

Viabilidad de los Procedimientos de Aprendizaje Automático Automatizado para el Estudio de las Evidencias de la Validez de Criterio: El Caso del VIA-IS

Feasibility of Automated Machine Learning Procedures for the Study of Criterion Validity Evidence: The Case of VIA-IS

Mercedes Ovejero¹, Jesús M. Alvarado² y Jaime Ballesteros³

Resumen

El enfoque tradicional de la validez de criterio se basa en el uso de correlaciones y regresión lineal empleados en una única muestra. Los avances en la algoritmia del aprendizaje automático supervisado han permitido el desarrollo de herramientas automatizadas (AutoML) que suponen un ahorro de tiempo y de coste computacional de la exploración de la capacidad predictiva de las puntuaciones de un test sobre criterios relevantes para el modelo teórico. El presente trabajo aplicó el pipeline de AutoML al estudio de la capacidad predictiva del cuestionario VIA-IS de fortalezas personales sobre la satisfacción general con la vida. Los resultados muestran que el AutoML permite explorar de forma rápida y eficiente gran cantidad de algoritmos predictivos, obteniéndose mejores medidas generales de ajuste y que las fortalezas gratitud, curiosidad, esperanza, apreciación de la belleza y la bondad son las que contribuyen en mayor grado a pronosticar la satisfacción vital.

Palabras clave: validez de criterio, AutoML, VIA-IS, SWLS, aprendizaje automático

Abstract

The traditional approach to criterion validity is based on the use of correlations and linear regression models used on a single sample. Advances in supervised machine learning algorithms have enabled the development of automated tools (AutoML) that save time and computational cost when exploring the predictive capacity of test estimates on criteria relevant to the theoretical model. The present work applied the AutoML pipeline to the study of the predictive capacity of the VIA-IS personal strengths questionnaire on general satisfaction with life. The results show that AutoML allows a large number of predictive algorithms to be explored quickly and efficiently, obtaining better general measures of fit and that the strengths gratitude, curiosity, hope, appreciation of beauty and kindness are the ones that contribute the most to predict life satisfaction.

Keywords: criterion validity, AutoML, VIA-IS, SWLS, machine learning

¹ Doctora en Psicología. Científico de datos y Profesora asociada. Departamento de IA en Sermes CRO (Madrid, España) y Departamento de Metodología y Psicobiología de la Facultad de Psicología de la Universidad Complutense de Madrid, España. Facultad de Psicología. Campus de Somosaguas. Pozuelo de Alarcón. 28223. Madrid, España. Tel.: 913 94 31 78. Correo: mercheovejero@psi.ucm.es (Autora de correspondencia)

² Doctor en Psicología. Catedrático de Universidad. Facultad de Psicología, Universidad Complutense de Madrid. Campus de Somosaguas, 28223 Pozuelo de Alarcón, Madrid, España. Tel.: 913943055. Correo: jmalvara@ucm.es

³ Grado en Psicología. Gestor de datos. Departamento de Biometría. Sermes CRO, Madrid (España). Calle Rufino González, 14. 28037. Madrid, España. Tel.: 913 75 69 30. Correo: jaime_ballesteros@sermescro.com

Revista Iberoamericana de Diagnóstico y Evaluación – e Avaliação Psicológica. RIDEP · Nº66 · Vol.5 · 117-125 · 2022

ISSN: 1135-3848 print /2183-6051online

This work is licensed under CC BY-NC 4.0. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>

Introducción

Los test psicológicos son las herramientas principales utilizadas para cuantificar el nivel que tienen los individuos en los constructos psicológicos. La toma de decisiones en los procesos de evaluación psicológica depende de la calidad técnica de los test aplicados (Toro et al., 2022). Para el estudio psicométrico de esta calidad técnica de los diferentes instrumentos de evaluación, existen una serie de guías y buenas prácticas (por ejemplo, Ledesma et al., 2018, Toro et al., 2022) basadas en dos propiedades fundamentales, a saber, la fiabilidad y la validez (Kimberlin & Winterstein, 2008). Tanto la fiabilidad como la validez no son propiedades del test en sí mismo, sino de la interpretación de las puntuaciones (Messick, 1989).

La fiabilidad se refiere al grado en el que las puntuaciones de una prueba o cualquier procedimiento de medición tienen precisión además de arrojar el mismo resultado en evaluaciones sucesivas (Carmines & Zeller, 1979). Se considera una condición necesaria pero no suficiente para garantizar la validez (Sullivan, 2011; Warne, 2008).

En el caso de la validez, esta correspondería al nivel de evidencia y justificación teórica que sustenta la interpretación y uso de las puntuaciones obtenidas por un test. Este concepto engloba cinco aspectos fundamentales (AERA, APA y NCME, 2018), a saber, contenido, representación del constructo, estructura interna, relación con otras variables y consecuencias de su aplicación. El presente trabajo se centra en las evidencias de validez referida a la relación con otras variables, la cual pretende cuantificar el rendimiento explicativo del instrumento de medida y, por ello, tiene como uno de sus pilares fundamentales las evidencias vinculadas a la validez predictiva, mediante la que se pretende verificar la capacidad de pronóstico del test con respecto a un criterio relevante para el modelo teórico del constructo (Taherdoost, 2016). Generalmente, esta validez ha sido estudiada desde un enfoque basado en las correlaciones de Pearson y modelos lineales de regresión.

Los modelos de regresión son parte de las estrategias basadas en el aprendizaje automático supervisado, el cual tiene como objetivo estimar

una función capaz de predecir el valor que le corresponde a cada participante después de haber visto una serie de ejemplos. Habitualmente consta de dos fases, una primera en la que se crea el modelo utilizando un conjunto finito de datos (evaluaciones de los participantes) denominada fase de entrenamiento, y, la segunda, que es una fase de evaluación o testeo, que pretende evaluar la potencia del modelo en situaciones prácticas como, por ejemplo, estimar la puntuación que le corresponde en el criterio a los participantes de una muestra diferente de la original utilizada para elaborar el modelo (Trognon et al., 2022). En esta fase se estudiaría si el modelo construido tras el entrenamiento sigue ofreciendo estimaciones adecuadas. Por ello, el aprendizaje automático supone ir más allá de la conceptualización estática de la mayoría de los modelos de validez predictiva, permitiendo estudiar la calidad predictiva del modelo en nuevos participantes pertenecientes a la población diana.

Entre los principales desafíos de la aplicación de los modelos de aprendizaje automático supervisado destacan que la gran cantidad de modelos predictivos y su calibración podrían traducirse en un tiempo y coste computacional por invertir. Estas situaciones quedan abordadas gracias a los pipelines de trabajo (los cuales se definen como un código común que genera un modelo para un problema de clasificación o de regresión) basados en el aprendizaje automático automatizado (en adelante, AutoML). El AutoML es el proceso de automatizar las tareas de aplicar el aprendizaje automático a problemas del mundo real (Hutter et al., 2019) surgiendo como una solución basada en la inteligencia artificial para el desafío cada vez mayor de aplicar el aprendizaje automático y modelos cada vez más complejos ya sea por su alta dimensionalidad o bien por su complejidad en la calibración óptima de los hiperparámetros (He et al., 2021). El AutoML utiliza técnicas que eliminan el trabajo manual que se utiliza a la hora de preparar las bases de datos, entrenar modelos de aprendizaje automático y calibrar sus hiperparámetros, aplicar el modelo en un entorno de producción además de permitir monitorizar el rendimiento del modelo seleccionado. Por tanto, cubre la elaboración completa de pipelines que abarcan desde el conjunto de datos sin procesar hasta el modelo final implementable (Karmaker et al., 2021).

En el presente trabajo se postula que las estrategias de AutoML pueden representar un enfoque prometedor para aportar evidencias de validez de criterio predictiva estudiando su estabilidad entre muestras. Para ello, se utiliza un ejemplo aplicado basado en el cuestionario VIA-IS de fortalezas personales. Este instrumento está basado en el modelo VIA de fortalezas personales de Peterson y Seligman (2004). Dicho instrumento ha mostrado tener buenos indicadores de fiabilidad de sus puntuaciones (Ovejero et al., 2018) y, además, se ha empleado ampliamente para analizar la relación de las fortalezas con diferentes indicadores, entre ellos, la satisfacción vital (Ovejero et al., 2016) aportando así ciertas evidencias de validez de criterio. Los análisis más utilizados para poner a prueba la hipótesis de la relación entre estos rasgos de personalidad y el bienestar subjetivo han estado basados en modelos lineales de carácter correlacional (por ejemplo, Azañedo et al., 2014; Baumann et al., 2020; Buschor et al., 2013, Kretschmar et al., 2022; Park et al., 2004) y predictivo (por ejemplo, Brdar et al., 2011; Peterson et al., 2007; Proyer et al., 2011). La evidencia empírica arroja una serie de resultados consistentes concluyendo que no todas las fortalezas tienen la misma capacidad predictiva siendo la curiosidad, la gratitud, la esperanza, y la capacidad de amar los rasgos más relacionados con la satisfacción con la vida (Buschor et al., 2013; Niemiec, 2013; Ovejero et al., 2016; Park et al., 2004; Peterson et al., 2007; Proyer et al., 2011; Shimai et al., 2006). Aunque los resultados apuntan a cierta generalización de los resultados, no están exentos de limitaciones, entre otras, que asumen que el uso de modelos lineales permite obtener los mejores resultados predictivos y que no se ha dividido el conjunto de datos en sets de entrenamiento y de testeo a pesar de que algunos estudios disponen de muestras de gran tamaño (por ejemplo, Linley et al., 2007). Esto implica que no es posible evaluar si el modelo elegido es el mejor dados los datos y la distribución de los mismos y, adicionalmente, no es posible estudiar la estabilidad de la calidad del modelo construido en muestras de la misma población de referencia. Ambas circunstancias pueden suponer un problema que afecta a la estabilidad de las evidencias de validez predictiva de las puntuaciones obtenidas por este instrumento.

En vista de lo anterior, en el presente trabajo se analizará si el AutoML aporta información adicional acerca de la validez predictiva de las puntuaciones del VIA-IS a la hora de pronosticar la satisfacción vital tal y como indica el modelo teórico en el que se sustenta (Peterson & Seligman, 2004). Además, se estudiará si el AutoML permite mejorar la capacidad predictiva de los constructos evaluados por el VIA-IS así como evaluar la estabilidad de las predicciones en un conjunto de datos diferente al usado en la construcción del modelo. Finalmente, se indagará si la hipótesis de las fortalezas más relacionadas con la satisfacción vital son también las que tienen un mayor peso en el mejor de los modelos.

Método

Participantes y procedimiento

La muestra de este estudio está compuesta por 1235 estudiantes universitarios (79.6% mujeres, 20.3% hombres, 0.2% de los participantes no dieron respuesta acerca de su sexo) que en el momento de la evaluación cumplieron los criterios de inclusión y ninguno de exclusión presentados en la Tabla 1. La media de edad fue de 20,88 años (DT=3.76).

En lo referido al procedimiento, se solicitó permiso al profesorado para poder explicar la finalidad de la investigación y repartir los cuestionarios al alumnado. Dicho profesorado y su estudiantado correspondiente fue seleccionado mediante un muestreo polietápico por ubicaciones de cada facultad y estratificado por titulación y curso. A continuación, todo el alumnado que participó en la investigación recibió información sobre el propósito de la investigación, además de solicitar su consentimiento informado acorde a la legislación vigente, asegurando que el tratamiento de los datos sería de forma colectiva, nunca de forma individual además de ser anónimos. Tras estas explicaciones, el alumnado completó los cuestionarios descritos en el siguiente apartado.

Instrumentos de evaluación

Se utilizaron preguntas de sociodemografía en la que los estudiantes respondieron cuestiones acerca de su sexo, edad, y qué carrera estaban estudiando. Tras estas preguntas se administraron

Tabla 1. Criterios de inclusión y exclusión

Inclusión	Exclusión
Alumnado que en el momento de la evaluación estaba matriculado en algún grado impartido en la Universidad Complutense de Madrid.	Alumnado de otros centros universitarios no pertenecientes a la Universidad Complutense de Madrid. Estudiantes que en el momento de la recogida de datos están matriculados en cursos superiores al grado. Estudiantes que forman parte de programas de movilidad del estudiantado (por ejemplo, Erasmus, etc.) que hayan permanecido menos de 6 meses realizando sus estudios en la Universidad Complutense de Madrid. Estudiantado con dificultad para la comprensión de los cuestionarios.

el cuestionario VIA de fortalezas personales (VIA-IS) y la Escala de Satisfacción General con la Vida (escala SWLS):

Cuestionario VIA de fortalezas personales (*Values In Action Inventory of Strengths o VIA-IS*, Peterson y Seligman, 2004) mide el grado en que un individuo posee cada una de las 24 fortalezas que propone el modelo VIA. Agrupadas en cada una de las 6 virtudes propuestas por Peterson y Seligman (2004), estos 24 rasgos son: Sabiduría y Conocimiento (Curiosidad, Amor por el conocimiento, Pensamiento crítico, Creatividad y Perspectiva), Coraje (Valentía, Perseverancia, Integridad, Vitalidad), Humanidad (Capacidad de amar y ser amado, Generosidad, Inteligencia emocional), Justicia (Civismo, Sentido de la justicia, Liderazgo), Moderación (Capacidad de perdonar, Humildad, Prudencia, Autocontrol) y Trascendencia (Apreciación de la belleza, Gratitud, Esperanza, Sentido del humor, Espiritualidad). El VIA-IS está compuesto por 240 ítems en una escala Likert de cinco puntos (desde uno: totalmente en desacuerdo hasta cinco: totalmente de acuerdo). Los datos del estudio de Ovejero et al. (2018) permiten comprobar que, en general, las fortalezas tienen una consistencia interna aceptable y están moderadamente relacionadas con variables de bienestar subjetivo como la satisfacción vital.

Escala de Satisfacción General con la Vida (del inglés, *Satisfaction With Life Scale*, Diener et al., 1985; Pavot, et al., 1991; Pavot & Diener, 2008, 2009; adaptación española de Vázquez, Duque y Hervás, 2013) consta de cinco ítems en escala tipo Likert de siete puntos (desde uno: completamente en desacuerdo hasta siete: completamente de acuerdo). La puntuación total supone la suma de cada uno de los ítems. Los análisis psicométricos

efectuados en muestras representativas españolas revelan una estructura unifactorial, así como una consistencia interna de .88 (Vázquez et al., 2013).

Análisis de datos

Los análisis se llevaron a cabo íntegramente en Python (versión 3.8.12). El código utilizado para el análisis de datos está a disposición bajo petición razonable a los autores del presente trabajo.

El análisis de datos comenzó con un estudio descriptivo de las variables de sociodemografía, las fortalezas y la satisfacción vital. Para ello se calcularon estadísticos de tendencia central, variabilidad y forma de la distribución en cada variable. Este análisis descriptivo se realizó utilizando la librería pandas (versión 1.3.5) (The Pandas Development Team, 2020) de Python.

Tras el análisis descriptivo se construyó el pipeline de AutoML utilizando el paquete de Python pycaret (versión 2.3) (Ali, 2020). Para el uso de pycaret se configuró una sesión en un entorno independiente dentro de Python. El primer paso consistió en el procesamiento de las variables que se caracterizó por la imputación de valores perdidos mediante la mediana de la distribución de la variable afectada y la estandarización de cada variable basada en el rango intercuartil. Esta metodología se utilizó dado que las distribuciones de las variables objeto de estudio presentan asimetrías y también datos atípicos. A continuación, en esta sesión de pycaret, la base de datos se dividió de forma aleatoria en dos partes con una semilla de replicación de resultados, un 70% de los datos se utilizaron para entrenar los modelos predictivos y el 30% para testarlos. De los diferentes modelos predictivos que dispone pycaret, en el presente trabajo se configuraron 17 modelos predictivos. Como método de validación se utilizó el método 10-fold CV, es decir un método de validación

Tabla 2. Estadísticos descriptivos

Variable	Media	DT	Mínimo	Máximo	Asimetría	Curtosis	ω
Curiosidad	36.90	5.22	18.00	50.00	-0.08	0.02	.784
Amor por el conocimiento	36.30	6.40	17.00	50.00	-0.21	-0.35	.829
Mentalidad abierta	37.90	5.03	20.00	50.00	-0.12	0.04	.778
Creatividad	35.40	6.10	13.00	50.00	-0.10	0.12	.851
Sabiduría	35.40	4.97	19.00	50.00	0.02	0.17	.763
Valentía	36.30	5.45	12.00	50.00	-0.43	0.56	.767
Persistencia	36.60	6.71	10.00	50.00	-0.42	0.05	.878
Autenticidad	39.40	4.63	19.00	50.00	-0.36	0.24	.731
Vitalidad	34.40	5.38	15.00	50.00	-0.15	0.21	.775
Amor	40.40	5.64	15.00	50.00	-0.83	0.98	.770
Bondad	41.80	4.56	18.00	50.00	-0.66	0.74	.776
Inteligencia social	36.40	4.78	20.00	50.00	-0.14	0.09	.722
Trabajo en equipo	38.10	4.82	17.00	50.00	-0.43	0.70	.724
Liderazgo	37.00	5.37	14.00	50.00	-0.32	0.59	.789
Imparcialidad	40.10	4.97	17.00	50.00	-0.62	0.77	.765
Autocontrol	32.80	5.79	14.00	50.00	-0.16	-0.13	.691
Prudencia	34.00	6.00	13.00	48.00	-0.25	0.06	.773
Modestia	34.80	5.49	10.00	50.00	-0.41	0.69	.747
Perdón	34.50	6.79	10.00	50.00	-0.32	0.05	.851
Espiritualidad	27.80	8.41	12.00	50.00	0.54	-0.26	.861
Sentido del humor	37.90	5.60	12.00	50.00	-0.44	0.49	.834
Esperanza	37.20	6.03	13.00	50.00	-0.61	0.64	.803
Gratitud	38.80	5.13	17.00	50.00	-0.44	0.19	.777
Apreciación de la belleza	36.60	6.31	13.00	50.00	-0.26	-0.07	.813
Satisfacción vital	24.10	5.64	5.00	35.00	-0.57	0.09	.843

Nota. DT: Desviación típica. ω : Omega de McDonald.

cruzada con 10 divisiones de la base de datos, utilizándose los indicadores descritos a continuación:

- Coeficiente de determinación (R^2): mide cómo de distantes están los datos de los valores predichos del modelo en comparación con los datos de la media.
- Error Absoluto Medio (*Mean Absolute Error* o MAE): se refiere a la magnitud de la diferencia entre la predicción de una observación y el valor real de esa observación.
- Error cuadrático medio (*Mean Square Error* o MSE): se define como la diferencia entre las predicciones de su modelo y la realidad, se eleva al cuadrado y se promedia en todo el conjunto de datos.
- Raíz cuadrada del error cuadrático medio o la desviación cuadrática media (*Root Mean Square Error* o RMSE): muestra cómo de distantes son las predicciones de los valores verdaderos usando la distancia euclidiana.
- Error cuadrático medio de los valores predichos transformados logarítmicamente y reales transformados logarítmicamente (*Root Mean Square Log Error* o RMSLE): agrega 1 a los valores reales y predichos antes de utilizar el logaritmo natural para evitar calcular el logaritmo natural de posibles valores iguales a 0.
- El error porcentual absoluto medio (*Mean Absolute Percentage Error* o MAPE): es una medida

estadística para definir la precisión en un conjunto de datos en particular interpretándose como la diferencia promedio entre el valor pronosticado y el valor real en porcentaje.

Tras estudiar los 17 modelos predictivos, se analizó en profundidad el mejor de los modelos construidos, estudiando los hiperparámetros mediante el procedimiento 10-fold CV, analizándose el rendimiento de los diferentes indicadores en cada paso, para promediar el resultado. Finalmente, se estudió la importancia de las variables con mayor R^2 y se construyeron los gráficos de ajuste del modelo definitivo para finalmente estudiar el mejor de los modelos en el conjunto de testeo.

Resultados

El análisis descriptivo de cada variable se expone en la Tabla 2. Se observa que las fortalezas y la satisfacción vital se caracterizan por tener puntuaciones con asimetría negativa, es decir, los participantes tienden a puntuar alto en estas variables. Además, se observa que la consistencia interna, evaluada mediante el coeficiente Omega de McDonald, es adecuada.

Como se ha indicado en la sección de análisis estadístico, para poder aplicar el AutoML, se han

Tabla 3. Resultados de la aplicación del procedimiento de AutoML

Modelo	MAE	MSE	RMSE	R ²	RMSLE	MAPE	Tiempo de ejecución (segundos)
Bayesian Ridge Regression	3.539	2.060	4.468	.344	.209	0.177	0.007
Elastic Net	3.544	2.175	4.481	.340	.210	0.178	0.009
OLS regression	3.553	2.202	4.483	.339	.209	0.177	0.774
Ridge regression	3.553	2.202	4.483	.339	.209	0.177	0.008
Huber regression	3.536	2.463	4.511	.331	.211	0.177	0.038
Lasso regression	3.575	2.570	4.526	.327	.212	0.180	0.009
Gradient boosting regressor	3.624	21.221	4.598	.306	.215	0.181	0.074
Random Forest	3.732	22.100	4.693	.275	.220	0.187	0.207
AdaBoost	3.729	22.250	4.708	.272	.220	0.186	0.061
Extra trees regressor	3.727	22.289	4.715	.268	.221	0.221	0.152
Orthogonal matching pursuit	3.804	23.193	4.809	.239	.226	0.226	0.008
Light gradient boosting machine	3.809	23.559	4.841	.226	.225	0.225	0.061
K-NN regression	3.953	25.125	4.997	.183	.233	0.233	0.011
Lasso least angle regression	4.497	31.406	5.586	-.002	.261	0.261	0.020
Dummy regressor	4.497	31.406	5.586	-.002	.261	0.261	0.011
Passive aggressive regressor	5.070	39.651	6.095	-.330	.285	0.285	0.010
Decision tree	5.272	46.121	6.781	-.528	.317	0.317	0.010

Nota. MAE: Mean Absolute Error; MSE: Mean Squared Error; RMSE: Root Mean Squared Error; R²: R cuadrado; RMSLE: Root Mean Squared Logarithmic Error; MAPE: Mean Absolute Percentage Error.

Tabla 4. Resultados tras la calibración de hiperparámetros de la Bayesian Ridge Regression

Set	MAE	MSE	RMSE	R ²	RMSLE	MAPE
Entrenamiento	3.539	20.054	4.467	.345	0.209	0.177
Testeo	3.794	22.836	4.779	.306	0.222	0.189

Nota. MAE: Mean Absolute Error; MSE: Mean Squared Error; RMSE: Root Mean Squared Error; R²: R cuadrado; RMSLE: Root Mean Squared Logarithmic Error; MAPE: Mean Absolute Percentage Error. Hiperparámetros calibrados mediante grid search y testeados mediante 10-fold CV. Hiperparámetros: alfa 1=1; alfa 2=1; lambda 1=.01; lambda 2=1x10⁻⁶; número de iteraciones=300; tolerancia=.001.

seguido los pasos asociados al preprocesamiento y estandarización de variables, así como el tratamiento de valores perdidos. Los modelos construidos en el subconjunto de entrenamiento y los resultados de ajuste de cada uno de ellos, se presentan en la Tabla 3.

El modelo que mejores resultados obtuvo fue el Bayesian Ridge Regression. El estudio y calibración de hiperparámetros de este modelo se expone en la Tabla 4. Se ajustaron los hiperparámetros mediante grid search y, una vez obtenidos dichos hiperparámetros, se pusieron a prueba mediante el procedimiento 10-fold CV, observándose que el R² en la muestra de entrenamiento fue igual a .345 y en testeo fue igual a .305. En lo que respecta a las variables más importantes para el modelo fueron la gratitud, la curiosidad, la esperanza, la apreciación de la belleza, la bondad, el amor, la sabiduría, la creatividad, la vitalidad y la imparcialidad mientras que el resto de fortalezas evaluadas en el VIA-IS tuvieron una importancia menor (ver Figura 1).

Discusión

En el presente trabajo se analizó la capacidad que tiene el pipeline basado en el AutoML para añadir información adicional acerca de la validez predictiva de las puntuaciones del VIA-IS a la hora de pronosticar las puntuaciones en la escala SWLS de satisfacción vital. Por otra parte, se analizó también si el AutoML permite mejorar la capacidad predictiva de los constructos evaluados por el VIA-IS y se exploró la estabilidad de las predicciones en un conjunto de datos diferente al usado en la construcción del modelo.

En el análisis de la validez de criterio, el modelo final construido con el pipeline de AutoML permite explicar un 34.5% de varianza de la satisfacción vital a partir de las puntuaciones obtenidas en las dimensiones evaluadas por el instrumento VIA-IS en el set de entrenamiento, siendo ligeramente inferior en el de testeo (30.5%). Con ello, se muestra que existe evidencia de que los constructos evaluados por el VIA-IS predicen en cierto grado las puntuaciones en el cuestionario de satisfacción vital (SWLS), tal y

como indica el modelo teórico de las fortalezas personales (Peterson & Seligman, 2004).

Adicionalmente, la rapidez en la ejecución y calibración de los hiperparámetros del mejor de los modelos permiten ver que el AutoML puede ser aplicado para resolver problemas psicométricos añadiendo una capa adicional de evidencia cuando se emplean en comparación con métodos tradicionales del análisis de la validez de criterio (Trognon et al., 2022). Por ello, el pipeline de AutoML ha permitido explorar diferentes modelos alternativos a la regresión por mínimos cuadrados ordinarios mostrando que las evidencias de la validez de criterio se pueden explorar de una forma rápida, ordenada y automática aportando evidencia sobre la estabilidad intermuestras de los resultados obtenidos como un valor adicional a los modelos tradicionalmente utilizados para la validez predictiva de las puntuaciones del VIA-IS en lo relativo a la satisfacción vital evaluada mediante la escala SWLS (Brdar et al., 2011; Peterson et al., 2007; Proyer et al., 2011).

En este trabajo también se consideró el análisis de las fortalezas más relacionadas con la satisfacción vital. Los resultados muestran que las variables más importantes para predecir las puntuaciones en la escala SWLS son la gratitud, la curiosidad, la esperanza, la apreciación de la belleza, la bondad, el amor, la sabiduría, la creatividad, la vitalidad y la imparcialidad apoyando parcialmente a la evidencia empírica disponible hasta la fecha, según la cual, hay cinco de estos rasgos que tienen mayor relación con la satisfacción vital, a saber, la curiosidad, la gratitud, la esperanza, y la capacidad de amar (Buschor et al., 2013; Niemiec, 2013; Ovejero et al., 2016; Park et al., 2004; Peterson et al., 2007; Proyer et al., 2011; Shimai et al., 2006). Por ello, los resultados del análisis de la segunda hipótesis aportan evidencia empírica de la validez de criterio predictiva de las fortalezas personales y su relación con la satisfacción vital, por tanto, han permitido aportar robustez al modelo propuesto originariamente por Peterson y Seligman (2004).

No obstante, aunque los hallazgos son esperanzadores dado que se expone que el AutoML es una herramienta adicional para explorar la validez de criterio, es esencial que tanto las puntuaciones de las dimensiones del test

evaluado como del criterio de interés tengan una fiabilidad adecuada (Vacha-Haase & Thompson, 2011). Es decir, como todo modelo y paradigma de aprendizaje automático, la calidad del dato es esencial y cualquier automatización de los procesos requiere siempre que los datos aportados estén limpios y que sean precisos y fiables (Rose & Fischer, 2011). Por otra parte, aunque gran parte de la algoritmia queda automatizada, es importante no perder de vista que se produce cierto efecto de caja negra que solo se puede mitigar mediante el conocimiento de los algoritmos estudiados y de los hiperparámetros característicos de cada uno de ellos (Watson et al., 2019). Además, para el uso de los modelos de AutoML es esencial disponer de instrumentos multidimensionales y de tamaños muestrales amplios.

Para finalizar, las futuras líneas de investigación relacionadas con el AutoML y la validez de criterio predictiva podrían extenderse a instrumentos de inteligencia, personalidad, psicopatología siendo áreas complejas cuyos instrumentos de evaluación son ampliamente utilizados para la elaboración de modelos y donde la capacidad predictiva del instrumento es esencial que sea de calidad dadas las diversas implicaciones en la toma de decisiones que derivan de sus aplicaciones.

Referencias

- Ali, M. (2020). *PyCaret: An open source, low-code machine learning library in Python*. Disponible en: <https://www.pycaret.org>.
- American Educational Research Association, American Psychological Association & National Council on Measurement in Education (AERA, APA, NCME) (2018). *Standards for educational and psychological testing*. American Psychological Association.
- Azañedo, C. M., Fernández-Abascal, E. G., & Barraca, J. (2014). Character strengths in Spain: Validation of the Values in Action Inventory of Strengths (VIA-IS) in a Spanish sample. *Clínica y Salud*, 25, 123-130. <https://doi.org/10.1016/j.clysa.2014.06.002>
- Baumann, D., Ruch, W., Margelisch, K., Gander, F., & Wagner, L. (2019). Character strengths

- and life satisfaction in later life: An analysis of different living conditions. *Applied Research in Quality of Life*, 15, 329-347.
<https://doi.org/10.1007/s11482-018-9689-x>
- Brdar, I., Anić, P., & Rijavec, M. (2011). Character strengths and well-being: Are there gender differences? En Bridar, I. (Ed.), *The human pursuit of well-being* (pp. 145-156). Springer.
- Buschor, C., Proyer, R. T., & Ruch, W. (2013). Self- and peer-rated character strengths: How do they relate to satisfaction with life and orientations to happiness? *Journal of Positive Psychology*, 8, 116-127.
<https://doi.org/10.1080/17439760.2012.758305>
- Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment*. SAGE.
- Diener, E. D., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment*, 49, 71-75.
https://doi.org/10.1207/s15327752jpa4901_13
- He, X., Zhao, K., & Chu, X. (2021). AutoML: A survey of the state-of-the-art. *Knowledge-Based Systems*, 212.
<https://doi.org/10.1016/j.knosys.2020.106622>
- Hutter, F., Kotthoff, L., & Vanschoren, J. (2019). *Automated Machine Learning. Methods, systems, challenges*. Springer.
- Karmaker, S. K., Hassan, M. M., Smith, M. J., Xu, L., Zhai, C., & Veeramachaneni, K. (2022). AutoML to date and beyond: Challenges and opportunities. *ACM Computing Surveys*, 54, 1-36.
<https://doi.org/10.1145/3470918>
- Kimberlin, C. L., & Winterstein, A. G. (2008). Validity and reliability of measurement instruments used in research. *American Journal of Health-System Pharmacy*, 65, 2276-2284.
<https://doi.org/10.2146/ajhp070364>
- Kretschmar, A., Harzer, C., & Ruch, W. (2022). Character strengths in adults and adolescents: Their measurement and association with well-being. *Journal of Personality Assessment*.
<https://doi.org/10.31234/osf.io/caemw>
- Ledesma, R. D., Ferrando, P. J., & Tosi, J. D. (2018). Uso del análisis factorial exploratorio en RIDEP. Recomendaciones para autores y revisores. *Revista Iberoamericana de Diagnóstico y Evaluación – e Avaliação Psicológica*, 52, 173-180.
<https://doi.org/10.21865/RIDEP52.3.13>
- Linley, P. A., Maltby, J., Wood, A. M., Joseph, S., Harrington, S., Peterson, C., ... & Seligman, M. E. (2007). Character strengths in the United Kingdom: The VIA Inventory of Strengths. *Personality and Individual Differences*, 43, 341-351.
<https://doi.org/10.1016/j.paid.2006.12.004>
- Messick, S. (1989). Validity. En R. Linn (Ed.), *Educational Measurement* (pp. 13-103). Macmillan.
- Niemiec, R. M. (2013). VIA character strengths: Research and practice (The first 10 years). En H. Henrik Knoop y A. Delle Fave (Eds.), *Well-being and cultures* (pp. 11-29). Springer.
- Ovejero, M. M., Brabete, A. C., & Alvarado, J. M. (2018). Reliability Generalization as a seal of quality of substantive meta-analyses: The case of the VIA Inventory of Strengths (VIA-IS) and their relationships to life satisfaction. *Psychological Reports*.
<https://doi.org/10.1177%2F0033294118779198>
- Ovejero, M. M., Cardenal, V., & Ortiz-Tallo, M. (2016). Fortalezas humanas y bienestar biopsicosocial: Revisión sistemática. *Escritos De Psicología - Psychological Writings*, 9, 4-14.
<https://doi.org/10.24310/espiescpsi.v9i3.13212>
- Park, N., Peterson, C., & Seligman, M. E. (2004). Strengths of character and well-being. *Journal of Social and Clinical Psychology*, 23, 603-619.
<https://doi.org/10.1521/jscp.23.5.603.50748>
- Pavot, W., & Diener, E. (2008). The Satisfaction with Life Scale and the emerging construct of life satisfaction. *The Journal of Positive Psychology*, 3, 137-152.
<https://doi.org/10.1080/17439760701756946>
- Pavot, W., & Diener, E. (2009). Review of the Satisfaction with Life Scale. En E. Diener (Ed.), *Assessing well-being* (pp. 101-117). Springer.
https://doi.org/10.1007/978-90-481-2354-4_5
- Peterson, C., Ruch, W., Beermann, U., Park, N., & Seligman, M. E. (2007). Strengths of character, orientations to happiness, and life satisfaction. *The Journal of Positive Psychology*, 2, 149-156.

- <https://doi.org/10.1080/17439760701228938>
Peterson, C., & Seligman, M. E. (2004). *Character strengths and virtues: A handbook and classification*. Oxford University Press.
- Proyer, R. T., Gander, F., Wyss, T., & Ruch, W. (2011). The relation of character strengths to past, present, and future life satisfaction among German-speaking women. *Applied Psychology: Health and Well-Being*, 3, 370-384.
<https://doi.org/10.1111/j.1758-0854.2011.01060.x>.
- Rose, L. T., & Fischer, K. W. (2011). Garbage in, garbage out: Having useful data is everything. *Measurement: Interdisciplinary Research & Perspective*, 9, 222-226
<https://doi.org/10.1080/15366367.2011.632338>
- Shimai, S., Otake, K., Park, N., Peterson, C., & Seligman, M. E. (2006). Convergence of character strengths in American and Japanese young adults. *Journal of Happiness Studies*, 7, 311-322.
<https://doi.org/10.1007/s10902-005-3647-7>
- Sullivan, G. M. (2011). A primer on the validity of assessment instruments. *Journal of Graduate Medical Education*, 3, 119-120.
<https://doi.org/10.4300/jgme-d-11-00075.1>
- Taherdoost, H. (2016). Validity and reliability of the research instrument. How to test the validation of a questionnaire/survey in a research. *SSRN Electronic Journal*.
<https://dx.doi.org/10.2139/ssrn.3205040>
- The Pandas Development Team (2020). *Pandas-dev/pandas: Pandas*.
<https://doi.org/10.5281/zenodo.3509134>.
- Toro, R., Peña-Sarmiento, M., Avendaño-Prieto, B. L., Mejía-Vélez, S., & Bernal-Torres, A. (2022). Análisis empírico del coeficiente alfa de Cronbach según opciones de respuesta, muestra y observaciones atípicas. *Revista Iberoamericana de Diagnóstico y Evaluación – e Avaliação*, 63, 17-30.
<https://doi.org/10.21865/RIDEP63.2.02>.
- Trognon, A., Cherifi, Y., Demange, L., & Prudent, C. (2022). Viability of machine-learning strategies to solve psychometric problems. *Research Square*.
<https://doi.org/10.21203/rs.3.rs-1314080/v1>
- Vacha-Haase, T., & Thompson, B. (2011). Score reliability: A retrospective look back at 12 years of reliability generalization studies. *Measurement and Evaluation in Counseling and Development*, 44, 159-168.
<https://doi.org/10.1177/0748175611409845>
- Warne, R. M. (2008). *Applied statistics: From bivariate through multivariate techniques*. Sage.
- Watson, D., Krutzinna, J., Bruce, I., Griffiths, C., McInnes, I., Barnes, M., & Floridi, L. (2019). Clinical Applications of Machine Learning Algorithms: Beyond the Black Box. *SSRN Electronic Journal*.
<https://doi.org/10.2139/ssrn.3352454>.