

INVESTIGACIONES

Modelos predictivos de la competencia pedagógica en docentes de EMTP mediante la minería de datos educacionales^{1, 2}

Predictive models of pedagogical competence in
Vocational Secondary School teachers through educational data mining

*Elisabet Díaz Costa^a, Marcela Romero Jeldres^a,
Tarik Faouzi Nadim^b, Camila Pardo Soto^c*

^aUniversidad Metropolitana de Ciencias de la Educación, Chile.
elisabet.diaz@umce.cl, marcela.romero@umce.cl

^bUniversidad de Santiago de Chile, Chile.
tarik.faouzi@usach.cl

^cUniversidad del BioBio, Chile.

RESUMEN

La investigación indaga en el uso de métodos predictivos mediante Minería de Datos Educacionales, utilizando árboles de decisión para determinar las necesidades de perfeccionamiento en competencias pedagógicas de profesores de EMTP y contribuir con investigaciones que fortalezcan la formación docente. La investigación se emprendió desde un diseño anidado concurrente de varios niveles con tres etapas. En este artículo se da cuenta de la etapa cuantitativa, siguiendo las fases que propone la minería de datos: Selección, exploración, limpieza, transformación, minería de datos, evaluación e interpretación de resultados, difusión y uso de modelos. Los resultados del trabajo se clasifican en dos partes. En la primera, se caracterizaron las variables latentes a través de valores de carga factorial con análisis confirmatorio. La segunda parte se centra en encontrar un modelo de predicción que establezca una fuerte relación entre la competencia pedagógica predicha y las variables latentes definidas por el análisis social, la acción profesional y el diagnóstico de la situación. Los hallazgos se ilustran utilizando datos sociales.

Palabras clave: Competencias Pedagógicas, Modelos Predictivos, Formación Técnica, Minería de Datos Educativos, Educación Técnico Profesional.

ABSTRACT

Research researches the use of predictive methods through Educational Data Mining, using decision trees to determine the development needs in pedagogical skills of Technical-Professional High Schools (TPHS) and contribute to research that strengthens teacher training. The research was undertaken from concurrent nested design of several three-stage levels. This article considers the quantitative stage, following the phases required by data mining: Selection, exploration, cleansing, transformation, data mining, evaluation and interpretation of results, dissemination and use of models. The results of the work are classified into two parts. In the first, latent variables were characterized through load values with confirmatory analysis. The second part focuses on finding

¹ Los resultados de esta publicación han sido posibles gracias al financiamiento otorgado por CONICYT, a través del Proyecto Fondecyt de Iniciación N.º 11140650.

² Esta investigación tributa al Objetivo 4: Educación de Calidad, de la Agenda 2030 de Desarrollo Sostenible.

a prediction model that establishes a strong relationship between predicted pedagogical competition and latent variables defined by social analysis, professional action, and situation diagnosis. Findings are illustrated using social data.

Key words: Pedagogical Competences, Predictive Models, Technical Training, Educational Data Mining, Professional, Vocational Secondary School.

1. ESCENARIOS DE LA FORMACIÓN TÉCNICO-PROFESIONAL EN CHILE

El primer establecimiento educacional de naturaleza técnico profesional en Chile, data del siglo XVIII. Según Servat (2017), el objetivo estuvo focalizado en formar especialistas de nivel técnico en las áreas de geometría, aritmética y dibujo, con el fin de aportar recursos humanos para el impulso de la agricultura, el comercio y la minería. No obstante su desarrollo histórico, solo en la década del 60 se permitió considerar la modalidad Técnico Profesional, como una variante del proceso educativo formal (cuando la educación secundaria se diferencia en educación científico humanista y educación técnica). Por estos días, esta modalidad goza de una Política Nacional para la Formación Técnico-Profesional que incorpora los niveles formativos de Educación de Personas Jóvenes y Adultas (EPJA), Educación Media Técnico-Profesional (EMTP) y Educación Superior Técnico-Profesional. (Ministerio de Educación de Chile, 2016, p. 9).

La Comisión Económica para América Latina y el Caribe (CEPAL) en su agenda 2030 para el desarrollo sostenible, plantea que la Formación Técnico Profesional presenta oportunidades significativas para enfrentar el cambio en la estructura productiva hacia la innovación y el desarrollo tecnológico de América Latina y el Caribe, con fines hacia la erradicación de la pobreza extrema, la reducción de la desigualdad en todas sus dimensiones, el crecimiento económico inclusivo con trabajo decente para todos, ciudades sostenibles y cambio climático, entre otros, con lo que enfatiza su aporte a los procesos educativos de las personas.

No obstante ello, diversos estudios recientes dan cuenta que a nivel regional y nacional la Formación Técnico Profesional debe enfrentar las desigualdades estructurales y restricciones, que afectan en especial a las y los jóvenes en su transición hacia el mercado laboral, así como las barreras, sesgos de género y patrones discriminatorios que aún persisten en este ámbito educativo, y que tiende a situarla como una opción de segunda categoría, asociada fundamentalmente a la movilidad social de personas según su origen, situaciones que luego se reproducen en el mundo del trabajo remunerado. (Romero-Jeldres *et al.*, 2021; Muñoz, 2019; Sepúlveda, 2017, 2016; Ministerio de Educación de Chile, 2016; Larrañaga *et al.*, 2013; Sevilla, 2012).

A nivel nacional, y a partir de los datos que maneja el Ministerio de Educación de Chile (2016), se ha podido establecer que el 39% de los estudiantes de tercero y cuarto año de Enseñanza Media o Secundaria, estudia alguna especialidad técnico-profesional; y el 44% de los estudiantes de Educación Superior estudia en un Centro de Formación Técnica (CFT) o Instituto Profesional (IP). Sepúlveda, (2017) agrega en relación con la distribución de estudiantes de la modalidad técnico profesional que cerca de 165.000 estudiantes cursan algunas de las 35 especialidades que se imparten en el nivel de 3° y 4° año medio y que de los más de 900 establecimientos educacionales que ofrecen esta modalidad formativa, el 60% se encuentran en el decil de mayor vulnerabilidad.

Lo anterior deviene en un imperativo respecto del fortalecimiento del cuerpo docente con mayor especialidad para la formación en oficios en las especialidades técnicas o en la modalidad tecnológica de los sectores productivos correspondientes. Sin embargo, los informes del Ministerio de Educación de Chile (2009, 2006) y la Secretaría Ejecutiva de Formación Técnico-Profesional, (Ministerio de Educación de Chile, 2016), dan cuenta que no existe un catastro nacional de docentes por especialidades; que un alto porcentaje de egresados de sus mismas especialidades enseñan. Además, la información oficial disponible permite observar que en la gran mayoría de los casos los docentes no poseen calificaciones adicionales, ni experiencia en ámbitos productivos, siendo un porcentaje muy bajo el que compatibiliza la enseñanza con un desempeño laboral al interior de las empresas (Sepúlveda, 2017).

El estado actual de la Formación Técnico Profesional en Chile podría entenderse como una secuela de la inexistencia de programas de formación inicial de docentes en modalidad técnico-profesional, junto con la falta de incentivos para que las instituciones de educación superior las ofrezcan, y también, de medidas ministeriales que buscan suplir la escasez de profesores para determinadas especialidades como el Decreto Supremo N° 352/2003, de 9 de octubre que reglamenta el ejercicio de la función docente. Normas Generales, que autoriza a profesionales no docentes a ejercer docencia sin estar formados para ello.

La Comisión Asesora de Formación Técnico Profesional, estimaba que el 58% de los docentes que imparten clases en la EMTP carecen de un título pedagógico, y que en su mayoría, son técnicos de nivel superior los responsables de movilizar el aprendizaje de los jóvenes. (a la fecha no existe un catastro nacional de docentes). (Ministerio de Educación de Chile, 2009). Esta situación deja a un buen número de profesionales que forman en esta modalidad fuera de los procesos de evaluación sobre la base de los estándares disciplinarios y pedagógicos utilizados en los programas de Formación Inicial Docente, exceptuando aquellas realizadas en el marco de la Asignación de Excelencia Pedagógica (AEP), de carácter voluntario y disponible para un número menor de especialidades técnicas.

Recién en el año 2017, el Centro de Perfeccionamiento, de Experimentación e Investigaciones Pedagógicas (CPEIP), incorporó en el sistema de evaluación docente a los profesores de los establecimientos técnico-profesionales de dependencia municipal, dado que cerca de 6.000 docentes de formación diferenciada Técnico-Profesional se incorporarán en los tramos de desempeño que les correspondan en el Sistema de Desarrollo Profesional Docente (Ley 20.903/2016, de 4 de marzo que crea el sistema de Desarrollo Profesional Docente y Modifica otras normas), a implementarse en Chile entre los años 2016 a 2025, con el objetivo de promover la mejora continua de la docencia mediante la actualización de los conocimientos disciplinares, de los métodos de enseñanza, y de apoyar la progresión en la carrera docente del profesorado. Sin embargo, sin apoyos, ni formación pedagógica, los formadores del ámbito técnico profesional, les corresponderán, consecuentemente, ser evaluados y encasillados en los tramos más bajos de remuneraciones.

Preguntarse acerca de las estrategias y dominios en ámbitos pedagógico-didácticos exclusivas para las y los docentes que trabajan en EMTP, resulta complejo, por cuanto los modelos internacionales de formación para la docencia y la enseñanza de la Educación Técnico Profesional (ETP) revisados, tienen en su base distintas representaciones sobre qué conforma un buen docente y de qué capacidades se debe disponer para transmitir el conocimiento e impartir clases de calidad en el contexto EMTP (Hallet, 2006).

Por tanto, el objetivo de este trabajo investigativo es indagar en el uso de métodos predictivos, para determinar un modelo que permita detectar las necesidades de

perfeccionamiento en competencias pedagógicas de los docentes técnicos y así contribuir con investigaciones que fortalezcan la labor docente de esta área formativa, junto con darle valor y capacidad de acción al rol que juegan los sistemas de Educación y de Formación continua, para robustecer el desarrollo y la adaptabilidad de la ETP.

Focalizar en esta área, desde el campo pedagógico también ayuda a establecer mecanismos que ayuden a modificar el entendimiento respecto de los fines y propósitos de ella a nivel nacional y regional.

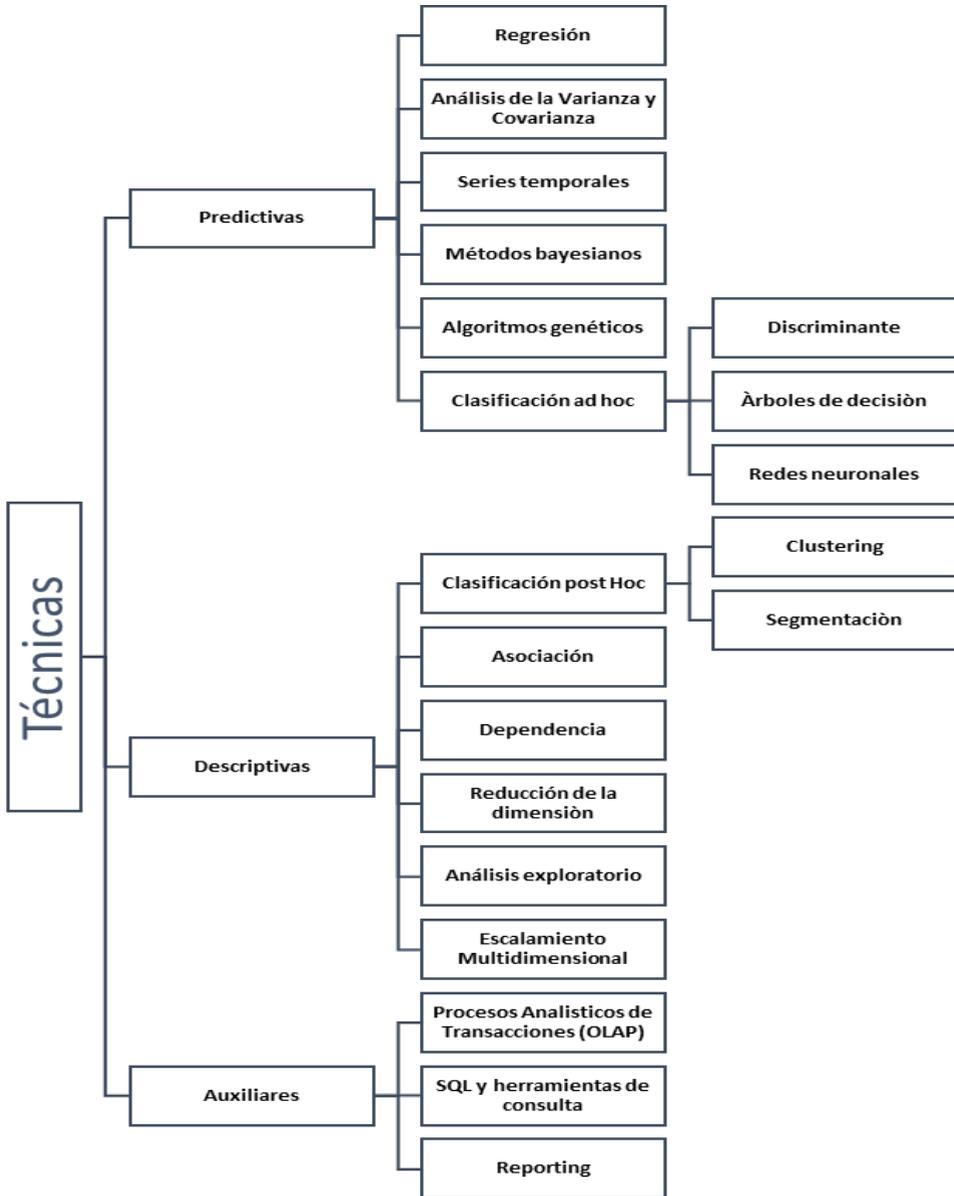
2. MINERÍA DE DATOS EN EDUCACIÓN

La Minería de Datos o Data Mining, es un cuerpo de conocimientos posicionados dentro de las Ciencias de la Información y derivado del enfoque denominado gestión del conocimiento cuyo objetivo es desarrollar el uso preeminente de los datos (Ponjuán Dante, 2015; Manek *et al.*, 2016). La minería de datos incluye la extracción de información inteligente desde un conjunto de datos utilizando metodologías científicas con el fin de examinar semiautomáticamente una gran cantidad de bases de datos para descubrir patrones valiosos (Sukhija *et al.*, 2016). Lo anterior implica el uso de algoritmos para identificar patrones y tendencias dentro de un conjunto de datos (Adekitan y Noma-Osaghae, 2019).

Pérez López y Santín González, (2007, pp. 8-9) clasifican las técnicas de Minería de Datos en: técnicas predictivas, donde las variables pueden clasificarse inicialmente en dependientes e independientes (similares a las técnicas del análisis de la dependencia o métodos explicativos del análisis multivariante); en técnicas descriptivas, en las que todas las variables tienen inicialmente el mismo estatus (similares a las técnicas del análisis de la interdependencia o métodos descriptivos del análisis multivariante) y, en técnicas auxiliares. Señalan que tanto las técnicas predictivas como las técnicas descriptivas están enfocadas al descubrimiento del conocimiento implícito en los datos, en cambio las técnicas auxiliares, serían herramientas de apoyo enfocadas en la verificación.

En la ilustración 1, se muestra una clasificación de las técnicas de Minería de Datos planteadas por los autores:

Ilustración 1. Técnicas de Minería de Datos.



Fuente: Adaptado de Pérez López y Santín González (2007, p. 9).

En ese encuadre, surge la Minería de Datos Educativos (MDE) como un área de investigación interdisciplinaria emergente que se ocupa del desarrollo de métodos para explorar y analizar datos que se originan en un contexto educativo, por medio de enfoques computacionales (Manek *et al.*, 2016).

Galindo y García, (2010), señalan que se ha incrementado el interés en utilizar MDE, centrándose en el desarrollo de métodos de descubrimiento para comprender mejor a los estudiantes y el entorno en el que aprenden. Plantea ventajas respecto de facilitar el estudio de las diferencias entre grupos de profesores o clases difíciles de estudiar estadísticamente con métodos tradicionales y desagrega una variedad de métodos empleados en la MDE, los cuales se pueden observar en la Tabla 1.

Tabla 1. Principales enfoques de la minería de datos educacionales

Categorías del método	Objetivo del método	Principales aplicaciones
Predicción	Desarrollo de un modelo que pueda inferir una variable a partir de la combinación de variables y factores.	<ul style="list-style-type: none"> • Detección de comportamiento del estudiante. • Detección del valor de salida en casos donde no existan datos etiquetados.
Agrupaciones	Encontrar conjuntos de datos que se agrupen naturalmente separando el conjunto completo en una serie de categorías.	<ul style="list-style-type: none"> • Descubrimiento de nuevos patrones de comportamiento de estudiantes. • Investigaciones de similitudes y diferencias entre escuelas.
Minería de relaciones	Descubrimiento de relaciones entre variables.	<ul style="list-style-type: none"> • Descubrimiento de asociaciones curriculares en secuencias de cursos. • Descubrimiento de estrategias pedagógicas que guíen en un proceso más efectivo de aprendizaje.
Descubrimiento mediante modelos	Modelado de un fenómeno mediante predicción, agrupamiento o ingeniería del conocimiento, es usado como componente en una futura predicción o minería de relaciones.	<ul style="list-style-type: none"> • Descubrimiento de relaciones entre comportamiento de estudiantes y sus características o variables contextuales. • Análisis de cuestiones de investigación para una amplia variedad de contextos.
Destilado de datos	Los datos son destilados para permitir a un humano identificar o clasificar rápidamente propiedades de los datos.	<ul style="list-style-type: none"> • Identificación humana de patrones en el aprendizaje de los alumnos, comportamiento, colaboración. • Etiquetado de datos para su uso en desarrollos posteriores de modelos predictivos.

Fuente: Adaptado de Galindo y García (2010).

2.1. ESTUDIOS USANDO MDE

Los estudios usando MDE, se focalizan en temáticas, asociadas preferentemente en la enseñanza aprendizaje. Por ejemplo, el estudio de Salas-Rueda, (2020), que construye diversos modelos predictivos sobre las percepciones de uso de Facebook y Twitter en el contexto educativo. El autor utiliza la técnica árbol de decisión por cuanto le permite con más claridad identificar las relaciones entre las variables y sugiere el uso de regresión logística y redes neuronales para el análisis del impacto de futuras investigaciones. En investigaciones similares respecto del uso del podcast en el campo educativo, se identifican tres modelos predictivos que indican que esta herramienta mejora las condiciones de enseñanza aprendizaje (Salas *et al.*, 2020). Otras investigaciones asociadas con el uso de la aplicación GeoGebra en las actividades escolares para la enseñanza-aprendizaje sobre las matemáticas, Salas y Salas (2019) dan cuenta que la técnica árbol de decisión, permite identificar las condiciones favorables para el proceso de enseñanza-aprendizaje del contenido matemático, en relación con la edad, el género y la carrera de los participantes.

Una segunda temática estudiada, dicen relación con investigaciones que utilizan modelos predictivos para estudiar el riesgo del fracaso académico. Los resultados planteados por Ferreira de Araújo (2019), se consideraron útiles para ayudar a los maestros a tomar decisiones pedagógicas. Sus resultados favorecieron modelos predictivos del tipo caja blanca o fenomenológicos. Otros resultados se asocian con la deserción donde Urteaga *et al.* (2020), aplican técnicas de MDE a los registros de interacciones entre estudiantes y plataformas con el fin de predecir el abandono académico. El algoritmo que les produjo modelos más eficientes fue una red neuronal entrenada con retropropagación.

Las investigaciones revisadas, y sus temáticas respectivas, permiten transferir la utilización de MDE para la identificación de dominios de competencia, específicamente, la competencia pedagógica en uso que poseen el profesorado que se desempeña en EMTP. Ello haría posible pensar que se pueden obtener modelos predictivos que posteriormente puedan apoyar procesos de perfeccionamiento continuo o bien encontrar nuevas relaciones entre las variables de estudio y así generar informes para el análisis de tendencias.

No obstante, se hace preciso atender que estos estudios pueden presentar limitaciones si no se atiende al tipo de muestreo o al uso de softwares o plataformas con fines de generalización y, por tanto, es necesario señalar que la MDE, depende estrictamente de la naturaleza y el tamaño de los datos en estudio (Almarabeh, 2017), y como tal, debe garantizar la privacidad y el consentimiento junto con modalidades adecuadas para gestionarlos.

3. METODOLOGÍA

Esta investigación se emprendió desde un modelo mixto de investigación, implementado desde un diseño anidado concurrente de varios niveles (Hernández *et al.*, 2010), cuyo propósito era indagar las competencias para la acción profesional que predominan en los profesionales que realizan docencia en EMTP. Con un modelo predominante CUAN/ cual, la opción metodológica, buscó dar cuenta de un mayor entendimiento del problema de investigación integrando por un lado las variables que influyen en la acción profesional y, por otra parte, recoger la experiencia humana de las y los docentes que se desempeñan

en EMTP en contextos mayormente desfavorecidos. Para ello se diseñaron tres etapas: teórica, cuantitativa y cualitativa. En la etapa dos, se desarrolló un instrumento que anidó una pregunta abierta que permitió recoger las voces de las y los docentes en tres ámbitos. Ello implicó recolectar datos cuantitativos y cualitativos en varios niveles, obteniendo información diferente, junto con hacer una variación de análisis. (Romero y Faouzi, 2020; Romero, 2020). En este artículo se da cuenta de la etapa cuantitativa, siguiendo las fases propuestas por Pérez López y Santín González, (2007, p. 9) cuando la investigación requiere de la Minería de Datos: selección, exploración, limpieza, transformación, minería de datos, evaluación e interpretación de resultados, difusión y uso de modelos. Los resultados de esta etapa permiten la integración de conjunto con las otras etapas implementadas en el diseño de investigación.

3.1. FASE 1. PREPARACIÓN DE DATOS, SELECCIÓN, LIMPIEZA Y TRANSFORMACIÓN

Las fuentes de datos se obtuvieron del Ministerio de Educación de Chile, Base de Datos Oficiales (2014). Ellas aglutinaban un universo de 234.168 profesionales que se desempeñaban como profesorado del contexto EMTP. Para la fase de selección, limpieza y transformación, lo primero que se realizó fue limitar la población teniendo como referencia la organización comunal de la Dirección Provincial Santiago Sur. Ello permitió cubrir los docentes de 54 Liceos de EMTP con una población de 1042 profesores (directores y docentes), repartidos en: 101 directivos, 173 docentes de especialidad y 768 docentes de formación general.

Una vez realizada la selección, se estratificó usando un muestreo aleatorio simple, que correspondió a 151 docentes de especialidad y/o directores pertenecientes a los 54 liceos existentes de la organización comunal señalada que también realizaban docencia. Para el cálculo del tamaño muestral se tuvo en cuenta lo planteado por Boomsma (1983, p. 119) en el trabajo de simulación donde sugiere que el estimador de chi-cuadrado no es preciso para muestras menores de 50 y recomienda 100 o más casos. Por su parte Anderson y Gerbing (1984) en sus simulaciones también encuentran las mayores desviaciones de cuadrado de una variante chi-cuadrado en muestras más pequeñas (<100). Lo anterior es correspondiente con lo señalado por Hulland *et al.* (1996, p. 184), que señalan que la mayoría de los estudios recomiendan un tamaño de muestra de al menos 100, pero también sugieren una muestra de 200 para modelos más complejos.

Para obtener información de la muestra estratificada, se aplicó presencial e individualmente un cuestionario de auto-aplicación de respuestas anónimas y confidenciales diseñado y validado para este fin por Romero-Jeldres y Faouzi, (2018), con el propósito de identificar las competencias pedagógicas predominantes de los y las docentes de EMTP. En todos los casos se tuvieron en cuenta los aspectos éticos de la investigación en Ciencias Sociales.

Una vez capturados y vaciados los datos, se tuvo una primera vista filtrable que permitiera aplicar las técnicas de minería de datos. Ellas concentraron 11 atributos asociados a las características sociodemográficas de los docentes técnicos, y 54 variables latentes, asociadas a las competencias pedagógicas (Ver Tabla 2).

Tabla 2. Atributos finales utilizados para el proceso de minería de datos

Atributo		Valores posibles
Características sociodemográficas	Comuna donde realiza docencia	1. San Ramón, 2. La Granja, 3. Buin, 4. Paine, 5. San Bernardo, 6. La Cisterna, 7. Lo Espejo, 8. El Bosque, 9. Calera de Tango, 10. Otra comuna.
	Género	Hombre (1), Mujer (2)
	Edad	≤ 24, ≥80
	Desempeño	1. Docente de formación diferenciada técnica 2. Docente de formación general 3. Directivo con docencia 4. Directivo sin docencia
	Tipo de establecimiento	1. Particular 2. Particular subvencionado 3. Municipal 4. De Administración delegada
	Título profesional	Abierto
	Especialización en Educación	Abierto
	Sectores y especialidades de la formación técnico-profesional donde realiza docencia	1. Maderero, 2. Agropecuarios, 3. Alimentación, 4. Construcción, 5. Metalmecánica, 6. Electricidad, 7. Marítimo, 8. Minero, 9. Gráfico, 10. Confección, 11. Administración, 12. Salud y Educación, 13. Química e Industria, 14. Tecnología y comunicaciones, 15. Hotelería y Turismo.
	Horas de contratación	Abierto
	Años de experiencia en educación	Abierto
	Años de Experiencia en empresas	Abierto
Competencias pedagógicas	Análisis de la Sociedad (4 variables latentes)	1 =Lo desconocía (se desconoce el ítem o la cuestión propuesta),
	Diagnóstico de la Sociedad (4 variables latentes)	2 = No soy capaz (no se es capaz o no se sabe realizar la cuestión propuesta),
	Autorreflexión (3 variables latentes)	3= Sí, pero con ayuda (se es capaz de realizar la cuestión propuesta con ayuda de otra persona o indicar que se tienen dificultades para realizar la cuestión propuesta
	Acción Profesional (20 variables latentes)	por sí solo), 4= Sí, siempre (se es capaz de realizar la cuestión propuesta),
	Competencia Social (14 variables latentes)	5= Sí y lo sabría explicar (se es capaz de realizar la cuestión propuesta y de explicar a otra persona cómo debe realizarse.
	Vocación Laboral (9 variables latentes)	

Fuente: Elaboración Propia.

3.2. FASE 2. MINERÍA DE DATOS

Para realizar la fase de minería de datos, se utilizó el análisis factorial confirmatorio (AFC) por ser un tipo de modelo de ecuaciones estructurales (SEM) que se identifica específicamente con modelos de medición, es decir, permiten determinar las relaciones entre las variables o factores latentes y sus medidas o indicadores observados (Pearl, 2000; Brown, 2006). Mientras que SEM es término genérico que permite averiguar si existen relaciones entre las variables latentes y el fenómeno en estudio o la variable latente endógena.

Del mismo modo, el hallazgo de un modelo predictivo requiere del uso de los árboles de decisión. En esta fase de la minería se usa esta técnica por cuanto intenta separar los datos y agrupar las muestras en las clases a las que pertenecen, identifica una etiqueta para construir los árboles del conjunto de entrenamiento, maximiza la pureza de los grupos tanto como sea posible, cada vez que crea un nuevo nodo del árbol, divide su conjunto en dos subconjuntos. Por supuesto, al final del árbol, plantea una respuesta clara identificada por clases o valoraciones de tipo Likert.

A continuación, se presentan los resultados y la verificación de los modelos.

3.3. FASE 3. EVALUACIÓN E INTERPRETACIÓN DE RESULTADOS

En esta fase se utilizaron los diferentes métodos estadísticos que permitieron la creación de los modelos predictivos. En primer lugar, la utilización del método AFC con datos categóricos reales.

Definimos la ecuación de medición como

$$x = M\varphi + \varepsilon, (1)$$

donde x es un vector de variables observadas, M es una matriz con cargas factoriales, φ es un vector de variables latentes y ε es un vector de singularidad.

Resulta importante señalar que las cargas de factores son las pendientes de regresión para predecir los indicadores o variables observables a partir de la variable latente. De este modo, la ecuación (1) establece una relación reflexiva, cuya variable latente explica las variables observables con magnitud

$$M_{ij}, i=1, \dots, n, j=1, \dots, m.$$

Por lo anterior se hizo necesario definir un estimador para evaluar todos los parámetros del modelo (1). Brown (2006), recomienda el estimador ponderado de media mínima y varianza (WLSMV) como una alternativa a la Máxima verosimilitud (ML) para datos ordinales.

Ahora bien, Romero y Faouzi (2018), definieron los modelos que mejor representaban las variables latentes (factores) del estudio para profesores y directores de EMPT. En la Figura 1, se muestra la representación gráfica del subconstructo (factor) diagnóstico de la situación, calculado a partir de las siguientes ecuaciones:

$$P^j = M^j \varphi^j + \varepsilon^j, j=1, \dots, 10,$$

Donde j-ésima ecuación es el modelo de medición. Por ejemplo, para j=1 se tiene que

$$P^1 = (P3 \ P23 \ P34 \ P86 \ P88)^T \quad M^1 = (.65, .56, .74, .76, .74)^T \quad \varepsilon^1 = (\varepsilon_1^1 \ \varepsilon_2^1 \ \varepsilon_3^1 \ \varepsilon_4^1 \ \varepsilon_5^1)^T \text{ y } \varphi^j$$

es el sub-constructo (factor) diagnóstico de la situación (Ver figura 1).

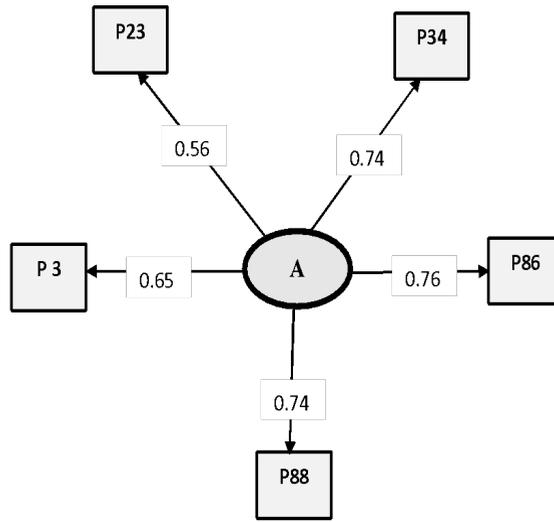


Figura 1. Modelo de medición de la variable latente (factor) diagnóstico de la situación.

Fuente: Romero y Faouzi (2018).

El uso de las variables latentes definidas a partir del análisis confirmatorio proveyó a los investigadores de un conjunto de nuevas variables (factores) que aglutinaban las sentencias del instrumento original, haciendo más plausible la interpretación del modelo predictivo tras la utilización del árbol de decisión. Sin embargo, esta ventaja comparativa, presentaba una dificultad relevante y es que, para su generación, es condición necesaria contar con el valor de las cargas factoriales o respuesta asumida que poseen las variables que participan del fenómeno a predecir, en este caso particular, de las nuevas variables (factores) establecidas.

Para resolver la imposibilidad de contar con un valor de carga factorial para los factores latentes descubiertos en el AFC por cada sujeto, y así poder usarlos como variables en la generación del modelo predictivo, se calculó la carga factorial para cada variable latente mediante la técnica k-means y su valoración usando el índice de la razón de validez de contenido. En efecto, por cada variable latente (factor) se agruparon los sujetos mediante la técnica k-means, en función de las respuestas (valoración) a los ítems que conformaron dicha variable latente. Luego de conocer la distribución de la valoración que hicieron los sujetos para cada agrupación arrojada por k-means en la variable latente, se procedió a calcular el índice de acuerdo (VCR) considerando como esencial la valoración más frecuente entre las respuestas de la agrupación de k-media y como no esencial el resto de las valoraciones.

$$CVR = \frac{n - N/2}{N/2},$$

donde n es la cantidad total de la valoración más frecuente dentro de un grupo (definido por k -means) y N es el número total de los sujetos que participan de esa agrupación.

Luego, se averigua si el valor encontrado del CVR es mayor que el valor crítico del CVR.

Con el CVR calculado para la valoración más frecuente de cada agrupación de k -medias, se procedió a calcular la significancia del valor hallado a través del CVR crítico en la tabla definida por los autores Baghestani *et al.*, (2019). (Ver la Tabla 4 en el Anexo).

Si el valor de la carga factorial hallado para la respuesta más frecuente resultaba significativo, se asume que la respuesta de cada sujeto a la variable latente era el valor más frecuente de la agrupación de k -medias. Este procedimiento fue seguido para cada una de las cuatro variables latentes definidas en el estudio previo (Romero y Faouzi, 2018). Por otra parte, para la construcción del árbol de decisión, se atendió a que la partición recursiva y el árbol de regresión son técnicas que permiten definir la entropía de Shannon para mostrar cómo este índice puede ser útil para definir el árbol calculado en todo el conjunto de datos. Es importante señalar que la entropía de Shannon es una medida absoluta que proporciona un número entre 0 y 1, independientemente del tamaño del conjunto. Ello se encuentra disponible en todo el software de minería de datos. Al usar R, se hace necesario programar el método, pero es bastante simple.

Así, para obtener el error de validación cruzada, se dividen aleatoriamente los datos en K pliegues. Reiteramos el siguiente proceso, extrayendo las submuestras, es decir, realizando el modelo de entrenamiento con $(K-1)$ pliegues, y calculando la tasa de error para el K -ésimo pliegue. El error por la validación cruzada es la media de estas tasas de error recopiladas. A partir de esta descripción, transcribimos las operaciones en R y creamos aleatoriamente una columna que indica los individuos que pertenecen a los pliegues.

Sobre la base de la descripción anterior, se definió el modelo predictivo que establece la relación causal directa entre la competencia pedagógica y las variables latentes/ sociodemográficas. Este modelo se identifica mediante la figura 2.

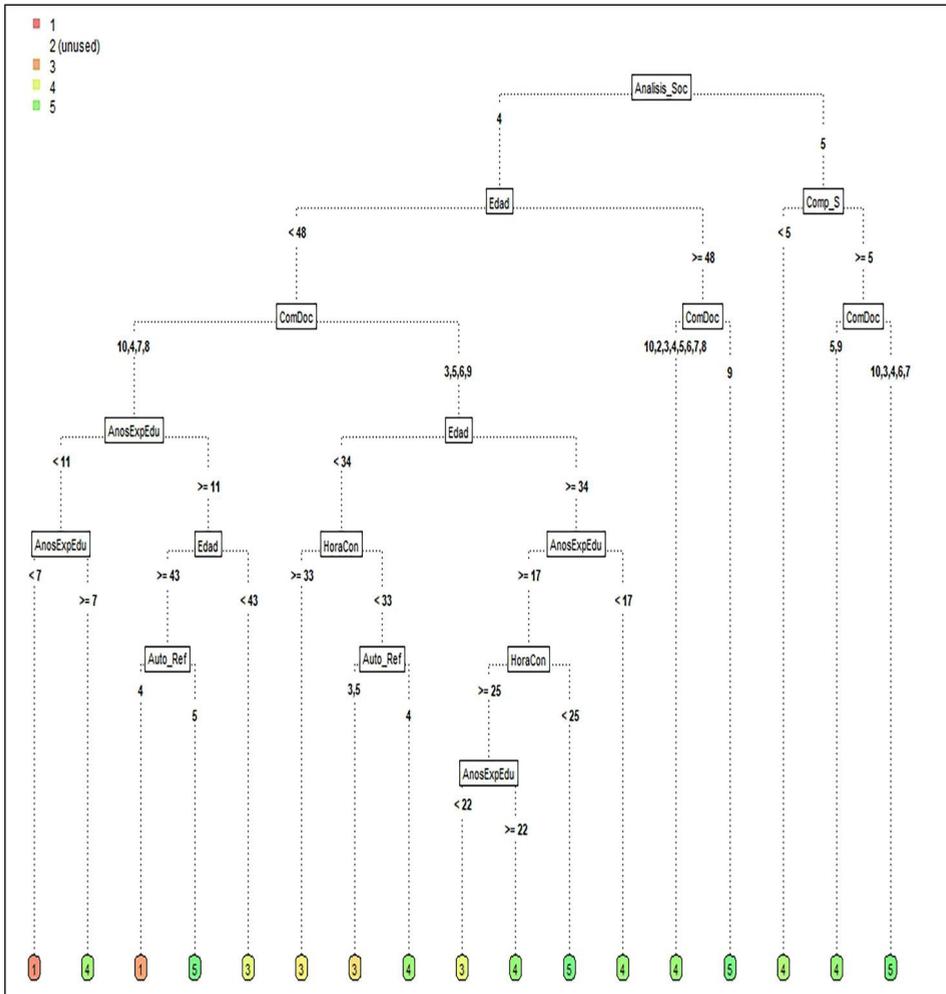


Figura 2. Modelo predictivo de la competencia pedagógica a través de árbol de decisión.

Fuente: Elaboración propia.

La Figura 2, entrega un esquema jerárquico que indica las condiciones por lo cuáles un sujeto pertenece a una clase (cinco clases). Es posible reconocer la autopercepción competencial en el árbol de decisión atendiendo a las conceptualizaciones que tiene las categorías (Ver Tabla 2).

En atención al modelo predictivo a través de árbol de decisión, presentado en la Figura 2, la investigación muestra si las y los profesores de EMTP poseen competencias pedagógicas.

La interpretación del árbol de decisión nos permite señalar que:

- Las variables que discriminan mejor están asociadas con la competencia pedagógica análisis de la sociedad y competencia social, seguidas de las variables sociodemográficas edad, comuna donde desarrollan la docencia y horas de contrato.
- Cuando el valor de las variables latentes, análisis de la sociedad y competencia social es igual a 5 (Sí y lo sabría explicar; Se es capaz de realizar la cuestión propuesta y se es capaz de explicar a otra persona cómo debe realizarse) y los docentes realizan la docencia en la comuna de Buin, Paine, La Cisterna, Lo Espejo y otra comuna, se puede deducir que poseen competencias pedagógicas.
- Cuando el valor de la variable análisis de la sociedad es 4 (Sí, siempre; Se es capaz de realizar la cuestión propuesta) se distinguen tres casos:
 1. Si la edad de las y los docentes de EMTP, es mayor a 48 años, entonces, la o el profesor(a) tiene competencias pedagógicas si realiza docencia en la comuna de Calera de Tango; de otro modo las o los profesores solo son capaces de realizar la cuestión propuesta si realizan docencia en las comunas de La Granja, Buin, Paine, San Bernardo, La Cisterna, Lo Espejo, El Bosque u Otra comuna.
 2. Si la edad de las y los docentes de EMTP, está entre 43 y 48 años con la variable autorreflexión igual a 5 (Sí y lo sabría explicar; Se es capaz de realizar la cuestión propuesta y se es capaz de explicar a otra persona cómo debe realizarse) y la comuna donde se realiza la docencia es Paine, Lo Espejo, El Bosque u Otra comuna, entonces, la o el profesor tiene competencias pedagógicas.
 3. Si la edad de las y los docentes de EMTP, es menor a 48 años, tienen experiencia en educación, menor a 7 años y realizan la docencia en Paine, Lo Espejo, El Bosque u Otra comuna, no poseen competencias pedagógicas.

Una vez generado el modelo se evalúa su capacidad predictiva.

Tabla 3. Error por validación cruzada (VC) del modelo de competencia pedagógica

Modelo	Error de precisión por VC	K
Modelo propuesto	0.19	10

Fuente: Elaboración Propia.

La Tabla 3, presenta el error del modelo por validación cruzada. Es decir que el error de medida representa un 19 %, por lo que se puede considerar que el modelo es aceptable.

Ahora bien, basado en lo anterior, en varias circunstancias la interacción entre variables mediante gráficos resulta de buena utilidad. Usando la función “plotmo” del paquete “plotmo”, se indagó en las interacciones entre las variables del modelo predictivo (Ver Figura 3).

Los gráficos muestran solo una porción delgada de los datos, pero de todos modos son útiles. Son más informativos cuando las variables que se grafican no tienen fuertes interacciones con las otras variables (Milborrow, 2016).

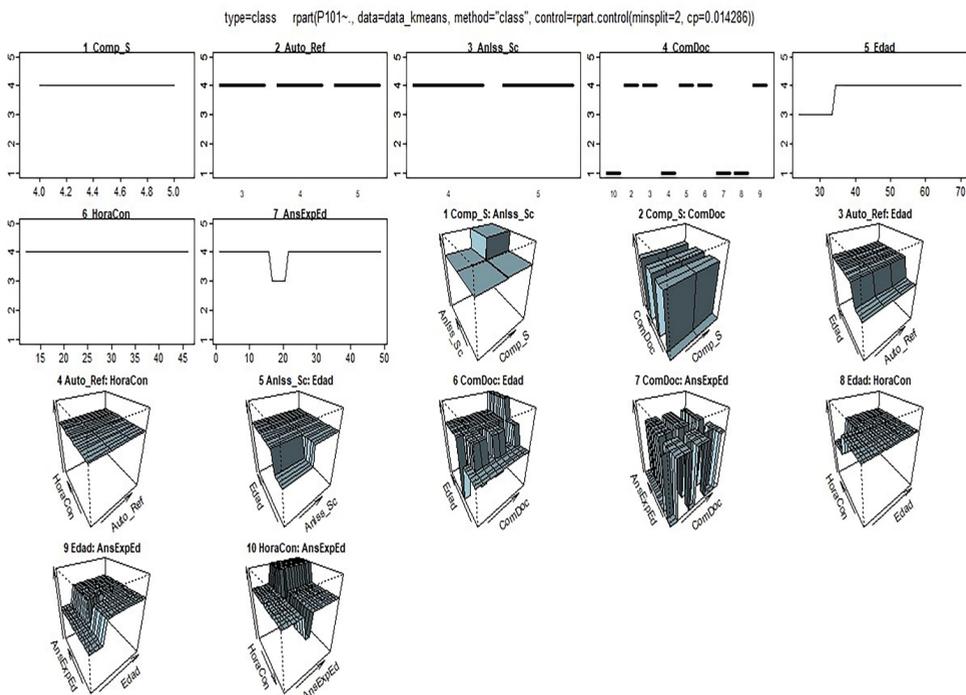


Figura 3. Clases predichas mediante interacciones entre variables del modelo predictivo.

Fuente: Elaboración Propia.

La imagen 4 de la Figura 3 muestra interacción entre las variables Edad y Auto-Reflexión, lo cual revela que, independientemente de los valores de ambas variables, la respuesta final en promedio de los profesores es igual 4, indicando que los profesores tienen un conocimiento pedagógico desde una perspectiva más bien teórica, mientras que la interacción entre las variables latentes análisis de la sociedad y competencia social, indica que cuando ambas variables tienen valores grandes entonces los profesores poseen la competencia pedagógica.

3.4. FASE 4. DIFUSIÓN Y USO DE MODELOS

En resumen, la investigación mostró que las y los docentes de EMTF presentan descendidas sus competencias pedagógicas cuando presentan las siguientes condiciones:

- La falta de dominio de competencias para el análisis de la sociedad resulta ser un factor determinante en el dominio de las competencias pedagógicas. Esta

dimensión en el modelo planteado por Romero y Faouzi (2018), tiene que ver con el conocimiento de las políticas nacionales de educación relacionadas con el currículum, la gestión educativa y la profesión docente, las normas de funcionamiento y convivencia, la implementación de actividades de acuerdo con el tipo y complejidad del contenido, así como con las políticas y metas del establecimiento.

- De igual modo, resulta relevante para entender el dominio de las competencias pedagógicas, el dominio de las competencias referidas a la competencia social, las que, según los autores, están relacionadas con el trabajo en equipo y las habilidades de cooperación, las habilidades de comunicación y las habilidades de gestión de conflictos.
- La edad, resulta otro factor importante asociado al dominio de competencias pedagógicas, constatándose que, a mayor edad, es mayor el manejo de competencias en uso que el manejo que tienen de las mismas profesoras/es más jóvenes, no obstante, ello está relacionado con los años de experiencia, lo cual hace pensar que un docente joven sin formación pedagógica formal debe ser mayormente apoyado si posee menos de siete años de experiencia docente.
- Finalmente, hay que señalar que la interacción entre las variables Edad y Auto-Reflexión (autoconcepto profesional y preocupación/afectación por/de sí mismo), revela que independientemente de los valores de ambas variables, la respuesta final indica que en promedio el profesorado de EMTP tiene un conocimiento pedagógico, pero que este conocimiento posee un carácter declarativo, lo que dificulta ser explicado o enseñado.

4. CONCLUSIONES

La primera conclusión de la investigación dice relación con que es posible afirmar que el uso de las técnicas de MDE, especialmente las utilizadas en este estudio, permiten determinar el nivel de dominio de las competencias pedagógicas en uso que poseen las y los docentes de EMTP sin formación pedagógica y que se desenvuelven profesionalmente en contextos de difícil desempeño, haciendo posible focalizar una oferta de perfeccionamiento ajustada a sus necesidades.

Dado que este instrumento se diseñó conforme a las demandas de las políticas educativas chilenas vigentes, atendiendo al Marco para la Buena Enseñanza y al Sistema de Desarrollo Profesional Docente, bien podría usarse en entornos sin formación pedagógica, tales como la Formación Inicial Docente, o en la docencia en Educación Superior. En todos los casos, de comprobarse el dominio de competencias, puede servir de base de un procedimiento confiable para el reconocimiento de aprendizajes previos (RAP).

Respecto de la aplicación de técnicas de MDE aquí reseñadas y del uso del índice VCR para obtener la carga factorial en constructos latentes, ha sido demostrada su utilidad con lo que se proyecta su uso en áreas de retención educacional o rendimiento académico.

Finalmente, se deja abierta la posibilidad al diseño de una interfaz que permita a cualquier docente evaluar sus competencias pedagógicas en uso y de esta manera, definir sus necesidades de perfeccionamiento, cuestión especialmente necesaria para las y los docentes que trabajan en contextos con alta vulnerabilidad social.

REFERENCIAS BIBLIOGRÁFICAS

- Adekitan, A. I. y Noma-Osaghae, E. (2019). Data mining approach to predicting the performance of first year student in a university using the admission requirements. *Education and Information Technologies*, 24(2), 1527-1543. <https://doi.org/10.1007/s10639-018-9839-7>
- Almarabeh, H. (2017). Analysis of Students' Performance by Using. *I. J. Modern Education and Computer Science*, 9(8), 9-15. <http://www.mecs-press.net/ijmecs/ijmecs-v9-n8/IJMECS-V9-N8-2.pdf>
- Anderson, J. C. y Gerbing, D. W. (1984). The effects of sampling error on convergence, improper solutions and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika*, (49), 155-173. <https://doi.org/10.1007/BF02294170>
- Baghestani, Ahmad Reza, Ahmadi, Farzane, Tanha, Azadeh y Meshkat, Mojtaba (2019). Bayesian Critical Values for Lawshe's Content Validity Ratio, Measurement and Evaluation in Counseling and Development, 52(1), 69-73, <https://doi.org/10.1080/07481756.2017.1308227>
- Boomsma, A. (1983). On the Robustness of LISREL (maximum likelihood estimation) against small sample size and nonnormality. [Doctoral dissertation, University of Groningen] <https://research.rug.nl/en/publications/on-the-robustness-of-lisrel-maximum-likelihood-estimation-against>
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. Guilford.
- Decreto Supremo N° 352/2003, de 9 de octubre que reglamenta el ejercicio de la función docente. Normas Generales. Diario Oficial de la República de Chile.
- Ferreira de Araújo, F. (2019). *Uma abordagem computacional de previsão de desempenho acadêmico de estudantes em cursos online de programação*. [Tese Doutorado em Sistemas e Computação, Universidade Federal de Campina Grande]. <http://dspace.sti.ufcg.edu.br:8080/jspui/handle/riufcg/12715>
- Galindo, Á. J. y García, H. (2010). Minería de Datos en la Educación. <https://www.it.uc3m.es/jvillena/irc/practicas/10-11/08mem.pdf>
- Hallet, W. (2006). *Didaktische Kompetenzen. Lehr –und Lernprozesse erfolgreich*. Auflage. Klett Lernen und Wissen GmbH.
- Hernández, R., Fernández, C. y Baptista, P. (2010). *Metodología de la Investigación (Vol. 5)*. Mac Graw Hill.
- Hulland, J., Chow, Y. H. y Lam, S. (1996). Use of causal models in marketing. *International Journal of Research in Marketing*, 13(2), 181-197. [https://doi.org/10.1016/0167-8116\(96\)00002-X](https://doi.org/10.1016/0167-8116(96)00002-X)
- Larrañaga, O., Cabezas, G. y Dussailant, F. (2013). Informe completo del Estudio de la Educación Técnico Profesional. Programa de las Naciones Unidas para el Desarrollo - Chile. Área de Reducción de la Pobreza y la Desigualdad. http://www.undp.org/content/dam/chile/docs/pobreza/undp_cl_pobreza_etp_2013.pdf
- Ley 20.903/2016, de 4 de marzo que crea el sistema de Desarrollo Profesional Docente y Modifica otras normas. Ministerio de Educación. <http://bcn.cl/2f72c>
- Manek, S., Vijay, S. y Kamthania, D. (2016). Educational data mining - a case study. *International Journal of Information and Decision Sciences (IJIDS)*, 8(2), 187 - 201. <https://doi.org/10.1504/IJIDS.2016.076517>
- Milborrow, S. (2016). Plotting rpart trees with the rpart. plot package. <http://www.milbo.org/rpart-plot/prp.pdf>
- Ministerio de Educación de Chile, Estrategia Nacional de Formación Técnico-Profesional (2016). Ministerio de Educación de Chile. <http://www.tecnico-profesional.mineduc.cl/wp-content/uploads/2018/03/Estrategia-Nacional-de-Formaci%C3%B3n-T%C3%A9cnico-Profesional.pdf>
- Ministerio de Educación de Chile, Base de Datos Oficiales (2014). Centro de Estudios. <http://centroestudios.mineduc.cl/index.php?t=96&i=2&cc=2519&tm=2>
- Ministerio de Educación de Chile, Bases para una política de formación técnico-profesional en Chile (2009). Informe Técnico. Ministerio de Educación. <https://bibliotecadigital.mineduc.cl/handle/20.500.12365/2257>

- Ministerio de Educación de Chile, Política Nacional de Formación Técnico-Profesional (2006). Ministerio de Educación. <https://bibliotecadigital.mineduc.cl/bitstream/handle/20.500.12365/2168/mono-9000.pdf?sequence=1&isAllowed=y>
- Muñoz Rojas, C. (2019). Educación Técnico-Profesional y autonomía económica de las mujeres jóvenes en América Latina y el Caribe. Serie Asuntos de Género, N°155(LC/TS.2019/26). Comisión Económica para América Latina y El Caribe (CEPAL). https://www.cepal.org/sites/default/files/publication/files/44628/S1900198_es.pdf
- Pearl, J. (2000). Causal inference without counterfactuals: Comment. *Journal of the American Statistical Association*, 95(450), 428-431. https://ftp.cs.ucla.edu/pub/stat_ser/r269-reprint.pdf
- Pérez López, C. y Santín González, D. (2007). Minería de datos. Técnicas y herramientas. Thomson Ediciones Paraninfo, S.A.
- Ponjuán Dante, G. (2015). La gestión del conocimiento desde las ciencias de la información: responsabilidades y oportunidades. *Revista Cubana de Información en Ciencias de la Salud*, 26(3), 206-216. <http://scielo.sld.cu/pdf/ics/v26n3/rci02315.pdf>
- Romero-Jeldres, M., Mardones, T. y Müller, V. (2021). Representaciones discursivas de los estudiantes de Educación Media Técnico Profesional en el discurso docente. *ALPHA. Revista de Artes, Letras y Filosofía*, 52.
- Romero-Jeldres, M. (2020). Competencias pedagógicas. Hacia la construcción de una didáctica para la Educación Media Técnico Profesional. *Revista de Estudios y experiencias en Educación (REXE)*, 19(40), 53-69. <http://dx.doi.org/10.21703/rexe.20201940romero3>
- Romero-Jeldres, Marcela y Faouzi Nadim, Tarik. (2020). Modelo estructural de Competencia Profesional Didáctica para profesores-técnicos no pedagogos. *Magis. Revista Internacional de Investigación en Educación*, 13(27), 1-22. <https://doi.org/10.11144/Javeriana.m13.mecp>
- _____. (2018) Validación de un modelo de competencias pedagógicas para docentes de Educación Media Técnica. *Revista Educación y Educadores*, 21(1).114-132. <https://doi.org/10.5294/edu.2018.21.1.6>.
- Salas- Rueda, R. A. (2020). Percepciones de los estudiantes sobre el uso de Facebook y Twitter en el contexto educativo por medio de la ciencia de datos y el aprendizaje automático. *Píxel-Bit. Revista de Medios y Educación*, (58), 91-115. <https://doi.org/10.12795/pixelbit.74056>
- Salas, R., Castañeda, R., Ramírez, J. y Gamboa, F. (2020). Análisis sobre el uso de Podcast en la Escuela Nacional de Trabajo Social considerando la ciencia de datos y el aprendizaje automático. *Revista Gestión de Personas y Tecnología. Facultad Tecnológica. USACH*, 13(37). <https://www.revistagpt.usach.cl/es/an%C3%A1lisis-sobre-el-uso-de-podcast-en-la-escuela-nacional-de-trabajo-social-considerando-la-ciencia-de>
- Salas-Rueda, R. A. y Salas-Rueda, R. D. (2019). Uso de la ciencia de datos y el aprendizaje automático para analizar la aplicación GeoGebra en el proceso educativo. *Digital Education Review*, (36), 117-151. <https://doi.org/10.1344/der.2019.36.117-151>.
- Sepúlveda, L. (2016). Trayectoria educativo-laborales de jóvenes estudiantes de educación técnica en Chile. ¿Tiene sentido un sistema de formación para el trabajo en la educación secundaria? *Páginas de Educación*, 9(2), 49-84. http://www.scielo.edu.uy/scielo.php?pid=S16887468201600200003&script=sci_abstract
- _____. (2017). Educación Técnica Profesional en el tiempo presente: Nudos críticos y desafíos de futuro. *Cuaderno de Educación*, (77), 1-9. <https://repositorio.uahurtado.cl/bitstream/handle/11242/10501/Leandrosepulveda.pdf?sequence=1&isAllowed=y>
- Servat, B. (2017). Origen, Trayectoria y Efectividad de la Formación de Enseñanza Media Técnico Profesional en Chile. *História DA Educação*, 21(52), 111-135. <https://doi.org/10.1590/2236-3459/67121>
- Sevilla, M. P. (2012). Educación Técnica Profesional en Chile: Antecedentes y claves de diagnóstico. Centro de Estudios. Ministerio de Educación. [http://portales.mineduc.cl/usuarios/mineduc/doc/201204241130130.DiagnosticoEducacionTP CentrodeEstudiosMINEDUC.pdf](http://portales.mineduc.cl/usuarios/mineduc/doc/201204241130130.DiagnosticoEducacionTP%20CentrodeEstudiosMINEDUC.pdf)

- Sukhija, K., Jindal, M. y Aggarwal, N. (2016). Educational data mining towards knowledge engineering: a review state. *International Journal of Management in Education (IJMIE)*, 10(1), 65-76. <https://doi.org/10.1504/IJMIE.2016.073362>
- Urteaga, I., Siri, L. y Garófalo, G. (2020). Predicción temprana de deserción mediante aprendizaje automático en cursos profesionales en línea. *RIED. Revista Iberoamericana de Educación a Distancia*, 23(2). <https://doi.org/10.5944/ried.23.2.26356>

ANEXO

Tabla 4. Valores críticos del índice CVR (nivel de significancia $\alpha=.05$)

N	CVR crítico	N	CVR crítico
2	--	22	.36
3	--	23	.39
4	1.00	24	.42
5	1.00	25	.36
6	1.00	26	.38
7	.71	27	.33
8	.75	28	.36
9	.78	29	.31
10	.60	30	.33
11	.64	31	.35
12	.50	32	.31
13	.54	33	.33
14	.57	34	.29
15	.47	35	.31
16	.50	36	.28
17	.41	37	.30
18	.44	38	.32
19	.47	39	.28
20	.40	40	.30
21	.43		

