

Clasificación no supervisada de la capacidad financiera usando datos de la ENIF 2024

Josué Orión Nava Ponce

Junio 2025

- La capacidad financiera no tiene una medición directa, sino que es una variable latente multidimensional.
- Usamos datos de la ENIF 2024 (del INEGI) para construir una clasificación.
- Como no hay una variable que mida directamente la capacidad financiera, se implementará una estrategia para lograr la clasificación.
- Definimos capacidad financiera como *un constructo multidimensional que refleja la habilidad de una persona u organización para administrar eficientemente sus ingresos y gastos, cumplir con sus obligaciones económicas y tomar decisiones informadas que le permitan realizar inversiones, afrontar compromisos de corto, medio y largo plazo, y estar preparada ante eventualidades económicas adversas.*

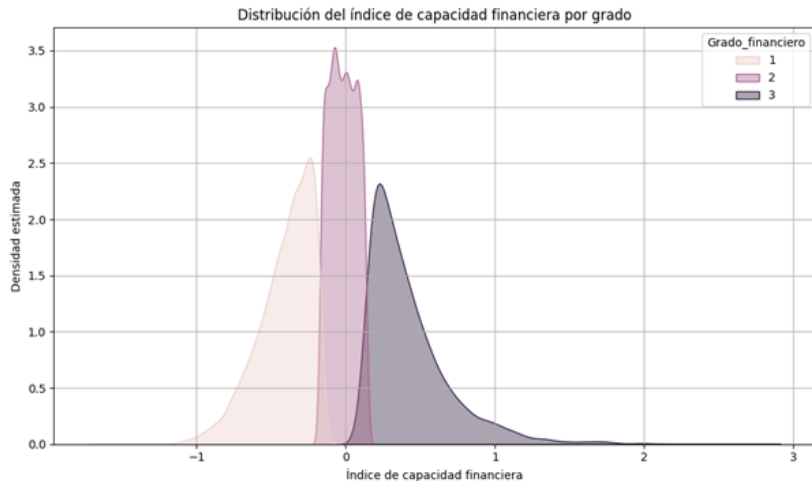
1. Exploración inicial. Selección de los conjuntos de datos a usar (información sobre las personas y las viviendas). LEFT JOIN de estos conjuntos de datos.
2. Exploración de las variables consideradas hasta ahora: nominales en su gran mayoría. Un total de 416 variables.
3. 1er. filtro de variables. Creación de una lista de palabras clave a partir de la definición de capacidad financiera, y filtración de las variables con coincidencias en sus descripciones.
4. 2do. filtro de variables. Eliminación justificada de variables con más de 20 % de datos faltantes (suelen ser particularidades a partir de condiciones de preguntas base).
5. 3er. filtro de variables. Eliminación de aquellas variables que parecen no aportar suficiente información conceptual. Selección manual.
6. Eliminación de filas con muchos datos faltantes (solo una fila en realidad). Esto deja un conjunto de datos completo (sin datos faltantes), y por lo tanto no es necesario imputar.

7. Análisis del tipo de variables. 85 categóricas nominales, 3 numéricas y 1 del identificador.
8. Transformación de las variables nominales usando OneHotEncoder y estandarización de todas con StandardScaler.
9. Análisis de Componentes Principales.
10. Creación del índice de capacidad financiera usando 40 componentes principales, que que explican casi el 68 % de varianza total.
11. Discretización del índice en tres categorías.
12. Exportación del conjunto de datos que contiene el identificador de la persona llavemod, su índice y su grado de capacidad financiera calculado.

Encabezado del resultado final

llavemod	Índice_financiero	Grado_financiero
101101	-0.662753	1
210101	-0.686279	1
303102	-0.338333	1
401101	-0.292944	1
503101	0.246904	3
603102	0.342867	3
704101	0.173962	3
803102	-0.151144	2
904102	-0.031295	2
1004101	0.242732	3
1102102	0.666169	3
1201102	-0.144713	2
1303104	-0.313014	1
1408102	0.528913	3
1504101	0.065670	2
1609104	-0.120863	2
1704101	0.003241	2
1816102	0.485603	3
1904105	-0.180112	1
2008101	0.172696	3

Visualización del índice



Conclusión

El **índice de capacidad financiera** es una variable continua construida mediante Análisis de Componentes Principales (PCA), que resume información de variables numéricas y categóricas nominales relacionadas con el comportamiento, acceso y conocimiento financieros.

Con fines prácticos, este índice fue discretizado en tres categorías para clasificar a las personas según su capacidad financiera. Representa, en una dimensión, un nivel estimado de dicha capacidad, pero no permite concluir directamente que valores altos implican un mejor comportamiento o mayor acceso.

Esto se debe a que el índice es una combinación ponderada de variables dummy obtenidas por codificación One-Hot, las cuales indican solo la presencia o ausencia de ciertas características. Una interpretación más precisa requeriría revisar en detalle las variables que componen cada componente principal.

¿Qué sigue y cómo mejorar?

- 1 Revisar la interpretación de los componentes principales y/o explorar análisis factorial como alternativa.
- 2 Analizar las variables dummy creadas para interpretar mejor el índice y su discretización.
- 3 Aplicar técnicas de *clustering* (ej. K-Means) y comparar con el índice y su clasificación.
- 4 Realizar una selección de variables más rigurosa: evaluar semánticamente su relación con la capacidad financiera y conservar solo las conceptualmente relevantes.
- 5 Explorar otras formas de clasificación y comparar resultados; si son consistentes, se fortalece la validez del enfoque adoptado.
- 6 Considerar una estrategia alternativa basada en la lectura manual y selección de variables ordinales relevantes, aunque esto requiere más tiempo y esfuerzo.