, esta promesa es justificante de los aclamados programas de becas universitarias de la capital colombiana. Sin embargo, en la práctica esa ilusión se rompe. Incluso los jóvenes con títulos se enfrentan a contratos precarios, a trabajos por fuera de la seguridad social y a trayectorias laborales que no corresponden con el esfuerzo educativo. La hipótesis que guía este trabajo es clara. Si en Chapinero, donde abundan las universidades y los jóvenes educados, la informalidad sigue siendo alta, entonces no estamos frente a un problema de falta de educación. Estamos frente a un problema estructural del mercado laboral.

Este estudio busca ponerlo en evidencia con datos y con modelos. Desde el lado de la oferta, mostramos con la GEIH cómo la educación ayuda pero no alcanza para blindar a los jóvenes contra la informalidad. Desde el lado de la demanda, usamos la EDL para mostrar cómo las propias empresas de Chapinero piden educación universitaria y al mismo tiempo ofrecen contratos que perpetúan la precariedad.

2. Pregunta de investigación.

La pregunta central es directa. ¿Qué factores explican que un joven profesional en Bogotá caiga en la informalidad laboral en Chapinero?

3. Objetivos

1. Estimar con la GEIH 2024 la probabilidad de informalidad juvenil en Bogotá y demostrar el peso real de la educación frente a otros determinantes.

2. Segmentar con la EDL 2020–2021 a las empresas de Chapinero según lo que piden y lo que ofrecen en contratación.

3. Unir ambas miradas para construir el marco oferta–demanda que demuestre que la raíz del problema es estructural.

3. Datos y alcance

Trabajo con dos fuentes públicas. La primera es la GEIH del DANE, corte diciembre 2024. Se hará uso de los módulos de características generales, fuerza de trabajo y hogar, previamente documentadas en el repositorio. Limitamos la muestra a Bogotá, jóvenes de 18 a 28 años y en condición de ocupados, connotación impuesta por la naturaleza de los datos. Con esas variables se construye la medida de informalidad y los predictores de educación, sexo, edad, horas trabajadas, estrato y sector económico. La segunda es la Encuesta de Demanda Laboral 2020–2021, levantada por la Secretaría de Desarrollo Económico y el Observatorio de Desarrollo Económico de Bogotá. A través de la base es posible materializar la delimitación en Chapineo, con datos de empresas que permiten ver sector, tamaño, vacantes, nivel educativo solicitado y modalidad de contratación.

Lejos de ser un problema, la diferencia temporal entre las dos bases son una prueba de persistencia. Si en 2021 las empresas de Chapinero ya mostraban precarización en sus prácticas, y en 2024 los jóvenes siguen cayendo en la informalidad a pesar de sus títulos, es posible determinar una confirmación de que la informalidad laboral en poblaciones jóvenes con títulos universitarios no se trata de un fenómeno coyuntural sino de un problema estructural.

4. Definiciones y variables

En la GEIH, la variable objetivo es la informalidad. Se define como un joven que no cotiza ni a salud ni a pensión. Pretendese computar las variables de cotización P5090 correspondiente a Salud y P5100 que responde a Pensión. Esa es la medida más clara y reconocida para identificar informalidad en los microdatos del DANE. Los predictores son directos. Edad, sexo, nivel educativo alcanzado, horas trabajadas, sector económico y estrato del hogar. Opcionalmente también se puede incorporar ingreso y posición ocupacional si se

cuenta con el módulo de ocupados, pero dadas las restricciones operacionales y temporales se opta por mantenerse en esa ambiciosa línea.

En la EDL, las variables clave son sector económico, tamaño de empresa, existencia de vacantes, nivel educativo que exigen y la modalidad de contratación. A partir de estas variables se construyen dos targets derivados de empresas que formalizan frente a las que no, y empresas que exigen educación universitaria frente a las que no lo hacen. Esto con el objetivo de demostrar que la paradoja de la informalidad en Chapinero no solo es de oferta, sino de demanda también.

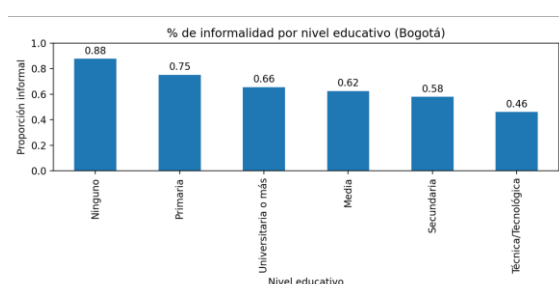
5. Metodología

La metodología de este trabajo se desarrolló en dos grandes fases, cada una asociada a una fuente de datos distinta. La primera fase se centra en la oferta laboral juvenil, utilizando la GEIH del DANE para diciembre de 2024. La segunda fase se concentra en la demanda empresarial, con la Encuesta de Demanda Laboral (EDL) 2020–2021 aplicada en Bogotá y con representatividad para Chapinero.

En la primera fase, se construyó una base de análisis a partir de la unión de los módulos de características generales, fuerza de trabajo y hogar de la GEIH. El procedimiento fue el siguiente: i) importación de archivos con estandarización de nombres de variables, ii) unión de los módulos a través de las llaves DIRECTORIO, ECUENCIA_P y ORDEN, iii) filtrado de la muestra para conservar únicamente a jóvenes de 18 a 28 años residentes en Bogotá (DPTO=11) y en condición de ocupados, iv) definición de la variable objetivo “informal” como aquella persona que no cotiza a salud ni a pensión, y v) selección de predictores relacionados con educación, edad, sexo, horas trabajadas, estratosocioeconómico y sector económico. Una vez curada la base, se aplicaron dos modelos: una regresión logística, que permite interpretar los efectos marginales de la educación y otras variables, y un Random Forest, que aporta flexibilidad y la posibilidad de evaluar la importancia de cada factor mediante valores SHAP.

En la segunda fase, se trabajó con la EDL, filtrando únicamente las empresas localizadas en Chapinero (código 2). Se construyeron indicadores de formalización y de exigencia educativa, y sobre esta base se aplicó un análisis de clustering con K-Means para identificar perfiles de empresas, complementado con un árbol de decisión que permite derivar reglas claras sobre las condiciones de contratación. En conjunto, la metodología integra técnicas supervisadas y no supervisadas, asegurando tanto interpretabilidad como capacidad de descubrimiento. Aunque las dos bases provienen de años diferentes, se consideran complementarias, pues permiten demostrar la persistencia de la paradoja aallizada.

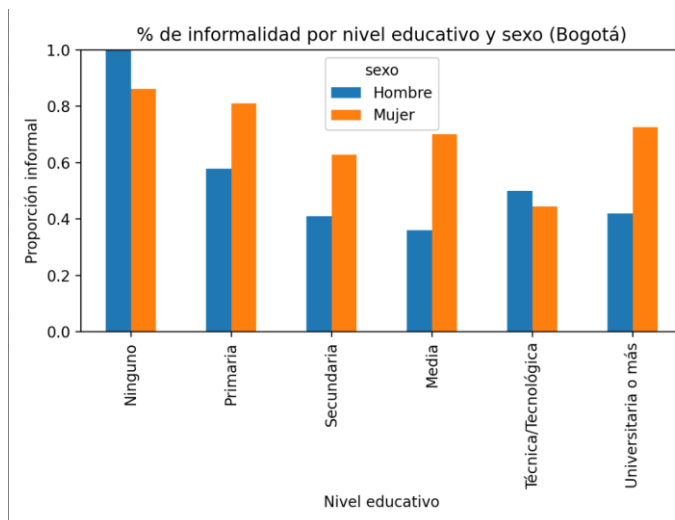
6. Resultados y Discusión



1. % de informalidad por nivel educativo (Bogotá)

El primer gráfico muestra la proporción de jóvenes informales según el nivel educativo alcanzado. Lo que se ve es contundente: la informalidad disminuye a medida que sube la educación, pero nunca llega a cero. De hecho, incluso en “universitaria o más” el indicador está alrededor del 66%. Lo más llamativo es que los niveles técnicos y tecnológicos tienen menor informalidad que la educación universitaria, cerca del 46%. Esto confirma un desajuste claro entre lo que produce el sistema educativo y lo que realmente demanda el mercado laboral.

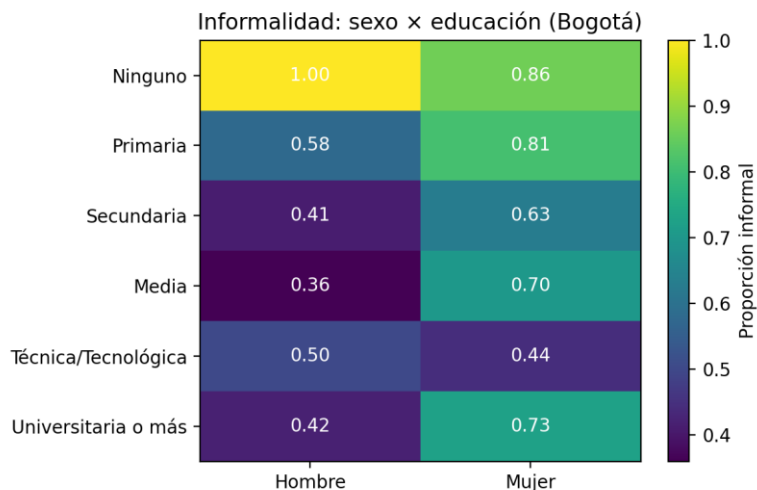
La lectura aquí es doble: la educación sí importa, pero no garantiza la formalización. Y además, en algunos casos, los oficios de inserción rápida (técnicos) protegen más que un título universitario.



2. % de informalidad por nivel educativo y sexo

El segundo gráfico, que cruza educación y sexo, abre un debate todavía más fuerte. En todos los niveles las mujeres muestran más informalidad que los hombres. Por ejemplo, en el nivel de media, la diferencia es brutal: los hombres tienen alrededor del 36% mientras que las mujeres llegan casi al 70%. El dato es incómodo, pero hay que decirlo: la brecha de género no desaparece con más educación, incluso se amplía en algunos tramos.

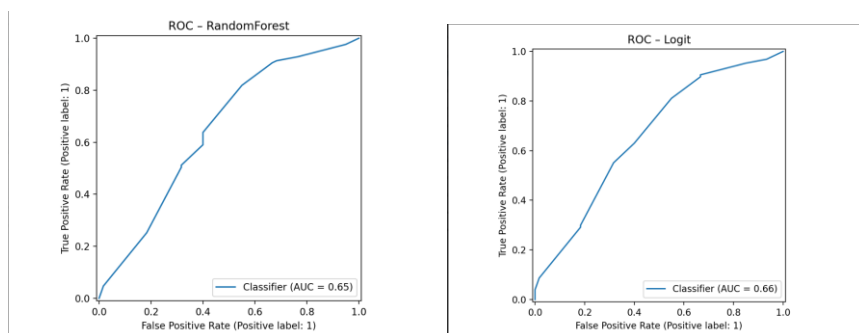
Esto implica que la política pública no puede limitarse a aumentar cobertura educativa. Hay un problema de segmentación laboral y de discriminación que sigue operando aunque los títulos se consigan.



3. Heatmap sexo × educación

El mapa de calor resume lo anterior con claridad visual. A simple vista, los cuadros más oscuros están en las mujeres, y los más claros en los hombres, salvo pequeñas excepciones. Los extremos son reveladores: los hombres sin educación están en informalidad total (100%), mientras que las mujeres universitarias siguen con un 73%. El mensaje que deja este gráfico es casi doloroso: ni el máximo nivel educativo es capaz de blindar a las mujeres de la informalidad en Bogotá.

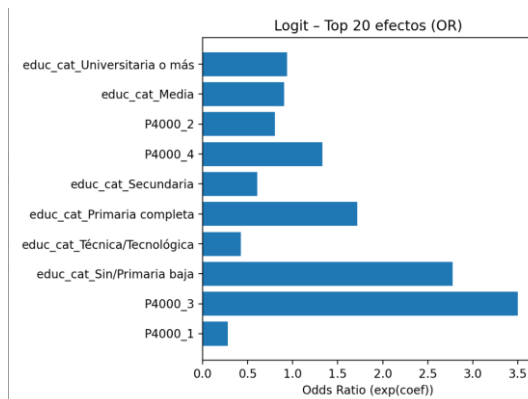
El resultado refuerza la idea de que la hipótesis central se cumple: no es la educación lo que explica el fenómeno, sino la estructura del mercado laboral.



4. Curvas ROC: Logit y Random Forest

Los modelos predictivos confirman lo anterior. Tanto la regresión logística como el Random Forest tienen un AUC cercano a 0.65–0.66. Esto significa que sí logran separar los grupos (formales e informales), pero no con una precisión alta. En términos simples: hay patrones claros, pero también factores invisibles que los modelos no capturan porque no están en la encuesta.

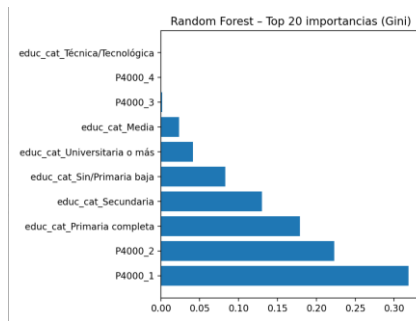
El logit permite interpretar coeficientes de manera sencilla, mientras que el RF aporta robustez a interacciones más complejas. Ambos coinciden en que la educación y el sexo tienen peso, pero no son suficientes para explicar todo.



5. Odds Ratios (Logit)

En la regresión logística los efectos se ven reflejados en los Odds Ratios. Queda claro que estar en primaria completa, secundaria o incluso universitaria cambia las probabilidades, pero ninguno de esos niveles garantiza la formalidad. Lo interesante es que la educación técnica muestra un efecto más protector, lo que se conecta con los descriptivos iniciales.

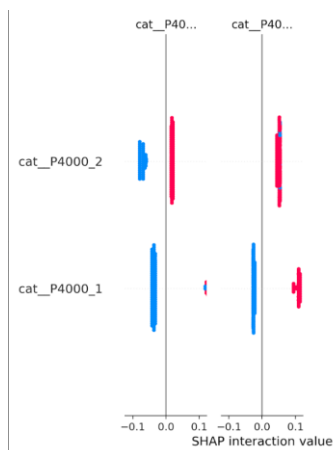
El error sería pensar que solo con más universidad se resuelve el problema. Los ORs demuestran que los títulos cuentan, pero cuentan menos de lo que se cree.



6. Importancias Gini (Random Forest)

El gráfico de importancias del Random Forest reafirma el papel central de las variables de sexo y educación. Sexo aparece como uno de los predictores más fuertes, seguido de primaria, secundaria y universitaria. La educación técnica queda más abajo, pero con un rol diferenciado.

La clave es que el modelo, aunque no perfecto, está mostrando la misma paradoja que vimos en los descriptivos: la educación universitaria no es garantía, y el sexo (ser mujer) es un factor de riesgo en sí mismo.



7. SHAP Summary

Finalmente, el análisis de SHAP permite ver la dirección y magnitud de los efectos. De nuevo, el sexo divide de forma evidente las predicciones de informalidad, y las categorías educativas se dispersan en diferentes direcciones. Lo valioso del SHAP es que pone en evidencia que los resultados del RF no son un “caja negra”, sino que se pueden interpretar en

la misma línea que los ORs. Los títulos importan, pero no blindan. Lo estructural pesa más que lo individual.

La hipótesis de este proyecto se confirma: en Chapinero y en Bogotá, incluso en los contextos de mayor concentración universitaria, la informalidad persiste. No es un problema de falta de educación. Es un problema estructural del mercado laboral colombiano, que combina discriminación de género, desajuste entre formación y demanda, y una economía que sigue precarizando a los jóvenes.

La promesa de la profesionalización, esa idea de que “si estudias tendrás trabajo formal”, no se cumple en los datos. Los resultados son una alerta: necesitamos políticas que ataquen directamente las prácticas de contratación y las condiciones de empleo, no solamente la oferta educativa.

Bibliografías:

Rincón-Báez, W. U. y Solear-Hurtado, A. J. (2015). Perspectiva socioeconómica de los vendedores informales de Chapinero, en Bogotá, Colombia. *Cooperativismo y Desarrollo*, 23(107), xx-xx. doi: <http://dx.doi.org/10.16925/co.v23i107.1255>