

Factores influyentes en la cotización de pensiones en Colombia

Abstract

Se realiza una comparación de metodologías para la predicción de ser cotizante a pensión o no, se basa principalmente en métodos de clasificación apoyados por la identificación de un modelo sugerido por la literatura y la teoría económica. Los resultados muestran que el mejor método para predecir es el lasso, pero realizando primero una técnica de submuestreo de upsampling debido a la naturaleza desbalanceada de los datos. Asimismo, variables como el tener o no contrato de trabajo, pertenecer a una organización sindical o gremial, y tener una segunda ocupación afectan positivamente el ser cotizante o no como lo sugiere la teoría económica. Por el otro lado, variables como el ingreso laboral y la cantidad de horas trabajadas a la semana se relacionan negativamente con el ser cotizante, lo cual se explica a partir del costo de oportunidad de trabajar en el sector formal. Finalmente, a partir de los resultados se recomienda el desarrollo de políticas públicas enfocadas en la contratación laboral y el fomento de cultura en pensiones que cambien la visión de costo laboral a una perspectiva de largo plazo. De esa forma, se puede modificar los incentivos de los posibles cotizantes y mejorar esta parte de la problemática.

Introducción

En Colombia actualmente hay una preocupación con las pensiones. Por un lado, el sistema pensional es un sistema regresivo que aumenta las inequidades sociales (Cárdenas, 2020). Lo anterior indica la necesidad de una reforma pensional. Sin embargo, la otra cara de la problemática tiene que ver con la cotización o no de las personas y las consecuencias que esta trae al sistema. Esto se debe a que, según el ex ministro de hacienda Mauricio Cárdenas, la cobertura pensional llega a tan solo 1 de cada 4 colombianos. Además, tan solo el 50.2% de trabajadores cotiza para pensiones (DANE, 2021). Por consiguiente, gran parte de la población termina participando del régimen no contributivo (o subsidiado), lo cual pone en jaque a la sostenibilidad financiera del sistema pensional. Entonces, a pesar de necesitar una reforma pensional que vuelva más equitativo al sistema, también se ve una clara problemática a nivel de cotizaciones de la población. Dado eso, con la necesidad de que los trabajadores aporten a los fondos, nace la siguiente pregunta: ¿Cuáles son los determinantes socioeconómicos para que una persona cotice para pensiones?

Ahora bien, a lo largo del siguiente artículo, la problemática a abordar será encaminada a analizar qué aspectos de las personas hacen posible su cotización pensional. La idea, en general, es identificar cuáles son estas variables para poder hacer recomendaciones de política pública. Hoy en día, el debate se centra en la otra cara de la moneda y asimismo es la información que se encuentra en la literatura. La mayoría de los artículos o repositorios hablan de la necesidad de una reforma, como el repositorio de Fedesarrollo Sistema pensional en Colombia ¿por qué tiene baja cobertura? ¿es regresivo e inequitativo?, pero ninguno se enfoca o es amplio en las dinámicas sociales y económicas que evitan que una persona cotice. Lo anterior bajo el supuesto de que, a pesar de una reestructuración completa del sistema, una persona puede mantener las dinámicas preexistentes a la reforma e imposibilitar su cotización. Por tanto, el trabajo a continuación profundiza esta parte de la problemática que se ha dejado de lado.

Para finalizar, los resultados encontrados se hallaron a través de un modelo predictivo upsampling Lasso, en donde con un AUC de 0.79 y con una precisión mayor al 84% se confía en la predicción. A través de este método, se concluyó que las variables de la metodología son claves para determinar si una persona toma la decisión de cotizar y afectan diferencialmente dicha decisión. De esa forma, se recomienda ampliar las políticas públicas orientadas a estas características; robusteciendo, si así sucede, la posible reforma pensional.

Datos

Como se dijo anteriormente, la problemática parte desde la situación socioeconómica y sus respectivas dinámicas de cada persona, es por eso que la base de datos escogida es la Gran Encuesta Integrada de Hogares (GEIH). En específico, se seleccionó la parte relacionada con las personas ocupadas en el mercado laboral, debido que son estos los que deberían cotizar pensiones. Esta base de datos cuenta con 165 variables de las cuales se escogieron 7 respectivas a la situación laboral y económica de las personas. Estas fueron: el tipo de contrato laboral, el tiempo que lleva en el actual trabajo y el tipo de trabajo (si es estacional, permanente u otro), dado que si se trata de empleo informal o formal debe existir una repercusión en la forma y en la cantidad de cotización dada la estabilidad laboral; las horas trabajadas a la semana, la existencia de una segunda ocupación, la afiliación a una asociación gremial o sindical y los ingresos laborales; explicando la cantidad de recursos de la persona y sus restricciones salariales correspondientes; y, por último, la variable de interés que enmarca si la persona efectivamente cotiza o no. Por último, vale aclarar que se borraron las observaciones referentes a las personas ya pensionadas, dado que estas, se sobreentiende, no son relevantes para la investigación.

Asimismo, la base final consta de 5058 observaciones y un total de 7 variables relevantes. Con respecto a la variable de interés, la variable de si la persona cotiza o no a la pensión tiene una

media de 0.84 que hace referencia a que el 84% de la base cotiza, el comportamiento de la variable se puede ver en el histograma que se encuentra en el compilado de las estadísticas descriptivas que se encuentra en la sección de anexos. En cuanto a las variables de predicción, el comportamiento de las horas de trabajo muestra que se trabajan casi 50 horas. Se observa que los cotizantes están en un rango de ingresos laborales mayores que los no cotizantes, fuera de una observación atípica de los no cotizantes que devenga un ingreso mucho mayor que cualquier otra observación y es no cotizante. De la misma manera, se observa que las personas que pertenecen a un sindicato o gremio tienen un rango mayor de ingresos laborales en relación a los que no pertenecen a este tipo de organizaciones.

Modelo

La metodología que se utilizó para el desarrollo del problema se enfoca en los modelos de clasificación, buscando encontrar el método que ofrezca los mejores resultados en términos de precisión predictiva, así como en poder de la predicción. Por tanto, se realizaron 7 modelos diferentes con el propósito de compararlos y así escoger aquel que realice las predicciones de forma correcta (o más correcta), los modelos son los siguientes: lasso-logit, ridge-reg, logit-reg, upsampling para lasso, upsampling para ridge, downsampling para lasso y downsampling para ridge.

Las variables que se utilizan para los modelos de clasificación son las mismas que se presentaron anteriormente, estas tienen como unidad de análisis las personas entrevistadas por la GEIH para el mes de noviembre del 2021. Retomando las variables, estas se escogen a partir del sustento teórico que relaciona el ser cotizante a pensión con las características del trabajo que se desempeña. Lo anterior se correlaciona mucho con el hecho de que una persona sea o no cotizante, y, por consiguiente, funcionan como predictores importantes para este trabajo.

Cabe aclarar que el criterio principal por el cual se escoge un modelo por encima de otro es el AUC, el cual da indicios acerca que la capacidad predictiva del modelo y, al ser solo un número, ofrece sencillez al momento de comparar distintas metodologías.

Resultados

Continuando, al momento de correr los modelos se observa que aquel que realiza una mejor predicción debido a su desempeño es el modelo de upsampling para lasso, este tiene un AUC de 0.796 (gráfico 1 es la curva de ROC correspondiente a este modelo), una precisión de 0.849 y una tasa de 11.1% de falsos negativos. Al momento de entender por qué este modelo presento los mejores resultados, se considera que un factor importante es que dentro de la base de datos al observar cuantas observaciones efectivamente cotizan o no a pensiones, se ve un claro desbalance en los datos ya que el porcentaje que si cotiza es bastante mayor al que no cotiza, causando desbalance dentro de los datos y esto afectando directamente a la predicción. Por esta razón, tiene sentido que al corregir esta situación por medio del upsampling se lleguen a mejores resultados que haciendo la predicción con desbalance. Asimismo, tiene sentido que los resultados de los modelos que corrigen la situación de desbalance primero tienen un AUC mayor que los modelos que no lo corrigen, que en este caso da como mejor modelo al dicho anteriormente.

Finalmente, se opta por mostrar los AUC de todos los modelos para demostrar que el modelo de upsampling lasso es el que presenta mejores resultados. Los AUC son los siguientes: upsampling para lasso (0.796), downsampling para lasso (0.783), upsampling para ridge (0.77), downsampling para ridge (0.767), lasso-logit (0.7214), logit-reg (0.721) y, por último, ridge-reg (0.58).

En cuanto a las variables, en el grafico 2 y en la tabla 1 se muestran los coeficientes de cada variable para la predicción, esto permite ver cómo afecta cada variable a la probabilidad de ser

cotizante a pensión y, por lo tanto, en la predicción. Se observa que todas las variables afectan el modelo excepto por la variable categórica que corresponde a si el trabajo es permanente, estacional u otro (P6780), de resto todas las variables influyen en el hecho de ser cotizante a pensiones o no. Tanto la variable de tener contrato (P6440) como la de estar afiliado a una asociación gremial o sindical (P7180) son las que más influyen en ser cotizante, relacionándose positivamente con la variable dependiente. Esto se debe a que, la primera, tener un contrato de trabajo, es garantía de cotización por la estabilidad laboral y la segunda, ser afiliado a estas organizaciones, genera presiones contantes y poder de mercado que afianza la seguridad laboral; es decir, fortalece la cotización. En suma, otra variable que se comporta de la misma forma es si la semana pasada la persona tuvo segundo trabajo u ocupación, pero esta tiene un coeficiente mucho menor por lo que no parece existir mayor relevancia; no obstante, se cree que esto se debe a que un segundo empleo puede aumentar cotizaciones por medio de la posibilidad de formalización de este. Por último, la variable de horas trabajadas a la semana y la de ingreso laboral se relacionan negativamente con ser cotizante o no, esto puede reflejar el costo de oportunidad de estar en la parte formal del mercado laboral. Es decir, el hecho de ser informal implica la no obligatoriedad de cumplir con normas regulatorias y seguridad social, por lo que todo el ingreso va directamente al trabajador, haciendo que el mayor ingreso laboral se correlacione con el no cotizar.

En pocas palabras, los resultados son concluyentes con respecto a la predicción de la cotización. Efectivamente se encuentran 3 variables con relación positiva y otras 2 de forma negativa. Adicionalmente, con la comparación de modelos, precisamente hay confianza en los resultados y se puede encaminar la investigación. Por lo tanto, bajo estas variables se puede concluir y recomendar políticas públicas enmarcadas en la problemática del sistema pensionala.

Conclusiones y recomendaciones

En Colombia hay un claro problema con el sistema pensional que se divide en 2 partes; la estructuración del sistema y la falta de cotización de los trabajadores. Dado eso, a pesar de que sean problemas intrínsecos del sistema que se correlacionan entre sí, la solución única no puede ser una reforma pensional a la estructura. Si bien esta soluciona problemas de forma, no tiene en cuenta al comportamiento del cotizante. Por tanto, se realizó una metodología predictiva que pudiera determinar cómo variables claves determinan la decisión de cotización.

Los resultados de la metodología brindan una confianza alta al tener un AUC de 0.796 y una precisión del 84.9%. Por medio de este, se permitió entender que existe una relación positiva entre la decisión de cotización y variables como el tipo de contrato del trabajador, su afiliación a grupos gremiales o sindicatos o, en menor cuantía, la existencia de un segundo empleo. Adicionalmente, se encontró una relación negativa entre la variable de interés y las horas trabajadas conjuntamente al ingreso laboral; entendiendo un costo de oportunidad entre tener mayor ingreso de forma informal o aumentar los costos laborales por medio de la formalización.

Dado lo anterior, se abre la recomendación, en primer lugar, de empezar a desarrollar espacios de discusión que se centren en esta parte de la problemática, tanto en espacios de debate legislativo como en argumentación académica que permita robustecer estos resultados preliminares. En suma, es recomendable el desarrollo de políticas públicas que faciliten tipos de contratación más estables que realmente generen incentivos a la cotización. Por último, dado los resultados de ingresos laborales, también se recomienda la adopción de campañas educativas que cambien la perspectiva de la seguridad social, de tal forma que no se vea cómo un aumento de costos sino como una visión de largo plazo.

Daniel Ráquira 201914059
Santiago Becerra 201911587

Datos y habilitación de código

Los datos y el código utilizado se encuentran disponibles en el siguiente repositorio:

https://github.com/danielraquira/Final_BD.git

Referencias

Cárdenas, M (2006). Introducción a la economía colombiana. Capítulo XX.

DANE (2022). Mercado Laboral. Recuperado de:

<https://www.dane.gov.co/index.php/estadisticas-por-tema/mercado-laboral>

Expansión (s.f). Pirámide poblacional de Colombia: Recuperado de:

<https://datosmacro.expansion.com/demografia/estructura-poblacion/colombia>

Fedesarrollo (2022). Sistema pensional en Colombia: ¿por qué tiene baja cobertura?, ¿es regresivo e inequitativo? Recuperado de:

<https://www.fedesarrollo.org.co/sites/default/files/enlosmediosimpreso/actualicesecom03junio2022.pdf>

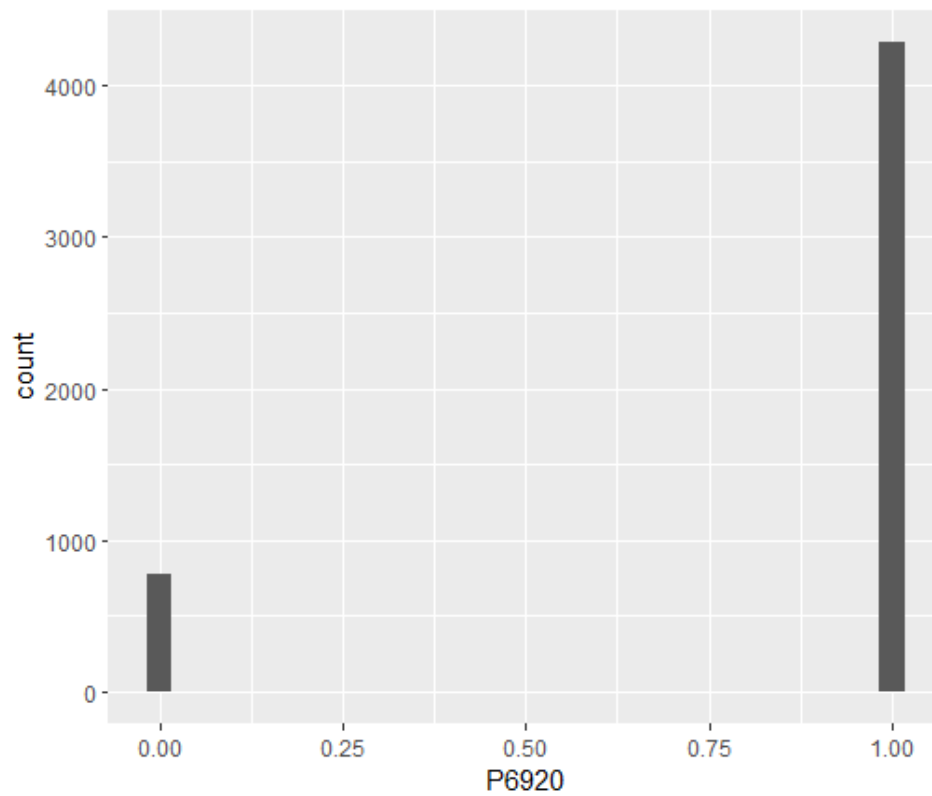
Semana (2021). Solo el 50 % de los trabajadores colombianos cotiza a pensión. Fuente: DANE.

Recuperado de: <https://www.semana.com/economia/macroeconomia/articulo/solo-el-50-de-los-trabajadores-colombianos-cotiza-a-pension/202145/>

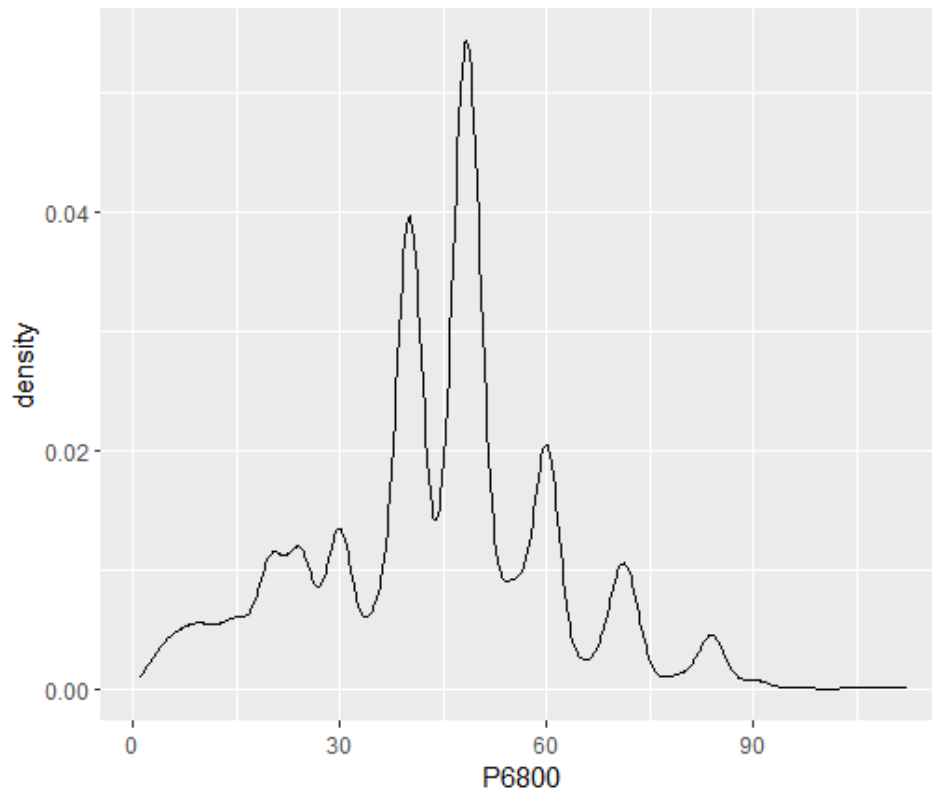
Anexos

Copilado de estadísticas descriptivas:

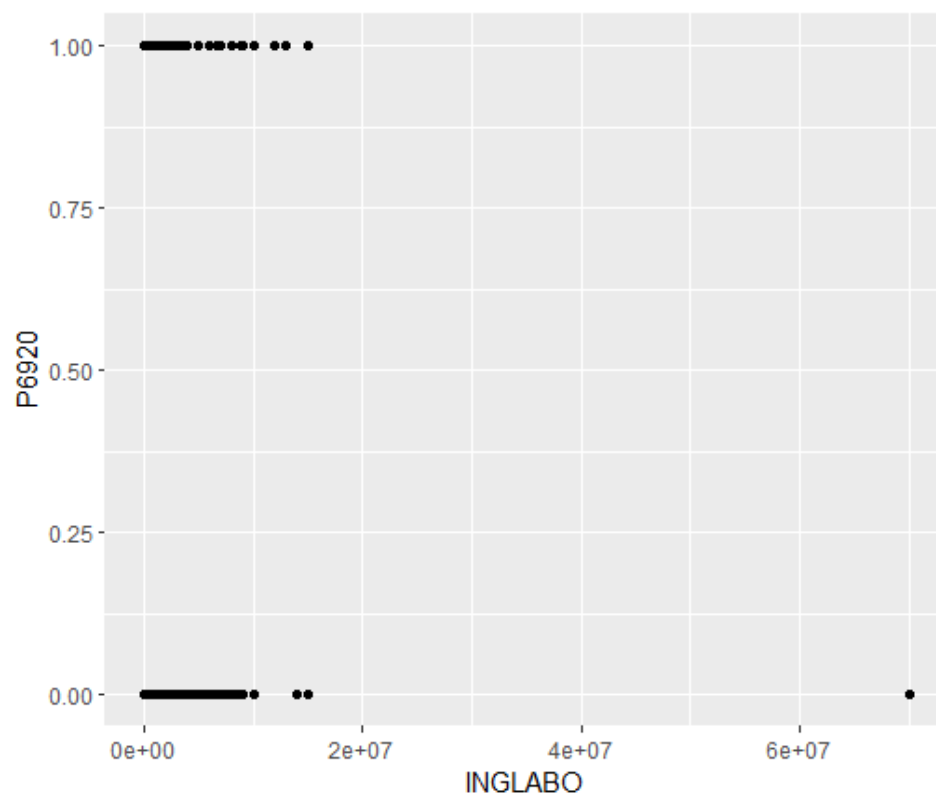
- *Comportamiento de la variable cotizante:*



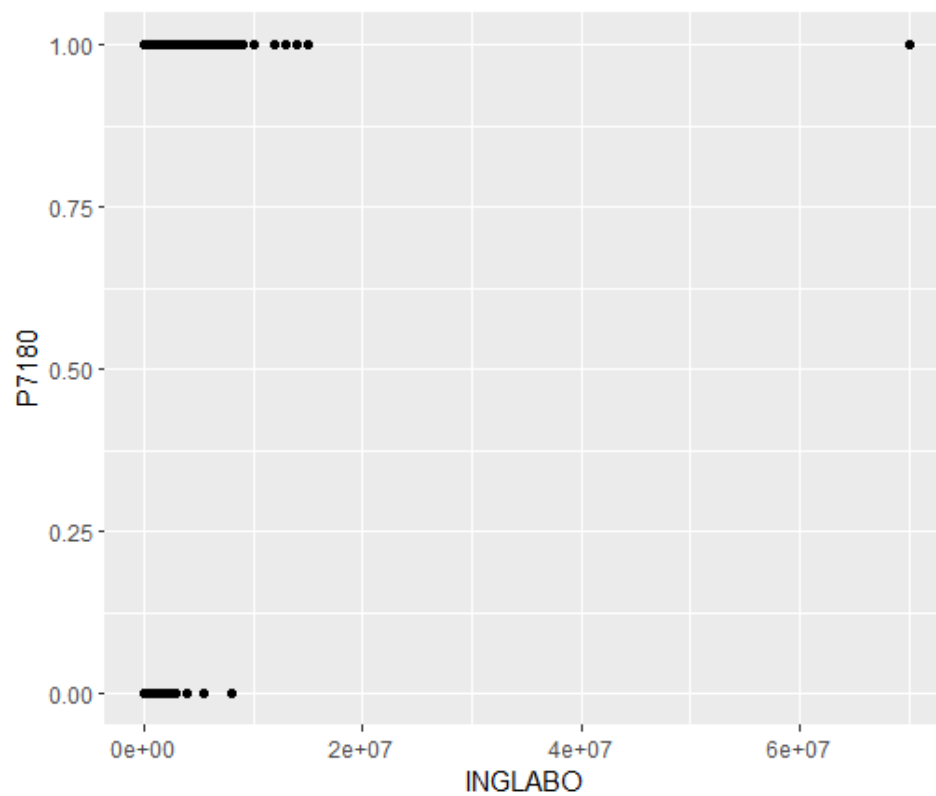
- *Comportamiento de las horas trabajadas en la última semana:*



- *Relación ingreso laboral y si es cotizante o no:*



- *Relación si pertenece a un sindicato o gremio e ingreso laboral:*



Gráfica 1. Curva de ROC para upsampling lasso.

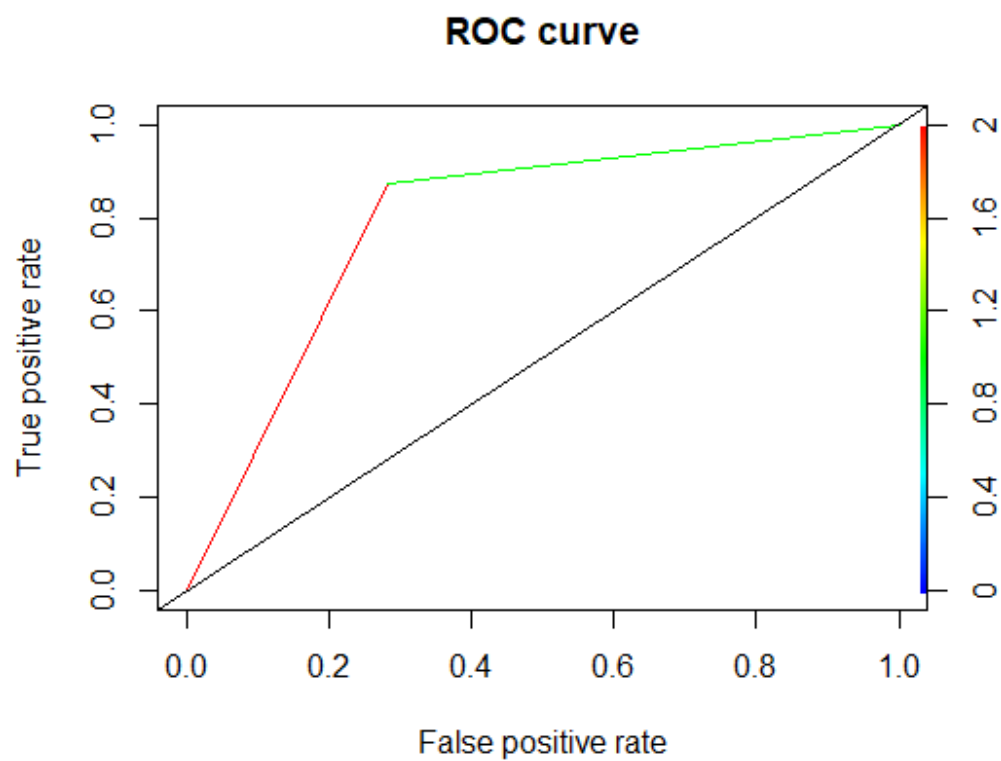


Gráfico 2. Coeficientes del mejor modelo.

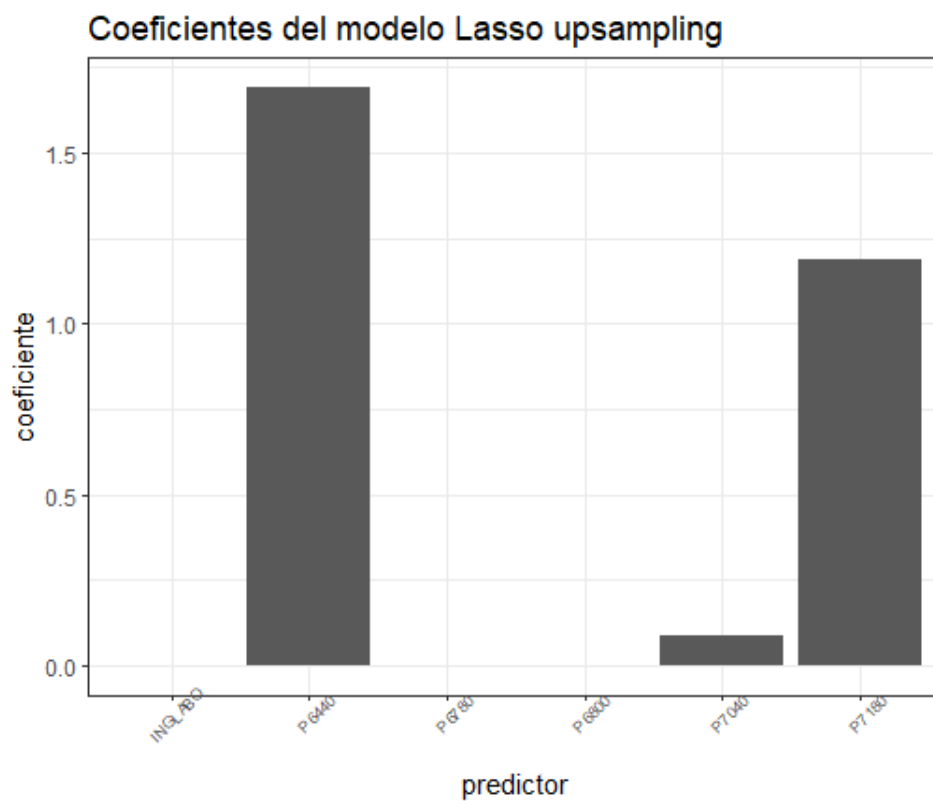


Tabla 1. Coeficientes del mejor modelo.

Predictor	Coeficiente
1 (Intercept)	-3.87
2 (Intercept)	0
P6440	1.69
P6780	0
P6800	-0.00160
P7040	0.0894
P7180	1.19
INGLABO	-0.00000142