



# IDENTIFICACIÓN DE PATRONES DE EVASIÓN TRIBUTARIA EN COLOMBIA MEDIANTE TÉCNICAS DE APRENDIZAJE NO SUPERVISADO



# INTEGRANTES



- Edgardo Sánchez
- Juan Medina
- Manuel Alarcón

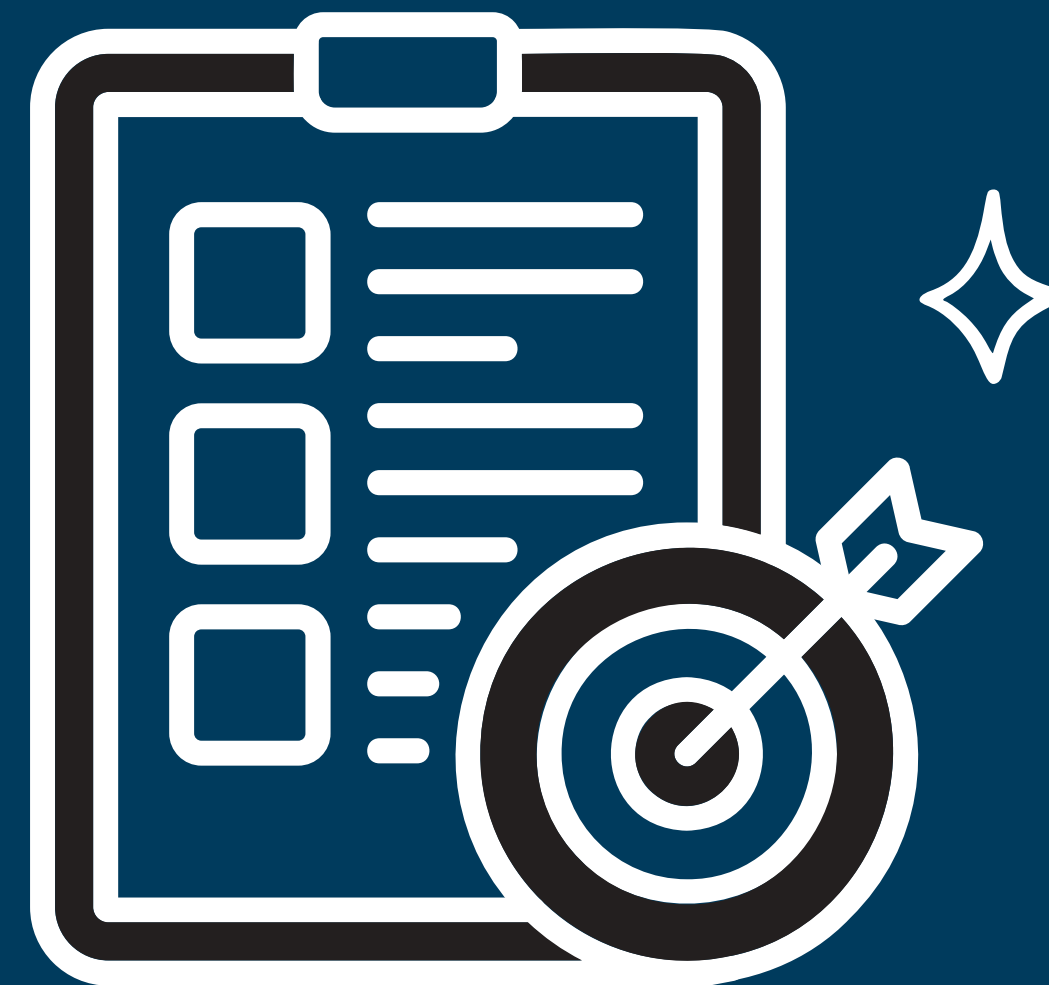


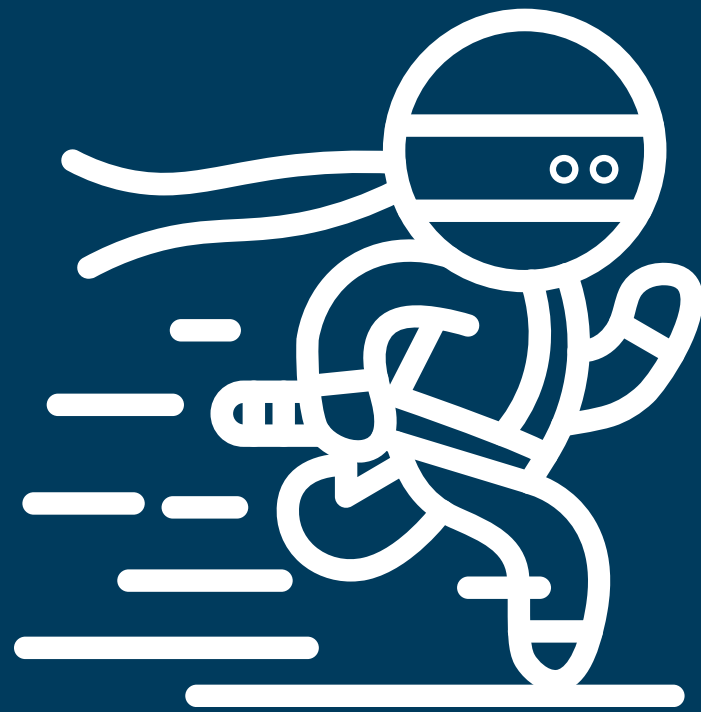


# OBJETIVO

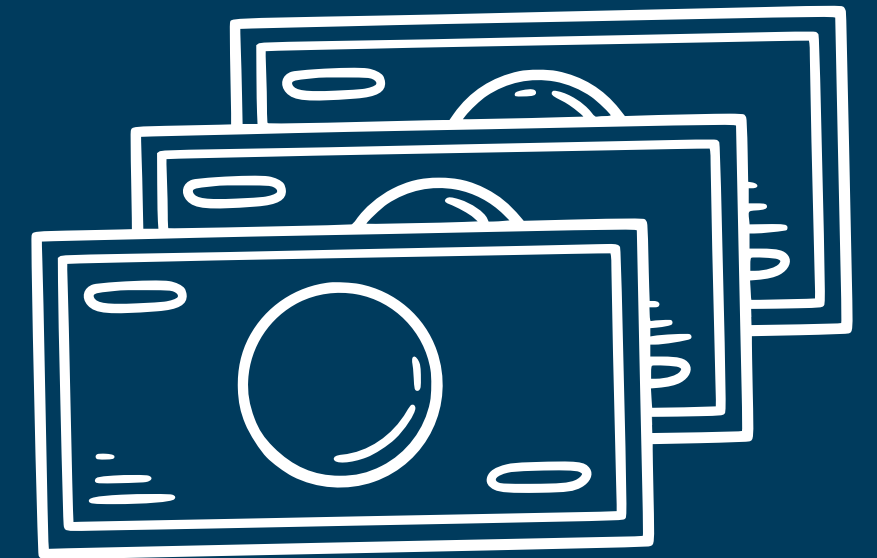


Analizar datos abiertos de fuentes oficiales y académicas para identificar clústeres y patrones de riesgo de evasión tributaria en Colombia, utilizando variables proxy que representen las principales causas de evasión.



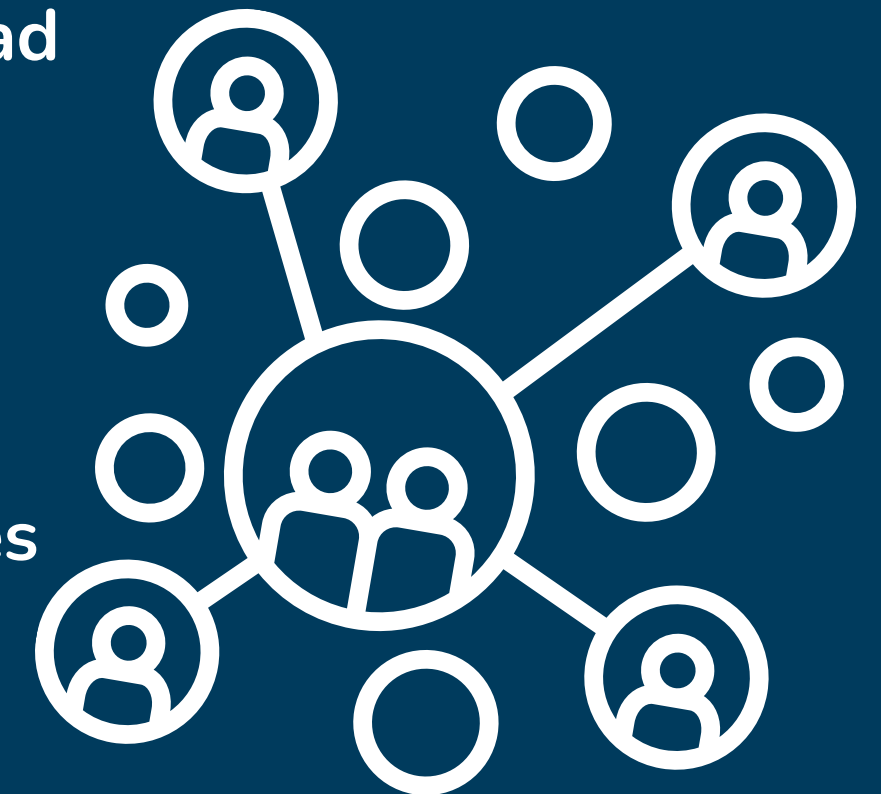


# DESCRIPCIÓN



Este proyecto busca explorar, a través de técnicas no supervisadas de aprendizaje automático, cómo factores socioeconómicos, de confianza institucional y de capacidad contributiva se relacionan con la evasión tributaria en Colombia.

Se utilizarán K-means, PCA y Apriori para detectar clústeres de contribuyentes, reducir dimensionalidad y encontrar reglas de asociación que revelen combinaciones de factores vinculados al incumplimiento fiscal.





# DESAFÍOS

- Acceso y limpieza de bases de datos dispersas (DIAN, DANE, Transparencia, etc.).
- Construcción de proxies representativos para variables intangibles como “confianza en el gobierno” o “cultura tributaria”.
- Posible falta de microdatos detallados sobre sanciones o evasión directa.
- Integración de diferentes fuentes de información (fiscal, socioeconómica y percepción ciudadana).

# ENTREGA DE VALOR

- Identificar perfiles de riesgo de evasión tributaria con base en datos públicos.
- Proveer evidencia útil para diseñar políticas públicas más focalizadas en aumentar el cumplimiento tributario.
- Generar un marco replicable para futuros estudios sobre evasión y cumplimiento fiscal.



DANE



# STAKEHOLDERS



- DIAN: interesada en mejorar el control y las estrategias de fiscalización.
- DANE: aporta datos socioeconómicos y de percepción social.
- Ministerio de Hacienda: beneficiario indirecto por mayor eficiencia recaudatoria.
- Ciudadanos y empresas: actores evaluados en términos de su cumplimiento fiscal.
- Academia: análisis que enriquece el campo de la economía pública y la ciencia de datos aplicada.



# TÉCNICAS

- K-means clustering: para agrupar contribuyentes o regiones según perfiles de riesgo de evasión.
- PCA (Análisis de Componentes Principales): para reducir la dimensionalidad de las variables proxy y obtener factores latentes.
- Apriori (reglas de asociación): para identificar combinaciones frecuentes de factores que explican patrones de incumplimiento.





# FUENTES DE DATOS



- DIAN: estadísticas de recaudo, sanciones y comercio exterior ([Datos Abiertos DIAN](<https://www.dian.gov.co/atencionciudadano/Paginas/Datos-Abiertos.aspx>)).
- DANE: Encuesta Pulso Social, GEIH (informalidad laboral), ingresos y gastos de los hogares ([DANE](<https://www.dane.gov.co/index.php/estadisticas-por-tema>)).
- Transparencia Internacional / Transparencia por Colombia: Índice de percepción de corrupción ([transparency.org](<https://www.transparency.org/en/cpi/2024>)).
- LAPOP – AmericasBarometer: Datos de confianza institucional y cultura tributaria ([LAPOP](<https://www.vanderbilt.edu/lapop/>)).
- Banco Mundial – CPIA :Indicadores de rendición de cuentas y corrupción ([World Bank Data](<https://data.worldbank.org/indicator/IQ.CPA.TRAN.XQ>)).

# VARIABLES

1. CARGA TRIBUTARIA RELATIVA =  $\text{IMPUESTOS RECAUDADOS} / \text{PIB REGIONAL (DIAN + DANE)}$ .
2. CAPACIDAD ECONÓMICA =  $\text{INGRESO PROMEDIO PER CÁPITA POR REGIÓN (DANE)}$ .
3. CONFIANZA EN EL GOBIERNO =  $\text{PORCENTAJE DE ENCUESTADOS CON CONFIANZA ALTA EN INSTITUCIONES (PULSO SOCIAL – DANE, LAPOP)}$ .
4. PERCEPCIÓN DE CORRUPCIÓN =  $\text{PUNTAJE DEL ÍNDICE DE PERCEPCIÓN DE CORRUPCIÓN (TI/TRANSPARENCIA POR COLOMBIA)}$ .
5. INFORMALIDAD LABORAL =  $\% \text{ DE OCUPADOS SIN SEGURIDAD SOCIAL (GEIH – DANE)}$ .
6. HISTORIAL DE TRADICIÓN TRIBUTARIA =  $\text{EVOLUCIÓN DE CONTRIBUYENTES REGISTRADOS EN RUT (DIAN)}$ .
7. SANCIONES EFECTIVAS =  $\text{NÚMERO DE SANCIONES TRIBUTARIAS APLICADAS POR AÑO (DIAN)}$ .
8. COMPLEJIDAD NORMATIVA =  $\text{NÚMERO DE REFORMAS TRIBUTARIAS APROBADAS POR AÑO (REPOSITORIOS ACADÉMICOS + EMI)}$ .





