Peligros y ventajas del diseño de instrumentos de medición auxiliado con estrategias de selección de atributos automáticas

Mario E. Marin[1], Julio C. Ponce[1], Ángel E. Muñoz[1], Alejandro Padilla[1]

[1] Universidad Autónoma de Aguascalientes, Centro de Ciencias Básicas, Aguascalientes, Ags., México

marin@ia.org.mx,{jcponce,aemz,apadilla}@correo.uaa.mx

Resumen. Se presenta una estrategia para ayudar en el diseño de instrumentos de medición en Ciencias Sociales por medio de selección de atributos automática utilizando eliminación recursiva de atributos con validación cruzada y comparándola con el uso de filtrado por correlación y por pruebas univariadas, regresión paso a paso, y eliminación recursiva. Para esto se usan conjuntos de datos producidos por instrumentos de medición prediseñados con rigor metodológico y estadístico, permitiendo generar un nuevo instrumento de medición con menor número de variables, pero conservando el poder predictivo con respecto a alguna variable de interés. Se presentan los peligros y ventajas de este acercamiento, así como un ejemplo utilizando microdatos de la Encuesta Nacional de Uso de Tiempo de INEGI, 2014 y 2019, para generar un nuevo instrumento con menos variables que los conjuntos de microdatos de estas encuestas, pero con poder predictivo mayor con respecto a variables de satisfacción en el uso del tiempo en ámbitos académicos y laborales.

Palabras clave: Aprendizaje máquina, selección de atributos, selección de modelo automatizada, instrumentos de medición, bienestar, uso de tiempo.

Dangers and advantages of design of measurement instruments using automated feature selection tools

Abstract. A strategy is presented to aid in the design of measurement instruments in Social Sciences by means of automated attribute selection using recursive feature elimination with cross-validation and comparing it to other methods such as filter feature selection using correlations and univariate tests, stepwise regression, and recursive elimination. For this, datasets generated by pre-designed instruments with methodological and statistical rigor are used, allowing for the generation of a new measurement instruments with fewer variables that preserve predictive power for some variable of interest. The dangers and advantages of this approach are presented, as well as an example using the National Time Use Survey of INEGI microdata, in its 2014 and 2019 editions, to generate a new instrument with fewer variables than the microdata sets of these surveys, but with better predictive power for variables of satisfaction in the use of time in academic and work settings.

Keywords: Machine learning, attribute selection, automatic model selection, measurement instruments, wellbeing, time use.

1. Introducción

Prácticamente todos los conjuntos de datos utilizados en el contexto de Datos Masivos (*Big Data*) y aprendizaje máquina son obtenidos mediante instrumentos de medición, que son herramientas que permiten asignar numerales a objetos o eventos de acuerdo a ciertas reglas [1], con mayor o menor grado de diseño metodológico. El diseño de instrumentos de medición es una importante disciplina dentro de la práctica de la investigación científica en Ciencias Sociales. Es un proceso que no puede ser resuelto de manera sencilla y sin un seguimiento riguroso de pasos en una metodología de la investigación asignando arbitrariamente números a fenómenos [2]. Generalmente se recomienda tener en mente un objetivo de estudio antes de comenzar siquiera el primer paso de decidir qué tipo de instrumento de medición se va a diseñar [3], incluso cuando ya se cuentan con datos e indicadores relevantes generados por instrumentos de medición preexistentes [2]. El proceso de diseño de instrumentos de medición se apega al proceso de aplicar el método científico en donde la formulación de las hipótesis precede al diseño de mediciones, experimentos y predicciones.

Por otro lado, el desarrollo de la analítica de *Big Data* y la minería de datos ha creado la posibilidad de generar y analizar grandes cantidades de datos que pueden ser preexistentes a las hipótesis. En la práctica, muchos investigadores y analistas que ejercen la minería de datos proponen hipótesis basadas en patrones encontrados en datos preexistentes a estas o incluso generados en tiempo real [4] y entrenan herramientas de aprendizaje máquina, que juegan el rol de conocimiento inferido [5], para predecir variables en base a datos fuera de la muestra original, y a veces generados por distintos instrumentos de medición. Existe controversia al respecto de si este tipo de labor investigativa tiene el mismo nivel de validez que con el enfoque mencionado en el párrafo anterior [4], o si este es un enfoque más ingenieril y aplicado. Sin embargo, existen prácticas como la validación de modelos [6] y el preregistro de hipótesis [7] que ayudan en este respecto, y ciertamente las prácticas de analítica de *Big Data* y la minería de datos se han incorporado a la labor investigativa [4], [5], por lo que es importante dar cara tanto a sus ventajas como a sus peligros.

El objetivo del presente trabajo se enfoca, en el cruce de las áreas de minería de datos e investigación, en averiguar si es posible utilizar un proceso de selección de modelo automatizado mediante el análisis de conjuntos de datos masivos preexistentes para auxiliar en el diseño de instrumentos de medición con rigor metodológico. Esto valiéndose de técnicas automatizadas de selección de atributos para así facilitar la creación de instrumentos de medición que sirvan para recopilar nuevos conjuntos de datos, de menor número de dimensiones que el original, pero con similar o mayor poder predictivo respecto a una variable de interés.

El presente texto se organiza así: después de esta introducción en donde se presentan temas, conceptos y objetivos; la sección 2 trata acerca de la definición, objetivos, y tipos de instrumentos de medición y de estrategias de selección de atributos. La sección 3 presenta las desventajas y peligros de utilizar enfoques automáticos de selección de atributos para generar instrumentos de medición basados

en conjuntos de datos de alta dimensionalidad que produzcan conjuntos de datos de menor dimensionalidad. La sección 4 propone un proceso de 6 pasos con 6 condiciones que deben ser cumplidas para evitar la mayor parte del riesgo y desventajas presentadas en la sección anterior. La sección 5 presenta un problema concreto basado en los microdatos de dos ediciones de la Encuesta Nacional de Uso de Tiempo (ENUT) de INEGI. La sección 6 aplica el proceso propuesto en la sección 4 al problema concreto presentado en la sección anterior y presenta los resultados obtenidos. Finalmente, la sección 7 presenta las conclusiones obtenidas del ejercicio de la sección anterior, así como puntos que quedan pendientes para trabajo futuro.

2. Teoría: Diseño de instrumentos de medición y selección de modelos y atributos

Los instrumentos más conocidos que los investigadores usan para recopilar datos acerca de la realidad, y evaluar sus hipótesis, son los de medición. En Ciencias Sociales pueden ser diseñados para medir tanto datos numéricos como nominales, para obtener conocimiento acerca de realidades personales o sociales de la muestra a la que se pretende investigar [8]. La gran mayoría de los conjuntos de datos relevantes en Ciencias Sociales existen porque previo a la recopilación de estos se tuvo un proceso de diseño de instrumento de medición para utilizar éste en la recopilación de los datos. Los instrumentos de medición o diseños de investigación pueden ser encuestas, entrevistas, experimentos, observaciones, recuperación de archivos textuales o de datos, o una combinación de cualquiera de los cinco primeros [3]. El proceso de elección de tipo de instrumento y su diseño es un proceso delicado que debe ser hecho a conciencia, tomando en cuenta tanto el tipo de datos que se quiere recabar como la naturaleza de la muestra. No está dentro de los alcances del presente trabajo entrar en detalles acerca de este procedimiento, pero es importante dejar en claro que el uso de conjuntos de datos preexistentes para la investigación siempre debe tomar en cuenta y alinearse con las características del instrumento original con el cual fueron recabados los datos. Ningún conjunto de datos existe en un vacío metodológico incluso si fueron recabados de manera automatizada.

La selección de modelos es el proceso de elegir un modelo matemático apropiado de entre una clase de modelos [9]. En el caso atendido por este trabajo se estaría hablando de un proceso de selección de modelos automatizado. Una de las herramientas para selección de modelos son las estrategias de selección de atributos que son procedimientos por los cuales se atiende el problema, cada vez más común, de que muchos conjuntos de datos de relevancia investigativa y en los que podrían aplicarse herramientas de aprendizaje máquina tienen una gran cantidad de atributos o dimensiones, lo que dificulta su procesado, disminuye el rendimiento de herramientas de regresión y aprendizaje máquina, y aumenta el uso de recursos computacionales en su procesado [4]. La selección de atributos tiene como objetivo elegir un subconjunto de atributos relevantes de entre los originales, descartando los atributos irrelevantes, redundantes o ruidosos [10].

Existen varios acercamientos de selección de atributos que pueden clasificarse de acuerdo con la disponibilidad de datos etiquetados o por el tipo de estrategia de búsqueda. En el caso que atañe a este trabajo se consideran casos automatizados en

donde las etiquetas están disponibles y en donde la predicción es una tarea de clasificación. Por el tipo de estrategia de búsqueda se hablará de herramientas de filtrado con las cuales los atributos se seleccionan en función de las características de los datos sin utilizar algoritmos de aprendizaje; y los de envoltorio, los cuales utilizan un algoritmo de aprendizaje específico para evaluar la calidad de las características seleccionadas [10]. Asimismo, se verán estrategias de regresión paso a paso. Se eligieron estos métodos por su popularidad en la comunidad de minería de datos.

3. El problema, los peligros y las ventajas

Una tarea común en los contextos de *Big Data* y aprendizaje máquina es la clasificación, en donde se entrena un algoritmo clasificador con un conjunto de datos etiquetado con las clases, y después este clasificador se pone a la tarea de predecir las clases de datos no etiquetados. Muchas veces se dispone de grandes conjuntos de datos de acceso público o de trabajos investigativos anteriores, pero para avanzar en el trabajo investigativo actual se requiere recabar nuevos datos de muestras diferentes de la misma población. Muchas veces es deseable que estos nuevos conjuntos de datos tengan un menor número de atributos y esto lleva a la pregunta de cómo elegir un subconjunto de atributos que conserve la mayor cantidad de poder predictivo del conjunto de datos originales o que lo supere. A esto se le conoce como el problema de la selección de modelo en minería de datos, y parece un problema para los métodos automatizados de selección de atributos.

En el contexto de minería de datos, la literatura propone soluciones como la regresión paso a paso, o los métodos de selección de atributos por filtrado [11]. Sin embargo, existe también en la literatura una crítica consistente del uso de este tipo de técnicas incluso para predicción de datos dentro de un subconjunto de la misma muestra [4], [12], que advierten de peligros tales como sobreajuste y tendencia a llegar a falsas conclusiones. Recientemente, Smith llevó a cabo una simulación Montecarlo en donde generó conjuntos de datos con entre 10 y 1000 variables explicativas candidatas cuyos valores fueron tomados aleatoriamente de distribuciones normales. Aleatoriamente se eligieron cinco variables explicativas y en base a estas se creó un valor objetivo. A continuación, se aplicó el método de selección de atributos de regresión paso a paso y se encontró que incluso a partir de conjuntos de datos con 100 variables explicativas candidatas la probabilidad de que una variable seleccionada sea realmente explicativa es prácticamente igual a que sea una variable sin relación real con la variable objetivo. La situación empeora conforme mayor es el número de variables explicativas candidatas [4], contrario a lo que podemos encontrar en parte de la literatura en *Big Data* y la intuición de algunos analistas en cuanto a que "más datos, es mejor". Esta crítica se extiende al uso de otras técnicas de selección de atributos más sofisticadas, aunque el mismo Smith reconoce que técnicas como los ensambles de árboles de decisión regularizados y la eliminación recursiva de atributos con validación cruzada tienen mayor éxito [4].

Sin embargo, los principales obstáculos en el uso de estas técnicas son con la teoría estadística. Muchas de las técnicas de selección de atributos sufren del problema de tener hipótesis post hoc, es decir, intentar probar la hipótesis con el mismo conjunto de datos que la sugiere [13]. Esto es claro en las técnicas de regresión paso a paso o

por filtrado, en donde se selecciona un subconjunto de atributos basado en pruebas estadísticas con la misma muestra con la que después se pretende hacer predicciones. Este uso de pruebas univariadas también conlleva violar varios de sus supuestos con su uso en conjunto, además de que éstas suponen hipótesis preespecificadas. Asimismo, la aplicación de estas técnicas puede ser un pretexto para evitar pensar en el problema de selección de modelo como tal [12].

Sin embargo, se puede argumentar que estos problemas son menos relevantes en el caso de predicción con nuevos datos, aunque dependiendo de los objetivos últimos del proyecto investigativo el uso de estrategias de selección de atributos automatizadas podría introducir restricciones en el tipo de conclusiones a las que se puede llegar. Por sí solas, las técnicas de selección de atributos automáticas son insuficientes para asegurar que el modelo sugerido por éstas lleve a un instrumento de medición apropiado para recabar nuevos datos que conserven el poder predictivo respecto a la variable de interés. Algo de lo que el investigador únicamente se enteraría cuando evaluara el rendimiento del nuevo conjunto de datos. Sin embargo, siempre y cuando no se caiga en la falsa confianza de querer solo "justificar con los números" el nuevo modelo, es posible apoyarse de las herramientas de selección de atributos automatizadas, lo que conlleva las ventajas asociadas a automatizar una parte del proceso de diseño de instrumento de medición alta en investigación teórica en el dominio de la aplicación, si se toman en cuenta algunas condiciones.

4. Una propuesta de solución

Existe un fuerte argumento de que el verdadero "estándar de oro" para la selección de atributos es el conocimiento experto en el dominio de la aplicación [4]. Esto está en línea con la literatura de diseño de instrumentos de medición en donde la operacionalización de las conceptos es un proceso que debe ser llevado a cabo de manera juiciosa, justificando con evidencia previa y teoría la inclusión o exclusión de cada variable que operacionalice un concepto que se pretende medir [2]–[4], [14].

Sin embargo, para estos propósitos, las estrategias de selección de modelo automatizadas ya han sido estudiadas y consideradas válidas si se cumplen ciertas condiciones en materia estadística. Además existe evidencia de que la inclusión de ciertos atributos puede no ser derivable de teoría y conocimiento experto y requerir evidencia empírica [15]. Asimismo, existen muchas críticas a la forma en que comúnmente se operacionalizan los conceptos en Ciencias Sociales, pero se reconoce que pese a todos los obstáculos metodológicos y epistemológicos, al final los datos pueden tener correlaciones con comportamientos reales y validez predictiva incluso si la fidelidad de los datos continua siendo un tema de discusión epistemológica o investigativa: a final de cuentas, una teoría es más que una generalización inductiva, más bien debe describir, explicar y predecir y no habiendo una diferencia profunda entre estos tres [14] existe valor en un modelo que tenga poder predictivo práctico siempre y cuando se respeten sus limitaciones estadísticas o inferenciales.

Atender a los pasos del proceso de diseño de instrumentos de medición y a las limitaciones de las herramientas estadísticas es precisamente lo que garantiza el poder predictivo, por lo que existen ciertas condiciones bajo las cuales se pueden utilizar herramientas de selección de atributos automatizadas para el presente objetivo:

1) El conjunto de datos original fue recabado por medio de un instrumento de medición cuyo diseño ya considera la teoría estadística y el conocimiento experto necesario. Esto no exime de estudiar el instrumento y su teoría.
2) Se tienen por lo menos dos muestras distintas para generar el subconjunto de atributos con uno y validar el rendimiento con otro. A falta de dos muestras distintas, un sustituto tolerable, con menor garantía de validez, sería separar en dos subconjuntos una sola muestra con gran cantidad de registros. Dependiendo del atributo a predecir puede ser necesario que se cuente con varios miles o decenas de miles de registros para este propósito.
3) Deben evitarse, en la medida de lo posible, usar estrategias de selección de atributos que violen ampliamente los supuestos de las pruebas estadísticas utilizadas, incluso si sus resultados parecen ser buenos. Esto incluye técnicas de filtrado por pruebas estadísticas y la regresión paso a paso.
4) Se debe tener un conocimiento general del dominio de la aplicación y acceso a conocimiento experto por medio de fuentes bibliográficas o en persona.
5) Se debe tener un conocimiento general del tema de diseño de instrumentos de medición y acceso a conocimiento experto.
6) Se debe contar con el tiempo necesario para hacer modificaciones relevantes al modelo sugerido por los métodos de selección de atributos automáticos.

Cumplidas estas condiciones pueden seguirse los siguientes seis pasos:

1) Asegurarse que los conjuntos de datos ya hayan sido limpiados y preprocesados de manera adecuada, así como probados para poder predictivo, preferentemente con validaciones cruzadas.
2) Aplicar una o varias estrategias de selección de atributos con los parámetros adecuados para obtener el número de atributos que se considera deseable para el nuevo conjunto de datos.
3) Comparar rendimientos de distintas estrategias, validando mediante entrenamiento con la muestra no utilizada para las estrategias de selección de atributos y comparando con el uso de los conjuntos de datos completos.
4) Aplicar conocimiento experto o juicio para elegir una de las estrategias de selección de atributos. Alternativamente, seleccionar el subconjunto de atributos producto de la técnica de eliminación recursiva de atributos con validación cruzada, la cual ha demostrado mayor eficacia y validez [4].
5) Elaborar un instrumento de medición de prueba, ciñéndose al diseño del instrumento original puesto que este impone limitaciones a respetarse.
6) A partir de este momento, considerar los siguientes pasos en el desarrollo del instrumento de medición [3], [8] exclusión de atributos o inclusión de nuevos atributos, u operacionalización de nuevos conceptos que sean necesarios y no estén incluidos en el instrumento original.

5. Un problema concreto

A continuación, se presenta un ejemplo de aplicación relevante al problema. Se tiene interés en crear un conjunto de datos de bajo número de atributos, pero con alto poder predictivo con respecto a variables de presencia/ausencia de satisfacción en el uso del tiempo en ámbitos laborales y académicos. La satisfacción es un componente

del constructo del bienestar subjetivo [16] utilizado ampliamente en muchas y variadas áreas de investigación social y económica [17], [18] para medir el bienestar de las personas de acuerdo a su propia percepción. La satisfacción se puede medir tanto en general, como en dominios específicos, tales como la vida familiar, la situación económica, la vida afectiva, etc. El uso del tiempo es uno de estos posibles dominios y, en concreto, el uso del tiempo en el trabajo o los estudios un par de dominios definidos más específicamente. Estas variables son relevantes porque están correlacionadas con otras variables de bienestar subjetivo [19] y objetivo [18], así como con comportamientos relevantes de administración del tiempo en ambientes laborales y académicos ligados a condiciones de alto bienestar subjetivo y alto desempeño importantes tanto para individuos como para organizaciones [20], [21].

Si se quisiera desarrollar un instrumento que midiera variables que fueran predictivas respecto a la satisfacción en el uso del tiempo en un tipo de actividades, se tendría que desarrollar una teoría de que tipo de cosas están ligadas a esta variable objetivo, como se obtendrían los datos, que tipos de herramientas se utilizarían, etc. Todo esto es un proceso largo y oneroso que requiere de la participación de equipos multidisciplinarios y la consulta de literatura y estado del arte del área. Aquí se ve la ventaja del acercamiento parcialmente automatizado propuesto en este texto.

Si se pudiera hallar un par de conjuntos de datos, obtenidos de distintas muestras, que incluyan las variables de interés, que hayan sido generados por medio de un mismo instrumento que ya haya seguido el procedimiento metodológico adecuado de diseño de instrumento y recolección de datos, y que dichos conjuntos de datos tengan poder predictivo suficiente para que alguna herramienta clasificadora tenga un nivel de desempeño juzgado satisfactorio, se podría cumplir la primera y segunda condición y se abre la posibilidad de trabajar en cumplir las siguientes cuatro.

6. Una solución concreta, y resultados

Para llevar a cabo la tarea propuesta en la sección anterior, se tomaron los microdatos de la Encuesta Nacional de Uso del Tiempo (ENUT), diseñada y llevada a cabo por el Instituto Nacional de Estadística y Geografía (INEGI) en México, en sus ediciones 2014 y 2019. Una consulta a la documentación de ambas encuestas [22], [23] permite comprobar el buen diseño de ambos instrumentos de medición y que son comparables, por lo que se puede utilizar un conjunto de datos para seleccionar atributos y otro para entrenar. En este caso se hará la selección de 30 atributos con el conjunto de datos de la ENUT-2014 (439 atributos con 42,118 registros,) y se harán pruebas del poder predictivo con validación cruzada con el conjunto de datos de la ENUT-2019 (432 atributos con 71,404 registros) utilizando únicamente los subconjuntos de 30 atributos generados con el conjunto de datos de la ENUT-2014.

En la Tabla 1 se puede ver el poder predictivo de ambos conjuntos de datos completos con respecto a las variables de presencia de satisfacción en el uso del tiempo en el trabajo y en el estudio. Estas son variables binarias. Se aplicó un proceso de submuestreo eliminando aleatoriamente registros de la clase más común hasta que ambas quedaron equilibradas, se eliminaron atributos correspondientes a metadatos y todos los atributos relacionados con bienestar subjetivo dada su alta correlación con las variables de interés y su poco común disponibilidad en el contexto de esta

investigación, y se aplicó una validación cruzada estratificada con 5 dobleces. Con la intención de ver si los cambios hechos a los conjuntos de datos cambian los desempeños de distintos clasificadores, se compararon los siguientes: bosque aleatorio con 1500 árboles y un máximo de 100 niveles, máquinas de soporte vectorial (MSV) con una forma de función de decisión de uno contra uno, red neuronal profunda con 1000 capas ocultas de 241 neuronas, y regresión logística multinomial con el algoritmo de optimización *Broyden–Fletcher–Goldfarb–Shanno*.

Tabla 1. Métricas de desempeño, usando validación cruzada estratificada con 5 dobleces, en clasificación para atributos de presencia de satisfacción en el uso del tiempo para actividades académicas y laborales de los conjuntos de datos completos de la ENUT-2014 y ENUT-2019; utilizando 4 diferentes clasificadores.

Año	Atributo	Clasificador	*Accuracy*	*Precision*	*F1 Sc.*	Tiempo C.
2014	Actividades académicas o de estudio	Bosque A.	**0.9495**	**0.9521**	**0.9494**	18.05s
		MSV	0.4977	0.4973	0.4874	46.57s
		Red Neuronal	0.8634	0.8758	0.8612	61.45s
		R. Logística.	0.8952	0.8981	0.8950	**18.78s**
2014	Actividades laborales o económicas	Bosque A.	**0.8149**	**0.8639**	**0.8085**	0.66s
		MSV	0.5032	0.5037	0.4913	367.47s
		Red Neuronal	0.5654	0.6172	0.4745	248.23s
		R. Logística.	0.73772	0.73856	0.73746	**1.63s**
2019	Actividades académicas o de estudio	Bosque A.	**0.9583**	**0.9607**	**0.9583**	32.66s
		MSV	0.9128	0.9136	0.9128	37.65s
		Red Neuronal	0.9376	0.9378	0.9375	171.18s
		R. Logística.	0.9177	0.9182	0.9177	**0.94s**
2019	Actividades laborales o económicas	Bosque A.	**0.8142**	**0.8630**	**0.8077**	130.77s
		MSV	0.7783	0.7896	0.7761	1006.16s
		Red Neuronal	0.7507	0.7514	0.7505	1816.61s
		R. Logística.	0.7368	0.7380	0.7365	**3.51s**

Como podemos ver en la Tabla 1, ambos conjuntos de datos tienen alto desempeño en la predicción de las dos variables de interés, por lo que son adecuados para la tarea de seleccionar un subconjunto de atributos aplicando las herramientas automáticas de selección de atributos. No sorprende que el subconjunto de datos de la ENUT-2019 tenga mayor desempeño ya que éste tiene más registros que el otro. Asimismo, podemos ver que el clasificador con el mejor desempeño es claramente el bosque aleatorio, a pesar de que la regresión logística tiene el menor tiempo de cómputo.

Hecho esto, se evaluaron distintas técnicas de selección de atributos, incluyendo algunas ya mencionadas como no recomendables, como filtrado por mayor índice de correlación Pearson, filtrado por pruebas univariadas f-test, y regresión paso a paso hacia adelante y hacia atrás. Se utilizaron dos que son recomendables por la literatura consultada: eliminación recursiva de atributos y eliminación recursiva de atributos con validación cruzada. En el caso de las dos últimas técnicas, que son de tipo envoltorio, se incorpora el clasificador de bosque aleatorio dado que fue el que tuvo el mejor desempeño de acuerdo con los datos de la Tabla 1. En la Tabla 2 podemos ver el tiempo de cómputo tomado por cada estrategia de selección de atributos. Todos los

cómputos fueron hechos en una computadora con procesador Intel Core i5-9300H a 2.40Ghz y 32GB de RAM. Se utilizó computo paralelo cuando fue posible.

Tabla 2. Tiempos de cómputo de estrategias de selección de atributos.

Estrategia de selección de atributos	**Actividades académicas**	**Actividades laborales**
Filtrado por correlación	**00:00:01**	**00:00:01**
Filtrado por pruebas univariadas (f-test)	**00:00:01**	**00:00:01**
Regresión paso a paso hacia adelante	00:21:10	00:37:03
Regresión paso a paso hacia atrás	00:05:42	00:05:39
Eliminación recursiva	05:40:28*	03:51:27*
Eliminación recursiva, validación cruzada	21:40:10*	23:05:21*

Formato: hh:mm:ss *Utilizando cómputo paralelo.

Como se puede ver en la Tabla 2, el tiempo de cómputo de las técnicas de filtrado es prácticamente instantáneo, mientras que la regresión paso a paso toma algunos minutos; sin embargo, utilizar eliminación recursiva toma varias horas, y casi un día para eliminación recursiva con validación cruzada (con k = 5). El conjunto de datos de la ENUT-2014 con algunos cientos de dimensiones y menos de 50,000 registros es relativamente pequeño comparado con los conjuntos de datos típicos en *Big Data*, por lo que podemos ver porque técnicas de filtrado y la regresión paso a paso son mucho más utilizadas que técnicas con mayor nivel de validez pero que toman mucho más tiempo de cómputo en el contexto de *Big Data*. En el caso de auxiliar en el diseño de instrumentos de medición, este proceso se realiza una sola vez y puede justificarse el uso de estrategias de selección de atributos con varias horas o incluso días de tiempo de cómputo, excepto en casos extremos.

En la Tabla 3 se puede ver una comparación de desempeño para clasificar por presencia de satisfacción en el uso de tiempo en actividades académicas y laborales tanto probando con la misma muestra de la ENUT-2014 como con la de la ENUT-2019. En ambos casos se usa validación cruzada con 5 dobleces. En todos los casos el tiempo de cómputo es menor que utilizando el conjunto de datos completo.

Tabla 3. Comparación de desempeño para clasificación por presencia de satisfacción de uso del tiempo en actividades académicas y laborales usando subconjuntos de entrenamiento obtenidos por distintas estrategias de selección de atributos, distintos clasificadores y entrenando con la misma muestra utilizada para selección de atributos (2014) y con una nueva muestra (2019)

Método de selección	**Clasificador**	**Act. Académicas**		**Act. Laborales**	
		***Accuracy* (2014)**	***Accuracy* (2019)**	***Accuracy* (2014)**	***Accuracy* (2019)**
Filtrado por correlación	Bosque A.	0.9576	0.9628	0.8079	0.8070
	MSV	0.9437	0.9516	0.8099	0.8039
	Red Neuronal	0.9427	0.9276	0.8082	0.8128
	R. Logística.	0.9154	0.9247	0.7790	0.7818
Filtrado, univariado (f-	Bosque A.	0.9561	0.9641	0.8102	0.8056
	MSV	0.9382	0.9518	0.8083	0.8051

test)	Red Neuronal	0.9432	0.9527	0.8113	0.8115
	R. Logística.	0.9191	0.9216	0.7799	0.7767
Regresión paso a paso hacia adelante	Bosque A.	**0.9601**	**0.9642**	**0.8191**	**0.8155**
	MSV	0.8997	0.9292	0.8034	0.7999
	Red Neuronal	0.9439	0.9582	0.8111	0.8144
	R. Logística.	0.9154	0.9246	0.7499	0.7425
Regresión paso a paso hacia atrás	Bosque A.	0.9583	0.9667	0.8159	0.8156
	MSV	0.8925	0.9412	0.8000	0.8019
	Red Neuronal	0.9437	0.9603	0.8051	0.8086
	R. Logística.	0.9202	0.9266	0.7355	0.7346
Eliminación recursiva	Bosque A.	0.9575	0.9619	0.8088	0.8074
	MSV	0.9399	0.9482	0.7927	0.7953
	Red Neuronal	0.8922	0.9532	0.7726	0.7797
	R. Logística.	0.9110	0.9203	0.7348	0.7335
Eliminación recursiva con validación cruzada	Bosque A.	0.9573	0.9639	0.8135	0.8119
	MSV	0.9390	0.9472	0.7936	0.7937
	Red Neuronal	0.9480	0.9514	0.7696	0.7812
	R. Logística.	0.9156	0.9202	0.7339	0.7372

Como podemos ver en la Tabla 3 y la Tabla 1, las diferencias en desempeño en cuanto a las variables de satisfacción en el uso del tiempo en actividades académicas y laborales son pequeñas tanto entrenando con los conjuntos de datos completos como con los subconjuntos sugeridos por las estrategias de selección de atributos. Asimismo, las diferencias también son pequeñas utilizando los subconjuntos sugeridos tanto entrenando con datos con la misma muestra con que se seleccionaron los atributos como entrenándolos con una muestra diferente, todo lo que da validación al procedimiento. No solo eso, sino que en todos los casos excepto uno, el desempeño usando la muestra de validación (ENUT-2019) no es menor al obtenido utilizando la muestra original (ENUT-2014) con cualquiera de los cuatro clasificadores probados, esto es atribuible a la mayor cantidad de registros en la nueva muestra, y que estas son comparables entre sí con el mismo diseño metodológico. Y la excepción, como veremos más adelante, se trata del único caso en que se aplicó un criterio no automatizado potencialmente mejorable. Sin embargo, mayor desempeño no necesariamente es el único criterio relevante. El mejor desempeño para predicción de satisfacción en el uso del tiempo en actividades académicas y laborales es visto utilizando regresión paso a paso hacia adelante con bosque aleatorio, aunque el acercamiento para selección de atributos más recomendable por la teoría es eliminación recursiva de atributos con validación cruzada cuya diferencia en desempeño con la regresión paso a paso es mínima y por lo tanto aceptable.

Comparando los valores de la Tabla 3 y la Tabla 1 es claro que, aunque la diferencia en rendimiento entre bosque aleatorio y los otros clasificadores se reduce considerablemente al utilizar conjuntos de datos con menor dimensionalidad, este primero sigue siendo la mejor opción en general dado que solo se dio un caso en donde las máquinas de soporte vectorial la superaron ligeramente.

Respecto a la excepción mencionada de un rendimiento inferior utilizando la muestra de validación de la ENUT-2019 con respecto a utilizar la misma muestra de

la ENUT-2014, se da al utilizar bosque aleatorio para predecir satisfacción en el uso del tiempo en actividades laborales con el conjunto de atributos seleccionado por eliminación recursiva con validación cruzada. Es importante notar que dicho método, a diferencia de los otros, sugiere 30 atributos para predicción de satisfacción en el uso del tiempo en actividades académicas que es el número de atributos que se busca. Sin embargo, la misma estrategia sugiere 181 atributos para predicción de satisfacción en el uso del tiempo en actividades laborales, lo que apunta a que el poder predictivo para esta variable está más distribuido entre las variables explicativas candidatas. Tanto para poder seguir comparando directamente como por razones prácticas, y pagando un costo pequeño en desempeño por ello, se seleccionaron los 30 atributos para predecir satisfacción en el uso del tiempo en actividades laborales de la siguiente forma: de entre los 181 sugeridos en el caso de la eliminación recursiva con validación cruzada se seleccionaron los 27 atributos que dicha estrategia sugirió tanto para predicción de satisfacción en el uso del tiempo en actividades académicas como laborales y se seleccionaron aleatoriamente 3 atributos más de entre los restantes atributos sugeridos para predicción de satisfacción en el uso del tiempo en actividades laborales y que no fueron sugeridos para predicción de satisfacción en el uso del tiempo en actividades académicas. Esto es un acercamiento útil, aunque mejorable, para esta situación específica en que el mismo nuevo instrumento de medición se pretende utilizar para recolectar datos para predecir ambas variables, con lo que al final de este trabajo se contaría con 33 atributos sugeridos para dicho instrumento debido a las coincidencias entre las sugerencias para predecir satisfacción en el uso del tiempo en actividades académicas y laborales con 30 atributos. Una mejor selección de los tres atributos no coincidentes, pero ya respaldada con conocimiento experto, podría mejorar el desempeño en la predicción de satisfacción en el uso del tiempo en actividades laborales utilizando como punto de partida el subconjunto de atributos sugerido por eliminación recursiva de atributos con validación cruzada.

En la Tabla 4 está la lista de atributos que se sugiere que el nuevo instrumento incluya para predecir satisfacción en el uso del tiempo en actividades académicas y laborales, es decir, los datos que se necesitarían de un respondiente para predecir si está satisfecho con su uso de tiempo en actividades académicas y laborales.

Tabla 4. Subconjunto de atributos de la ENUT [22], [23] con mayor desempeño, producido por eliminación recursiva con validación cruzada, para predicción de satisfacción en el uso del tiempo en actividades académicas y laborales.

Número total de habitaciones en la vivienda	Tiempo dedicado a limpiar o recoger el interior de la vivienda (F.S.)
Número de personas que habitan la vivienda	Tiempo dedicado a doblar/acomodar/guardar ropa (E.S.)
Edad en años	Estado civil: soltero/no soltero
Asiste a la escuela (edad 5 a 24 años)	Tiempo dedicado a hacer deporte o ejercicio físico (E.S.)
Tiempo dedicado a trabajar (E.S.)	Tiempo dedicado a ver televisión sin hacer otra actividad (E.S.)
Tiempo dedicado a dormir (incluye siesta) (E.S.)	Tiempo dedicado a revisar el correo; consultar redes sociales o (E.S.)
Tiempo dedicado a dormir (incluye	Tiempo dedicado a revisar el correo;

siesta) (F.S.)	consultar redes sociales o chatear (F.S.)
Tiempo dedicado a comer (E.S.)	Vivienda tiene bomba de agua**
Tiempo dedicado a comer (F.S.)	Se dedica exclusivamente a estudiar
Tiempo dedicado a arreglo e higiene personal; ir al baño (E.S.)	Tiempo dedicado a proteger bienes y vivienda (E.S.)
Tiempo dedicado a arreglo e higiene personal; ir al baño (F.S.)	Tiempo dedicado a tomar cursos o clases (F.S.)*
Tiempo dedicado a tomar cursos o clases (E.S.)	Tiempo dedicado a tareas; prácticas escolares o estudiar (F.S.)*
Tiempo dedicado a tareas; prácticas escolares o estudiar (E.S.)	Tiempo dedicado a trasladarse de y hacia la escuela (F.S.)*
Tiempo dedicado a trasladarse de y hacia la escuela (E.S.)	Tiempo dedicado a organizar o repartir los quehaceres del hogar (E.S.)**
Tiempo dedicado a cocinar o preparar alimentos o bebidas (E.S.)	Tiempo dedicado a consultar información, navegar en Internet (E.S.)
Tiempo dedicado a servir comida; recoger; lavar; secar y acomodar trastes (E.S.)	Tiempo dedicado a asistir o participar en actividades/celebraciones religiosas (E.S.)**
Tiempo dedicado a limpiar o recoger el interior de la vivienda (E.S.)	

**Sugerido solo para predicción en el contexto de actividades académicas.*
***Sugerido solo para predicción en el contexto de actividades laborales.*
E.S. = Entre semana. F.S.= Durante el fin de semana.

El trabajo de diseño de un instrumento de medición no acaba en este punto, sino que continua con un análisis de los atributos sugeridos por el procedimiento tanto en el marco teórico como de aplicación del instrumento. De ser relevante se continua con más pruebas utilizando diferentes clasificadores, descartando algunos atributos e incluyendo otros no contenidos en el conjunto de datos original en base a sugerencias de la literatura o de expertos; asimismo, dependiendo de la naturaleza de los atributos algunos pueden transformarse o agruparse, como por ejemplo en el caso del ejemplo presentado: distintas variables correspondientes a distintas actividades a las que se asignó tiempo pueden englobarse en una sola, por ejemplo, tiempo dedicado a "actividades domésticas", o "actividades sociales". Sin embargo, este trabajo ya forma parte de los siguientes pasos del proceso de diseño de instrumentos de medición y a partir de este momento requiere de un trabajo metodológico propio.

7. Conclusiones y Trabajo a Futuro

El presente trabajo ofrece un ejemplo que muestra evidencia de que es posible utilizar el análisis de conjuntos de datos preexistentes para auxiliar en el diseño de instrumentos de medición con rigor metodológico, valiéndose de técnicas automatizadas de selección de atributos utilizadas para selección de modelos. En la sección 4 se ofrecen seis pasos y seis condiciones que deben cumplirse para poder validar los resultados obtenidos con este acercamiento, principalmente el uso de dos muestras diferentes obtenidas mediante instrumentos de medición con la misma metodología y el respaldo de una cantidad razonable de conocimiento experto en el

dominio de la aplicación y en el diseño de instrumentos de medición. Se mencionan las ventajas: que las herramientas utilizadas en el contexto de *Big Data* y minería de datos permiten acelerar y facilitar el proceso de diseño de instrumentos de medición, pero se alertan de los peligros y se propone una forma de evitarlos.

En base a los resultados en la Tabla 1 y la Tabla 3, se puede ver que los modelos de 30 atributos seleccionados por las estrategias de selección de atributos automatizadas quedan validados al entrenarlos con una muestra diferente, la de la ENUT-2019, con la cual las métricas de *accuracy* no solo son altas sino que en este caso, usando el mejor clasificador, nunca son menores al desempeño obtenido al entrenar con los conjuntos de datos originales completos de 432 y 439 atributos para la ENUT-2019 y ENUT-2014 respectivamente. Esto es evidencia de que es posible utilizar un instrumento de medición preexistente para diseñar uno de menor número de atributos con similar o mejor poder predictivo para una o varias variables de interés, con las ventajas que esto conlleva a comparación de empezar el diseño desde un principio.

La evidencia presentada apunta a que cumplir las seis condiciones y seguir los seis pasos descritos en la sección 4 evita problemas tanto estadísticos como metodológicos para este procedimiento al tiempo que permite automatizar una importante parte del proceso de diseño de instrumentos de medición, lo que es una ventaja grande en el contexto de minería de datos e investigación.

Sin embargo, debe considerarse que este acercamiento no exime al investigador o analista de pensar en y estudiar los conceptos operacionalizados por el instrumento original, y que este procedimiento de automatización necesariamente traslada las bondades, límites y posibles errores del instrumento original al nuevo de maneras que necesitan ser analizadas independientemente durante los subsecuentes pasos de diseño del instrumento de medición que no es agotado por el procedimiento presentado en este trabajo. Atención a las situaciones que se presentan en estos siguientes pasos y un análisis del desempeño obtenido con las muestras recabadas utilizando el nuevo instrumento de medición son precisamente trabajo futuro pendiente.

Referencias

1. S. S. Stevens, “On the Theory of Scales of Measurement”, Science (80-.)., vol. 103, núm. 2684, pp. 677 LP – 680, jun. 1946, doi: 10.1126/science.103.2684.677.
2. F. Martínez Rizo, “Los indicadores como herramientas para la evaluación de la calidad de los sistemas educativos”, Sinéctica, núm. 35, pp. 1–17, 2010, doi: 2007-7033.
3. P. W. Vogt, D. C. Gardner, y L. M. Haeffele, “General Introduction, Design, Sampling, and Ethics”, en When to use what research design, New York, NY: The Guilford Press, 2012, pp. 1–7.
4. G. Smith, “Step away from stepwise”, J. Big Data, vol. 5, núm. 1, dic. 2018, doi: 10.1186/s40537-018-0143-6.
5. U. Fayyad, G. Piatetsky-Shapiro, y P. Smyth, “From Data Mining to Knowledge Discovery in Databases”, AI Mag., vol. 17, núm. 3, p. 37, 1996, doi: 10.1609/aimag.v17i3.1230.
6. S. I. Gass y M. C. Fu, Eds., “Validation”, en Encyclopedia of Operations Research and Management Science, Boston, MA: Springer US, 2013, p. 1597.

7. J. M. Fressoli y V. Arza, “Los desafíos que enfrentan las prácticas de ciencia abierta”, Teknokultura, vol. 15, núm. 2, pp. 429–448, 2018, doi: 10.5209/tekn.60616.
8. E. Mejía Mejía, “Instrumentos de acopio de datos”, en Técnicas e instrumentos de investigación, 1a ed., núm. 9972-834-08–05, Lima, Perú: Centro de Producción Editorial e Imprenta de la Universidad Nacional Mayor de San Marcos, 2005, pp. 13–94.
9. C. Sammut y G. I. Webb, Eds., “Model Selection”, en Encyclopedia of Machine Learning and Data Mining, Boston, MA: Springer US, 2017, p. 845.
10. S. Wang, J. Tang, y H. Liu, “Feature Selection”, en Encyclopedia of Machine Learning and Data Mining, C. Sammut y G. I. Webb, Eds. Boston, MA: Springer US, 2017, pp. 503–511.
11. K. J. Cios, R. W. Swiniarski, W. Pedrycz, y L. A. Kurgan, “Feature Extraction and Selection Methods”, en Data Mining: A Knowledge Discovery Approach, Boston, MA: Springer US, 2007, pp. 133–33.
12. F. Harrel, “Problems with stepwise regression”, Stata | FAQ, 1996. [En línea]. Disponible en: https://www.stata.com/support/faqs/statistics/stepwise-regression-problems/. [Consultado: 06-mar-2021].
13. P. W. Vogt, “Post Hoc Theorizing”, Dictionary of Statistics & Methodology, 2005. [En línea]. Disponible en: https://methods.sagepub.com/reference/dictionary-of-statistics-methodology.
14. E. E. Roskam, “Operationalization, a superfluous concept”, Qual. Quant., vol. 23, núm. 3–4, pp. 237–275, 1989, doi: 10.1007/BF00172446.
15. J. L. Castle, J. A. Doornik, y D. F. Hendry, “Evaluating Automatic Model Selection”, J. Time Ser. Econom., vol. 3, núm. 1, 2011, doi: 10.2202/1941-1928.1097.
16. E. Diener, “Subjective Well-Being”, Psychol. Bull., vol. 95, núm. 3, pp. 542–575, 1984, doi: 10.1037/0033-2909.95.3.542.
17. E. Diener, E. M. Suh, R. E. Lucas, y H. L. Smith, “Subjective well-being: three decades of progress”, Psychological Bulletin, vol. 125. pp. 276–302, 1999.
18. OECD, “Why does subjective well-being matter for well-being?”, en How’s Life?: Measuring well-being, Paris: OECD Publishing, 2011, pp. 266–267.
19. I. Boniwell, “Satisfaction with time use and its relationship with subjective well-being”, The Open University, 2006.
20. B. Aeon y H. Aguinis, “It’s About Time: New Perspectives and Insights on Time Management”, Acad. Manag. Perspect., vol. 31, núm. 4, pp. 309–330, 2017, doi: 10.5465/amp.2016.0166.
21. B. J. C. Claessens, W. van Eerde, C. G. Rutte, y R. A. Roe, “A review of the time management literature”, Pers. Rev., vol. 36, núm. 2, pp. 255–276, 2007, doi: 10.1108/00483480710726136.
22. INEGI, Encuesta Nacional sobre Uso del Tiempo 2014 (ENUT): documento metodológico. Aguascalientes: Instituto Nacional de Estadística y Geografía, 2015.
23. INEGI, Encuesta Nacional sobre Uso del Tiempo 2019, Diseño Conceptual. Aguascalientes: Instituto Nacional de Estadística y Geografía, 2020.