

Métodos de aprendizaje automático en los estudios prospectivos desde un ejemplo de la financiación de la innovación en Colombia

Machine learning methods in prospective studies after an example of financing innovation in Colombia

Ana Milena Padilla-Ospina¹
Javier Enrique Medina-Vásquez²
Javier Humberto Ospina-Holguín³

Recibido: octubre 15 de 2019

Aceptado: abril 03 de 2020

Resumen

El propósito de este artículo es hacer una breve introducción a cinco métodos avanzados de predicción de aprendizaje automático, que pueden ser de utilidad para el desarrollo de estudios prospectivos: la regresión logística, las máquinas de vectores de soporte, las máquinas de gradiente potenciado, los bosques aleatorios y las redes neuronales. Además, se explica qué metodología se puede llevar a cabo para asegurar la robustez y validar dichos modelos de predicción. A manera de ejemplo, se presenta cómo el uso de estos métodos permitió identificar las variables financieras más importantes para predecir el desarrollo de actividades de innovación en pymes colombianas. Los resultados del uso de estos métodos pueden permitir la generación de pronósticos al corto y mediano plazo, que sirvan para la realización de estudios prospectivos con métodos más amplios, como la construcción de escenarios, con el propósito de generar propuestas basadas en la evidencia, como hoja de ruta para la planeación y política pública al largo plazo.

Palabras clave: regresión logística, máquinas de vectores de soporte, máquinas de gradiente potenciado, bosques aleatorios, redes neuronales.

Abstract

The purpose of this article is to make a brief introduction to five advanced machine learning prediction methods which may be useful for the development of prospective studies: logistic regression, support vector machines, gradient powered machines, random forests and neural networks. In addition, it is explained what methodology can be carried out to ensure robustness and validate these prediction models. As an example, it is presented how the use of these methods allowed to identify the most important financial variables to predict the development of innovation activities in Colombian SMEs. The results of the use of these methods may allow generating short and medium-term forecasts that serve to facilitate prospective studies with broader methods, such as the construction of scenarios, with the purpose of generating evidence-based proposals as a roadmap for long-term planning and public policy.

Keywords: logistic regression, support vector machines, gradient powered machines, random forests, neuronal networks.

¹ Administradora de Empresas, Doctora en Administración, Universidad del Valle, Cali, Colombia. E-mail: ana.milena.padilla@correounivalle.edu.co. ORCID: 0000-0003-3859-8741

² Psicólogo, Doctor en Ciencias Sociales, Universidad del Valle, Cali, Colombia. E-mail: jmedinavasquez@hotmail.com

³ Físico, Doctor en Administración, Universidad del Valle, Cali, Colombia. E-mail: javier.ospina@correounivalle.edu.co. ORCID: 0000-0002-0103-3280

1. Introducción

La prospectiva ha sido una herramienta efectiva para la planeación estratégica y la generación de escenarios en el desarrollo de políticas públicas en un ámbito nacional; sin embargo, para que una política pública sea efectiva, se deben realizar estudios previos en el país de contexto, con el fin de establecer cuál es la mejor estrategia para lograr las metas y resultados esperados.

Ya el Premio Nobel Herbert Simon (1976), subrayaba la importancia de la racionalidad sustantiva relacionada con los resultados óptimos de una política (en contraste con la racionalidad procedimental relacionada con quién lleva a cabo la política y cómo). La racionalidad sustantiva está íntimamente relacionada con el conocimiento científico, como lo explica Andrews:

(...) El conocimiento producido científicamente contribuye a la racionalidad sustantiva. Tal conocimiento describe los fenómenos y explica los factores causales. Proporciona la base fáctica para tomar mejores decisiones. Satisface criterios científicos de validez y fiabilidad, y justifica afirmaciones autorizadas de conocimiento. El conocimiento científico es un insumo clave que contribuye al 'mejor' resultado. (Andrews, 2007)

El análisis racional imbuido en el conocimiento científico, es: lógico, válido, confiable, y empíricamente probado, además de estar basado en hechos (Andrews, 2007). Así, la información de alta calidad y la evidencia son partes clave de la realización de buenas políticas públicas, de modo que las políticas modernas exigen que los gobiernos hagan uso de la evidencia disponible en la formulación y ejecución de sus políticas (Geurts, 2011). Se ha sugerido incluso que un rasgo deseable de las políticas del siglo XXI es que estén basadas en la evidencia (Geurts, 2011).

Una manera de incorporar tanto la evidencia como la racionalidad al proceso de decisión de la política

pública, es mediante los modelos cuantitativos. El análisis de política frecuentemente implica el uso de métodos cuantitativos sofisticados para examinar cómo los problemas de política son impactados por numerosas variables, tanto las debidas a las intervenciones de la política como a los factores contextuales (Yang, 2007). Los métodos cuantitativos ayudan a establecer una relación causal generalizable entre el diseño y el resultado de la política, permiten evaluar de antemano la magnitud de los efectos de la política, y permiten encontrar mejores alternativas de política (Yang, 2007). Publicaciones tan importantes como el *Journal of Policy Analysis and Management* o el *Review of Policy Research*, están colmadas de estudios de políticas que utilizan análisis cuantitativos avanzados (Yang, 2007). Además, diversos métodos cuantitativos para explicar y predecir las políticas públicas hacen parte de los métodos de investigación sancionados en la práctica de la administración pública (Yang & Miller, 2008).

Entre las diversas técnicas cuantitativas, la modelación es una de las más utilizadas en estudios de políticas (Yang, 2007). Las políticas basadas en modelos son preferibles para formular una política basada en la evidencia, ya que es muy difícil simular, analizar y evaluar políticas que no estén basadas en modelos (Geurts, 2011). Geurts (2011), afirma que: "el proceso de hacer políticas puede mejorarse considerablemente al modelar los candidatos de política y examinar los modelos resultantes", lo que: "hace posible integrar técnicas de modelación formal en las etapas tempranas de la decisión". Geurts (2011), alega que: "[el] ambiente dirigido por modelos conduce a mejores políticas, al permitir examinar la consistencia y análisis (de escenarios) 'y si'", al permitir "identificar los posibles efectos de las políticas propuestas" y "al permitir la comparación del resultado realizado con los efectos esperados".

La modelación puede incorporarse fácilmente al esquema racional-comprensivo de las políticas públicas (Frey, 2011). En este esquema, los políticos sociales se aproximan a la realización de

políticas, como a un proceso racional donde se considera toda la información sobre el problema de la política, y todas las opciones de política, para luego seleccionar las opciones que mejor satisfacen la meta de la política (Cochran, Mayer, Carr & Cayer, 2009). En lugar de ser el “producto de un proceso político caótico y aleatorio”, las políticas se suponen con una lógica intrínseca (Schneider & Ingram, 1997, citados por Sydney, 2007). Existe incluso un modelo tradicional de etapas para el proceso de las políticas públicas, el cual involucra la identificación del problema, el desarrollo de alternativas, la evaluación de las alternativas, la selección de la solución, la implementación de la solución y la evaluación del impacto de la solución implementada (Wittmer & McGowan, 2007). Procesos de políticas públicas similares, han sido descritos en la literatura por analistas como: Bardach, MacRae y Wilde, Mintzberg, Miser y Quade, Patton y Sawicki, Segal y Brzuzy, Stokey y Zeckhauser, el Urban Institute o Weimer y Vining (Frey, 2011).

En este trabajo se proponen varios modelos cuantitativos que pueden emplearse en la segunda, tercera y cuarta etapa del proceso de políticas públicas, en un ejemplo particular asociado al proyecto de cómo desarrollar la innovación mediante la intervención en variables relacionadas a la financiación en las pymes de una nación en desarrollo, como Colombia. Mediante la modelación aquí propuesta, es posible considerar un gran número de variables financieras identificadas en la literatura, sobre las que la política podría intentar incidir para ayudar al desarrollo de alternativas. Igualmente, se puede evaluar qué impacto real tiene cada alternativa y predecir de mejor manera la generación de la innovación de las pymes.

2. Marco teórico y metodología

Identificar las variables financieras que más influyen en el desarrollo de la innovación en las pymes colombianas corresponde a un problema de clasificación, debido a que se debe realizar una predicción de una variable dependiente

dicotómica (innova o no innova), a partir de un conjunto de predictores. Para lograr un modelo de clasificación adecuado, se compararon un modelo de clasificación lineal (regresión logística), un modelo de clasificación no lineal (máquinas de vectores de soporte), dos modelos de clasificación basados en reglas y árboles (bosques aleatorios y máquinas de potenciación del gradiente), y un modelo no lineal de clasificación (redes neuronales). A continuación, se realiza una breve descripción de cada uno de los métodos.

2.1 Modelo lineal de clasificación: regresión logística

El modelo de regresión logística, es uno de los métodos más básicos para la solución de problemas de clasificación (Kuhn & Johnson, 2013). La regresión logística es un modelo con un intercepto y parámetros de pendiente para los términos del modelo, tal como en la regresión lineal; sin embargo, la regresión lineal convencional, no aplica para modelar la probabilidad de innovación ya que si se quiere predecir la probabilidad de que una empresa innove, la probabilidad del evento en un modelo lineal convencional no estaría necesariamente entre 0 y 1. Para superar esta limitación, la regresión logística modela linealmente no la probabilidad P , sino el logaritmo de la razón $\frac{P}{1-P}$ (razón llamada odds en inglés) (Kuhn & Johnson, 2013).

2.2 Modelo no lineal de clasificación: máquinas de vectores de soporte

Una máquina de vectores de soporte clasifica un conjunto de datos mediante hiperplanos. Un hiperplano es la generalización de un plano a varias dimensiones (Ho, 2012; Kuhn & Johnson, 2013; Vapnik, 2000). La idea que se persigue al usar la máquina de vectores de soporte, es la de hallar un hiperplano que divida el conjunto total de empresas entre las empresas que innovan y las que no innovan. Para escoger un hiperplano particular,

la máquina de vectores de soporte intenta hallar el hiperplano que maximiza el “margen”, el cual es, en términos laxos, la distancia del hiperplano a ciertas empresas claves del conjunto, los así llamados “vectores de soporte”.

Sin embargo, cuando la frontera entre los datos es fundamentalmente no lineal, no es posible separar a las empresas innovadoras de las no innovadoras con un hiperplano (que es lineal). En este caso, la máquina de vectores de soporte añade más dimensiones al problema. En un espacio con más dimensiones, sí hay hiperplanos que separan los datos en sus respectivas categorías. La optimización necesaria para hallar la mejor máquina de vectores de soporte se convierte entonces en la de un producto de vectores de los puntos transformados y es equivalente a realizar una transformación no lineal en el espacio original (el espacio sin añadir dimensiones). (A esta transformación no lineal se le denomina función de kernel).

2.3 Modelos de árboles y reglas: bosques aleatorios

El “bosque aleatorio”, es un conjunto de árboles de decisión que combina las predicciones de distintos árboles y puede utilizarse como método de clasificación. El bosque aleatorio utiliza una variación del “boosting”, llamada “boosting aggregation” o “bagging” (Gu et al., 2018). En esta técnica primero se elaboran árboles de base, construidos cada uno con una muestra bootstrap aleatoria de los datos y se escoge la moda de las clases predichas como pronóstico.

Los árboles resultantes tienden a ser profundos, i.e., de múltiples niveles, y tienden a sobreajustar los datos, es decir, a ser poco generalizables, pero el procedimiento de hallar la moda de los pronósticos de cada uno de los árboles de decisión generados mejora el rendimiento de la predicción (cf. Gu et al., 2018). De hecho, el error de generalización de un bosque converge al aumentar el número de árboles y depende, a su vez, de cada árbol de decisión y de la correlación existente entre estos (Breiman, 2001).

En principio, los árboles que resultan del procedimiento anterior suelen ser correlacionados, lo que empeora el pronóstico general, pero una técnica denominada “dropout” permite descorrelacionar los árboles al considerar solo un subconjunto de los predictores que han sido establecidos aleatoriamente al momento de dividir una rama potencial del árbol (Gu et al., 2018). Esta técnica permite la reducción de la correlación promedio entre las predicciones y mejora el comportamiento de la predicción global en comparación con el “bagging” estándar. Para modelar un bosque aleatorio, se suelen optimizar parámetros de ajuste como la profundidad de los árboles y el número de muestras “bootstrap” mediante la validación cruzada; esto es, usando una muestra de validación no vista previamente para garantizar la generalización del método (cf. Gu et al., 2018).

2.4 Modelo no lineal de árboles y reglas: máquina de gradiente potenciado

Una “máquina de gradiente potenciado” o “gradient boosting machine”, es una técnica en que se parte de un árbol extremadamente poco profundo, que se espera tenga un gran sesgo (diferencia entre el valor esperado y realizado), y que sea un predictor muy débil (cf. Gu et al., 2018). Posteriormente, se entrena un segundo árbol simple de la misma profundidad pero que intenta predecir los residuos de la predicción del primer árbol.

Para los modelos de clasificación, cada hoja del nodo indica el valor de clase. Como no todas las observaciones llegan a ser certeramente clasificadas, se le da menor peso a aquellas observaciones que fueron clasificadas correctamente y mayor a las que no lo fueron, lo que tiene como consecuencia que se incremente el peso para las clasificaciones incorrectas y se reduzca para las clasificaciones correctas. Luego, el proceso de generar un nuevo árbol con los residuos del árbol anterior se repite y los pesos de las observaciones se ajustan hasta que no haya mejora en el error,

es decir, hasta que no se logre un error menor (Malakar, 2018).

Una vez no mejore el error, se deberán combinar todos los árboles para generar la predicción final. Del promedio de los pesos de los diferentes árboles de predicción, se obtiene el valor de clase final. Usualmente los pesos están inversamente proporcional a la tasa de error, así, a menor error mayor peso y a mayor error, menor peso.

2.5 Modelos no lineales de clasificación: redes neuronales

Una red neuronal intenta emular grosso modo una red de neuronas del cerebro. En una red neuronal, las “neuronas” están organizadas en capas (Ho, 2012; Kuhn & Johnson, 2013). La predicción final está modelada por variables intermedias (las variables o unidades ocultas), es decir, neuronas que sirven de intermediarias. Puede haber una o múltiples capas intermedias. Típicamente, cada neurona de una capa está conectada hacia adelante con cada neurona de la capa que le sigue, pero con ninguna otra neurona más. Cada unidad oculta arroja una combinación lineal de algunos o de todos los predictores; sin embargo, esta combinación lineal está transformada al llegar al resultado final, comúnmente por una función logística.

El aprendizaje de la red neuronal ocurre mediante un mecanismo de retroalimentación hacia adelante, donde el error de la salida de los datos de entrenamiento se usa para ajustar los pesos de las neuronas (los datos de entrenamiento son los datos de las empresas originales que se clasifican). Este proceso de ajuste se propaga hacia las capas anteriores, por lo que se habla de “propagación hacia atrás” (back-propagation).

2.6 Medidas de adecuación del modelo

En el ejemplo que se presenta, se toma como atributo positivo el que la empresa examinada innove. Para cuantificar la bondad o adecuación

del modelo de clasificación alcanzado, se utilizan varias medidas en los datos fuera de la muestra, es decir, en la muestra de prueba (de 82 nuevas empresas). Las principales medidas utilizadas, fueron: la exactitud, el coeficiente kappa de Cohen, y el área bajo la curva ROC.

La exactitud se refiere al porcentaje de empresas correctamente predichas (Manning, Raghavan & Schütze, 2010; Olson & Delen, 2008; Sokolova, Japkowicz, & Szpakowicz, 2006). En un escenario de predicción cada vez más perfecto, el valor de la exactitud tiende a uno (Olson & Delen, 2008), siendo cero el peor valor posible. La exactitud es el criterio general de adecuación utilizado en este trabajo para cada modelo de clasificación, los demás criterios son complementarios.

El coeficiente kappa de Cohen tiene en cuenta si el ajuste entre las predicciones y la realidad se debe al azar o no (Carletta, 1996). El coeficiente compara una clasificación de base con la clasificación realizada (Manning et al., 2010). El coeficiente kappa de Cohen es un buen indicador de ajuste en muestras desbalanceadas. Aunque sin un sustento empírico, ni teórico, Landis y Koch (1977) clasifican los valores del coeficiente kappa de Cohen menores a cero como pobres (peores que la clasificación ingenua al azar) y aquellos en los intervalos [0-0.20] como leves, [0.21-0.40] como justos, [0.41-0.60] como moderados, [0.61-0.80] como sustanciales y [0.81-1] como acuerdos casi perfectos (Carletta, 1996).

El área bajo la curva ROC, también se conoce como exactitud balanceada (Sokolova et al., 2006) y “es equivalente a la probabilidad de que el clasificador vaya a ranquear una instancia positiva [empresa innovadora] elegida al azar más alto que una instancia negativa [empresa no innovadora] elegida al azar” (Fawcett, 2004). Claramente el valor de esta área se encuentra entre 0 y 1, siendo el caso del modelo de clasificación con un valor de 1 el que corresponde a una predicción perfecta (Olson & Delen, 2008).

2.7 Metodología

Para identificar las variables financieras más importantes para la innovación de las pymes colombianas, se trabajó con un marco muestral compuesto por la información contable de las pymes de la base de datos EMIS (2017) durante el año 2016. Para identificar el tamaño de la empresa, se tuvieron en cuenta los parámetros establecidos en el artículo 2 de la Ley 905 de 2004 sobre el valor de los activos (Ley 905, 2004), utilizando el valor del salario mínimo mensual legal vigente del año 2016 (Banco de la República de Colombia, 2017).

A través de cuestionarios, se recolectó la información relacionada con las variables financieras y de innovación. Se utilizó un muestreo estratificado separado para las empresas que innovaban y para las empresas que no innovaban. En este muestreo, cada estrato reflejó la proporción aproximada de las pymes a encuestar por sector, tamaño y ciudad. El entrenamiento y la validación de los modelos de predicción, se hizo con los datos recolectados con la muestra estratificada (que constó de 157 cuestionarios completos). Adicionalmente, se recolectaron 82 cuestionarios por un muestreo de conveniencia para conformar la muestra de prueba, es decir, para verificar la generalización de los resultados con datos por fuera de la muestra utilizada en la construcción de los modelos.

Antes de construir los modelos de predicción, fue necesario construir las variables independientes y dependiente que se utilizaron, lo que se logró mediante la ingeniería de rasgos. Se utilizaron técnicas como el análisis de componentes principales, el promedio y el reajuste, para lograr rasgos predictivos óptimos. Una vez establecidas las variables o rasgos, se desarrollaron los modelos de predicción. El proceso de desarrollar cada modelo de predicción de clasificación conllevó varios pasos. En primer lugar, en la construcción de cada modelo se consideraron varios tipos de modelos como candidatos. En este estudio, se calcularon modelos de predicción de regresión logística, de máquinas de vectores de soporte, de redes

neuronales, de bosques aleatorios y de máquinas de gradiente potenciado.

Establecidos los tipos de modelos candidatos a conformar el modelo de predicción óptimo, se utilizó la optimización bayesiana (Bayesian optimization) para hacer competir estos modelos. La optimización bayesiana es un tipo de optimización global basada en la inferencia bayesiana, que utiliza procesos gaussianos para intentar encontrar el valor máximo de una función desconocida, en el menor número de iteraciones posible. La función a optimizar en cada modelo fue la exactitud.

La generalización se potenció por medio de la validación cruzada, es decir, se verificó que la optimización del parámetro de adecuación del modelo (i.e., la exactitud) fuera válido en un subconjunto de los datos originales, llamado datos de validación. La validación cruzada evita el sobreajuste de los modelos de predicción, fenómeno según el cual se logran modelos casi perfectos dentro de la muestra, pero de muy pobre desempeño en datos que el modelo no ha visto antes. Los resultados del parámetro de adecuación reportados en este trabajo, son, sin embargo, los que corresponden a los datos de prueba, datos que no han sido vistos previamente durante la construcción del modelo, ni durante su entrenamiento, ni durante su validación.

Es importante describir cómo cada tipo de modelo depende de hiperparámetros que también se deben optimizar, pues impactan el parámetro principal de adecuación del modelo. Los hiperparámetros son parámetros del modelo que determinan cómo funciona el tipo de modelo particular. Por ejemplo, en un modelo de red neuronal, el número de conexiones (llamadas también “parámetros” coincidentalmente), la función de activación (logística, rampa o tangente hiperbólica) o el tipo de red (convolucional o prealimentada “feedforward”) corresponden a hiperparámetros cuyas combinaciones definen cada una un tipo de modelo distinto, que puede generar un parámetro de adecuación general diferente (mejor o peor).

La optimización bayesiana utilizada, optimizó el parámetro de adecuación general del modelo alterando tanto el tipo de modelo como los conjuntos de hiperparámetros que definen cada modelo particular. La técnica bayesiana es particularmente adecuada para este tipo de optimización de una función de alto costo de evaluación, como la descrita, donde se busca el equilibrio entre la exploración (evaluar la mayor cantidad de modelos), y la explotación (hallar los mejores modelos).

Una vez seleccionado el tipo de modelo más adecuado y sus hiperparámetros, de acuerdo con la optimización bayesiana, se realiza una exploración de otros conjuntos de hiperparámetros vecinos para lograr un conjunto de mejores modelos del tipo seleccionado. Finalmente, se combinan estos modelos pre-óptimos, siguiendo la amplia evidencia que existe de que la combinación de modelos de predicción usualmente supera a los mejores modelos de predicción individuales.

Por esta razón, se desarrolló un único modelo combinado a partir de nueve modelos originales de redes neuronales. Este modelo combinado utiliza como probabilidad de innovación el promedio de las probabilidades que arroja cada uno de los nueve modelos pre-óptimos. El modelo combinado, resultado de la combinación de pronósticos o predicciones, mejoró sustancialmente la calidad de la adecuación de los modelos originales, tal y como se esperaba.

Para el mejor modelo de clasificación (redes neuronales), se combinaron nueve modelos pre-óptimos con una cantidad de conexiones entre 90 y 130, en intervalos de cinco en cinco parámetros. Para cada uno de los nueve modelos se calcularon las distintas medidas de adecuación, además de la principal, la exactitud en los datos de prueba. Se examinaron así en los datos de prueba también el coeficiente kappa de Cohen y el área bajo la curva ROC. Además, se pudo hallar la importancia de las variables del modelo resultante al combinar todos los nueve modelos.

La importancia de las variables corresponde a una medida que permite la jerarquización final de las variables independientes o rasgos predictivos, del más relevante al menos relevante, para lograr la adecuación óptima del modelo. El objetivo principal del trabajo de ejemplo era, en definitiva, averiguar, mediante el cálculo de la importancia de las variables, cuáles son las variables más importantes para determinar si una empresa innova o no. En este contexto, una variable es más importante que otra, cuando al permutar al azar los valores de esta variable entre las distintas empresas, el ajuste del modelo (la exactitud) decae más que al permutar los valores de otra variable entre las distintas empresas.

3. Resultados y discusión

3.1 Resultados

Tras utilizar la optimización bayesiana con los cinco tipos de modelos, a saber: la regresión logística, las máquinas de vectores de soporte, las máquinas de gradiente potenciado, los árboles aleatorios y las redes neuronales, fue claro que, para tiempos razonables medios de entrenamiento, las redes neuronales lograban modelos con una exactitud mayor (y con un área bajo la curva ROC (AUC) mejor). A partir de esta conclusión, se decidió optimizar posteriormente mediante optimización bayesiana solo los hiperparámetros de los modelos de redes neuronales.

Después de experimentos preliminares, se realizó una búsqueda grid o de rejilla del número de parámetros (o conexiones neuronales), yendo desde una red neuronal de 90 parámetros a una de 130 parámetros, en intervalos de cinco en cinco, para seleccionar las características de la mejor red neuronal. Es decir, para un número de parámetros establecido, se realizó la optimización bayesiana de los demás hiperparámetros en cada caso. El máximo número de rondas de entrenamiento fue de 40.000 para cada modelo, y los modelos fueron entrenados en unidades de procesamiento gráficos (GPUs).

Las redes finales seleccionadas por la optimización bayesiana, fueron todas redes de convolución, con: una función de activación de tangente hiperbólica, un parámetro de regularización L_2 de 0.1 y dropouts de entre 0.2861 y 0.63505, (siendo más comunes los valores alrededor de 0.6), usando una validación cruzada bajo optimización ADAM y con función de pérdida de entropía cruzada. La profundidad de las redes (el número de capas ocultas), seleccionada por la optimización bayesiana fue consistentemente de dos, por lo que se privilegiaron redes no muy profundas. El resumen de los resultados de los nueve modelos alcanzados, se aprecia en la Tabla 1.

Tabla 1. Medidas de adecuación de los nueve modelos de redes neuronales de predicción de la presencia de innovación.

Núm. de conex. o parám.	Exactitud	Área bajo la curva ROC	Kappa de Cohen
90	0.63	0.63	0.08
95	0.71	0.66	0.18
100	0.71	0.58	0.15
105	0.73	0.6	0.07
110	0.76	0.64	0.32
115	0.56	0.53	0.03
120	0.67	0.7	0.24
125	0.67	0.7	0.24
130	0.77	0.58	0.09

Según la Tabla 1, los mejores modelos de redes neuronales que predicen la innovación, son los de 110, 120 y 125 parámetros. Los tres modelos obtuvieron un valor alto en exactitud: 0.76 para el de 110 parámetros, y 0.67 para los otros dos modelos; lo cual sugiere que estos tres modelos están muy cerca de predecir la presencia de la innovación; sin embargo, cabe anotar que el 78% de las empresas del conjunto de prueba innovan, así que la exactitud lograda no supera la exactitud del modelo ingenuo, en que todas las empresas innovan. El modelo combinado posterior sí logró superar el 78%, como se indica más adelante. No

obstante, el comportamiento de estos modelos en términos del AUC es razonable y se observa que el modelo con 110 parámetros obtuvo un valor de 0.64, mientras que para los modelos con 120 y 125 parámetros el valor del AUC fue de 0.70. Aunque se espera que este valor sea cercano a uno, lo que indicaría que el modelo tiene una precisión perfecta en la predicción de la presencia de la innovación, un valor de 0.7 se considera como aceptable en la literatura (Mandrekhar, 2010).

Por otra parte, la kappa de Cohen calcula el ajuste de la predicción considerando el desbalance de la muestra. Para el modelo de 110 parámetros la kappa de Cohen fue de 0.32, mientras que para los modelos de 120 y 125 parámetros fue de 0.24, valores que sugieren que los modelos tienen un ajuste de predicción leve a pesar de estar desbalanceada la muestra (Lee et al., 2018). Los peores modelos de predicción de la innovación, son los de 90 y 115 parámetros, debido a los valores obtenidos en la exactitud (0.63 y 0.56), AUC (0.63, 0.53), kappa de Cohen (0.08, 0.03).

A partir de los nueve modelos pre-óptimos alcanzados, se combinaron a continuación los pronósticos para averiguar si un modelo combinado podía superar a todos los modelos individuales. En efecto, así fue, y las medidas de adecuación de este modelo combinado, se aprecian en la Tabla 2.

Tabla 2. Medidas de adecuación de ajuste del modelo combinado, a partir de los nueve modelos de redes neuronales de predicción de la presencia de la innovación.

Medidas de adecuación	resultado
Exactitud	0,792683
Área bajo la curva	0,72309
Kappa de Cohen	0,356417

Las medidas de adecuación del modelo combinado (Tabla 2), son en general superiores a las de los modelos individuales (Tabla 1). La exactitud alcanzada de 0.79 está por encima de la exactitud que se lograría con el modelo ingenuo de que todas las empresas innovan (ya que el 78% de ellas

innova), y el *AUC* de 0.72 del modelo combinado resulta aceptable de acuerdo con la literatura (Mandrekar, 2010) y es mayor al de cualquiera de los modelos individuales. El kappa de Cohen de 0.35, por su parte, el cual considera el ajuste de la predicción teniendo en cuenta el desbalance de la muestra (Lee et al., 2018), muestra un ajuste leve, mayor al máximo ajuste logrado por los modelos individuales.

En la Tabla 3, se observa la importancia de las variables en el modelo combinado. Todas las variables resultaron importantes en este modelo, de acuerdo con el criterio de ostentar un valor de importancia de la variable menor a uno, calculada

esta importancia como la exactitud del modelo con la variable permutada entre la exactitud del modelo original. Las primeras diez variables más importantes para la presencia de la innovación en las pymes colombianas, según el modelo final, son: las fuentes de financiación de corto plazo, la dificultad de las fuentes de financiación, el costo de capital, la preferencia por las fuentes de financiación, el sector mayor, la educación de los empleados, el costo de ajuste, el perfil del gerente, la edad y las fuentes de financiación a largo plazo. Entre las variables que aparecen como menos importantes, se encuentran: las características del crédito, la aversión al riesgo del gerente y las características para acceder al crédito.

Tabla 3. Importancia de las variables del modelo de predicción combinado construido a partir de los nueve modelos de redes neuronales.

Posición	Variable
1	Fuentes de financiación de corto plazo
2	Dificultad de las fuentes de financiación
3	Costo de capital
4	Preferencia por las fuentes de financiación
5	Sector mayor
6	Educación de los empleados
7	Costo de ajuste
8	Perfil del gerente
9	Edad
10	Fuentes de financiación de largo plazo
11	Riesgo moral y costos de agencia
12	Costo esperado de quiebra
13	Tamaño por empleados
14	Características para acceder al crédito
15	Aversión al riesgo del gerente
16	Características del crédito

Nota: Una variable es más importante si permutar sus valores entre las diferentes empresas hace disminuir más abruptamente la exactitud del modelo original que al permutar los valores de otra variable.

3.2 Discusión

En los estudios prospectivos, se ha hecho uso de un abanico de herramientas cualitativas y cuantitativas por igual, cuya aplicación depende en gran parte, del tipo de estudio a realizar. La mayoría de los métodos prospectivos, tiene como objetivo: la

observación de cambios sociales, la identificación de tendencias, de eventos y los propósitos de los actores sociales en un periodo de tiempo. En los estudios prospectivos, se entiende que el futuro es producto de la interacción de diferentes actores y factores sociales que pueden producir cambios en las tendencias (Medina-Vásquez, Becerra & Castaño, 2014).

Popper (2008) realiza un inventario de los métodos que más se utilizan en la realización de estudios prospectivos, y los categoriza en cuatro dimensiones: métodos basados en la creatividad, en la experticia, en la evidencia y en la interacción. La mayoría de los métodos relacionados con creatividad son cualitativos, mientras que la mayoría de los métodos semicuantitativos y cuantitativos están más relacionados con la evidencia, la experticia y la interacción.

Los métodos basados en la evidencia tienen como objetivo recoger tendencias e información cuantitativa, mientras que los métodos basados en la experticia se enfocan en consultar a personas con conocimientos especializados, directos o suficientes sobre un asunto particular, relacionado con las decisiones que se deben adoptar sobre un tema. Los métodos basados en la interacción tienen como objetivo hacer uso de la participación activa de los ciudadanos, para identificar situaciones estratégicas a tratar. Finalmente, los métodos basados en la creatividad permiten explorar alternativas que no habían sido pensadas y generar un pensamiento estratégico sobre las ideas no convencionales que surgen de estos ejercicios (Medina-Vásquez et al., 2014; Popper, 2008).

El conjunto de métodos de prospectiva está en continua evolución, donde la aplicación del aprendizaje de máquinas contribuye a esta dinámica. Existen, por ejemplo, estudios que trabajan con métodos de minería de datos y Exploratory Modeling and Analysis (EMA) para el análisis de tecnologías futuras (TFA) (de Miranda-Santo, Coelho, dos Santos, & Fellows-Filho, 2006; Kim, Hwang, Jeong, & Jung, 2012; Kwakkel & Pruyt, 2013). Estudios

más recientes, como el de Lee, Kwon, Kim y Kwon (2018), proponen el uso de redes neuronales para la prospectiva tecnológica basada en el análisis de patentes. Usando una muestra de empresas coreanas, estos últimos autores construyen dos indicadores cuantitativos, capaces de identificar las tendencias emergentes en tecnología.

Zhang, Porter, Chiavetta, Newman y Guo (2019), sostienen que los métodos que usan indicadores basados en el aprendizaje de máquinas, las redes neuronales y otros similares, promueven la capacidad de análisis de datos, que conducen a medidas de tecnología emergente avanzada. Los autores referencian algunos ejemplos del uso de estos métodos para la realización de estudios de prospectiva tecnológica, tales como: Porter, Garner, Carley y Newman (2019), Kose y Sakata (2018), Wang, Porter, Wang y Carley (2019), Jeong, Park y Yoon (2019), Ma, Abrams, Porter, Zhu y Farrell (2019), Moerhle y Caferoglu (2019), Berg, Wustmans y Bröring (2019).

La mayoría de las aplicaciones que ha tenido hasta ahora el aprendizaje de máquina en los estudios prospectivos, han estado enfocadas a la prospectiva tecnológica. Pero no cabe duda de que se puede aplicar a otras temáticas, como los estudios enfocados a la prospectiva territorial y sobre todo a aquellos temas de decisiones, que implican la construcción o la reforma de políticas públicas, en especial en temas de alta complejidad, de diversas variables y de comportamientos no lineales.

4. Conclusiones

Este trabajo tuvo como objetivo demostrar cómo los métodos del aprendizaje de máquinas permiten generar insumos para la política pública basada en la evidencia, construida a través de técnicas cuantitativas de modelación.

Se calcularon cinco modelos del estado del arte de la predicción, a saber: regresión logística, máquinas de vectores de soporte, máquinas de

gradiente potenciado, bosques aleatorios y redes neuronales. Esto con el fin de identificar cuáles son las variables financieras más importantes para que las pymes colombianas desarrollen actividades de innovación, como un ejemplo de benchmark de aplicación de estos métodos.

Los resultados sugieren que para que las pymes colombianas decidan innovar, parecen predominar variables asociadas a las fuentes y usos del financiamiento pero no tanto a las características de la empresa y del crédito, mientras que las variables asociadas al gerente son secundarias. Estos resultados pueden considerarse como insumo para el desarrollo de estudios prospectivos en el tema del desarrollo de innovación en las pymes colombianas; sin embargo, debe tenerse en cuenta que los resultados presentados son a corto plazo, en el caso de que se vayan a utilizar estos resultados con otros métodos prospectivos o con otros métodos para la generación de políticas públicas.

Se concluye que el uso de estos métodos puede ser de gran utilidad en el campo de la prospectiva y de los estudios enfocados al mejoramiento de las políticas públicas, ya que permiten generar modelos muy cercanos a los acontecimientos reales, siempre y cuando se incluyan la mayor cantidad de variables que puedan afectar un fenómeno y se calculen los modelos de manera rigurosa. Adicionalmente, con este estudio se brinda una nueva forma de abordar temas complejos y no lineales, que requieran jerarquizar las variables que influyen en determinado fenómeno.

Referencias

Andrews, C. J. (2007). Rationality in policy decision making. En F. Fischer, G. J. Miller, & M. S. Sidney (Eds.), *Handbook of public policy analysis: Theory, politics, and methods*, 161–171. Boca Ratón, FL: CRC Press.

Banco de la República de Colombia (2017). Salario mínimo legal de Colombia. Recuperado de: <http://obiee.banrep.gov.co>

Berg, S., Wustmans, M., & Bröring, S. (2019). Identifying first signals of emerging dominance in a technological innovation system: A novel approach based on patents. *Technological Forecasting and Social Change*, 146 (C), 706–722.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.

Carletta, J. (1996). Assessing agreement on classification tasks: The kappa statistic. *Computational Linguistics*, 22 (2), 249–254.

Cochran, C. E., Mayer, L. C., Carr, T. R., & Cayer, N. J. (2009). *American public policy: An introduction*. Boston, MA: Cengage Learning.

De Miranda-Santo, M., Coelho, G. M., dos Santos, D. M., & Fellows-Filho, L. (2006). Text mining as a valuable tool in foresight exercises: A study on nanotechnology. *Technological Forecasting and Social Change*, 73(8), 1013–1027.

EMIS (2017). *EMIS*. Recuperado de: <https://auth.emis.com/module.php/>

Fawcett, T. (2004). ROC graphs: Notes and practical considerations for researchers. *Machine Learning*, 31(1), 1–38.

Frey, D. J. (2011). Policy analysis in practice: Lessons from researching and writing a “statenote” for Education Commission of the States. *Capstone Collection*, 2425.

Geurts, T. (2011). *Public Policy Making: The 21st Century Perspective*. Amsterdam: Apeldoorn: Beinformed.

- Gu, S., Kelly, B., & Xiu, D. (2018). Empirical asset pricing via machine learning (No. w25398). National Bureau of Economic Research.
- Ho, R. (2012). Big data machine learning: Patterns for predictive analytics. *DZone Refcardz*, 2014. Recuperado de: <http://refcardz.dzone.com/refcardz/machine-learning-predictive>
- Jeong, Y., Park, I., & Yoon, B. (2019). Identifying emerging Research and Business Development (R&BD) areas based on topic modeling and visualization with intellectual property right data. *Technological Forecasting and Social Change*, 146, 655–672.
- Kim, J., Hwang, M., Jeong, D.-H., & Jung, H. (2012). Technology trends analysis and forecasting application based on decision tree and statistical feature analysis. *Expert Systems with Applications*, 39 (16), 12618–12625.
- Kose, T., & Sakata, I. (2018). Analysis of technology convergence in robotics and technological portfolios among robot-related organizations. *2018 Portland International Conference on Management of Engineering and Technology (PICMET)*, 1–12. IEEE.
- Kuhn, M., & Johnson, K. (2013). Applied predictive modeling. New York: Springer.
- Kwakkel, J. H., & Pruyt, E. (2013). Exploratory Modeling and Analysis, an approach for model-based foresight under deep uncertainty. *Technological Forecasting and Social Change*, 80(3), 419–431.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174.
- Lee, C., Kwon, O., Kim, M., & Kwon, D. (2018). Early identification of emerging technologies: A machine learning approach using multiple patent indicators. *Technological Forecasting and Social Change*, 127, 291–303.
- Ley 905. (2004). Por medio de la cual se modifica la Ley 590 de 2000 sobre promoción del desarrollo de las micro, pequeña y mediana empresa Colombiana y se dictan otras disposiciones. Colombia: Congreso de la República de Colombia.
- Ma, J., Abrams, N. F., Porter, A. L., Zhu, D., & Farrell, D. (2019). Identifying translational indicators and technology opportunities for nanomedical research using tech mining: The case of gold nanostructures. *Technological Forecasting and Social Change*, 146, 767–775.
- Malakar, G. (2018). *Introduction to Gradient Boosting algorithm*. Recuperado de: <https://youtu.be/ERDgauhTHk>
- Mandrekar, J. N. (2010). Receiver operating characteristic curve in diagnostic test assessment. *Journal of Thoracic Oncology*, 5 (9), 1315–1316.
- Manning, C., Raghavan, P., & Schütze, H. (2010). Introduction to information retrieval. *Natural Language Engineering*, 16 (1), 100–103.
- Medina-Vásquez, J. E., Becerra, S., & Castaño, P. (2014). *Prospectiva y política pública para el cambio estructural en América Latina y el Caribe*. Santiago de Chile: CEPAL.
- Yang, K., & Miller, G. J. (2007). *Handbook of Research Methods in Public Administration*. Boca Ratón: CRC Press.
- Moerhle, M. G., & Caferoglu, H. (2019). Technological speciation as a source for emerging technologies. Using semantic patent analysis for the case of camera technology. *Technological Forecasting and Social Change*, 146, 776–784.
- Olson, D. L., & Delen, D. (2008). *Advanced data mining techniques*. Berlin: Springer.
- Popper, R. (2008). How are foresight methods selected? *Foresight*, 10(6), 62–89.

- Porter, A. L., Garner, J., Carley, S. F., & Newman, N. C. (2019). Emergence scoring to identify frontier R&D topics and key players. *Technological Forecasting and Social Change*, 146, 628–643.
- Simon, H. A. (1976). *Administrative Behavior: A Study of Decision-Making Processes in Administrative Organization*. Nueva York: Harper & Rowe.
- Sokolova, M., Japkowicz, N., & Szpakowicz, S. (2006). Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation. *Proceedings of the ACS Australian joint conference on artificial intelligence*, 1015–1021. Berlín: Springer.
- Sydney, M. S. (2007). Policy formulation: Design and tools. In F. Fischer, G. J. Miller., & M. S. Sidney (Eds.), *Handbook of public policy analysis: Theory, politics, and methods* 79–88. Boca Ratón, FL: CRC Press.
- Vapnik, V. (2000). *The nature of statistical learning theory* (2a Ed.). New York: Springer.
- Wang, Z., Porter, A. L., Wang, X., & Carley, S. (2019). An approach to identify emergent topics of technological convergence: A case study for 3D printing. *Technological Forecasting and Social Change*, 146, 723–732.
- Wittmer, D.P., & McGowan, R.P. (2007). Five conceptual tools for decision-making. En Jack Rabin, W. B. Hildreth, & G. J. Miller (Eds.), *Handbook of Public Administration* 315–342. Boca Ratón, FL: CRC Press.
- Yang, K. (2007). Quantitative methods for policy analysis. En J. Rabin, B. Hildreth, & G. J. Miller (Eds.), *Handbook of Public Policy Analysis: Theory, Politics, and Methods* 349–367. Boca Ratón: CRC Press.
- Zhang, Y., Porter, A., Chiavetta, D., Newman, N. C., & Guo, Y. (2019). Forecasting technical emergence: An introduction. *Technological Forecasting and Social Change*, 146, 626–627.