Estudio de la pobreza basada en Ingresos de la encuesta hogares 2022 de INE Bolivia

Jan Franco Figueredo Roncal

2024-06-18

Índice

Introducción

La sociedad en su conjunto depende de los ingresos que provienen de sus actividades,rentas , alquileres, ingresos asalariados y otros, que les brindan la oportunidad de acceder una mejor calidad de vida según su distribución de los mismos. La pobreza es un fenómeno siempre presente, en mayor o menor medida, en todas las sociedades, es constante el estudio realizado que busca entender de mejor manera este fenómeno a fin de proporcionar herramientas de políticas para disminuir estas brechas.

El estudio de la pobreza basada en ingresos es un enfoque fundamental para entender y abordar la desigualdad económica en la sociedad. La pobreza la definiremos como la falta de recursos suficientes para satisfacer una canasta básica de bienes y servicios necesarios para una vida digna, puede ser medida de diversas formas, pero el enfoque basado en ingresos es uno de los más utilizados debido a su claridad y facilidad de aplicación.

Existen varios métodos para medir la pobreza, dentro los mas reconocidos a nivel internacional se encuentran el de las necesidades Básicas Insatisfechas, el de la Línea de la Pobreza y el Método Integrado, que es una combinación de los dos anteriores.

En este artículo se estudia el método de la línea de pobreza por estar directamente relacionada con los ingresos laborales. En particular, además de contextualizar la problemática de la pobreza, se indaga la relación que tiene la pobreza monetaria con las características laborales de los miembros del hogar en Bolivia.

El análisis de la pobreza, y su relación con los ingresos laborales y el trabajo, se desarrolla para la gestión 2022, haciendo uso de las Encuestas de Hogares provistas por el Instituto Nacional de Estadística de Bolivia; las cuales contienen estimaciones de las líneas de pobreza y pobreza extrema.

El proceso de análisis que se realiza en la búsqueda de patrones, correlaciones y tendencias significativas en grandes conjuntos de datos mediante el uso de técnicas de análisis estadístico, aprendizaje automático y bases de datos se denomina minería de datos.

La minería de datos tiene una variedad de métodos según el tipo de análisis que se requiera; para este estudio utilizaremos métodos de clasificación que provienen de la analítica predictiva.

Una de las herramientas más utilizadas y especializada en este tipo de estudios estadísticos es el lenguaje R definido como un lenguaje de programación y un entorno para la informática estadística, asimismo el Rstudio es un programa que permite mejorar la interfaz del usuario de brindar herramientas que ayuden a su manejo (TORGO, 2011).

Este estudio se divide en 5 partes comenzando con la recopilación de datos, en segunda el diseño de la base de datos, la limpieza y transformación de datos, como cuarta parte el diseño del método de clasificación y finalmente con las conclusiones de datos.

Objetivos

Objetivo Principal

Analizar los factores socioeconómicos, demográficos y geográficos que contribuyen a la pobreza mediante métodos de clasificación de la minería de datos.

Objetivos específicos

- Seleccionar las variables más relevantes de la encuesta hogares 2022 proporcionada por el INE.
- Utilizar algoritmos de clasificación como árboles de decisión, redes neuronales, máquinas de soporte vectorial, entre otros, para segmentar la población en diferentes grupos según su nivel de pobreza.
- valuar las relaciones entre los factores identificados y la pobreza, diferenciando entre correlación y causalidad.
- Crear visualizaciones gráficas y mapas geográficos que representen la distribución de la pobreza y sus factores determinantes.

- Validar los modelos de clasificación y predicción utilizando conjuntos de datos independiente.

Motivación

El estudio de la pobreza basada en ingresos es de suma importancia debido a su impacto en la calidad de vida y el desarrollo socioeconómico. A pesar de los numerosos estudios existentes, se busca identificar lagunas significativas en la comprensión de los factores socioeconómicos, demográficos y geográficos que contribuyen a la pobreza.

Este artículo pretende abordar estas lagunas mediante la aplicación de métodos de clasificación de minería de datos, lo cual permitirá identificar patrones y tendencias que no son evidentes con análisis tradicionales. Los resultados de este estudio no solo contribuirán al conocimiento académico, sino que también proporcionarán información valiosa para la formulación de políticas más efectivas y focalizadas en la reducción de la pobreza.

Marco Teórico

La Pobreza

La pobreza se define como una exclusión derivada de la falta de los recursos requeridos para acceder a las condiciones materiales de existencia de una sociedad según su configuración histórica. Lo que se considera necesario es, a la vez, el núcleo de privación de cuya satisfacción depende la subsistencia y el conjunto de necesidades que aluden a la dignidad e igualdad del ser humano dotado de capacidades para integrarse a la sociedad (de_la_provincia_de_Buenos_Aires, 2010).

Desde fines del siglo XIX, la visión de la pobreza ha fluctuado en torno a tres conceptos. En primer lugar, está la idea de subsistencia que concibe como pobres a las familias que no obtienen el mínimo necesario para mantener tan solo la capacidad física de supervivencia del individuo. En los años setenta, la definición de pobreza desde la perspectiva de subsistencia comenzó a ser cuestionada por limitar las necesidades humanas a necesidades físicas antes que sociales. Segundo, en este contexto, comenzó a influir la noción de necesidades básicas en la definición de pobreza. Las necesidades básicas suponen una extensión de la idea de subsistencia, al considerar dos componentes: i) requerimientos mínimos de una familia para consumo privado (alimentos, techo, abrigo, ciertos muebles y equipamiento doméstico), y ii) servicios comunitarios esenciales, como agua potable, saneamiento, transporte público, salud, educación e infraestructura cultural. (Stezano, 2021).

Finalmente, en la última parte del siglo XX, las ciencias sociales generan una nueva formulación del significado de pobreza: el de privación relativa. La relatividad se refiere aquí a los recursos y a las condiciones sociales y materiales, y atiende al fenómeno de creciente dinamismo de las sociedades modernas. Este fenómeno hace inconducente el limitarse a estándares de pobreza estáticos previamente establecidos De este modo, mientras que la noción de pobreza absoluta marca una línea mínima requerida para la subsistencia, la idea de la pobreza relativa muestra que las necesidades de vida son fluctuantes, no fijas, se adaptan conforme a los procesos de transformación de la sociedad (Townsend, 2007).

Dos enfoques sobre la pobreza: pobreza absoluta y pobreza relativa

El enfoque de la pobreza absoluta parte del supuesto de que las necesidades son independientes de la riqueza de los demás y el que no sean satisfechas revela una condición de pobreza en cualquier contexto. La pobreza absoluta se define sin referencia al contexto social o las normas, sino en términos de necesidades físicas simples de subsistencia, no sociales (Spicker, 2007). En contraste, la visión de pobreza relativa supone que las necesidades surgen desde la comparación con los demás y que la pobreza depende del nivel general o promedio de riqueza (Feres, 2001). El concepto de pobreza relativa la define en términos de su relación con los estándares que existen en la sociedad. Esto solía entenderse principalmente en términos de desigualdad (Spicker, 2007), como un estándar que se aplica al segmento inferior de la distribución de ingreso (Roach, 1972).

La pobreza relativa tiene dos elementos constitutivos definitorios. El primero alude a su definición social. Townsend (2007) define la pobreza como una forma de privación relativa, es decir, como la insuficiencia o carestía (no como ausencia) en las dietas, servicios, normas y actividades comunes en la sociedad. De modo que, en tanto la pobreza depende de la riqueza general y esta no es constante en el tiempo, el estándar para identificar a los pobres requiere definirse según cierto nivel de ingreso (Boltvinik, 2020).

Desde una perspectiva histórica, la pobreza ha sido relacionada con el ingreso, noción que ha permanecido en el centro del significado del concepto. Sin embargo, la delimitación y medición precisa del ingreso es compleja. Cuando las personas carecen del ingreso y otros recursos para lograr las condiciones de vida que les permiten desempeñar sus papeles e involucrarse en relaciones conforme a papeles y estatus socialmente reconocidos, pueden reconocerse en una situación de pobreza Nolan (2011). El segundo elemento constitutivo de la pobreza relativa es el uso de métodos comparativos, es decir, se disciernen las condiciones de pobreza al contrastar entre los pobres y los miembros de la sociedad que no son pobres. Esto identifica la pobreza con desventajas y también con la desigualdad, de forma que una persona se define como pobre en relación con determinada situación de desventaja económica y social con respecto al resto de personas de su entorno. Esta concepción de la pobreza lleva a entender que la diferencia entre pobres y no pobres depende del nivel de desarrollo de la sociedad específica analizada y no puede trasladarse a otra diferente (INE, 2005).

Medición de la pobreza centrada en la riqueza

La medición de la pobreza considerada solo a partir del nivel individual de ingreso económico inhibe cualquier perspectiva comparada en el análisis de la pobreza (Sen, 2006). En este sentido, crece la importancia de hacer mediciones sensibles a particiones grupales Duclos (2004).

La visión económica ha marcado al ingreso como el indicador central de bienestar y, por ende, ha situado a las políticas públicas de fortalecimiento del ingreso como el punto clave de cualquier política para reducir pobreza y desigualdad: el paradigma de la pobreza de ingreso Ravallion (2003). El índice de desarrollo humano (IDH), por su parte, reconoce la necesidad de incluir variables complementarias al ingreso, por lo que es una suma ponderada de tres componentes (ingresos, alfabetización y esperanza de vida), que evalúa el nivel de vida de los individuos y las poblaciones de manera explícitamente multidimensional. Esto ha representado importantes avances, pero también dilemas conceptuales no resueltos: ¿qué significan movimientos en dirección contraria de estos tres componentes y cómo agregarlos? (Arcelus, 2006).

El método de la linea de Pobreza (LP)

La línea de pobreza representa un valor monetario en que se consideran dos componentes: el costo de adquirir una canasta básica de alimentos y el costo de los demás bienes y servicios, expresado sobre la base de la relación entre el gasto total y el gasto en alimentos.(CEPAL, 2018)

La línea de pobreza basada en la canasta de alimentos se denomina habitualmente “absoluta”, en contraposición a las líneas de “pobreza relativa”, puesto que se construye a partir de los requerimientos calóricos y nutricionales que aseguran un adecuado funcionamiento físico de la persona. A su vez, la determinación de la línea de pobreza sobre la base del comportamiento de un grupo de referencia introduce un criterio de adecuación al nivel de vida que existe en cada país y época. En consecuencia, la línea de pobreza definida de esta manera incluye implícitamente el costo de bienes y servicios necesarios para atender a los requerimientos de participación social. Debido a esta característica del método, es necesario que las líneas de pobreza sean sometidas a actualizaciones con el fin de adaptarlas a los cambios en el nivel de desarrollo, los hábitos de consumo y el sistema de precios.(CEPAL, 2018)

El ingreso total del hogar se define entonces como la suma del ingreso primario y las transferencias corrientes.

Mineria de Datos

La minería de datos, también denominada descubrimiento de conocimiento en datos (KDD, por sus siglas en inglés), es el proceso de descubrir patrones y otra información valiosa en grandes conjuntos de datos. Dada la evolución de la tecnología de depósito de datos y el crecimiento de los big data, la adopción de técnicas de

minería de datos se ha acelerado rápidamente en las últimas dos décadas, y las empresas las utilizan para transformar sus datos sin procesar en conocimiento útil. Sin embargo, a pesar de que la evolución continua de dicha tecnología permite tratar los datos a gran escala, su escalabilidad y automatización todavía suponen un reto para los líderes.(IBM, 2024)

La minería de datos mejora la toma de decisiones organizativas y ejecutivas a traves del análisis de datos. Las técnicas de minería de datos estan basados en dos principales funciones una de ellas se relaciona a la clasificaciono agrupamiento de datos y la otra al que llamamos aprendisaje automatico o Machine Lerning. Estos métodos se utilizan para poder organizar y filtrar los datos, buscando revelar informacion que no se ve a simple vista, buscar patrones inusuales que permiten realizar diversos analisis para la toma de decisiones.

Es importante mencionar que debe existir resultados graficos o visualizaciones que permitan entender los resultados de los analisis realizados.

Proceso de minería de datos

El proceso de minería de datos consiste en una serie de pasos que incicia en la recopilacion de información su proceso de analisis entregar gráficos de visualizacion, que nos proporciona resumenes de sociedad de datos. Como ya hemos mencionado, las técnicas de minería de datos se utilizan para generar descripciones y previsiones sobre un conjunto de datos de destino. Los científicos de datos describen los datos mediante la observación de patrones, asociaciones y correlaciones. Así mismo, clasifican y agrupan en clúster los datos por medio de métodos de clasificación y regresión, e identifican valores atípicos para los casos de uso, como la detección de correo no deseado.(IBM, 2024)

El proceso de mineria de datos según IBM (2024) de define en cuatro pasos.

1. Definir los objetivos de negocio: Este proceso es de suma importacion debido a que enfoca el estudio y se pretende obtener informacion que permita mejorar el funcionamiento de la empresa en cuestion.
2. Preparar los datos: una vez definido el alcance del problema o la estrategia, es más sencillo identificar qué conjunto de datos ayudará a dar respuesta a las preguntas u objetivos planteados por la empresa. Una vez que recopilan los datos relevantes, estos sufren un proceso de limpiesa para eliminar lo que no sirve, como los duplicados, los valores que faltan y los valores atípicos. Los científicos de datos intentan retener los predictores más importantes para garantizar la precisión óptima dentro de cualquier modelo.
3. Crear modelos y realizar minería de patrones: en función del tipo de análisis que se desea realizar, se debe investigar las relaciones de datos que sean de interés, como los patrones secuenciales, las reglas de asociación o las correlaciones.
4. Evaluación de resultados e implementación de conocimientos: una vez agregados los datos, los resultados deben evaluarse e interpretarse. Para finalizar los resultados, estos deben ser válidos, útiles y comprensibles. Cuando se cumplen estos criterios, las organizaciones pueden utilizar este conocimiento para implementar nuevas estrategias y lograr los objetivos previstos.(IBM, 2024).

Técnicas de minería de datos

La minería de datos funciona utilizando diversos algoritmos y técnicas para transformar grandes volúmenes de datos en información útil. Estos son algunos de los más habituales:

Reglas de asociación: una regla de asociación es un método basado en reglas para detectar relaciones entre variables en un conjunto de datos determinado. Estos métodos se utilizan con frecuencia para los análisis de cesta de la compra, que permiten a las empresas comprender mejor las relaciones entre los diferentes productos. Entender los hábitos de consumo de los clientes permite a las empresas desarrollar mejores estrategias de venta cruzada y motores de recomendaciones.

Redes neuronales: las redes neuronales, que se utilizan principalmente para los algoritmos de deep learning, procesan los datos de entrenamiento imitando la interconectividad del cerebro humano a través de capas de nodos. Cada nodo está formado por entradas, ponderaciones, un sesgo (o umbral) y una salida. Si ese valor de salida excede un umbral determinado, “dispara” o activa el nodo, pasando datos a la siguiente capa de

la red. Las redes neuronales aprenden esta función de correlación a través del aprendizaje supervisado y se ajustan con base en la función de pérdida, a través del proceso de pendiente de gradiente. Cuando la función de coste es igual o casi igual a cero, podemos confiar en la precisión del modelo para obtener la respuesta correcta.(**IBM?**)

Árbol de decisiones: esta técnica de minería de datos utiliza métodos de clasificación o regresión para clasificar o prever resultados potenciales en función de una serie de decisiones. Como su propio nombre indica, utiliza una visualización en forma de árbol para representar los posibles resultados de estas decisiones.

K vecino más cercano (KNN): el algoritmo K vecino más cercano, que también se denomina algoritmo KNN, es un algoritmo no paramétrico que clasifica los puntos de datos en función de su proximidad y asociación con otros datos disponibles. Este algoritmo presupone que los puntos de datos similares se encuentran cerca unos de otros. En consecuencia, busca la distancia entre puntos de datos, generalmente mediante distancia euclídea, y luego asigna una categoría basada en el promedio o la categoría más frecuente.

Aplicaciones de la minería de datos

Las técnicas de minería de datos gozan de una amplia adopción entre los equipos de inteligencia empresarial y de analítica de datos, y les permiten extraer conocimientos para su organización y sector.

Los campos mas utilizados son Ventas y marketing, Educación, Detección de fraude y otros.

Descripcion de la base de datos

La base de datos fue descargada del portal del Instituto Nacional de Estadistica de Bolivia, la misma corresponde a la encuesta hogares de la gestion 2022.

Asi tambien se descargó un archivo y se convirtio en una base para registro de actividades economicos.

Se utiliza los archivos pobre=p0,folio,depto, area,jefe=s01a_05,aestudio,yhog,antNeg=s04b_11aa,tipoAct=s04b_12,caeb_op de la base personas eh22p y de la base equipamiento eh22e.

Metodología

En cuanto se refiere a la metodología citar:

Recopilacion de datos:

Fuentes de Datos: Se utilizaron datos de encuestas de hogares 2022 realizada por el Instituto Nacional de Estadística de Bolivia.

Periodo: El análisis abarcó datos utilizados es de la gestión 2022, se toma un periodo en específicos siendo que se pretende realizar el estudio de indicadores y no su evolución en el tiempo.

Definición de Pobreza por Ingresos:

Umbral de Pobreza: Se establecieron umbrales de pobreza basados en el costo de una canasta básica de alimentos y otros bienes y servicios esenciales. Este umbral varía según el país y la región debido a las diferencias en el costo de vida.

Pobreza o No Pobre: Se diferenciaron dos niveles de pobreza: pobre , definida como ingresos insuficientes para cubrir únicamente la canasta básica de alimentos, y No pobre, que incluye además otros bienes y servicios esenciales.

Procesamiento de Datos:

Limpieza y Normalización: Los datos se limpiaron para eliminar inconsistencias y se normalizaron para asegurar la comparabilidad entre diferentes departamentos y áreas, además que se creó variables que identifiquen el tipo de actividad que pertenecen.

Imputación de Datos Faltantes: Se aplico la técnica de imputación Listwise deletion que consiste en una técnica fácil de implementar y entender, ya que simplemente elimina los registros incompletos y garantiza que el análisis se realice siempre con los mismos datos completos, lo que puede facilitar la interpretación y la comparación de resultados.

Métodos de Análisis:

Estadísticas Descriptivas: Se realizaron análisis descriptivos para entender la distribución de ingresos y la incidencia de la pobreza en diferentes grupos demográficos.

Regresión Logística: Se utilizó la regresión logística para identificar los factores socioeconómicos y demográficos que tienen una relación significativa con la probabilidad de estar en situación de pobreza.

Validación del modelo de Clasificacion:

Entrenamiento y Prueba: Divide el conjunto de datos en un conjunto de entrenamiento (usualmente 70%) y un conjunto de prueba (30%). El modelo se entrena en el conjunto de entrenamiento y se evalúa en el conjunto de prueba..

Presentación de Resultados:

Tablas y Gráficos: Los resultados se presentaron mediante tablas y gráficos detallados que ilustran la incidencia de la pobreza y las brechas de ingresos.

#Analisis Para realizar la mineria de datos como se indicó anteriormente se realiza el procesamiento en bruto de las bases ah22p y ah22e, como se muesra a continuacion:

```
load("eh22.Rdata")
Seleccion de los datos a utilizar.
Utilizaremos la base eh22p, e22e, para este estudio.
persona <- eh22p
persona$s04b_12 <- as.factor(persona$s04b_12)

personas <- persona %>% dplyr::select(pobre=p0,folio,depto,area,jefe=s01a_05,aestudio,
             yhog,antNeg=s04b_11aa,tipoAct=s04b_12,caeb_op)
personas <- personas%>% filter((jefe==1))
personas$tipoAct <- as.factor(personas$tipoAct)
personas <- personas %>%dplyr::filter((tipoAct == "1" | tipoAct == "3"))
```

```
head(personas,5)
```

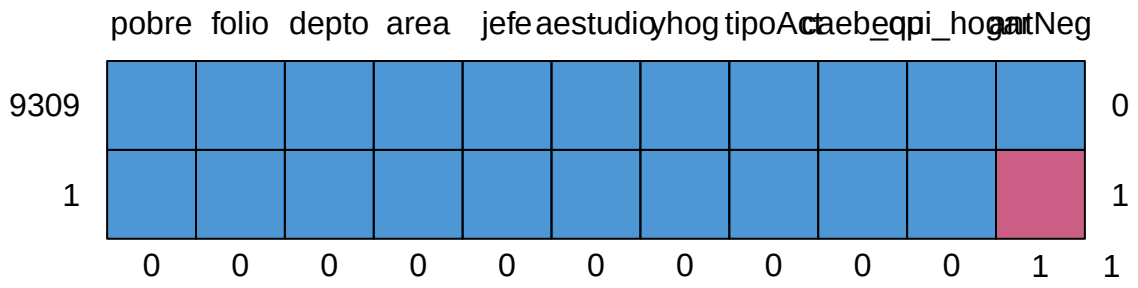
```
## # A tibble: 5 x 10
##   pobre     folio depto   area    jefe    aestudio   yhog antNeg tipoAct caeb_op
##   <dbl+lbl> <chr> <dbl+l> <dbl+l> <dbl+l>    <dbl>  <dbl>  <dbl> <fct>   <dbl+l>
## 1 0 [No Po~ 111-~ 1 [Chu~ 1 [Urb~ 1 [1. ~       12  1559.      2 3       7 [Tra~
## 2 0 [No Po~ 111-~ 1 [Chu~ 1 [Urb~ 1 [1. ~        3 11062.      5 1       5 [Con~
## 3 0 [No Po~ 111-~ 1 [Chu~ 1 [Urb~ 1 [1. ~       17  4197       5 3       6 [Ven~
## 4 1 [Pobre] 111-~ 1 [Chu~ 1 [Urb~ 1 [1. ~       12  5250      15 3       6 [Ven~
## 5 1 [Pobre] 111-~ 1 [Chu~ 1 [Urb~ 1 [1. ~        7  3694.      8 3       6 [Ven~
```

```
Se busca incluir a los gastos de equipamiento.
```

```r
bdequi <- eh22e %>% dplyr::select(folio,s08b_4) %>%
  dplyr::group_by(folio)%>% summarise(equi_hogar=sum(s08b_4,na.rm = TRUE))
aux <- left_join(personas, bdequi, by = "folio")



Se crea la base de enrenamiento y prueba.

set.seed(605)
aux<-createDataPartition(bd$pobre, p=0.7 , list = F)
bdtrain<-bd[aux,]
bdtest<-bd[-aux,]
```

Verificando existencia de datos vacios.

```r
md.pattern(aux)
```




```
##      pobre folio depto area jefe aestudio yhog tipoAct caeb_op equi_hogar
## 9309     1     1     1    1    1        1    1       1       1          1
## 1        1     1     1    1    1        1    1       1       1          1
##          0     0     0    0    0        0    0       0       0          0
##      antNeg
## 9309      1 0
## 1         0 1
##           1 1
```

Eliminando los datos

```
bd <- na.omit(aux)
bd<-bd %>% dplyr::select(-folio, -jefe) %>% to_factor()
bd<-bd %>% mutate(pobre=(pobre=="Pobre"))
```

1. Se aplicara los modelos logit/probit

```
m1<-glm(pobre ~ . , data=bdtrain, family = binomial(link="logit"))
m2<-glm(pobre ~ . , data=bdtrain, family = binomial(link="probit"))
```

2. Identificar las variables significativas

```
step(m1)
step(m2)
```

3. Construir el modelo con variables significativas (evitar colinealidad alta)

```
m3<-step(m1)
```

```
## Start:  AIC=5466.28
## pobre ~ depto + area + aestudio + yhog + antNeg + tipoAct + caeb_op +
##     equi_hogar
##
##              Df Deviance    AIC
## <none>            5396.3 5466.3
## - aestudio    1   5398.4 5466.4
## - equi_hogar  1   5408.9 5476.9
## - antNeg      1   5410.7 5478.7
## - tipoAct     1   5412.7 5480.7
## - area        1   5452.2 5520.2
## - depto       8   5471.2 5525.2
## - caeb_op    20   5516.5 5546.5
## - yhog        1   7124.5 7192.5
```

```
m4<-step(m2)
```

```
## Start:  AIC=5463.66
## pobre ~ depto + area + aestudio + yhog + antNeg + tipoAct + caeb_op +
##     equi_hogar
##
##              Df Deviance    AIC
## <none>            5393.7 5463.7
## - aestudio    1   5396.3 5464.3
## - equi_hogar  1   5404.3 5472.3
## - antNeg      1   5405.6 5473.6
## - tipoAct     1   5409.5 5477.5
## - area        1   5448.7 5516.7
## - depto       8   5472.4 5526.4
## - caeb_op    20   5508.8 5538.8
## - yhog        1   7124.0 7192.0
```

4. Predecir la clase de pertenencia en la base de test ($prob > 0.5$)

```
prob_l<-predict(m3, bdtest, type="response")
prob_p<-predict(m4, bdtest, type="response")
```

5. Comparar lo observado y lo predicho

```
bdtest<-bdtest %>% mutate(yl=(prob_l >0.5),
                  yp=(prob_p >0.5))
```

6. Generar la matriz de confusión

```
tl<-table(bdtest$pobre, bdtest$yl)
tp<-table(bdtest$pobre, bdtest$yp)
confusionMatrix(tl)
```

```
## Confusion Matrix and Statistics
##
##
##         FALSE TRUE
##   FALSE  1707  217
##   TRUE    363  505
##
##                Accuracy : 0.7923
##                  95% CI : (0.7767, 0.8072)
##     No Information Rate : 0.7414
##     P-Value [Acc > NIR] : 1.955e-10
##
##                   Kappa : 0.4917
##
##  Mcnemar's Test P-Value : 1.736e-09
##
##             Sensitivity : 0.8246
##             Specificity : 0.6994
##          Pos Pred Value : 0.8872
##          Neg Pred Value : 0.5818
##              Prevalence : 0.7414
##          Detection Rate : 0.6114
##    Detection Prevalence : 0.6891
##       Balanced Accuracy : 0.7620
##
##        'Positive' Class : FALSE
##
```

```
confusionMatrix(tp)
```

```
## Confusion Matrix and Statistics
##
##
##         FALSE TRUE
##   FALSE  1705  219
##   TRUE    363  505
##
##                Accuracy : 0.7915
##                  95% CI : (0.776, 0.8065)
##     No Information Rate : 0.7407
##     P-Value [Acc > NIR] : 2.036e-10
##
##                   Kappa : 0.4903
##
##  Mcnemar's Test P-Value : 3.075e-09
##
```

```
##             Sensitivity : 0.8245
##             Specificity : 0.6975
##          Pos Pred Value : 0.8862
##          Neg Pred Value : 0.5818
##              Prevalence : 0.7407
##          Detection Rate : 0.6107
##    Detection Prevalence : 0.6891
##       Balanced Accuracy : 0.7610
##
##        'Positive' Class : FALSE
##
```

En las tablas se pueden evidenciar que existe una diferencia nada significativa en comparacion de los modelo: - el modelo logit nos muestra una acuracidad de 0.7923. - el modelo probit nos resulta 0.7915

7. Efectos marginales

```
logitmfx(pobre ~ . , data=bdtrain)
probitmfx(pobre ~ . , data=bdtrain)
```

Conclusiones y recomendaciones

La mineria de datos puede identificar los factores que más contribuyen a la pobreza, como el nivel educativo, el acceso a compras de archivos y otros, la infraestructura, el desempleo, y la composición del hogar. Estos factores pueden variar según la región y el contexto.

Estas conclusiones permiten a los formuladores de políticas y a los actores sociales comprender mejor la naturaleza de la pobreza en una región y tomar medidas informadas para reducirla y mejorar la calidad de vida de la población afectada.

Dentro las recomendaciones estan un estudio mas detallado y la apliación de otras herramientas que permitan identificas grupos con caracteristicas similares.

Referencias

Arcelus, B. S. y. G. S., F. (2006). «The human development index adjusted for efficient resource utilization»,. *nequality, Poverty and Well-being, M. McGillivray (ed.), Nueva York, Palgrave MacMillan y United Nations University.*

Boltvinik, J. y. A. D. (2020). Medición de la pobreza de México: análisis crítico comparativo de los diferentes métodos aplicados. Recomendaciones de buenas prácticas para la medición de la pobreza en México y América Latina. *Comisión Económica para América Latina y el Caribe (CEPAL).*

Bourguignon, F. (2006). *"From income to endowments: the difficult task of expanding the income poverty paradigm* (segunda, Ed.). Stanford University Press, Cambridge University Press.

CEPAL. (2018). Medición de la pobreza por ingresos. *Comisión Económica para América Latina y el Caribe (CEPAL).*

de_la_provincia_de_Buenos_Aires, D. (2010). Métodos de medición de la pobreza: conceptos y aplicaciones en América Latina. *Entrelíneas de la política económica,.*

Duclos, J. E. y. D. R., J. (2004). Polarization: concepts, measurement and estimation". *Econometrica.*

Feres, J. C. y. X. M. (2001). Enfoques para la medición de la pobreza: breve revisión de la literatura,División de Estadística y Proyecciones Económicas, Comisión Económica para América Latina y el Caribe (CEPAL). *https://dds.cepal.org/infancia/guia-para-estimar-la-pobreza-infantil/bibliografia/capitulo-I/Feres.*

IBM. (2024). ¿Qué es la minería de datos? *https://www.ibm.com/es-es/topics/data-mining.*

INE. (2005). La pobreza y su medición: presentación de diversos métodos de obtención de medidas de pobreza [. *https://www.ine.es/daco/daco42/sociales/pobreza.pdf.*

Nolan, B. y. M. I. (2011). Economic inequality, poverty, and social exclusion". *The Oxford Handbook of Economic Inequality - Oxford, University Press.*

Ravallion, M. (2003). The debate on globalization, poverty and inequality: Why measurement matters". *International Affairs*.

Roach, J. L. y. J. K. R. (1972). Poverty: Selected Readings, Harmondsworth. *Penguin*.

Spicker, S. A. y. D. G., P. (2007). Pobreza: un glosario internacional, Buenos Aires, Consejo Latinoamericano de Ciencias Sociales. *(CLACSO), http://biblioteca.clacso.edu.ar/ar/libros/clacso/crop/glosario/glosario.pdf*.

Stezano, F. (2021). Enfoques, definicionesy estimaciones de pobrezay desigualdad en América Latinay el Caribe. *Naciones Unidas CEPAL*.

TORGO, L. (2011). *Data mining with R: Learning with case studies* (PRIMERA, Ed.). Boca Ratón, Fla.: CRC.

Townsend. (2007). «Introducción a Grupo de Expertos en Estadísticas de Pobreza», Compendio de mejores prácticas en la medición de la pobreza. *Grupo de Río, ttps://dds.cepal.org/infancia/guide-to-estimating-child-poverty/bibliografia/capituloII/Grupo*.

Zhang, X. y. R. K. (2001). What difference do polarization measures make? *Journal of Development Studies, 37*.