



Ingeniería, investigación y tecnología

ISSN: 1405-7743

Universidad Nacional Autónoma de México, Facultad de Ingeniería

Navarro-Acosta, J. Alejandro; Soto-Mendoza, Valeria; Saucedo-Zendejo, Felix Raymundo; Guajardo-Espinoza, José María; Rivera-Morales, María Teresa  
Evaluación psicológica de profesores y alumnos mexicanos durante la pandemia de COVID-19 mediante técnicas de *Machine learning*  
Ingeniería, investigación y tecnología, vol. XXII, núm. 4, e1805, 2021, Octubre-Diciembre  
Universidad Nacional Autónoma de México, Facultad de Ingeniería

DOI: <https://doi.org/10.14482/INDES.30.1.303.661>

Disponible en: <https://www.redalyc.org/articulo.oa?id=40471804002>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

UNAM  redalyc.org

Sistema de Información Científica Redalyc  
Red de Revistas Científicas de América Latina y el Caribe, España y Portugal  
Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto



## Evaluación psicológica de profesores y alumnos mexicanos durante la pandemia de COVID-19 mediante técnicas de *Machine learning*

### Machine learning approaches for psychological assessment of mexican professors and students during COVID-19 pandemic

Navarro-Acosta J. Alejandro

Universidad Autónoma de Coahuila

Centro de Investigación en Matemáticas Aplicadas, México

Correo: [alejandro.navarro@uadec.edu.mx](mailto:alejandro.navarro@uadec.edu.mx)

<https://orcid.org/0000-0002-2979-8269>

Soto-Mendoza Valeria

Universidad Autónoma de Coahuila

Centro de Investigación en Matemáticas Aplicadas, México

Correo: [vsoto@uadec.edu.mx](mailto:vsoto@uadec.edu.mx)

<https://orcid.org/0000-0001-8171-8994>

Saucedo-Zendejo Felix Raymundo

Universidad Autónoma de Coahuila

Centro de Investigación en Matemáticas Aplicadas, México

Correo: [fesaucedo@uadec.edu.mx](mailto:fesaucedo@uadec.edu.mx)

<https://orcid.org/0000-0001-8434-6129>

Guajardo-Espinoza José María

Universidad Autónoma de Coahuila

Unidad Saltillo, Facultad de Psicología, México

Correo: [jguajar@uadec.edu.mx](mailto:jguajar@uadec.edu.mx)

<https://orcid.org/0000-0003-4728-4648>

Rivera-Morales María Teresa

Universidad Autónoma de Coahuila

Unidad Saltillo, Facultad de Psicología, México

Correo: [teresa.rivera@uadec.edu.mx](mailto:teresa.rivera@uadec.edu.mx)

<https://orcid.org/0000-0002-8660-9786>

#### Resumen

En la presente obra se describe la realización de un ejercicio de validación de resultados de una prueba psicológica aplicada a maestros y alumnos en estado de aislamiento provocado por la pandemia debida al COVID-19 en el estado de Coahuila, México. El objetivo de este trabajo es aplicar técnicas de *machine learning* para validar un instrumento que mide las emociones y los sentimientos negativos, así como el sesgo cognitivo o desviación de pensamiento sobre la educación y la pandemia en situación de aislamiento. Para el cumplimiento del objetivo se aplicó un instrumento en formato electrónico que se diseminó en el estado de Coahuila, fue respondido por los usuarios y se generó una base de datos, la cual después de su preprocesamiento fue analizada mediante la combinación de *Random forest* (RF) y *Support Vector Machines* (SVM); donde se obtuvo como resultado la pertinencia o no de algunos de los reactivos en las pruebas, lo que dio una validez interna al instrumento. Los resultados experimentales muestran que la metodología propuesta es capaz de seleccionar las variables predictoras más relevantes. De esta manera, se obtienen resultados satisfactorios en la clasificación y predicción de diagnósticos psicológicos globales y segmentados por características de los respondientes. Por otro lado, aunque las técnicas implementadas son robustas y confiables, estas presentan limitaciones en cuanto a la observación de los otros tipos de validez: la de constructo y la externa, entre otras; lo cual pudiera limitar su utilización. Si bien, en el campo de la psicometría existen diversas estrategias clásicas, la metodología propuesta basada en la combinación de técnicas de *machine learning* para el análisis y validación de este tipo de pruebas, favorece el crecimiento de opciones para mejorar los diagnósticos y en consecuencia el tratamiento de padecimientos psicológicos.

**Descriptores:** Psicometría, validez, clasificación, predicción, *machine learning*.

#### Abstract

This work describes the validation of the results of a psychological test applied to teachers and students in isolation due to the COVID-19 pandemic in the state of Coahuila, Mexico. The objective of this work is to apply machine learning techniques to validate an instrument that measures negative emotions and feelings, as well as cognitive bias or deviation of thinking about education and the pandemic in isolation. For the fulfillment of the objective, an instrument was applied in electronic format that was disseminated in the state of Coahuila, the users respond and the database is generated, which, after its pre-processing, is analyzed using the combination of Random Forest (RF) and Support Vector Machines (SVM); obtaining as a result the relevance or not of some of the items from the tests, thereby giving an internal validity to the instrument. The experimental results show that the proposed methodology is capable of selecting the most relevant predictor variables. In this way, satisfactory results were obtained in the classification and prediction of psychological diagnoses. On the other hand, although the implemented techniques are robust and reliable, they present limitations in terms of the observation of the other types of validity: construct, external, among others; which could limit its use. Although, in the field of psychometry there are various classic strategies, the proposed methodology based on the combination of machine learning techniques for the analysis and validation of this type of tests, favors the growth of options to improve diagnoses and consequently the treatment of psychological ailments.

**Keywords:** Psychometry, validity, classification, prediction, machine learning.

## INTRODUCCIÓN

A finales de diciembre de 2019 en la ciudad China de Wuhan, médicos reportaron pacientes con neumonía viral causada por un patógeno desconocido. En enero 30 de 2020, la Organización Mundial de la Salud anunció la emergencia sanitaria debido a un nuevo coronavirus (SARS-CoV-2) y posteriormente catalogó a la enfermedad como pandemia, debido a su rápida propagación entre los países. Para mayo de 2020, la cifra de casos positivos confirmados ascendía a los 3,581,884 mientras que las defunciones eran alrededor de 248,558. Para el mes de junio de 2020, la enfermedad ya se había extendido por más de 185 países y afectado a más de 7,145,800 personas (Tuli, 2020). Debido a esto, la mayoría de los países tomaron acciones para frenar la propagación de la enfermedad.

Por su parte, en México, el 28 de febrero el gobierno de la Ciudad confirmó el primer infectado de SARS-CoV-2 y para el 16 de marzo, la Secretaría de Salud anunció la implementación de la jornada de sana distancia. Es decir, se suspendieron todas las actividades catalogadas como no esenciales y se adelantó el periodo vacacional. En un principio esta jornada culminaría el 19 de abril de 2020, sin embargo, esto se postergó hasta el 30 de junio del mismo año. Factores como el confinamiento, la incertidumbre, el miedo, entre otros; han causado estragos psicológicos en las personas. De acuerdo con Liang *et al.* (2020) este tipo de epidemias llevan a las personas a experimentar problemas psicológicos como trastorno de estrés postraumático, angustia psicológica, depresión y ansiedad. Según la teoría del sistema inmunológico conductual, es probable que las personas desarrollen emociones negativas y una evaluación cognitiva negativa para la autoprotección (Li *et al.*, 2020).

La emoción es un componente de la psicología humana conformado por elementos fisiológicos que se expresan de forma instintiva y de aspectos cognitivos, así como socioculturales. Bisquerra *et al.* (2000) menciona que las emociones son un estado complejo del organismo caracterizado por una excitación o perturbación que predispone a una respuesta organizada, las cuales se generan habitualmente como respuesta a un acontecimiento externo e interno. Lawler & Thye (1999) definen las emociones como “estados evaluativos, sean positivos o negativos, relativamente breves, ya que tienen elementos fisiológicos, neurológicos y cognitivos”.

Por otro lado, los sentimientos son producto de la observación por parte de la mente, los cambios generados por las emociones (Pallarés, 2010). Son estados afectivos de baja intensidad y de larga duración. Por lo

tanto, emociones y sentimientos se conjuntan en el estado de ánimo, lo cual hace a un individuo reaccionar de forma pública (emociones) y de forma privada (sentimientos).

En lo que respecta al sesgo cognitivo, es una especie de efecto psicológico, el cual producirá desviaciones en el procesamiento psicológico, resultando en distorsiones, juicios incorrectos y explicaciones ilógicas, generalmente llamados fenómenos irracionales, los cuales ocurren con base en la interpretación de las cosas. De acuerdo con Kahneman (2012) el sesgo cognitivo es “el efecto psicológico que produce una desviación en el procesamiento de la percepción que causa una alteración al procesar esta información”.

En el área de psicología, uno de los pilares que tiene junto con la observación en campo y la entrevista, es la psicometría. Dicha estrategia es una herramienta que se ha utilizado desde el inicio de la disciplina como ciencia.

Aunado a las matemáticas, la psicometría ha estado avanzando cada vez más en la precisión de las observaciones de los fenómenos psicológicos en los diferentes estadios de desarrollo del ser humano con diversos rasgos de emociones, sentimientos y conductas, consideradas normales y anormales estadísticamente hablando. Del mismo modo, las matemáticas y la estadística también han hecho lo propio, dado el avance de los sistemas de cómputo y aplicaciones que manejan una gran cantidad de datos. En los últimos años se han desarrollado estrategias que mejoran el análisis, la clasificación y la predicción de fenómenos, y en este rubro, fenómenos psicológicos.

En la actualidad, el *machine learning* es una de las estrategias más empleadas para analizar varias fuentes de datos, incluidos mensajes de texto, comunicaciones en línea, redes sociales, artículos web, etcétera, ya que estos pueden ser muy útiles para analizar diversos aspectos de la pandemia que se relacionan con el comportamiento de las personas.

La presente investigación, considerada empírica e interdisciplinaria (psicología, matemáticas, informática), tiene como objetivo analizar, con estrategias de *machine learning*, un instrumento, respondido por alumnos y maestros, que mide las emociones y los sentimientos negativos, el sesgo cognitivo (desviación de pensamiento) sobre la educación y la pandemia en situación de aislamiento provocado por el COVID-19 en el estado de Coahuila, México.

## TRABAJO RELACIONADO

La pandemia por COVID-19 constituye una de las mayores crisis de salud pública a nivel mundial con cerca de 16,500,000 casos confirmados y 655,000 muertes en

todo el mundo (European Centre for Disease Prevention and Control, 2020). México es uno de los países más afectados por este virus con cerca de 400,000 casos confirmados y 46,000 muertes a finales de julio de 2020 (Secretaría de Salud, 2020). Esta pandemia no solo trajo consigo el riesgo de padecer enfermedades graves o morir a causa de la infección, sino también acarreó consecuencias adversas para la salud mental como ansiedad, depresión y estrés (Rajkumar, 2020). La continua propagación del virus, las medidas de aislamiento y contención de la pandemia, así como la suspensión de las clases presenciales en escuelas, colegios y universidades en todo el país constituyen un cambio significativo en la realidad de los estudiantes, lo que tiene una influencia directa en su salud mental.

Recientemente se ha reportado el impacto psicológico de la epidemia en el público en general: Pacientes, personal médico, niños, adultos mayores y estudiantes universitarios, quienes enfrentan la pandemia (Özdin *et al.*, 2020; González *et al.*, 2020a; Tull *et al.*, 2020; Huarcaya, 2020; Zolotov *et al.*, 2020; Elmer *et al.*, 2020). Así, Cao *et al.* (2020) reportaron que cerca de 24.9 % de estudiantes universitarios en China experimentaron algún nivel de ansiedad y que los efectos económicos, en la vida diaria, así como los retrasos en las actividades académicas causados por la pandemia se asociaron positivamente con los síntomas de ansiedad. Similarmente, Odriozola *et al.* (2020) reportaron un porcentaje similar de ansiedad en estudiantes españoles en conjunto con un 34 % de ellos, experimentando depresión causada por la repentina interrupción de actividades escolares, la consecuente separación de sus amistades y la preocupación constante por la salud y el trabajo de sus familias

Por otro lado, Ardan *et al.* (2020) encontraron que cerca de 40.3 % de estudiantes universitarios en Indonesia experimentaron algún nivel de ansiedad causado principalmente por el distanciamiento social y la preocupación por el contagio durante sus actividades diarias. Mientras que Rakhmanov & Dane (2020) reportaron un aumento en los niveles de ansiedad correlacionado con una disminución en los niveles de aprendizaje en estudiantes universitarios en África. Estos resultados sugieren la necesidad de monitorear la salud mental de los alumnos durante la pandemia. Sin embargo, a la fecha hay pocos estudios sobre el estado de salud mental de los estudiantes universitarios que enfrentan la pandemia en México (González *et al.*, 2020b; Cortés *et al.*, 2020).

La salud mental es un indicador del bienestar emocional, psicológico y social, por lo tanto, múltiples factores contribuyen a generarle problemas. Estos problemas de salud mental conducen a condiciones como el estrés,

ansiedad o depresión. Es por ello que la naturaleza de los algoritmos de *machine learning* e inteligencia artificial se pueden aprovechar para predecir la aparición de problemas de salud mental y pueden servir como una herramienta predictiva y de monitoreo ante situaciones de crisis como las que se viven hoy en día (Srividya *et al.*, 2018).

Por ejemplo, Srividya *et al.* (2018) aplicaron varios algoritmos de *machine learning* como *support vector machine*, *random forest*, árboles de decisión, clasificador *naïve bayes*, clasificador K vecinos más cercanos y regresión logística para identificar el estado de salud mental en estudiantes de nivel medio superior, estudiantes universitarios y trabajadores.

Similarmente, Ahuja & Banga (2019) utilizaron algoritmos de *machine learning* como una herramienta para detectar la prevalencia de estrés en estudiantes, considerando los algoritmos de clasificación *Random Forest*, *Naïve Bayes*, *SVM* y regresión lineal. Más aún, Shen *et al.* (2020) desarrollaron una aplicación basada en el algoritmo *random forest* para detectar de manera efectiva y precisa la probabilidad de un intento de suicidio en estudiantes de medicina en China, encontrando una exactitud de 90 % en las predicciones. Es por ello que la aplicación de algoritmos de *machine learning* e inteligencia artificial constituyen una herramienta prometedora para el diagnóstico de la salud mental y la detección oportuna de posibles peligros (Graham *et al.*, 2019; Shatte *et al.*, 2019; Mohammad *et al.*, 2018).

En consecuencia, esta herramienta novedosa ha sido utilizada para determinar y predecir consecuencias psicológicas durante la pandemia de COVID-19. Así por ejemplo, Samuel *et al.* (2020) aplicaron dos métodos de clasificación de *machine learning* para analizar texto relacionado con el COVID y determinar información sobre la progresión del sentimiento de miedo, observando una exactitud de clasificación de 91 % para textos cortos con el método *Naïve Bayes* y de 74 % con el método de regresión logística. Similarmente, Yang *et al.* (2020), introdujeron un nuevo trabajo de análisis sentimental utilizando más de 105 millones de mensajes en redes sociales para evaluar el aumento y la caída de los sentimientos durante la pandemia de COVID-19 al cual llamaron SenWave, utilizando una red neuronal profunda para la clasificación de los sentimientos, con la cual encontraron que los sentimientos optimistas y positivos aumentan con el tiempo, pronosticando un deseo de buscar, juntos, un restablecimiento para un mejor mundo post COVID-19.

Por otro lado, Khattar *et al.* (2020) estudiaron las actividades, los estilos de aprendizaje y la salud mental de estudiantes de la India durante la crisis ocasionada

por esta pandemia para evaluar cómo se están adaptando a los nuevos estilos de aprendizaje en línea y cómo están manejando su nueva realidad a través de reglas de asociación.

Sin embargo, estas herramientas no han sido utilizadas todavía para analizar instrumentos que miden el estado de salud psicológica de las comunidades universitarias durante la pandemia en México. Por lo que esta propuesta es muy relevante.

### RECOLECCIÓN DE DATOS

Un grupo de psicólogos diseñó un instrumento para medir el nivel de incertidumbre que produce el estado de aislamiento provocado por la pandemia de COVID-19 en el estado de Coahuila. El objetivo general de este instrumento fue validar las categorías de diagnóstico (es decir, los baremos) emocionales, sentimentales y de sesgo cognitivo que se observan ante la incertidumbre educativa en situación de aislamiento por dicha pandemia en alumnos y maestros. El instrumento se hizo disponible para responderse en línea. Consta de tres componentes medidos durante el distanciamiento

social provocado por la pandemia de COVID-19: El de la observación de emociones, sentimientos y sesgo cognitivo en el estado de Coahuila, México. En la Tabla 1 se pueden observar las variables que miden cada componente y subcomponente. En el componente del sesgo cognitivo sobre la educación, se ajusta la redacción del reactivo según sea el respondiente alumno o profesor.

A los alumnos y profesores se les invitó a responder el instrumento vía correo electrónico institucional, iniciando el 25 de junio del 2020 hasta el 6 de julio del mismo año. Para garantizar la privacidad de los datos se expone el aviso y el cumplimiento del aviso de privacidad de la universidad (Anexo 1), con el cual se garantiza que los datos, aportados voluntariamente, serán utilizados únicamente para fines de investigación. Todas las variables se midieron con una escala decimal, la cual mide el grado en que el respondiente presenta (en el momento de dar respuesta) cada uno de los atributos a observar. La escala va desde el cero hasta el 10, donde el cero implica la ausencia del atributo a medir y el 10, la máxima expresión del mismo.

Tabla 1. Descripción de componentes y variables del cuestionario aplicado

Componente	Subcomponente	Variables
Emociones		Ira
		Disgusto
		Amenaza
		Miedo
		Inseguridad
Sentimientos		Felicidad
		Tristeza
		Abandono
		Decepción
		Odio
		Amor
Sesgo cognitivo	sobre la pandemia	Paz
		Preocupación
		Culpabilidad
		Tranquilidad
		Ya pasará
		No es grave
	sobre la educación	Siguiendo las instrucciones, no pasa nada
		Moriré
		Me enfermarse
		Se enfermará alguien querido
		Todo es mejor por internet
		Me apegó a la información científica
		Toda información es buena
		Las noticias confirman lo que pienso
		No aprenderé/No podré enseñar
		Reprobaré/No tendré evidencia del aprendizaje de los alumnos
		La educación a distancia es de baja calidad, reprobaré el semestre/La educación a distancia es de baja calidad, reprobarán mis alumnos
		Pasaré la contingencia, pero no aprenderé/Pasaré la contingencia, pero no enseñaré
		Lo que aprenda por medios electrónicos no me servirá/Lo que enseñe por medios electrónicos no le servirá a los alumnos
		Puedo aprender lo que siempre he querido/Puedo enseñar lo que siempre he querido
		Me enfocaré en aprender/Me enfocaré en enseñar
		Los maestros/alumnos saben lo que hacen en cuanto a la educación a distancia
		Puedo complementar lo visto en clase/Puedo mejorar mi enseñanza
		Solo estudiando saldré adelante/Haciendo mi trabajo saldremos adelante

Fuente: Elaboración propia

## BASE DE DATOS

Para observar los fenómenos de las emociones, los sentimientos y el sesgo cognitivo en situación de aislamiento causado por la pandemia de COVID-19, se envió un cuestionario a docentes y alumnos en el estado de Coahuila, quienes se distribuyeron de la manera siguiente: el 55.46 % de los encuestados, son alumnos (1132) en el estado de Coahuila y el resto, 44.54 % son maestros en dicha entidad (909). El 62.7 % de quienes respondieron el instrumento son mujeres (1280) y 37.2 % hombres (761), de un total de 2041 cuestionarios respondidos. Las valoraciones correspondientes a las emociones, sentimientos y sesgo cognitivo fueron realizadas por dos de los autores del presente trabajo quienes son profesionales de la salud mental dentro del ámbito educativo.

## PREPARACIÓN DE LOS DATOS

Se utilizó *Python* (versión 3.8.3) para el proceso de preparación o pre-procesado de los datos. En la Tabla 2 se detallan las adecuaciones realizadas para las variables de la base de datos. Todas las respuestas de las variables mostradas en la Tabla 1 fueron tratadas como tipo int.

Una vez que se realizaron las modificaciones, la base de datos pre-procesada constó con un total de 2041 filas, 68 columnas de variables predictoras (Tabla 1 y Tabla 2) y cinco columnas de variables objetivo categorizadas según los cinco baremos de diagnóstico (Tabla 2). Las variables objetivo indican el nivel de emociones negativas o sentimientos negativos que expresan los sujetos de observación. Por ejemplo, en emociones, una clasificación etiquetada como 'Muy alto', indica que el sujeto experimenta constantemente emociones negati-

Tabla 2. Descripción de las adecuaciones realizadas a las variables durante el proceso de preparación de los datos

Nombre	Tipo	Descripción de la adecuación realizada	
Sujeto de observación	Predictora	Se establecieron dos valores booleanos correspondientes a: <i>Profesores = True</i> <i>Alumnos = False</i>	
Nivel educativo que cursa o imparte	Predictora	Se aplicó la categorización siguiente: <i>Preescolar: 1</i> <i>Preparatoria: 4</i> <i>Primaria: 2</i> <i>Licenciatura: 5</i> <i>Secundaria: 3</i> <i>Posgrado: 6</i>	
Sexo	Predictora	Se establecieron dos valores booleanos correspondientes a: <i>Hombre = True</i> <i>Mujer = False</i>	
Edad	Predictora	Se generaron seis grupos de edad con base en los siguientes rangos: <i>13 - 17 años: 1</i> <i>23 - 35 años: 3</i> <i>51 - 64 años: 5</i> <i>18 - 22 años: 2</i> <i>36 - 50 años: 4</i> <i>65 - 73 años: 6</i>	
Tienes internet en casa	Predictora	Se establecieron dos valores booleanos correspondientes a: <i>No = False</i> <i>Si = True</i>	
Estado	Predictora	Se aplicó la categorización siguiente: <i>Coahuila de Zaragoza: 1</i> <i>Nuevo León: 3</i> <i>Michoacán de Ocampo: 2</i> <i>Zacatecas: 4</i> <i>nan: 5</i>	
Emociones durante cuarentena	Objetivo	Nivel de emociones negativas expresadas por el sujeto durante la cuarentena	
Sentimientos durante cuarentena	Objetivo	Nivel de sentimientos negativos experimentados por el sujeto durante la cuarentena	Se aplicó la categorización siguiente: <i>Muy bajo: 1</i> <i>Bajo: 2</i> <i>Regular: 3</i> <i>Alto: 4</i> <i>Muy alto: 5</i>
Sesgo cognitivo general	Objetivo	Nivel de justificaciones no objetivas que hace el sujeto en general	
Sesgo cognitivo epidemia	Objetivo	Nivel de justificaciones no objetivas que hace el sujeto sobre la pandemia	
Sesgo cognitivo educación	Objetivo	Nivel de justificaciones no objetivas que hace el sujeto sobre la educación	

Fuente: Elaboración propia

vas durante la pandemia; por el contrario, una clasificación etiquetada como 'Muy bajo', indica que el sujeto no experimenta o experimenta muy pocas emociones negativas durante la pandemia. En el caso del sesgo cognitivo, una clasificación etiquetada como 'Muy alto', indica que el sujeto genera una alta cantidad de explicaciones no objetivas sobre lo que ocurre respecto a la pandemia.

## MÉTODOS Y MÉTRICAS

En esta etapa se lleva a cabo la combinación de dos potentes algoritmos de *machine learning*. *Random Forest* y *Support Vector Machines*, dos técnicas que han brindado resultados eficientes en diversas tareas de clasificación. A continuación se describen sus bases teóricas y sus propiedades principales:

*Random Forest* (RF): Propuesto por (Breiman, 2001), es un popular método de aprendizaje automático, el cual se construye a partir de múltiples árboles de decisión no relacionados. Dicha combinación reduce el error en las tareas de clasificación y regresión gracias al uso del proceso popularmente conocido como *Bootstrap* (Niu *et al.*, 2020). Es decir, cada árbol del bosque se forma con un número distinto de muestras aleatorias con reemplazo, a esto se le conoce como *Bootstrapping*. Así mismo en cada nodo del árbol se seleccionan de manera aleatoria un número de variables explicativas para realizar la partición y se construye el árbol hasta determinado punto (profundidad).

Este procedimiento junto con el *Bootstrapping* constituye dos fuentes importantes de aleatoriedad en el algoritmo. Lo cual tiene como objetivo la reducción de la varianza mediante la reducción de la correlación entre los árboles (Pita González, 2017). Una vez construido el bosque cada árbol entrega como resultado a qué clase pertenece cada una de las observaciones de entrenamiento. Finalmente, *Random forest* lleva a cabo la predicción tomando en cuenta la clase más votada por los árboles, esto se conoce como la regla de voto mayoritario (Izquierdo & Zurita, 2020).

En este algoritmo deben ser considerados dos parámetros para lograr un buen desempeño en la clasificación. El primero es el número de árboles que conforman el bosque  $N_{trees}$  y el segundo es el tamaño del subconjunto  $N_j$ , es decir, el número de variables que serán consideradas como candidatas para la partición. En conformidad con Liaw & Wiener (2002), para fines de clasificación será de  $\sqrt{N}$ , donde  $N$  es el número de variables en el conjunto de datos  $X$  mientras que  $N_{trees}$  se recomienda que sean algunos miles de árboles.

Además de las grandes capacidades de predicción del algoritmo, una de las características que lo vuelven

robusto para tareas de análisis de datos, es que entrega la importancia de cada una de las variables de entrada. Es decir, mide cómo las variaciones de estas afectan a la respuesta. Esto se conoce en RF como *Feature importance* (FI). De esta manera se concluye que aquellas variables que influyan de forma más significativa en la variabilidad de la variable de salida son las que mejor explican el modelo.

Por lo tanto, estas variables podrán seleccionarse para realizar la predicción. A este procedimiento también se le conoce dentro del área del *machine learning* como *feature selection*. La importancia de la variable  $x_j$  esta dada por:

$$Importancia_j = \frac{1}{N_{trees}} \sum_{v \in S} G(x_j, v)$$

Donde  $S$  es el conjunto de nodos,  $x_j$  es considerada para particionar las muestras y  $G(x_j, v)$  es conocida como la ganancia RF de  $x_j$ . Por lo tanto, la ganancia se basa en la medida de impureza cuando las muestras son divididas en cada nodo. Diversos criterios se emplean para determinar la impureza con el fin de dividir los datos y, por lo tanto, para determinar el grado importancia de determinada variable (Wang *et al.*, 2016).

Por otro lado, *Support Vector Machine* (SVM) desarrollado por Vladimir Vapnik (Cortes & Vapnik, 1995) es un algoritmo de aprendizaje automático para la clasificación de conjuntos de datos que presentan alta dimensionalidad. En su forma estándar SVM está diseñado para llevar a cabo clasificaciones binarias. SVM construye un hiperplano dirigido en dirección tal que su distancia a los vectores (observaciones) más cercanos en cada una de las dos clases (margen) sea la máxima.

Los vectores más cercanos son conocidos como vectores de soporte (Deng *et al.*, 2012). Para casos donde los datos son linealmente separables, existe un plano que separa a las clases dado por:

$$w \cdot x + b = 0$$

donde  $w \in R^n$  es un vector que define el hiperplano, mientras  $b$  determina un umbral. Por lo tanto, se puede encontrar un hiperplano óptimo que separe en dos clases los datos resolviendo el siguiente problema de optimización cuadrático:

$$\min \frac{1}{2} \|w\|^2$$

Sujeto a:  $y_i (w \cdot x_i + b) \geq 1$ .

Donde  $w \in R^n$  es un vector normal (plano),  $b$  se define como el umbral y  $x_i \in R^d$ ,  $i = 1, 2, \dots, n$  son vectores de entrada (observaciones) con sus respectivas etiquetas (clases). Utilizando multiplicadores de Lagrange  $\alpha_i$ ,  $i = 1, 2, \dots, n$ , el problema se vuelve un problema dual de la forma:

$$\max L(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j$$

Sujeto a:  $\sum_{i=1}^n \alpha_i y_i = 0$ ,  $\alpha_i > 0$

Aplicando las condiciones de Karush-Kuhn-Tucker se tiene que  $\alpha_i [y_i (w \cdot x_i + b) - 1] = 0$ . Si los multiplicadores de Lagrange son mayores a cero, los vectores asociados a dichos multiplicadores serán vectores de soporte. Por lo tanto, la solución para el vector  $w$  será:

$$w^* = \sum_{i=1}^N \alpha_i y_i x_i$$

Donde  $N$  es el número de vectores de soporte. De esta manera el hiperplano óptimo se calcula de la siguiente manera:

$$f(x) = \text{sgn} \left( \sum_{i=1}^N \alpha_i y_i (x \cdot x_i) + b^* \right)$$

Donde  $\text{sgn}(\cdot)$  es la función tal que  $x = \begin{cases} +1 & f(x) > 0 \\ -1 & f(x) < 0 \end{cases}$

En casos donde los datos de entrada no son linealmente separables, SVM busca en espacios dimensionales mayores la separación lineal (hiperplano), (Erfani *et al.*, 2016). Este procedimiento se efectúa mediante la incorporación de funciones Kernel de la siguiente forma:

$$f(x) = \text{sgn} \left( \sum_{i=1}^N \alpha_i y_i \cdot K(x \cdot x_i) + b^* \right)$$

donde  $K$  representa a dicha función Kernel. Existen gran variedad de funciones que cumplen con este criterio, como el Kernel sigmoide, el polinomial, basados en Wavelets, etc.

En esta investigación se usará el Kernel Gaussiano debido a que es adecuado en casos donde los datos presentan alta dimensionalidad (Lee & Bum, 2018). Lo cual

es una característica de los datos analizados en esta investigación. Esta función se define a continuación:

$$K(x_i, x_j) = \exp \left( -\frac{\|x_i - x_j\|^2}{s^2} \right)$$

La metodología propuesta combina la capacidad de *Random Forest* de calcular el nivel de importancia de las variables de entrada respecto a la variable de salida, y la eficiencia de *Support Vector Machines* para llevar a cabo tareas de clasificación en conjuntos de datos de alta dimensionalidad. A continuación, se describe la secuencia de la metodología propuesta:

1. Una vez que los datos se han limpiado y estructurado son analizados empleando RF y SVM por separado y se compara la exactitud en la clasificación alcanzada por ambos modelos. Si RF alcanza mayor exactitud en la clasificación, este será el modelo que se usará para llevar a cabo la clasificación. De lo contrario pasamos a la segunda etapa:
2. De acuerdo con la etapa anterior SVM logró una mayor exactitud en la clasificación respecto a RF. En esta etapa se busca aumentar la capacidad de SVM mediante el proceso llamado *Feature Selection*. Es decir, aprovechando la capacidad de RF de brindar la importancia de las variables, se eligen las  $p - q$  más importantes detectadas por el modelo, donde  $p$  es el total de variables y  $q$  será elegido dependiendo del caso de estudio. Posteriormente se construye un nuevo conjunto de datos con el que SVM es entrenado nuevamente.
3. Por último, se compara el nivel de exactitud de SVM entrenado con el conjunto de datos completo respecto a la exactitud alcanzada cuando se entrena con el conjunto determinado mediante RF. Finalmente se utiliza el mejor modelo.

La exactitud (ACC) final de cada modelo se determina usando validación cruzada de 10 subconjuntos, esta se calcula de la siguiente manera:

$$ACC = \frac{TP + TN}{TP + FN + FP + TN}$$

Donde TP y TN son los datos que fueron clasificados en cada clase de manera correcta, es decir, los verdaderos positivos y negativos. Mientras que FP y FN son los datos clasificados incorrectamente. El resultado será un valor comprendido entre 0 y 1, donde 0 representa que el modelo erró al clasificar la totalidad de las observaciones. Mientras que un valor de 1 significa que el mo-

delo sería capaz de clasificar correctamente el 100 por ciento de los datos.

### EXPERIMENTOS

Los experimentos se ejecutaron en una computadora con sistema operativo macOS Catalina (versión 10.15.5) con procesador 3.1 GHz 6-Core Intel Core i5 y memoria de 16 GB 2667 MHz. Se utilizó Python, ya que incluye una gama amplia de algoritmos de aprendizaje automático. La base de datos pre-procesada se dividió de manera aleatoria en dos: un conjunto para el entrenamiento (80 %) del modelo mediante validación cruzada de 10 sub conjuntos y, otro conjunto, para su prueba (20 %). El conjunto de prueba se utilizó para simular la llegada de nuevos datos (sin clasificar) y probar el modelo.

Se realizaron diversos experimentos, primero analizando la base de datos pre-procesada de manera general, es decir, incluyendo todos los registros. Segundo, los datos pre-procesados fueron segmentados con base en diferentes criterios para identificar el nivel de importan-

cia de cada variable predictora para cada segmento. Los criterios de segmentación fueron de la siguiente manera:

1. El análisis incluyó todos los datos, es decir, sin segmentación (Tabla 3).
2. El análisis se realizó por tipo de sujeto de observación: alumnos (Tabla 4) y profesores (Tabla 5).
3. El análisis se realizó por sexo de los sujetos de observación: hombres (Tabla 6) y mujeres (Tabla 7).
4. El análisis se realizó combinando los criterios de los incisos 2 y 3: alumnos-hombres (Tabla 8), alumnos-mujeres (Tabla 9), profesores-hombres (Tabla 10) y profesores-mujeres (Tabla 11).

Los resultados de los experimentos se detallan en las siguientes tablas donde se muestran las cinco variables objetivo, los valores de exactitud obtenidos tanto con el conjunto de entrenamiento (Ent) como el conjunto de prueba (Pr), empleando la técnica de *Random Forest* (RF) y *Support Vector Machine*. Para esta última técnica, con datos originales (SVM) y con las 10 variables predictoras más importantes (SVM Top10).

Tabla 3. Resultados de experimentos con datos pre-procesados completos (sin segmentación)

Variable objetivo	RF Ent	RF Pr	SVM Ent	SVM Pr	SVM Top10 Ent	SVM Top10 Pr
Emociones durante cuarentena	0.811241	0.838631	0.844355	<b>0.870416</b>	0.873762	0.858191
Sentimientos durante cuarentena	0.851728	0.848411	0.888512	<b>0.887531</b>	0.880525	0.867971
Sesgo cognitivo general	0.806386	0.801956	0.893394	<b>0.914425</b>	0.811873	0.831296
Sesgo cognitivo epidemia	0.776927	0.821516	0.856599	0.870416	0.939339	<b>0.93643</b>
Sesgo cognitivo educación	0.778775	0.826406	0.864619	0.853301	0.954687	<b>0.93643</b>

Fuente: Elaboración propia

Tabla 4. Resultados de experimentos con datos pre-procesados segmentados por sujeto de observación (Alumnos)

Variable objetivo	RF Ent	RF Pr	SVM Ent	SVM Pr	SVM Top10 Ent	SVM Top10 Pr
Emociones durante cuarentena	0.776813	0.828194	0.813297	0.814978	0.853101	<b>0.876652</b>
Sentimientos durante cuarentena	0.811099	0.810573	0.854078	<b>0.854626</b>	0.850867	0.84141
Sesgo cognitivo general	0.790159	0.819383	0.861856	<b>0.876652</b>	0.816667	0.814978
Sesgo cognitivo epidemia	0.770073	0.788546	0.823175	0.814978	0.932527	<b>0.942731</b>
Sesgo cognitivo educación	0.768046	0.762115	0.829915	0.819383	0.926044	<b>0.933921</b>

Fuente: Elaboración propia

Tabla 5. Resultados de experimentos con datos pre-procesados segmentados por sujeto de observación (Profesores)

Variable objetivo	RF Ent	RF Pr	SVM Ent	SVM Pr	SVM Top10 Ent	SVM Top10 Pr
Emociones durante cuarentena	0.800571	0.813187	0.818436	0.835165	0.885807	<b>0.912088</b>
Sentimientos durante cuarentena	0.846024	0.862637	0.861149	<b>0.934066</b>	0.881754	0.89011
Sesgo cognitivo general	0.790868	0.802198	0.8582	<b>0.912088</b>	0.808771	0.785714
Sesgo cognitivo epidemia	0.786701	0.747253	0.822508	0.818681	0.916115	<b>0.906593</b>
Sesgo cognitivo educación	0.748364	0.82967	0.817142	0.851648	0.917542	<b>0.928571</b>

Fuente: Elaboración propia

Tabla 6. Resultados de experimentos con datos pre-procesados segmentados por sexo (Hombres)

Variable objetivo	RF Ent	RF Pr	SVM Ent	SVM Pr	SVM Top10 Ent	SVM Top10 Pr
Emociones durante cuarentena	0.781175	0.79085	0.814044	0.797386	0.866721	<b>0.849673</b>
Sentimientos durante cuarentena	0.809208	0.797386	0.850219	0.843137	0.879836	<b>0.875817</b>
Sesgo cognitivo general	0.777923	0.764706	0.833934	<b>0.843137</b>	0.822268	0.79085
Sesgo cognitivo epidemia	0.736858	0.718954	0.774727	0.777778	0.901339	<b>0.954248</b>
Sesgo cognitivo educación	0.753279	0.738562	0.794372	0.836601	0.921066	<b>0.928105</b>

Fuente: Elaboración propia

Tabla7. Resultados de experimentos con datos pre-procesados segmentados por sexo (Mujeres)

Variable objetivo	RF Ent	RF Pr	SVM Ent	SVM Pr	SVM Top10 Ent	SVM Top10 Pr
Emociones durante cuarentena	0.798943	0.800781	0.82629	0.816406	0.867314	<b>0.863281</b>
Sentimientos durante cuarentena	0.855502	0.796875	0.872168	<b>0.847656</b>	0.877955	0.839844
Sesgo cognitivo general	0.805663	0.765625	0.874986	<b>0.859375</b>	0.822263	0.835938
Sesgo cognitivo epidemia	0.789977	0.792969	0.845726	0.839844	0.93458	<b>0.917969</b>
Sesgo cognitivo educación	0.775452	0.742188	0.839882	0.804688	0.937483	<b>0.90625</b>

Fuente: Elaboración propia

Tabla 8. Resultados de experimentos con datos pre-procesados segmentados por sujeto de observación y por sexo (Alumnos-Hombres)

Variable objetivo	RF Ent	RF Pr	SVM Ent	SVM Pr	SVM Top10 Ent	SVM Top10 Pr
Emociones durante cuarentena	0.782583	0.725275	0.782583	0.813187	0.890165	<b>0.824176</b>
Sentimientos durante cuarentena	0.820871	0.78022	0.809685	0.802198	0.870495	<b>0.857143</b>
Sesgo cognitivo general	0.76006	0.78022	0.804354	<b>0.835165</b>	0.83476	0.802198
Sesgo cognitivo epidemia	0.697072	0.703297	0.760435	0.692308	0.851276	<b>0.824176</b>
Sesgo cognitivo educación	0.752177	0.78022	0.776877	0.802198	0.889715	<b>0.901099</b>

Fuente: Elaboración propia

Tabla 9. Resultados de experimentos con datos pre-procesados segmentados por sujeto de observación y por sexo (Alumnos-Mujeres)

Variable objetivo	RF Ent	RF Pr	SVM Ent	SVM Pr	SVM Top10 Ent	SVM Top10 Pr
Emociones durante cuarentena	0.773232	0.757353	0.78404	0.794118	0.856162	<b>0.860294</b>
Sentimientos durante cuarentena	0.777003	0.742647	0.850572	<b>0.772059</b>	0.789764	0.742647
Sesgo cognitivo general	0.793401	0.764706	0.834007	0.786765	0.819293	<b>0.801471</b>
Sesgo cognitivo epidemia	0.77862	0.779412	0.806229	0.816176	0.929933	<b>0.933824</b>
Sesgo cognitivo educación	0.763939	0.705882	0.797003	0.830882	0.907778	<b>0.889706</b>

Fuente: Elaboración propia

Tabla 10. Resultados de experimentos con datos pre-procesados segmentados por sujeto de observación y por sexo (Profesores-Hombres)

Variable objetivo	RF Ent	RF Pr	SVM Ent	SVM Pr	SVM Top10 Ent	SVM Top10 Pr
Emociones durante cuarentena	0.7485	0.806452	0.735833	0.83871	0.813667	<b>0.870968</b>
Sentimientos durante cuarentena	0.861333	0.790323	0.878	<b>0.806452</b>	0.865667	0.774194
Sesgo cognitivo general	0.754833	0.725806	0.750833	<b>0.806452</b>	0.808333	0.790323
Sesgo cognitivo epidemia	0.731	0.645161	0.771667	0.725806	0.796667	<b>0.790323</b>
Sesgo cognitivo educación	0.723167	0.66129	0.750833	0.693548	0.8985	<b>0.854839</b>

Fuente: Elaboración propia

Tabla 11. Resultados de experimentos con datos pre-procesados segmentados por sujeto de observación y por sexo (Profesores-Mujeres)

Variable objetivo	RF Ent	RF Pr	SVM Ent	SVM Pr	SVM Top10 Ent	SVM Top10 Pr
Emociones durante cuarentena	0.742049	0.801653	0.804507	0.801653	0.846173	<b>0.842975</b>
Sentimientos durante cuarentena	0.829549	0.884298	0.852509	0.892562	0.898299	<b>0.900826</b>
Sesgo cognitivo general	0.76926	0.793388	0.841879	<b>0.834711</b>	0.827338	0.760331
Sesgo cognitivo epidemia	0.777466	0.702479	0.815051	0.743802	0.912628	<b>0.818182</b>
Sesgo cognitivo educación	0.758801	0.743802	0.785799	0.768595	0.902338	<b>0.876033</b>

Fuente: Elaboración propia

## DISCUSIÓN Y ANÁLISIS DE RESULTADOS

La técnica de RF empleada obtuvo para todos los experimentos la exactitud más baja, tanto para los datos de entrenamiento como los de validación. Sin embargo, con esta técnica se seleccionaron las variables predictoras más importantes para la clasificación. Se compararon los resultados de clasificación utilizando la técnica SVM considerando todas las variables predictoras contra el uso de únicamente las 10 variables predictoras más importantes (SVM Top10).

La técnica SVM Top10 presentó una exactitud superior que el SVM para los datos segmentados en la variable *Emociones durante cuarentena*. Caso contrario, para el caso sin segmentación (Tabla 3).

En cuanto a la variable *Sentimientos durante cuarentena*, los datos sin segmentación y segmentación por sujeto de observación muestran mejores resultados de exactitud utilizando la técnica SVM (Tabla 3, Tabla 4 y Tabla 5). En cuanto a la segmentación por sexo, los resultados de exactitud muestran un patrón inverso. Por ejemplo, en la Tabla 10 se muestran los resultados para Profesor-Hombre en donde la exactitud del SVM es mejor que la del SVM Top10; por el contrario, en la Tabla 11 donde se muestran los resultados para Profesor-Mujer la exactitud del SVM Top10 superó al SVM.

Para la variable *Sesgo cognitivo general* la técnica que obtuvo mejores resultados fue la SVM para todos los casos, excepto en la segmentación Alumno-Mujer que se muestra en la Tabla 9, donde SVM Top10 superó a la exactitud alcanzada por el SVM, tanto para los datos de entrenamiento como los de validación.

Como se puede observar en todas las tablas presentadas, las variables objetivo *Sesgo cognitivo epidemia* y *Sesgo cognitivo educación* la técnica SVM Top10 obtuvo los mejores valores de exactitud para los datos de entrenamiento y validación.

En la Tabla 12 se presentan las variables predictoras más y menos relevantes para cada una de las variables objetivo. Las variables predictoras más relevantes representan aquellas variables más importantes en las cuales el algoritmo se basa para la clasificación. Es decir, se puede realizar una selección de variables únicamente considerando aquellas más relevantes y obtener resultados satisfactorios para la clasificación. Por otro lado, las variables predictoras menos relevantes pueden ser omitidas para el análisis (e incluso del instrumento), ya que no son importantes para la clasificación que realiza el algoritmo. Con lo que se cumple la condición de parsimonia de los test, explicar lo máximo con los menos elementos posibles.

En la Figura 1 se muestra el nivel de importancia de las 20 variables predictoras (o reactivos) más importantes para la variable objetivo *Sesgo cognitivo epidemia* considerando los diferentes experimentos: sin segmentación (Figura 1a); con segmentación por alumnos (Figura 1b), por profesores (Figura 1c), por hombres (Figura 1d), por mujeres (Figura 1e), por alumnos-hombres (Figura 1f), por alumnos-mujeres (Figura 1g), por profesores-hombres (Figura 1h) y por profesores mujeres (Figura 1i). Las barras indican el nivel de importancia (eje x) del reactivo mostrado en la parte izquierda del gráfico (eje y).

La teoría clásica que ha sido utilizada de los test para fundamentar las pruebas más famosas de la psicometría, data desde 1904, cuando la desarrolló Charles Edward Spearman, por lo que hay que dejar muy en claro sus posibilidades y limitaciones actuales, sin quitarle el mérito de “clásica”, pues tiene ya en ejecución más de 100 años (Muñiz, 2010). Sin embargo, las necesidades de medir fenómenos que ocurren en la mente de los seres humanos, con el poder de cómputo actual y lo masivo de los datos a los que se puede tener acceso, esta propuesta de validación psicométrica se restringe a ciertos ámbitos muy concretos (pocos sujetos medidos, limitaciones para realizar los cálculos y la capacidad de respuesta de los sistemas de información utilizados, favorece a que investigadores propositivos, propongan nuevas metodologías de validación en el marco del Big data actual.

Tabla 12. Variables predictoras más y menos relevantes para cada variable objetivo

Variable objetivo	Variables predictoras más relevantes	Variables predictoras menos relevantes
Emociones durante cuarentena	<p>Qué nivel de Disgusto presentas ahora</p> <p>Qué nivel de Decepción presentas ahora</p> <p>Qué nivel de Tristeza presentas ahora</p> <p>Qué nivel de Miedo presentas ahora</p> <p>Qué nivel de Ira presentas ahora</p> <p>Qué nivel de Inseguridad presentas ahora</p> <p>Qué nivel de Abandono presentas ahora</p> <p>Qué nivel de Amenaza presentas ahora</p> <p>Qué nivel de Frustración presentas ahora</p>	<p>Laptop</p> <p>Tabletas</p> <p>Computadora personal</p> <p>Sexo</p> <p>Tienes internet en casa</p> <p>Sujeto de observación</p> <p>Nivel educativo cursa o imparte</p> <p>En qué medida te sientes a gusto con tus hermanos</p> <p>Teléfonos inteligentes</p>
Sentimientos durante cuarentena	<p>Qué nivel de Agobio presentas ahora</p> <p>Qué nivel de Frustración presentas ahora</p> <p>Qué nivel de Ansiedad presentas ahora</p> <p>Qué nivel de Desesperanza presentas ahora</p> <p>Qué nivel de Soledad presentas ahora</p> <p>Qué nivel de Paz presentas ahora</p> <p>Qué nivel de Tranquilidad presentas ahora</p> <p>Qué nivel de Preocupación presentas ahora</p> <p>Qué nivel de Odio presentas ahora</p> <p>Qué nivel de Decepción presentas ahora</p> <p>Qué nivel de Aburrimiento presentas ahora</p>	<p>Laptop</p> <p>Computadora personal</p> <p>Tabletas</p> <p>Sexo</p> <p>Nivel educativo cursa o imparte</p> <p>Tienes internet en casa</p> <p>Sujeto de observación</p> <p>Acceso a dispositivos</p> <p>Teléfonos inteligentes</p>
Sesgo cognitivo general	<p>Las noticias confirman lo que pienso</p> <p>Reprobaré</p> <p>Puedo aprender lo que siempre he querido</p> <p>Me puedo enfocar en aprender</p> <p>Dicen que la educación a distancia es de baja calidad, reprobaré el semestre</p> <p>Todo es mejor por internet</p> <p>No aprenderé</p> <p>Siguiendo las instrucciones, no pasa nada</p> <p>Los maestros saben lo que hacen en cuanto a la educación a distancia</p> <p>Pasará la contingencia, pero no aprenderé</p> <p>Me enfermaré</p> <p>Lo que aprenda por medios electrónicos no me servirá</p> <p>No es grave</p>	<p>Computadora personal</p> <p>Sexo</p> <p>Tienes internet en casa</p> <p>Sujeto de observación</p> <p>Tabletas</p> <p>Nivel educativo cursa o imparte</p> <p>Qué nivel de Abandono presentas ahora</p> <p>Laptop</p> <p>Teléfonos inteligentes</p> <p>Qué nivel de Soledad presentas ahora</p> <p>Qué nivel de Odio presentas ahora</p>
Sesgo cognitivo epidemia	<p>Todo es mejor por internet</p> <p>Las noticias confirman lo que pienso</p> <p>Me enfermaré</p> <p>Se enfermará alguien querido</p> <p>Siguiendo las instrucciones, no pasa nada</p> <p>No es grave</p> <p>Toda información es buena</p> <p>Me apego a la información científica</p> <p>Moriré</p> <p>Ya pasará</p>	<p>Computadora personal</p> <p>Sexo</p> <p>Tienes internet en casa</p> <p>Sujeto de observación</p> <p>Nivel educativo cursa o imparte</p> <p>Laptop</p> <p>Tabletas</p> <p>Qué nivel de Odio presentas ahora</p> <p>Qué nivel de Abandono presentas ahora</p>
Sesgo cognitivo educación	<p>Dicen que la educación a distancia es de baja calidad, reprobaré el semestre</p> <p>Pasará la contingencia, pero no aprenderé</p> <p>Reprobaré</p> <p>Me puedo enfocar en aprender</p> <p>No aprenderé</p> <p>Puedo aprender lo que siempre he querido</p> <p>Puedo complementar lo visto en clase</p> <p>Lo que aprenda por medios electrónicos no me servirá</p> <p>Los maestros saben lo que hacen en cuanto a la educación a distancia</p> <p>Solo estudiando saldré adelante</p>	<p>Tabletas</p> <p>Computadora personal</p> <p>Sexo</p> <p>Tienes internet en casa</p> <p>Sujeto de observación</p> <p>Qué nivel de Abandono presentas ahora</p> <p>Nivel educativo cursa o imparte</p> <p>Laptop</p> <p>Qué nivel de Culpabilidad presentas ahora</p> <p>Qué nivel de Odio presentas ahora</p>

Fuente: Elaboración propia

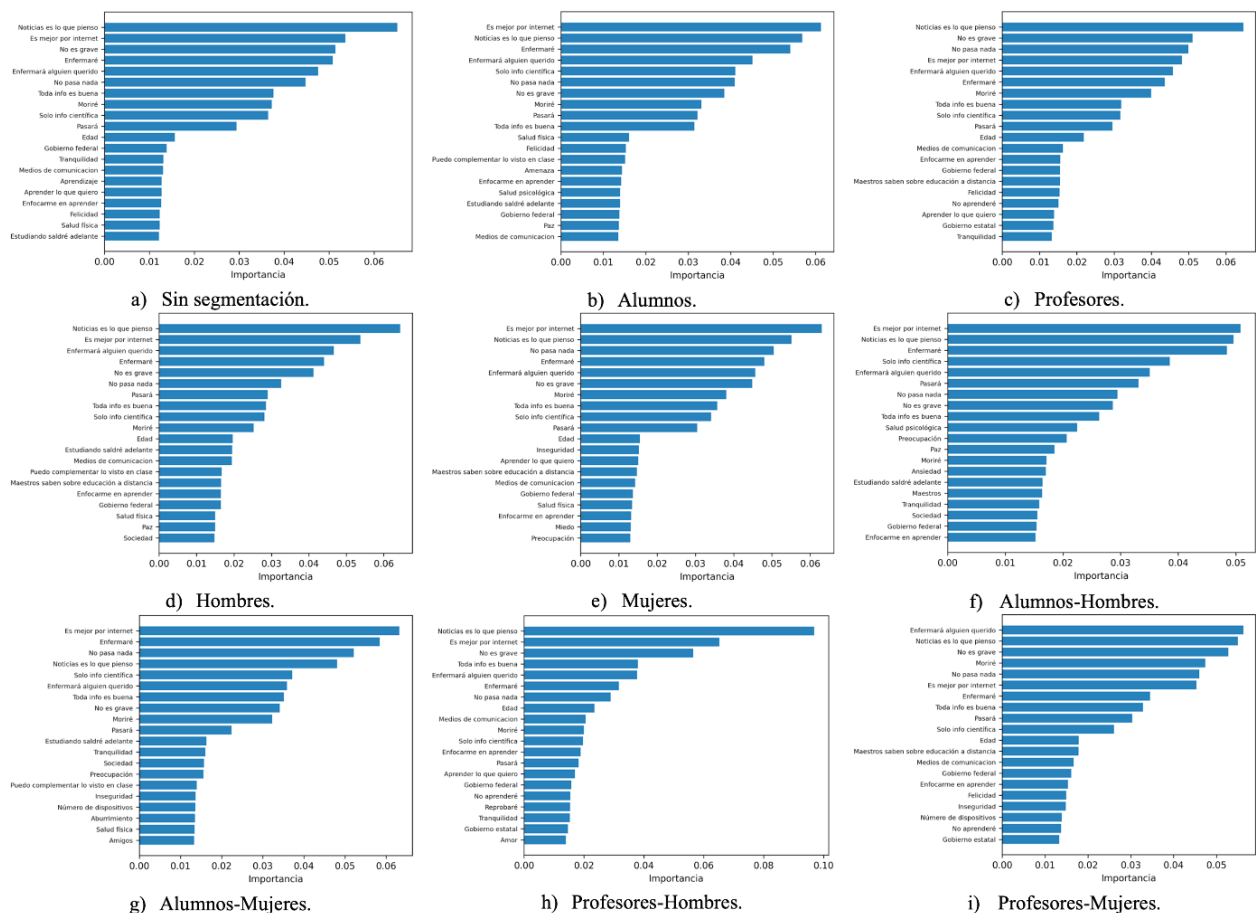


Figura 1. Reactivos más relevantes (top 20) para la variable sesgo cognitivo durante la epidemia

## CONCLUSIONES

De acuerdo con los procesos utilizados para dar respuesta al objetivo, y la evidencia mostrada; se puede concluir que la estrategia de *machine learning*, da una aproximación a la pertinencia de un reactivo en una prueba psicológica, es una estrategia que valida lo anterior y expone la porción mostrada del reactivo en la prueba, de acuerdo con las categorías de resultados obtenidos. Las variables objetivo a predecir por los elementos predictores se muestran mejor con la técnica *support vector machine*, que con la estrategia de *random forest*. Sin embargo, esta última es de gran utilidad para seleccionar las variables de mayor impacto en la clasificación de los datos.

De acuerdo con lo presentado en este trabajo se concluye que las técnicas de *machine learning* aplicadas a la psicometría, mejoran mucho la descripción de un instrumento psicológico. Estos procedimientos en conjunto con otras áreas del cómputo, como el *big data*, le

darán un empuje muy importante a la psicometría en entornos de datos masivos, tanto a la propia clasificación mencionada, como a la predicción de trastornos psicológicos.

## ANEXO 1

## AVISO DE PRIVACIDAD UADEC

Los datos personales que voluntariamente proporciona a esta Máxima Casa de Estudios de forma electrónica o personal, serán recabados, protegidos, incorporados y tratados en el Sistema de Datos Personales de esta Universidad Autónoma de Coahuila, ubicada en Boulevard Venustiano Carranza sin número, con fundamento en la Ley de Acceso a la Información Pública y Protección de Datos Personales para el Estado de Coahuila de Zaragoza.

Debiendo tomar en consideración que por datos personales debe entenderse cualquier información nu-

mérica, alfabética, gráfica, fotográfica, acústica o de cualquier otro tipo concerniente a una persona, identificada o identificable.

En el entendido que el responsable del sistema de sus datos personales es la autoridad que los recaba, los mismos serán utilizados única y exclusivamente para los fines propios de cada una de las dependencias universitarias que conforman ésta Institución. Asimismo, tales datos personales podrán ser proporcionados únicamente a la Autoridad que al efecto acredite contar con facultad para ello.

En caso de negativa a suministrar los datos personales requeridos por la autoridad y/o funcionario, interrumpirá indefinidamente el trámite y/o procedimiento a realizar en esta Universidad. La autenticidad y actualización de los datos proporcionados es responsabilidad tanto de Usted como de la Autoridad que los recaba.

Ahora bien usted podrá ejercer sus derechos de Acceso, Rectificación, Cancelación u Oposición de sus Datos Personales, de manera presencial ante la Unidad de Enlace de Transparencia, Acceso a la Información y Protección de Datos Personales o bien con la Dependencia correspondiente. Finalmente para cualquier cambio respecto a este Aviso de Privacidad esta Universidad se reserva el derecho a efectuar en cualquier momento modificaciones o actualizaciones al presente aviso y se compromete a informar a Usted, de manera oportuna, a través de los medios adecuados como lo es el portal de Transparencia de la Universidad (Universidad Autónoma de Coahuila, 2020).

## REFERENCIAS

- Ahuja, R. & Banga, A. (2019). Mental stress detection in university students using machine learning algorithms. *Procedia Computer Science*, 152, 349-353. <https://doi.org/10.1016/j.procs.2019.05.007>
- Ardan, M., Fadzilul-Rahman, F. & Bali-Geroda, G. (2020). The influence of physical distance to student anxiety on COVID-19, Indonesia. *Journal of Critical Reviews*, 7(17), 1126-1132.
- Bisquerra, R. (2000). *Educación emocional y bienestar*. Barcelona: Praxis.
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5-32.
- Cao, W., Fang, Z., Hou, G., Han, M., Xu, X., Dong, J. & Zheng, J. (2020). The psychological impact of the COVID-19 epidemic on college students in China. *Psychiatry research*, 287, 112934. <https://doi.org/10.1016/j.psychres.2020.112934>
- Cortés-Álvarez, N. Y., Piñeiro-Lamas, R. & Vuelvas-Olmos, C. R. (2020). Psychological effects and associated factors of COVID-19 in a Mexican sample. *Disaster Medicine and Public Health Preparedness*, 14(3), 413-424. <https://doi.org/10.1017/dmp.2020.215>
- Cortes, C. & Vapnik, V. (1995). Support vector networks. *Machine Learning*, 20(3), 273-297. <https://doi.org/10.1007/BF00994018>
- Deng, N., Tian, Y. & Zhang, C. (2012). *Support vector machines: Optimization based theory, algorithms, and extensions*. USA: CRC press.
- Elmer, T., Mepham, K. & Stadtfeld, C. (2020). Students under lockdown: Comparisons of students' social networks and mental health before and during the COVID-19 crisis in Switzerland. *PloS one*, 15(7), e0236337. <https://doi.org/10.1371/journal.pone.0236337>
- Erfani S. M. Rajasegarar, S., Karunasekera, S. & Leckie, C. (2016). High-dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning. *Pattern Recognition*, 58, 121-134. <https://doi.org/10.1016/j.patcog.2016.03.028>
- European Centre for Disease Prevention and Control. (2020). European Centre for Disease Prevention and Control. Recuperado el 08 de 2020 de <https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-covid-19-cases-worldwide>
- González-Sanguino, C., Ausín, B., Castellanos, M. A., Saiz, J., López-Gómez, A., Ugidos, C. & Muñoz, M. (2020a). Mental health consequences during the initial stage of the 2020 Coronavirus pandemic (COVID-19) in Spain. *Brain, Behavior, and Immunity*, 87, 172-176. <https://doi.org/10.1016/j.bbi.2020.05.040>
- González-Jaimes, N., Alejandro-Tejeda, A., Espinosa-Méndez, C. M. & Ontiveros-Hernández, Z. O. (2020b). Psychological impact on Mexican university students due to confinement during the Covid-19 pandemic. *Health Sciences*. <https://doi.org/10.1590/SciELOPreprints.756>
- Graham, S., Depp, C., Lee, E., E., Nebeker, C., Tu, X., Ho-Cheol K. & Jeste, D. V. (2019). Artificial intelligence for mental health and mental illnesses: an overview. *Current psychiatry reports*, 21(11), 1-18. <https://doi.org/10.1007/s11920-019-1094-0>
- Huarcaya-Victoria, J. (2020). Consideraciones sobre la salud mental en la pandemia de COVID-19. *Revista Peruana de Medicina Experimental y Salud Pública*, 37(2), 327-334. <https://doi.org/10.17843/rpmesp.2020.372.5419>
- Izquierdo-Verdiguier, E. & Zurita-Milla R. (2020). An evaluation of guided regularized random forest for classification and regression tasks in remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, 88, 102051. <https://doi.org/10.1016/j.jag.2020.102051>
- Kahneman, D. (2012). *Pensar rápido, pensar despacio*. Debate.
- Khattar, A., Rai-Jain, P. & Quadria, S. M. (2020). Effects of the disastrous pandemic COVID 19 on learning styles, activities and mental health of young indian students-A machine learning approach. 4th International Conference on Intelligent Computing and Control Systems (ICICCS), IEEE, 1190-1195.
- Lawler, E. J. & Thye, S. R. (1999). Bringing Emotions into Social Exchange Theory. *Annual Review of Sociology*, 25(1), 217-244. <https://doi.org/10.1146/annurev.soc.25.1.217>
- Li, S., Xue, J., Wang, Y. & Zhao, N. (2020). The impact of COVID-19 epidemic declaration on psychological consequences: A study

- on active weibo users. *International Journal of Environmental Research and Public Health*, 17(6), 2032. <https://doi.org/10.3390/ijerph17062032>
- Liang, L., Ren, H., Cao, R., Hu, Y., Qin, Z., Li, Ch. & Mei, S. (2020). The effect of COVID-19 on youth mental health. *Psychiatric Quarterly*, 91(3), 841-852.
- Liaw, A. & Wiener, M. (2002). Classification and regression by RandomForest. *R news*, 2(3), 18-22.
- Lee, S. & Bum, K. S. (2018). Time-adaptive support vector data description for nonstationary process monitoring. *Engineering Applications of Artificial Intelligence*, 68, 18-31. <https://doi.org/10.1016/j.engappai.2017.10.016>
- Mohammad, S., Bravo-Marquez, F., Salameh, M. & Kiritchenko, S. (2018). Semeval-2018 task 1: Affect in tweets. Proceedings of the 12th international workshop on semantic evaluation. New Orleans, Louisiana, 1-17. Recuperado de <https://www.aclweb.org/anthology/S18-1001>
- Muñiz-Fernández, J. (2010). Las teorías de los tests: teoría clásica y teoría de respuesta a los ítems. *Papeles del Psicólogo*, 31(1), 57-66.
- Niu, D., Wang, K., Sun, L. & Wu, J. (2020). Short-term photovoltaic power generation forecasting based on random forest feature selection and CEEMD: A case study. *Applied Soft Computing*, 93(1), 106389. <https://doi.org/10.1016/j.asoc.2020.106389>
- Odrizola-González, P., Planchuelo-Gómez, A., Iruetia, M. J. & de Luis-García, R. (2020). Psychological effects of the COVID-19 outbreak and lockdown among students and workers of a Spanish university. *Psychiatry Research*, 290, 113108. <https://doi.org/10.1016/j.psychres.2020.113108>
- Özdin, S. & Özdin, Ş. B. (2020). Levels and predictors of anxiety, depression and health anxiety during COVID-19 pandemic in Turkish society: The importance of gender. *International Journal of Social Psychiatry*, 66(5), 504-511. <https://doi.org/10.1177/0020764020927051>
- Pallarés, M. (2010). *Emociones y sentimientos*. Barcelona: Marge Books.
- Pita González-Campos, V. (2017). *Modelado mediante Random Forests de las emisiones de autobuses urbanos en función de los ciclos cinemáticos* (Grado en Ingeniería en Tecnologías Industriales) Proyecto Fin de Carrera / Trabajo Fin de Grado, Ingeniería de Organización, Administración de Empresas y Estadística. Madrid, España: Universidad Politécnica de Madrid.
- Rajkumar, R. P. (2020). COVID-19 and mental health: A review of the existing literature. *Asian journal of psychiatry*, 52, 102066. <https://doi.org/10.1016/j.ajp.2020.102066>
- Rakhmanov, O. & Dane, S. (2020). Knowledge and anxiety levels of African university students against COVID-19 during the pandemic outbreak by an online survey. *Journal of Research in Medical and Dental Science*, 8(3), 53-56.
- Samuel, J., Ali, N., Rahman, M., Esawi, E. & Samuel, Y. (2020). Covid-19 public sentiment insights and machine learning for tweets classification. *Information*, 11(6), 314. <https://doi.org/10.3390/info11060314>
- Shatte, A. B., Hutchinson, D. M. & Teague, S. J. (2019). Machine learning in mental health: a scoping review of methods and applications. *Psychological medicine*, 49(9), 1426-1448. <https://doi.org/10.1017/S0033291719000151>
- Secretaría de Salud. (2020). Boletín epidemiológico. Recuperado el 27 de junio de 2021 de Dirección General de Epidemiología. <https://www.gob.mx/salud/acciones-y-programas/historico-boletin-epidemiologico>
- Shen, Y., Zhang, W., Man-Chan, B. S., Zhang, Y., Meng, F., A Kenon, E., Wue, H. E., Luo, X. & Zhang, X. (2020). Detecting risk of suicide attempts among Chinese medical college students using a machine learning algorithm. *Journal of affective disorders*, 273, 18-23.
- Srividya, M., Subramaniam, M. & Natarajan, B. (2018). Behavioral modeling for mental health using machine learning algorithms. *Journal of medical systems*, 42(5), 88. <https://doi.org/10.1007/s10916-018-0934-5>
- Tuli, S., Tuli, R. & Singh-Gill, S. (2020). Predicting the growth and trend of COVID-19 pandemic using machine learning and cloud computing. *Internet of Things*, 11, 100222. <https://doi.org/10.1016/j.iot.2020.100222>
- Tull, M. T., Edmonds, K. A., Scamaldo, K. M., Richmond, J. R., Rose, J. P. & Gratz, K. L. (2020). Psychological outcomes associated with stay-at-home orders and the perceived impact of COVID-19 on daily life. *Psychiatry Research*, 289, 113098. <https://doi.org/10.1016/j.psychres.2020.113098>
- Universidad Autónoma de Coahuila (2020). Aviso de privacidad, transparencia. Derechos Arco. Recuperado de <http://www.transparencia.uadec.mx/sassit/docs/protecciondatospersonales.pdf>
- Wang, H., Yang, F. & Luo, Z. (2016). An experimental study of the intrinsic stability of random forest variable importance measures. *BMC bioinformatics*, 17(60), 1-18. <https://doi.org/10.1186/s12859-016-0900-5>
- Yang, Q. et al. (2020). Sen Wave: The public tweets sentimental analysis dataset SenWave for Covid-19 research.
- Zolotov, Y., Reznik, A., Bender, S. & Isralowitz, R. (2020). COVID-19 fear, mental health, and substance use among Israeli university students. *International Journal of Mental Health and Addiction*, 1-7. <https://doi.org/10.1007/s11469-020-00351-8>
- Cómo citar:** Navarro-Acosta, J. A., Soto-Mendoza, V., Saucedo-Zendejo, F. R., Guajardo-Espinoza, J. M. & Rivera-Morales, M. T. (2021). Evaluación psicológica de profesores y alumnos mexicanos durante la pandemia de COVID-19 mediante técnicas de Machine learning. *Ingeniería Investigación y Tecnología*, 22 (04), 1-15. <https://doi.org/10.22201/fi.25940732e.2021.22.4.026>