



Lic. en Ciencias del Comportamiento

Trabajo Práctico N° 3

Alumnas

Agustina Kiessner

Isabela Nicola

Clara Mollón

Profesores

María Noelia Romero

Tomas Enrique Buscaglia

Ignacio Anchorena

Asignatura

Ciencia de datos - Tutorial 3

Fecha de presentación

31 de Octubre de 2025

Parte A: Enfoque de validación

1.

Para empezar, se seleccionaron cinco predictores relevantes según la literatura para estimar la probabilidad de que una persona sea pobre. Las variables incluidas fueron ch04 (sexo), ch06 (edad), ch07 (estado civil), nivel_ed (nivel educativo) y estado (condición de actividad). Es importante destacar que el análisis de todo este trabajo fue realizado en la región del NOA. Con el objetivo de simplificar la metodología, las variables sexo, estado civil y condición de actividad se transformaron en *dummies*, dado que son categóricas y su codificación binaria facilita su inclusión en los modelos. Las demás variables se mantuvieron en su formato original, ya que edad es numérica y nivel educativo es una variable categórica ordinal, por lo que no resulta necesario convertirla en *dummy*. Asimismo, para ambos años se restringió la muestra a la población económicamente activa (PEA), es decir, personas ocupadas o desocupadas, excluyendo a los inactivos y a los menores de 10 años. Esta decisión se basa en la necesidad de mantener coherencia en el análisis. Las personas inactivas no están directamente expuestas a las mismas dinámicas económicas que determinan la pobreza entre quienes trabajan o buscan empleo. Incluirlas podría distorsionar los resultados, ya que sus ingresos y condiciones de vida dependen de factores distintos. Por eso, al limitar el estudio a la población económicamente activa (PEA) se logra un grupo más comparable y representativo de la relación entre las variables laborales y la pobreza.

Incluimos la variable sexo ya que se ha demostrado que las mujeres en Argentina enfrentan mayores desventajas estructurales en el mercado laboral: por ejemplo, la brecha salarial, donde suelen tener más empleos informales y menor acceso a trabajos de calidad. Además, elegimos la edad ya que los jóvenes suelen tener inserciones laborales precarias o menores ingresos, mientras que los adultos mayores pueden depender de jubilaciones o del trabajo no remunerado dentro del hogar, lo que aumenta su vulnerabilidad económica (Esquivel, 2014). Luego, el nivel educativo se incluyó ya que, según Chagalj & Lagomarsino (2019), constituye uno de los principales predictores de pobreza. La evidencia muestra que a mayor nivel de educación, menor es la probabilidad de que sea pobre dado que la educación mejora las oportunidades laborales y el acceso a empleos formales. Después, se agregó al estado civil debido a que determina la composición del hogar, la disponibilidad de ingresos, la carga de dependientes y la estabilidad financiera. Por último, la condición de actividad influye en la predicción de pobreza porque refleja directamente el acceso de los hogares a ingresos laborales. Esto se debe a que los desocupados presentan mayores probabilidades de ser pobres, dado que la falta o inestabilidad del empleo reduce los ingresos del hogar. Mientras que los ocupados tienden a ubicarse por encima de la línea de pobreza (Fernández, 2005). Los hogares con miembros desocupados perciben ingresos más bajos y enfrentan una mayor vulnerabilidad económica (Beccaria & Groisman, 2008).

Luego, se construyó una tabla comparando las medias de las variables seleccionadas en la matriz X entre la base de entrenamiento y la de testeo, tanto para 2025 (Tabla 1) como para 2005 (Tabla 2). Al analizar las diferencias de medias en ambos años, se observa que las variaciones son en general muy pequeñas, tanto en valores absolutos como relativos. Esto indica que el procedimiento de partición de los datos mantuvo una adecuada representatividad de ambas muestras, evitando sesgos sistemáticos. En 2025, las diferencias relativas son menores al 3%, lo que refleja una buena homogeneidad entre los conjuntos. En

particular, la edad y el “nivel_ed” presentan variaciones mínimas (0,78 % y 2,37 %, respectivamente), mientras que las categorías de estado civil y condición de actividad muestran una distribución muy similar. En 2005 se mantiene la misma tendencia donde las diferencias relativas son bajas, por debajo del 3 % en la mayoría de los casos, lo cual confirma que no existen desbalances relevantes entre las bases.

En definitiva, no se observan diferencias significativas entre las medias de las variables de entrenamiento y testeo. Ambas muestras pueden considerarse estadísticamente equivalentes, lo que garantiza la validez del entrenamiento del modelo.

Cabe aclarar que las variables de “estado civil” se toma como categoría de referencia a las personas “unidas”, y se crean variables binarias para las demás categorías: “casado/a”, “separado/a o divorciado/a”, “viudo/a” y “soltero/a”. Por ejemplo, Estado Civil_2 = 1 indica que el individuo está casado, mientras que Estado Civil_2 = 0 implica que pertenece a otra categoría de estado civil. Este procedimiento permite incluir variables categóricas en la regresión logística, comparando cada grupo respecto a la categoría base.

Tabla 1. Base de datos de 2025

| | Media train | Media test | Diferencia | Diferencia relativa % |
|-------------------------------|--------------------|-------------------|-------------------|------------------------------|
| Sexo | 0.425287 | 0.438045 | -0.012758 | -2.999858 |
| Edad | 37.427883 | 37.054101 | 0.373782 | 0.998672 |
| Estado Civil_2 | 0.362996 | 0.355148 | 0.007848 | 2.161892 |
| Estado Civil_3 | 0.067482 | 0.081152 | -0.013669 | -20.256314 |
| Estado Civil_4 | 0.023359 | 0.025305 | -0.001946 | -8.331256 |
| Estado Civil_5 | 0.358176 | 0.376963 | -0.018788 | -5.245358 |
| Nivel educativo | 3.588061 | 3.570681 | 0.017380 | 0.484389 |
| Condición de actividad | 0.109381 | 0.131763 | -0.022382 | -20.462330 |

Tabla 2. Base de datos de 2005

| | Media train | Media test | Diferencia | Diferencia relativa % |
|------------------------|--------------------|-------------------|-------------------|------------------------------|
| Sexo | 0.425287 | 0.438045 | -0.012758 | -2.999858 |
| Edad | 37.427883 | 37.054101 | 0.373782 | 0.998672 |
| Estado Civil_2 | 0.362996 | 0.355148 | 0.007848 | 2.161892 |
| Estado Civil_3 | 0.067482 | 0.081152 | -0.013669 | -20.256314 |
| Estado Civil_4 | 0.023359 | 0.025305 | -0.001946 | -8.331256 |
| Estado Civil_5 | 0.358176 | 0.376963 | -0.018788 | -5.245358 |
| Nivel educativo | 3.588061 | 3.570681 | 0.017380 | 0.484389 |

| | | | | |
|------------------------|----------|----------|-----------|------------|
| Condición de actividad | 0.109381 | 0.131763 | -0.022382 | -20.462330 |
|------------------------|----------|----------|-----------|------------|

2. En el código

Parte B: Modelo de Regresión Logística

3.

Realizamos una regresión logística usando X_{train} tanto para 2025 (Tabla 3) como para 2005 (Tabla 4). En el modelo estimado para el año 2025, se observa que las variables sexo, edad, nivel educativo, estado civil y condición de actividad inciden de manera diferenciada sobre la probabilidad de ser pobre. El nivel educativo presenta un coeficiente negativo (-0.5758), lo que indica que, a mayor nivel educativo, menor es la probabilidad de encontrarse en situación de pobreza. El *odds ratio* (0.5622) confirma esta relación por lo que cada incremento en el nivel educativo reduce en aproximadamente un 44% las chances de ser pobre, manteniendo constantes las demás variables. La condición de actividad tiene un efecto positivo y significativo donde el coeficiente de estado es 1.4464, con un *odds ratio* de 4.25, lo que implica que las personas desocupadas tienen más de cuatro veces las probabilidades de ser pobres en comparación con las ocupadas. Este resultado es coherente con la literatura sobre el vínculo entre desempleo e ingreso. Por otro lado, la variable sexo presenta un coeficiente positivo (0.2015), con un *odds ratio* de 1.22, lo que sugiere que ser mujer aumenta ligeramente la probabilidad de ser pobre, aunque el efecto es moderado. En cambio, la edad tiene un coeficiente negativo (-0.03), lo que indica que a mayor edad, disminuye levemente la probabilidad de pobreza, posiblemente reflejando mayor estabilidad laboral o acumulación de capital humano. En cuanto al estado civil, las categorías “separado/a”, “viudo/a” y “soltero/a” muestran coeficientes negativos, lo que sugiere que estas condiciones tienden a asociarse con una menor probabilidad de pobreza respecto a la categoría base (unido/a), aunque los efectos son más débiles y pueden no ser estadísticamente significativos. Finalmente, el intercepto elevado (3.1167) refleja una probabilidad base relativamente alta de pobreza en ausencia de las demás características explicativas, coherente con la magnitud del fenómeno en la población total.

En el año 2005, la dirección de los efectos es muy similar, aunque se observan diferencias en la magnitud de los coeficientes. El nivel educativo (-0.6241) mantiene un efecto fuertemente negativo, con un *odds ratio* de 0.5357, lo que significa que cada nivel adicional de educación reduce aproximadamente en un 46% las probabilidades de ser pobre, por lo tanto resulta ser una variable protectora frente a la pobreza a lo largo del tiempo. La condición de actividad también muestra una fuerte asociación positiva, indicando que las personas desocupadas en 2005 tenían más del doble de probabilidades de ser pobres que las ocupadas. Sin embargo, el efecto es algo menor que en 2025, lo que podría sugerir un agravamiento de la penalización por desempleo en el contexto reciente. El sexo vuelve a tener un efecto positivo (0.0588), aunque de menor magnitud, mostrando una brecha de género más reducida. La edad conserva un coeficiente negativo (-0.0353), consistente con el hallazgo de que los adultos mayores tienden a presentar menor vulnerabilidad relativa frente a la pobreza. En el caso de las variables de estado civil, los efectos son pequeños y alternan entre positivos y negativos, sin un patrón fuerte ni significancia. Finalmente, el intercepto

(3.1358) indica una probabilidad base similar a la del modelo 2025, sugiriendo que la estructura de pobreza general se mantiene, aunque la intensidad de algunos determinantes varía.

Tabla 3. Regresión logística con los datos de 2025

| | Coefficientes | Error Estándar | Odd-Ratio |
|-------------------------------|----------------------|-----------------------|------------------|
| Sexo | 0.2015 | 0.0864 | 1.2232 |
| Edad | -0.0301 | 0.0040 | 0.9703 |
| Estado Civil_2 | -0.0869 | 0.1290 | 0.9168 |
| Estado Civil_3 | 0.0396 | 0.1772 | 1.0404 |
| Estado Civil_4 | -0.0294 | 0.3270 | 0.9710 |
| Estado Civil_5 | -0.3931 | 0.1077 | 0.6750 |
| Nivel educativo | -0.5758 | 0.0359 | 0.5622 |
| Condición de actividad | 1.4464 | 0.2096 | 4.2479 |
| Intercept | 3.1167 | 0.2406 | 22.5722 |

Tabla 4. Regresión logística con los datos de 2005

| | Coefficientes | Error Estándar | Odd-Ratio |
|-------------------------------|----------------------|-----------------------|------------------|
| Sexo | 0.0588 | 0.0918 | 1.0606 |
| Edad | -0.0353 | 0.0042 | 0.9653 |
| Estado Civil_2 | -0.2392 | 0.1291 | 0.7872 |
| Estado Civil_3 | 0.1688 | 0.2014 | 1.1838 |
| Estado Civil_4 | -0.0373 | 0.3208 | 0.9634 |
| Estado Civil_5 | -0.3517 | 0.1259 | 0.7035 |
| Nivel educativo | -0.6241 | 0.0339 | 0.5357 |
| Condición de actividad | 0.8632 | 0.1413 | 2.3708 |
| Intercept | 3.1358 | 0.2176 | 23.0070 |

4.

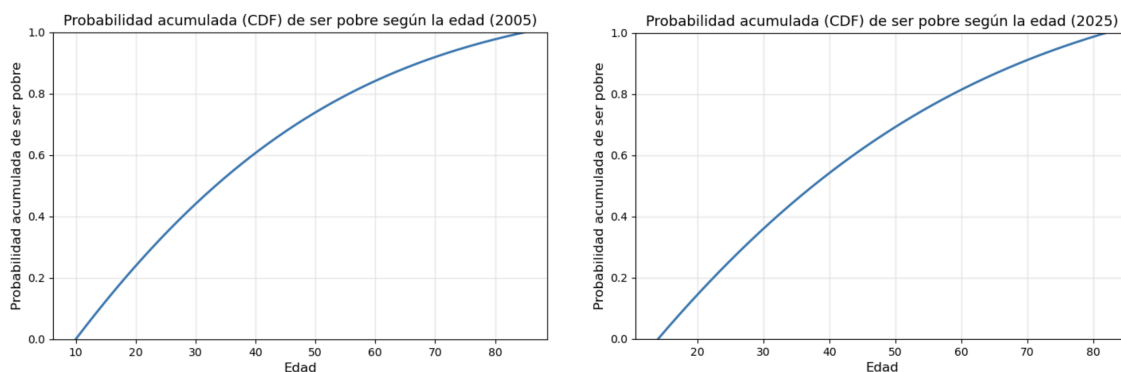
La Figura 1 muestra la probabilidad acumulada de ser pobre según la edad para los años 2005 (gráfico de la izquierda) y 2025 (gráfico de la derecha). En ambos casos, la curva tiene una forma creciente y suavemente cóncava hacia abajo, lo que indica que la probabilidad acumulada aumenta con la edad, aunque a un ritmo decreciente.

En 2025, la probabilidad acumulada crece de manera sostenida desde edades tempranas hasta alrededor de los 40 años, momento en el cual la pendiente comienza a reducirse. Esto refleja que las personas más jóvenes concentran una mayor probabilidad de

encontrarse en situación de pobreza, mientras que a medida que la edad avanza, esta probabilidad disminuye y se estabiliza en edades adultas y mayores. La forma suavemente curvada indica un efecto negativo moderado de la edad sobre la pobreza. En el modelo *logit*, el coeficiente asociado a la edad es negativo, por lo que el aumento en la edad reduce la probabilidad puntual de ser pobre, y la acumulada tiende a saturarse conforme se llega a edades más altas. Esto sugiere que en 2025 los jóvenes se enfrentan a condiciones económicas más vulnerables, posiblemente relacionadas con la inserción laboral precaria, el desempleo o la inestabilidad de ingresos, mientras que los adultos mayores presentan una menor probabilidad acumulada de pobreza. En 2005 se observa un patrón similar, aunque la curva se encuentra levemente desplazada hacia arriba en los tramos iniciales, lo que indica una mayor probabilidad acumulada de pobreza a edades tempranas en comparación con 2025. La pendiente disminuye con la edad, reflejando que la probabilidad marginal de ser pobre cae conforme aumenta la edad. Sin embargo, la curva de 2005 presenta una acumulación más acelerada en los primeros años, lo que sugiere que la juventud estaba más expuesta a la pobreza relativa a la población total que en la actualidad.

Comparando ambos períodos, podemos observar una mejora general en las condiciones socioeconómicas de los grupos jóvenes en 2025 respecto de 2005, evidenciada por la menor pendiente inicial y la curva acumulada más baja. En ambos años, la edad se mantiene como un factor protector frente a la pobreza, pero en 2025 el efecto parece algo más atenuado.

Figura 1



Parte C: Método de Vecinos Cercanos (KNN)

5.

Los resultados de clasificación presentados en la Tabla 5 muestran que, en ambos años, cuando el número de vecinos es muy bajo ($K = 1$), el modelo identifica una cantidad mayor de hogares como pobres (1.123 en 2025 y 1.002 en 2005). Esto ocurre porque con un K chico la frontera de decisión es muy sensible a las particularidades del conjunto de entrenamiento, generando un modelo con bajo sesgo pero alta varianza. A medida que K aumenta ($K = 5$ y $K = 10$), la frontera se suaviza, por lo que el modelo se vuelve más estable y disminuye la cantidad de hogares clasificados como pobres (985 y 648 en 2025, 914 y 802 en 2005). Este patrón refleja el *trade-off* entre sesgo y varianza, donde al considerar más vecinos el modelo gana generalidad aunque con mayor sesgo pero pierde sensibilidad a fluctuaciones locales, por lo que consigue una menor varianza.

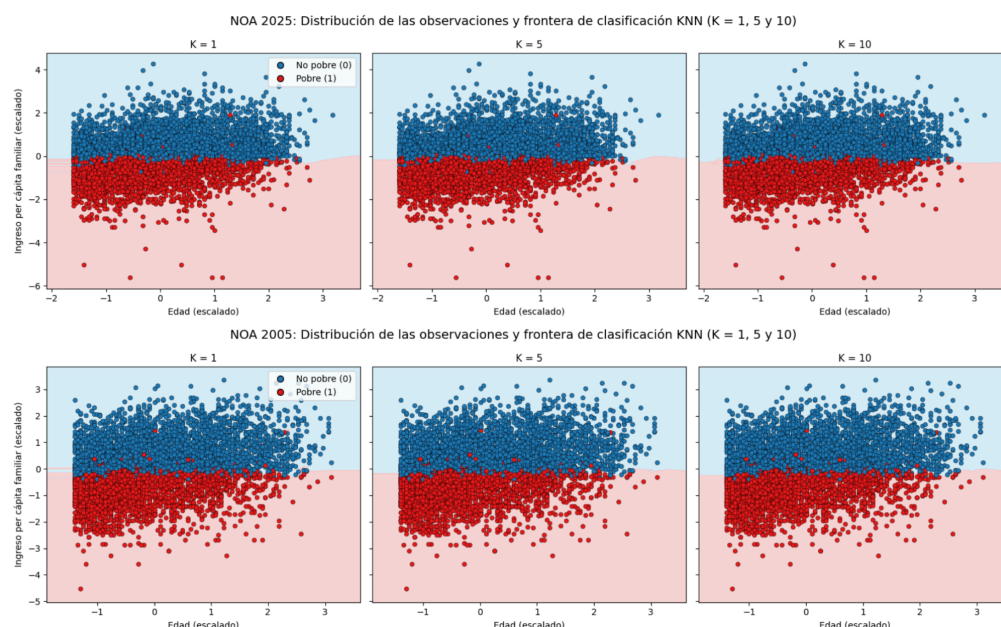
Para la construcción de la matriz de predictores se decidió estandarizar la variable ch06 (edad) mediante la función *StandardScaler*, ya que el algoritmo KNN se basa en el cálculo de distancias entre observaciones y es necesario que todas las variables numéricas estén en la misma escala para evitar que una de ellas domine la métrica de distancia. Asimismo, se generaron variables *dummies* para todas las categorías posibles de las variables ch04 (sexo), ch07 (estado civil), nivel_ed (nivel educativo) y estado (condición de actividad). Esta transformación permite que las variables categóricas puedan ser interpretadas correctamente por el modelo al expresarse en formato numérico binario. Se mantuvieron las mismas categorías fijas en ambos años para garantizar que las matrices “X_train_2005” y “X_train_2025” tuvieran una estructura idéntica, lo cual permite una comparación entre periodos. En los casos en que una categoría no aparece en la muestra, la columna correspondiente adopta valores iguales a cero, lo que no afecta la clasificación dado que esas variables permanecen constantes.

Tabla 5. Resultados de clasificación KNN por año y valor de K en la región NOA

| | | |
|--|---|--|
| [NOA 2025 K=1] No pobres (0): 1716 Pobres (1): 1123 | [NOA 2025 K=5] No pobres (0): 1854 Pobres (1): 985 | [NOA 2025 K=10] No pobres (0): 2191 Pobres (1): 648 |
| [NOA 2005 K=1] No pobres (0): 1695 Pobres (1): 1002 | [NOA 2005 K=5] No pobres (0): 1783 Pobres (1): 914 | [NOA 2005 K=10] No pobres (0): 1895 Pobres (1): 802 |

6.

Figura 3



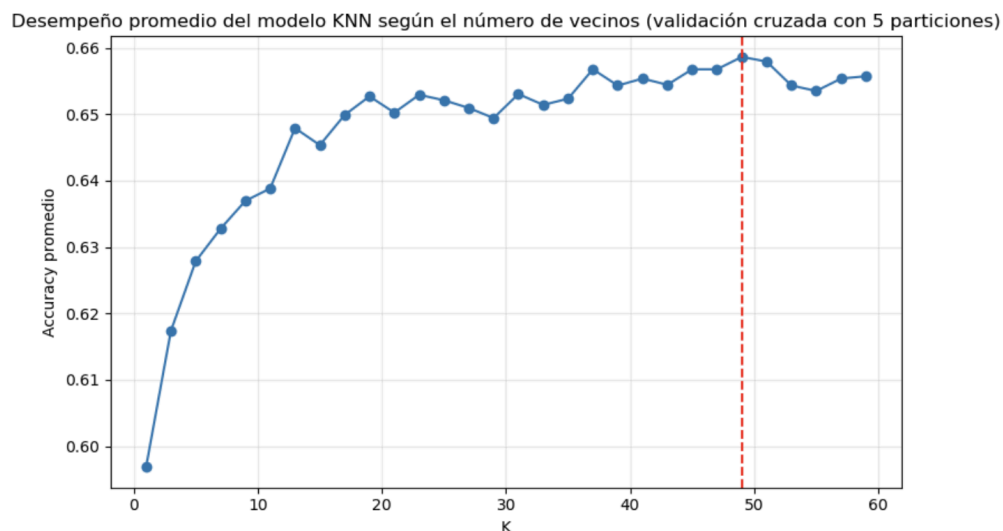
Nota. En este ejercicio se utilizó el ingreso per cápita familiar como variable continua principal, dado que el resto de las variables no presentaban una escala numérica adecuada para este tipo de gráficos. Se utiliza esta variable exclusivamente para visualización, el modelo final no usa ingreso.

La Figura 3 muestra la distribución de los hogares de la región NOA en los años 2005 y 2025 según su edad (eje horizontal) e ingreso per cápita familiar (eje vertical), junto con la frontera de clasificación obtenida por el modelo KNN para distintos valores de K (1, 5 y 10). En todos los casos, la separación entre pobres y no pobres se da principalmente por el nivel de ingreso per cápita familiar, mientras que la edad aporta muy poca información para distinguirlos. Por este motivo, la frontera de decisión luce muy similar entre los distintos valores de K y entre ambos años, mostrando una división casi horizontal donde los hogares con menores ingresos (zona inferior) se clasifican como pobres, y los de mayores ingresos (zona superior) como no pobres. A medida que aumenta el número de vecinos considerados (de K=1 a K=10), la frontera se suaviza levemente, lo que refleja una menor sensibilidad a las variaciones locales o casos atípicos, aunque el patrón general de separación se mantiene estable. Este comportamiento es consistente entre 2005 y 2025, lo que sugiere que la relación entre pobreza e ingreso no presenta grandes cambios estructurales en el período analizado.

Aunque se aplicó *StandardScaler* para estandarizar las variables, esta transformación solo corrige diferencias de escala, pero no la asimetría de los ingresos. Por ello, se utilizó la transformación $\log(1 + \text{ipcf})$, que permite reducir los valores extremos y mejorar tanto la visualización como el desempeño del modelo KNN..

7.

Figura 4



Para elegir el número óptimo de vecinos en el modelo KNN, se dividió la base “respondieron_2025” en cinco partes y se aplicó *cross-validation*. En cada iteración, el modelo se entrenó con cuatro partes y se evaluó en la restante, promediando el *accuracy* obtenido para cada valor de K entre 1 y 60.

La Figura 4 muestra cómo el *accuracy* promedio aumenta con K y luego se estabiliza, alcanzando su máximo en $K = 49$ con un *accuracy* de aproximadamente 0.6565. A partir de ese punto, el desempeño del modelo deja de mejorar, por lo que $K = 49$ se considera el valor óptimo. Este resultado indica que el modelo logra un buen equilibrio entre ajuste y generalización, evitando tanto el *overfitting* como el *underfitting*.

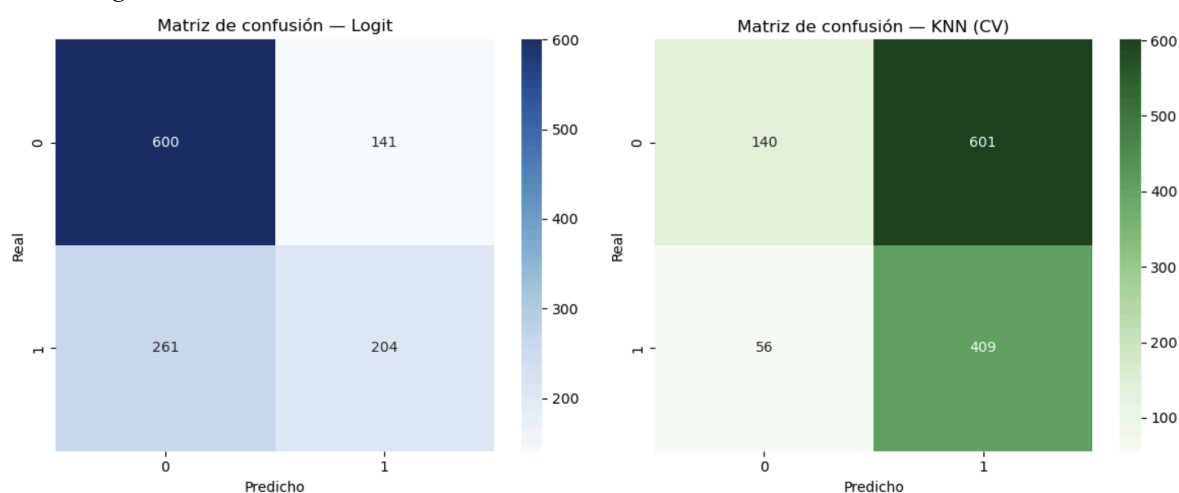
Parte D: Desempeño de modelos, elección y predicción afuera de la muestra

8.

En esta consigna, se calculó la matriz de confusión para comparar el desempeño de los modelos Logit y KNN con CV sobre la base test 2025. Luego, se generó la curva ROC para analizar la capacidad de discriminación de cada modelo y como métricas adicionales vistas en clase se calculó la medida de *accuracy* que indica la proporción de aciertos totales y sensibilidad (*recall*) que muestra la capacidad del modelo para identificar correctamente a los hogares pobres (evitando falsos negativos). Se eligieron estas métricas ya que resultan adecuadas al ser las más informativas para un problema de clasificación en contextos sociales como la predicción de pobreza, buscando maximizarla sin descuidar la precisión total.

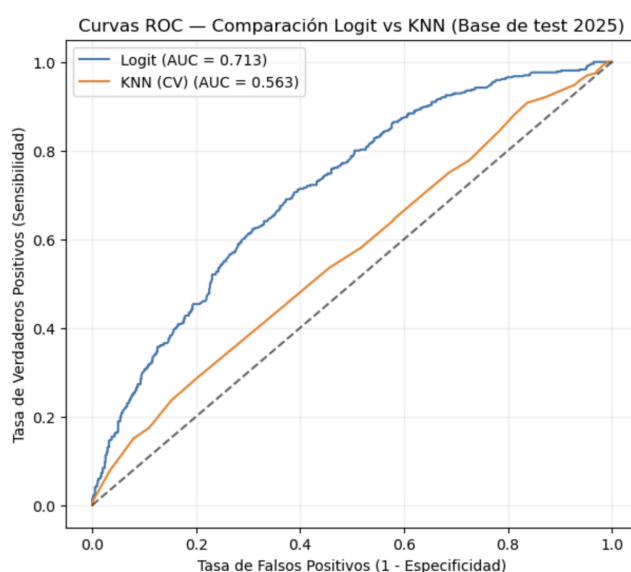
En la matriz de confusión (Figura 5) se observa que el modelo Logit logra 600 verdaderos negativos (VN), 141 falsos positivos (FP), 204 verdaderos positivos (VP) y 261 falsos negativos (FN). Esto significa que el modelo clasifica correctamente a la mayoría de los hogares no pobres, aunque tiende a subestimar la pobreza al no identificar una parte importante de los hogares pobres. Por su parte, el modelo KNN con CV muestra 140 VN, 601 FP, 409 VP y 56 FN, lo que implica que reduce considerablemente los falsos negativos e incrementa los verdaderos positivos, detectando con mucha mayor eficacia a los hogares pobres, aunque comete más errores clasificando como pobres a quienes no lo son.

Figura 5



En cuanto a la curva ROC y AUC (Figura 6), el modelo Logit (curva azul) presenta un AUC de 0.713, mientras que el KNN con CV (curva naranja) alcanza un AUC de 0.563. Esto indica que el Logit tiene una mayor capacidad de discriminación entre hogares pobres y no pobres, ya que su curva se encuentra más alejada de la diagonal que representa un modelo sin capacidad predictiva. En otras palabras, aunque el KNN logra una alta sensibilidad, su desempeño global al distinguir ambas clases es inferior.

Figura 6



En la Tabla 6, los resultados muestran que el modelo Logit presenta un *accuracy* de 0.667 y una sensibilidad de 0.439, es decir, identifica correctamente al 43,9% de los hogares pobres y mantiene un rendimiento general equilibrado. En cambio, el modelo KNN con CV obtiene una *accuracy* menor (0.455) pero una sensibilidad muy alta (0.880), lo que indica que logra detectar al 88% de los hogares pobres reales, aunque a costa de clasificar erróneamente a muchos hogares no pobres.

Tabla 6

| Modelo | Accuracy | Sensibilidad (Recall) |
|----------|----------|-----------------------|
| Logit | 0.667 | 0.439 |
| KNN (CV) | 0.455 | 0.880 |

9.

Como el Ministerio de Capital Humano busca identificar hogares pobres para dirigir recursos alimentarios, su principal objetivo es minimizar el error de no detectar a los pobres reales. En términos estadísticos, el error Tipo I consiste en predecir que un hogar es pobre cuando en realidad no lo es (falso positivo), lo cual implicaría entregar recursos a quien no los necesita. En cambio, el error Tipo II ocurre al predecir que un hogar no es pobre cuando sí lo es (falso negativo), lo que implicaría excluir a un hogar vulnerable del programa. Dado que el costo social de no asistir a un hogar pobre es mucho mayor que el de otorgar ayuda a uno que no lo necesita, el Ministerio debería priorizar minimizar el error Tipo II, es decir, maximizar la sensibilidad del modelo (la proporción de pobres correctamente identificados).

Al comparar los resultados, el modelo KNN con validación cruzada presenta una sensibilidad muy alta (0.880), lo que significa que detecta correctamente al 88% de los hogares pobres reales, reduciendo significativamente los falsos negativos (56 casos). Sin embargo, este modelo lo hace a costa de cometer muchos falsos positivos (601), lo que se refleja en una *accuracy* baja (0.455) y un AUC de 0.563, que indica menor capacidad para discriminar entre pobres y no pobres. Por otro lado, el modelo Logit logra un balance más

equilibrado, con una sensibilidad de 0.439, una *accuracy* de 0.667 y un AUC de 0.713, mostrando una mejor capacidad de discriminación general y una clasificación más precisa, aunque deja sin identificar a algunos hogares pobres.

En este contexto, el modelo KNN con CV sería más adecuado para el objetivo del Ministerio, ya que permite maximizar la detección de hogares vulnerables (minimizando el error Tipo II), aun cuando implique una menor precisión global y la entrega de recursos a algunos hogares no pobres. Este modelo resulta más consistente con un enfoque de política social inclusivo y preventivo, que prioriza evitar la exclusión de quienes realmente necesitan la ayuda alimentaria.

10.

Se utilizó el modelo KNN con validación cruzada como el modelo con mejor desempeño para predecir qué personas son pobres dentro de la base norespondieron_2025 y calcular qué proporción del total fueron clasificadas como pobres. El modelo utilizó las mismas variables explicativas que en el entrenamiento, y predijo la probabilidad de pobreza para cada individuo. Al aplicar el umbral de clasificación de 0.5, se identificaron 810 personas pobres de un total de 1035, lo que representa una proporción del 78,26% de los casos analizados. Esto significa que, según el patrón aprendido por el modelo, casi ocho de cada diez personas que no respondieron pueden ser consideradas vulnerables o con alta probabilidad de ser pobres. Esto es consistente con el comportamiento del modelo observado en las consignas anteriores, donde el KNN mostró una alta sensibilidad (0.88), priorizando la detección de la mayor cantidad posible de hogares pobres, aun a costa de una menor precisión global.

En términos de política pública, este hallazgo sugiere que una gran parte de la población no respondiente podría requerir atención prioritaria o verificación adicional, ya que es un grupo potencialmente vulnerable que no fue cubierto en la información original. De esta forma, el modelo permite orientar recursos y focalizar estrategias de intervención hacia quienes podrían estar en mayor riesgo de exclusión.

Bibliografía:

- Beccaria, L., & Groisman, F. (2008). *Informalidad y pobreza en Argentina. Investigación Económica*, 67(266), 135–169. Universidad Nacional de General Sarmiento (UNGS).
- Chagalj, C., & Lagomarsino, B. C. (2019). Predicción de la pobreza en Argentina usando Random Forest. *Universidad de San Andrés*.
- Esquivel, V. (2014). *La pobreza de ingreso y tiempo en Buenos Aires, Argentina: Un ejercicio de medición de la pobreza para el diseño de políticas públicas*. Programa de las Naciones Unidas para el Desarrollo (PNUD).
- Fernández, M. L. (2005). *Indicadores de pobreza e indigencia en el mercado laboral femenino*. Consejo de Investigaciones de la Universidad Nacional de Rosario (CIUNR).

Apéndice

Link al GitHub:

Diccionario para la base de datos 2005

https://www.indec.gob.ar/ftp/cuadros/menusuperior/eph/EPH_disenoreg_T1_2005.pdf

Diccionario para la base de datos 2025

https://www.indec.gob.ar/ftp/cuadros/menusuperior/eph/EPH_registro_1T2025.pdf