

DETERMINACIÓN DE VARIABLES PREDICTIVAS DE DESERCIÓN INICIAL PARA GENERAR UN SISTEMA DE ALERTA TEMPRANA. ANÁLISIS SOBRE UNA MUESTRA DE ESTUDIANTES BENEFICIARIOS DE LA BECA DE NIVELACIÓN ACADÉMICA EN UNA UNIVERSIDAD PÚBLICA EN CHILE

Cristian Contreras¹

RESUMEN

Se analizaron los resultados cuantitativos del programa Beca de Nivelación Académica, año 2016, para una muestra de 250 estudiantes beneficiarios de ingreso a primer año en una universidad pública en Chile. Se revisaron los datos estadísticos, utilizando las técnicas de clustering y árbol de decisiones, en búsqueda de información en las bases de datos institucionales que resultara relevante para la generación de un sistema de alerta temprana, que permitiera seleccionar a los estudiantes que presentaran un mayor riesgo de deserción inicial para un acompañamiento focalizado oportuno. Se concluyó que existe al menos una variable de caracterización al ingreso a primer año que permite discriminar el riesgo de deserción dentro de la muestra seleccionada correspondiente al puntaje obtenido en la prueba específica de matemáticas en la prueba de selección universitaria. Esta investigación tiene relevancia pues aporta evidencia predictiva que permite orientar los esfuerzos que hoy las diferentes universidades desarrollan para mejorar sus índices de permanencia, especialmente en relación con la reciente diversificación de la población que ingresa a la educación superior, a propósito de las políticas de acceso universal y gratuidad, y de equidad e inclusión.

Conceptos clave: democratización de la educación, deserción estudiantil, indicadores educativos, política educacional, diversificación de la educación.

¹ Universidad de Valparaíso, Valparaíso, Chile. Contacto: cristian.contrerasca@uv.cl

SETTING PREDICTOR VARIABLES TO GENERATE AN EARLY DROPOUT PREDICTION MODEL FOR FIRST-YEAR UNIVERSITY STUDENTS. ANALYSIS OF A SAMPLE OF FIRST-YEAR STUDENTS BENEFICIARIES OF THE ACADEMIC LEVELING SCHOLARSHIP IN A PUBLIC UNIVERSITY IN CHILE

ABSTRACT

The quantitative results of the program “Academic Leveling Scholarship” from year 2016 were analyzed, from a sample of 250 beneficiaries. Statistic datasets were checked using Decision Tree and Clustering techniques in the search of information in institutional data bases that will be relevant to develop an Early Dropout Prediction Model. This model will allow for selecting students with a higher risk of dropping out to receive a timely focused accompaniment. It concluded that there exists, at least, one characterization variable upon the first-year students that allows for identifying the dropout risk in the selected sample regarding the score obtained in the specific math test of the university selection process.

The relevance of this research is its contribution to predictive evidence to align efforts that different universities are making to improve their permanence indexes, especially regarding the recent diversification of the population that has been admitted to higher education due to universal access policy, free education, equity and inclusion.

Key concepts: Democratization of education, Dropping out, Educational indicators, Educational policy, Diversification of education.

Deserción y permanencia en la educación superior

En la actualidad existe consenso en describir el abandono en educación superior (ES) desde el modelo explicativo de Tinto (1975, 1989), ya que es el que más ha influido en los estudios de deserción (Castaño, Gallón, Gómez y Vásquez, 2004; Donoso y Schiefelbein, 2007; Díaz, 2008; Saldana y Barriga, 2010; Arriaga, Burillo, Carpeño y Casaravilla, 2011; Bernardo, Cervero, Esteban, Fernández, Solano y Agulló, 2018). Este concepto se define como un abandono voluntario que puede ser explicado por diferentes categorías de variables socioeconómicas, individuales, institucionales y académicas. Sin embargo, la forma de operacionalizarlas depende del punto de vista desde el cual se haga el análisis, esto es, individual, institucional y estatal o nacional. En general, se asume que el estudio de la deserción en la educación superior es extremadamente complejo, ya que implica no solo una variedad de perspectivas, sino que, además, una gama de diferentes tipos de abandono (Díaz, 2008; Arriaga et al., 2011; Bernardo et al., 2018). Las variables que presentan una mayor frecuencia en los estudios se agrupan en cuatro categorías: individuales, académicas, institucionales y socioeconómicas (Díaz, 2008).

Basándose en el modelo de Tinto (1975, 1989), que propone que el estudiante ingresa a la universidad con una gama de atributos, experiencias y antecedentes familiares que afectan directa o indirectamente su desempeño, Díaz (2008) propuso un modelo conceptual para la deserción estudiantil universitaria chilena que considera tres ejes principales:

- Factores académicos: rendimiento académico, carga académica, compromiso con la carrera, calidad de la carrera y actividades extracurriculares relacionadas con la formación profesional.
- Factores institucionales: procesos de admisión, calidad del cuerpo académico, apoyo financiero (becas y créditos), servicios estudiantiles y calidad del programa.
- Factores sociales: actitud hacia los pares, habilidades comunicacionales, valores culturales, apoyo familiar, apoyo de pares, estrato socioeconómico y expectativas laborales.

El autor describe una dinámica en la que el estudiante se moviliza dentro de estos factores, los que aparecen representados por ejes en su modelo, distribuyendo la fuerza de cada uno según la debilidad de los demás y forzando a los otros ejes a reconfigurarse, para así mantener el equilibrio dinámico que exige mantenerse dentro de la universidad. Si uno de los ejes se debilita, el foco se moverá. A través de esta dinámica, sostiene que la deserción y permanencia son el resultado del grado de motivación que poseen los estudiantes, en el cual la motivación se relaciona directamente con la integración académica y social del estudiante. La integración académica, por su parte, se ve afectada por características preuniversitarias e institucionales, las cuales poseen las siguientes variables:

- Características preuniversitarias: dependencia administrativa del colegio, rama educacional del establecimiento de origen, nota promedio de la enseñanza media, puntaje promedio de la Prueba de Selección Universitaria (PSU), edad de egreso de la enseñanza media, edad de ingreso a la universidad y preferencia por la carrera.
- Características institucionales: grados académicos del cuerpo académico, condición de acreditación de la carrera, carga académica, rendimiento académico, vinculación externa, becas y créditos de financiamiento de estudios, infraestructura, satisfacción de los servicios estudiantiles y de la relación académico-estudiante.

Por otro lado, la integración social se ve afectada por características familiares e individuales, cuyas variables son:

- Características familiares: número de integrantes del grupo familiar, nivel educacional de los padres, situación laboral de los padres, ingreso económico familiar, lugar de residencia (cercanía con la universidad) y valores personales (familiares y socioculturales).
- Características individuales: edad, género, estado civil, situación laboral, horas de trabajo, compromiso inicial con la carrera, compromiso con metas parciales, satisfacción con la relación de pares, calidad de salud, técnicas y habilidades de estudio.

Además, Díaz (2008) menciona que la integración académica y social es afectada por las expectativas laborales de los alumnos. El modelo conceptual propuesto asume que todas las características actúan en forma permanente sobre el alumno durante sus años de estudios, por lo que está sometido a una tensión continua entre estos factores, manteniendo así un equilibrio en la intención de permanecer en la carrera. Cuando se rompe este equilibrio, abandona la carrera, la universidad o la educación terciaria del todo (Díaz, 2008).

La tasa de retención de estudiantes en la ES, en especial la de primer año, es uno de los indicadores internacionales más utilizados para evaluar la eficiencia interna de las instituciones de educación superior (IES), considerando que la mayor parte de la deserción de estudiantes se da precisamente en ese periodo. Los niveles y variaciones de la tasa de retención y deserción de los alumnos al cabo de su primer año de estudios constituyen una de las grandes preocupaciones del sistema de educación superior en Chile, y del Ministerio de Educación (Mineduc) en particular, ello porque altos índices de retención no solo pueden ser indicadores de buena gestión de las IES, sino que también pueden traer consigo importantes ventajas para los estudiantes y su entorno familiar, así como para el propio Estado (SIES, 2015). De acuerdo con estimaciones basadas en estadísticas nacionales entregadas por el Servicio de Información de Educación Superior (SIES, 2015; 2016), la tasa de retención global de pregrado evaluada para 2014 fue cercana al 76,6%, mientras que la misma tasa para 2015 fue de 77,2%. Al considerar las áreas de estudio, las relacionadas con Salud y Agropecuaria son las con mayor tasa de retención al primer año para ambas cohortes, con 80,2% y 77,4% en 2014, y 81,8% y 79,8% en 2015, respectivamente. Por su parte, las áreas con menor retención son Humanidades y Artes para ambas cohortes, con 72,4% y 70,8% en 2014, y 72,8% para ambas áreas en 2015. Trece de las 20 carreras con mayor matrícula superan el 80% de retención de primer año en 2014, y 15 de las 25 carreras en la misma situación en 2015. Para ambas cohortes, Medicina es la carrera con mayor retención, con un 94,7% en ambos años. Las IES acreditadas tienen persistentemente mayor tasa de retención de primer año que aquellas instituciones no acreditadas. Para 2014, las instituciones acreditadas tuvieron una tasa de retención de 72,2% y

las instituciones sin acreditación un 53,7%. Para la cohorte del año 2015, las instituciones acreditadas muestran una tasa de 72,7% y las sin acreditación de 57,2%. En general, los matriculados en la ES provenientes de colegios particulares pagados tienen mayores tasas de retención que los particulares subvencionados y municipales, con tasas de 79,1%, 72,1% y 68,5%, respectivamente para 2014, y de 78,0%, 73,1% y 69,4%, respectivamente para 2015. Los matriculados provenientes de establecimientos científico-humanistas tienen mayores tasas de retención en las IES que los provenientes de establecimientos técnico-profesionales, con una tasa de 79,3% contra un 71,1% en 2014, y de 79,4% contra un 72,1% en 2015. En términos de género, consistentemente las mujeres muestran una tasa de retención de primer año más alta que la de los hombres, con un 73,1% versus 67,4% para la cohorte 2014, y de 74,3% versus 68,1% para la cohorte 2015.

Hasta 2011, en la mayoría de los estudios prevalece la caracterización de la población desertora y la estadística descriptiva del fenómeno a través de las distintas cohortes con un abordaje metodológico de tipo estático. Desde entonces la investigación nacional se ha volcado a la elaboración de modelos predictivos y a la construcción de sistemas de alerta temprana (SAT) que lleguen a ser capaces de predecir el abandono y así las IES puedan propiciar la implementación de las medidas necesarias para aumentar las tasas de retención, y a la elaboración de indicadores de eficiencia que permitan que el gasto que se destina a la retención tribute más directamente en los resultados obtenidos.

Iniciativas para favorecer el acceso a la educación superior

El sistema de educación superior chileno mantiene amplios niveles de inequidad debido a numerosos factores, tanto del propio sistema como externos. Sin perjuicio de lo anterior, las brechas de la tasa neta de asistencia a la educación superior, entre quintiles de ingreso socioeconómico, se han visto reducidas durante los últimos años (Zapata y Tejeda, 2016). En efecto, mientras en 2003 el quinto quintil de ingreso socioeconómico presentaba una tasa neta de asistencia a

la educación superior de cinco veces el primer quintil, al año 2013 la brecha se redujo y la tasa neta de asistencia del quinto quintil alcanzó a dos veces más que la del primero. Para 2016 la Encuesta de Caracterización Socioeconómica (CASEN) señalaba que la cobertura de la educación superior era de 27,4% para el quintil I, mientras que en el quintil V fue de 57,5%. Los datos presentados dan cuenta de un incremento gradual y sostenido en el tiempo de acceso de grupos que anteriormente no ingresaban a este nivel educativo (Espinoza, González y Latorre, 2009; Soto, Díaz y Chiang, 2017). El acceso masificado de estudiantes de origen vulnerable a las universidades chilenas demanda que estas instituciones desarrollen mecanismos de diversa índole, que garanticen equidad en los resultados de aprendizaje, retención y titulación oportuna (Soto et al., 2017). El crecimiento de la cobertura de acceso a la educación superior ha evidenciado que las cifras de abandono o deserción universitaria han aumentado considerablemente, acentuando las brechas de éxito y fracaso académico entre estudiantes con mayores y menores condiciones socioculturales (Basso, 2016; Castillo, Morales y Miranda, 2019).

A principios del presente siglo, aún se sostenía que la permanencia de los estudiantes dependía exclusivamente de ellos mismos, en consideración de sus intereses, expectativas y capacidades, y de su círculo familiar inmediato, liberando a las IES de la responsabilidad que podrían tener en el proceso formativo (Donoso, Donoso y Arias, 2011). Desde entonces, conceptualmente, la retención se ha enfrentado desde mediante enfoques parciales hasta modelos más complejos, que han permitido mayor visibilización pública de la temática y de su magnitud, y la identificación de sus causas, estrategias y planes de intervención (Altbach, 2016; Donoso et al., 2011). Sin embargo y debido a la progresiva integración de estudiantes de grupos sociales minoritarios, con características desconocidas por las IES, este ciclo ha requerido respuestas más eficientes que las estrategias tradicionales (Gairín-Sallán, 2015). Las políticas públicas pro retención han pasado desde propuestas generales a iniciativas crecientemente más específicas, buscando dar un mejor y más oportuno apoyo a los estudiantes que se han incorporado a la educación terciaria y que requieren de soluciones

diferentes a las convencionales, procesos que en algunos casos están en desarrollo y aún no se consolidan.

En una primera etapa, las políticas públicas destinadas a ampliar el acceso a la educación superior fueron orientadas al establecimiento de un sistema de financiamiento, asociado a programas de ayuda estudiantil como créditos y becas (Espinoza et al., 2009; Soto et al., 2017). Posteriormente se incluyó el uso del “ranking de notas” como una variable que se sumó al Sistema Único de Admisión (SUA), además del puntaje en la PSU y el promedio de notas de la enseñanza media (NEM). El ranking de notas, como factor, reconoce el esfuerzo de los estudiantes durante su trayectoria escolar, beneficiando a quienes presentan un buen desempeño en su contexto, mediante una bonificación en su puntaje PSU si se encuentran por sobre la media de su propio colegio (Del Valle, 2017). Comenzó a operar a partir de 2012 y ya ha evidenciado ser un buen predictor de éxito académico (CRUCH, 2017). En particular, se destaca el potencial de su aporte en dos ámbitos: primero, motivar a los estudiantes al estudio durante toda su enseñanza media, con las externalidades positivas que esto puede tener, tanto para el sistema escolar como posteriormente para el universitario; segundo, en el ámbito de la equidad, contribuir al acceso a la universidad de estudiantes de menor nivel socioeconómico, en especial aquellos pertenecientes a establecimientos educacionales de sectores más vulnerables (Larroucau, Ríos y Mizala, 2015; Del Valle, 2017).

Actualmente, las instituciones han comenzado a atender de manera focalizada los problemas académicos y de integración que afectan a los estudiantes más vulnerables, mediante programas de apoyo académico y de inducción a la vida universitaria, reconociendo que la deserción involucra otros factores además de los económicos (Soto et al., 2017). En 2007 fue por primera vez implementado el Propedéutico en la Universidad de Santiago de Chile, para luego ampliarse a una Red de Propedéuticos conformada, con el patrocinio de la UNESCO, por todas las universidades que en el país desarrollan programas con bases comunes (Figueroa, Maillard, Véliz, Toledo, Andaur, Quezada y González, 2016), y ya en 2016 contaba con 18 universidades (Orellana, Moreno y Gil, 2015; Blok, 2016). Este

programa apunta no solo al acceso inclusivo en tanto mecanismo de ingreso especial, sino también a la preparación de los estudiantes para la vida universitaria, de modo que, a pesar del dispar acceso a oportunidades que han tenido, logren transitar con éxito por la ES, de modo que la equidad no esté solo en el acceso sino también en el tránsito y en el egreso oportuno (Espinoza y González, 2015; Sánchez, Silva y Jouannete, 2011). En última instancia, cabe también destacar que el Propedéutico ha servido en Chile como base para el actual Programa de Acompañamiento y Acceso Efectivo a la Educación Superior (PACE), que opera como un trabajo conjunto entre el Ministerio de Educación (MINEDUC) y las universidades.

En 2014, como política pública, el gobierno chileno impulsó el Programa de Acceso a la Educación Superior, PACE, que ofrece preparación en la enseñanza media a estudiantes vulnerables —sin exigir rendimiento académico previo—, cupos especiales en algunas instituciones que participan voluntariamente del programa y cursos de nivelación (Zapata y Tejada, 2016). La primera promoción se incorporó al sistema de educación superior a partir de 2016, de manera que sus resultados son altamente esperados (Donoso-Díaz, Neira y Donoso, 2018), aunque preliminarmente se ha reportado un 85% de retención total al segundo año para la primera generación que comenzó el programa en 2014 e ingresó a la ES en 2016 (SIES, 2017). También en 2016 se inició la gratuidad en la educación superior, dejando atrás el sistema de becas de matrícula y arancel que respondía a un modelo subsidiario dirigido a individuos que cumplieran determinados requisitos. En este panorama reciente no existen evaluaciones relevantes que permitan identificar la progresión de los alumnos a partir de sus datos registrados en los múltiples sistemas informáticos y bases disponibles en las distintas IES, al menos no aún para estas poblaciones específicas de estudiantes que comienzan a hacer uso de estos recientes apoyos estudiantiles. Al respecto, una de las medidas recientes fue la investigación sobre políticas y prácticas de eficiencia interna y externa que mejoren los índices de retención de éstos y de todos los estudiantes. Esta iniciativa ha considerado aportes económicos y el compromiso de las instituciones por implementar estructuras organizacionales y prácticas multidimensionales que involucren a sus principales actores institucionales. Ello implica

soporte a los estudiantes en materia psicosocial y psicoeducativa, coordinación de esfuerzos para la retención de estudiantes, mejora en la evaluación y el reporte de indicadores de desempeño académico, un mejor clima para la diversidad cultural y el desarrollo de SAT para implementar medidas de apoyo oportunas y eficaces para estudiantes en riesgo de abandono (Donoso-Díaz et al., 2018). Lo cierto es que las IES, en forma creciente, enfrentan tensiones por su rápida expansión, la necesidad de sincronizarse con la calidad de la enseñanza y sus variables determinantes, entre ellas la retención de los estudiantes más vulnerables (Donoso-Díaz et al., 2018).

Iniciativas pro permanencia de estudiantes de educación superior: la nivelación académica

En Chile, desde 1999, el Programa de Mejoramiento de la Calidad y Equidad de la Educación Superior (MECESUP) forma parte de los esfuerzos del Estado para apoyar la transición de su economía actual a una basada en el conocimiento, incrementando la equidad y la efectividad de su sistema de ES. Los objetivos del programa MECESUP van dirigidos a proveer las condiciones necesarias que permitan mejorar la calidad de la oferta educacional de las universidades estatales y del sistema de educación superior en general, aumentar la competitividad internacional, sostener el desarrollo económico y social, y asegurar que ningún talento se pierda por diferencias en las oportunidades de aprendizaje. Concretamente el programa MECESUP pertenece a la División de Educación Superior del Ministerio de Educación de Chile y otorga financiamiento incremental a proyectos y planes de mejoramiento institucionales adjudicados en forma competitiva a través del Fondo de Innovación Académica (FIAC) y por negociación institucional con Convenios de Desempeño (CD). Por una parte, el FIAC es un instrumento de adjudicación competitiva de recursos, que busca incentivar actividades de fomento a la calidad e innovación académica en instituciones elegibles del sistema de educación terciaria, con características de coherencia, capacidad de respuesta, equidad y calidad (MINEDUC, 2011). Por otra parte, los CD constituyen contratos entre el Estado y las IES, por medio de los

cuales éstas comprometen desempeños notables que impliquen un significativo mejoramiento institucional que por sí solas no habrían podido lograr. Es decir, los CD proveen financiamiento sobre la base de resultados e impacto, permitiendo a las instituciones generar iniciativas capaces de enfrentar problemas estructurales de diferente tipo y focalizar sus esfuerzos institucionales en aspectos asociados al mejoramiento de la calidad de su quehacer. Desde su implementación inicial, en 1999, el programa MECESUP ha contado con el sostenido apoyo financiero del Estado chileno y del Banco Mundial. Ha financiado acciones de mejoramiento académico e infraestructura en las 25 universidades del Consejo de Rectores de Universidades Chilenas (CRUCH), contribuyendo a la ampliación del acceso a la educación superior de los estudiantes en todo el país y a un aumento significativo en la equidad y la calidad en el sistema. Una primera etapa se ejecutó entre 1999 y 2005, la segunda entre 2005 y 2011, y la tercera entre 2012 y 2016.

Una de las dimensiones consideradas para resolver la adjudicación de recursos es el impacto de los servicios prestados a los estudiantes y es precisamente en esta dimensión que aparece la Beca de Nivelación Académica (BNA). La beca se enmarca en las iniciativas que el Ministerio promueve para alcanzar una mayor equidad en la educación superior con enfoque en la permanencia; se enfoca específicamente en lograr que los estudiantes beneficiarios alcancen un adecuado desempeño y éxito académico en su transición a la educación superior, mediante la implementación de programas institucionales para nivelar sus competencias académicas. Este programa marca un punto de inflexión en el curso de las políticas públicas, revelando un cambio en la concepción del éxito y fracaso académico en la universidad, desde el fracaso como responsabilidad individual al fracaso como una responsabilidad colectiva de los estudiantes, las carreras y las IES (Donoso-Díaz et al., 2011; Soto et al., 2017). Este programa tiene dos vertientes en el MINEDUC: por una parte, el Departamento de Ayudas Estudiantiles adjudica las becas a estudiantes meritorios de los tres primeros quintiles de ingreso familiar, de establecimientos municipales o particulares subvencionados y con promedio igual o superior a 6.0. En este proceso la universidad no participa ni tiene injerencia. El MINEDUC

cruza la información disponible acerca de los estudiantes que ingresan a la universidad, los selecciona y envía la nómina a la institución para que se gestione su incorporación al proyecto (Soto et al., 2017). Por otra parte, el Departamento de Financiamiento Institucional convoca, selecciona y adjudica recursos públicos a programas de nivelación académica postulados por las IES. Esto conduce a que, una vez que se adjudican los programas, cada IES recibe estudiantes focalizados. En general, los estudiantes BNA son beneficiados con becas de arancel debido a su situación socioeconómica; no obstante, adjudicarse la beca de arancel no implica la obtención de una BNA, y viceversa. En síntesis, la BNA tiene un doble propósito: mejorar el rendimiento académico y la permanencia en la universidad de los estudiantes de origen vulnerable, y mejorar los indicadores institucionales de retención, particularmente en aquellas carreras con antecedentes de baja tasa de retención (Soto et al., 2017). Entre 2011 y 2018, cuarenta y cinco instituciones de educación superior se adjudicaron los recursos para la implementación de un total de 109 programas distintos, alcanzando un número estimado de 3.295 estudiantes beneficiarios.

Al tanto de este desafío —y avanzando en el desarrollo de la política de equidad e inclusión—, es que este trabajo se propuso aportar en la generación de evidencia respecto de variables predictivas del riesgo de deserción inicial en una muestra de estudiantes con alto riesgo de abandono. La muestra corresponde a estudiantes beneficiarios de la BNA en la Universidad de Valparaíso (UV), la que se adjudicó en 2016 atendiendo a las necesidades psicosociales de los nuevos estudiantes que la universidad convoca y en pos de la retención y el logro educativo de la diversidad de perfiles que hoy ingresan a la institución. Se propuso el desarrollo de un modelo estadístico que permitiera identificar factores relacionados al riesgo de abandono de los estudiantes en función de un conjunto de variables de ingreso, aplicando análisis de datos basado en *clustering* y árboles de decisión. Como principal herramienta para este estudio se utilizó IBM SPSS 24.

Caracterización de la muestra

En 2016, la UV presentó la BNA como un programa de intervención con una red de apoyo institucional, que buscó complementar la instalación de un modelo más amplio de acompañamiento para estudiantes en tránsito a la educación superior, o como parte de un modelo educativo global, guardando las proporciones de su alcance, generando un programa de nivelación de competencias básicas y acompañamiento integral para estudiantes de primer año, que entregó un soporte a las dimensiones académicas y psicosociales de la vida universitaria, con el objetivo de asegurar su inserción y éxito académico. Dentro de las acciones comprometidas se consideró mentorías de pares, tutorías académicas, talleres de competencias y de aprendizaje, y acompañamiento individual de los estudiantes. Estas estrategias se centraron en las líneas de trabajo de autogestión del aprendizaje, orientación vocacional, estrategias de aprendizaje, habilidades del pensamiento, manejo del tiempo, fortalecimiento de la autoestima, definición de metas, habilidades sociales, toma de decisiones, motivación y manejo del estrés.

Del universo total de estudiantes, la BNA seleccionó una muestra de 250 de ingreso a primer año en el campus Casa Central, focalizados desde el MINEDUC. Correspondiendo a 40 de Ingeniería Comercial, 31 de Ingeniería Civil Informática, 15 de Educación Parvularia, 61 de Ingeniería Civil Industrial, 8 de Gestión en Turismo y Cultura, 30 de Trabajo Social, 26 de Diseño, 19 de Ingeniería Civil Biomédica y 20 de Ingeniería Ambiental. Las carreras fueron seleccionadas sobre la base de criterios cuantitativos y cualitativos, vinculados a indicadores de progresión y logro, índice de vulnerabilidad escolar (IVE) y perfil de sus estudiantes. En términos cuantitativos, para 2015 es posible informar que el porcentaje de retención promedio de estas carreras fue de 74,6% y el de aprobación fue de 71,5%, lo cual se encuentra bajo la media institucional para el mismo periodo (80,4% y 81,9% respectivamente). En relación con el IVE de las carreras seleccionadas, éste también se encuentra bajo el promedio con un 48,3 contra un 49,1. El ranking, en tanto, promedia un total de 595,1, lo que también está por debajo de la media institucional de 650. En términos cualitativos, la selección de estas carreras se

asoció con los resultados en los diagnósticos institucionales que cada año la universidad realiza a la población de estudiantes que ingresa a primer año, los que evalúan las competencias de pensamiento lógico matemático, lectura comprensiva y autorregulación.

Tabla1
Cantidad de estudiantes por carrera

Nombre de carreras focalizadas	Facultad o unidad académica	Cantidad de estudiantes de sexo femenino	Cantidad de estudiantes de sexo masculino	Cantidad total de estudiantes
Educación Parvularia	Facultad de Medicina	15	0	15
Ingeniería Civil Biomédica	Facultad de Ingeniería	10	9	19
Ingeniería Civil Industrial	Facultad de Ingeniería	25	36	61
Ingeniería Civil Informática	Facultad de Ingeniería	28	3	31
Ingeniería Ambiental	Facultad de Ingeniería	11	9	20
Matemática	Facultad de Ciencias	0	0	0
Diseño	Facultad de Arquitectura	22	4	26
Gestión en Turismo y Cultura	Facultad de Arquitectura	7	1	8
Trabajo Social	Facultad de Derecho y Ciencias Sociales	25	5	30
Ingeniería Comercial	Facultad de Ciencias Económicas y Administrativas	22	18	40

Incluye distribución de estudiantes por sexo según carrera. Para el caso de la carrera de Matemáticas, ésta se seleccionó desde la Universidad, pero luego desde el Ministerio no se focalizaron estudiantes.

Seleccionadas las carreras por la Universidad, el Ministerio seleccionó a los estudiantes focalizados según los términos referenciados. A partir de la aplicación de una encuesta de caracterización, creada por el programa con el fin de profundizar en información respecto de distintas temáticas de los estudiantes beneficiarios, podemos señalar que un 75% de los estudiantes

proviene de la región de Valparaíso y un 25% de otras regiones. 85% de los estudiantes proviene de establecimientos científico-humanistas y 15% de técnico-profesionales. Un 78% de los estudiantes cuenta con gratuidad, 17% solo con becas ministeriales y 5% con crédito universitario. El 3% de los estudiantes es madre/padre y 6% de los estudiantes trabaja además de estudiar.

En relación con la variable dependiente de este estudio, según indica la tabla 2, al término del cuarto semestre el porcentaje global de retención fue de 74%. En relación con la media institucional, sobresalen bajos índices de retención para las carreras de Ingeniería Civil Informática, Ingeniería Ambiental e Ingeniería Comercial. Altos índices de retención se observan en la carrera de Ingeniería Civil Industrial.

Tabla 2
Cantidad de desertores por carrera

Nombre de la carrera	Cantidad de estudiantes que permanecen	Cantidad de estudiantes que desertan	Porcentaje de retención
Educación Parvularia	11	4	73,3%
Ingeniería Civil Biomédica	14	5	73,6%
Ingeniería Civil Industrial	55	6	90,1%
Ingeniería Civil Informática	17	14	54,8%
Ingeniería Ambiental	11	9	55%
Diseño	22	4	84,6%
Gestión en Turismo y Cultura	6	2	75%
Trabajo Social	24	6	80%
Ingeniería Comercial	25	15	62,5%
Total	185	65	74%

La cantidad total de estudiantes que permanecen es de 185, mientras que la de estudiantes que desertan es de 65. El porcentaje global de retención es de 74%.

La tabla 3 muestra que, al término del segundo año, el porcentaje de retención en la población femenina llega a 78,57% y en la masculina a 68,18%.

Tabla 3
Cantidad de desertores por sexo

Permanencia	Cantidad de estudiantes de sexo femenino	Cantidad de estudiantes de sexo masculino
Permanece	110	75
Abandona	30	35
Total	140	110
Porcentaje de retención	78,57%	68,18%

La tabla muestra el cruce entre la variable *permanencia* y la variable *sexo*.

La información en la tabla 4 muestra la comparación de las tasas de retención por carrera con las tasas de retención institucional para 2016 y 2017. En el gráfico 1 aparece la tabla de frecuencias para la variable *semestre*. Esto nos señala que 10 estudiantes (4% de la muestra total) desertaron durante el primer semestre, 19 (7,6%) durante el segundo, 21 (8,4) en el tercero y 16 (6,4%) en el cuarto. Del mismo gráfico también se puede obtener los resultados para la variable *periodo*, en los que 29 estudiantes abandonaron durante el 2016 y 37 lo hicieron en el 2017. Se observan resultados positivos en las carreras de Educación Parvularia e Ingeniería Civil Industrial cuando se comparan los resultados por carrera con los resultados institucionales. Cinco estudiantes han reingresado a la universidad, esto es, un total de 7% de los estudiantes que desertaron.

Tabla 4
Tasa de retención institucional para primer año

Carrera	Tasa de retención por carrera 2016	Tasa de retención por carrera 2017	Tasa de retención institucional 2016	Tasa de retención institucional 2017
Educación Parvularia	89%	87%	83%	84%
Ingeniería Civil Biomédica	81%	80%	83%	84%
Ingeniería Civil Industrial	93%	89%	83%	84%
Ingeniería Civil Informática	67%	79%	83%	84%
Ingeniería Ambiental	82%	83%	83%	84%
Diseño	83%	80%	83%	84%

28 DETERMINACIÓN DE VARIABLES PREDICTIVAS DE DESERCIÓN INICIAL PARA GENERAR UN SISTEMA DE ALERTA TEMPRANA. ANÁLISIS SOBRE UNA MUESTRA DE ESTUDIANTES BENEFICIARIOS DE LA BECA DE NIVELACIÓN ACADÉMICA EN UNA UNIVERSIDAD PÚBLICA EN CHILE - C. Contreras

Gestión en Turismo y Cultura	76%	78%	83%	84%
Trabajo Social	79%	86%	83%	84%
Ingeniería Comercial	81%	84%	83%	84%

Trayectoria de las tasas de retención institucional y por carreras para 2016 y 2017 (Anuario UV, 2017).

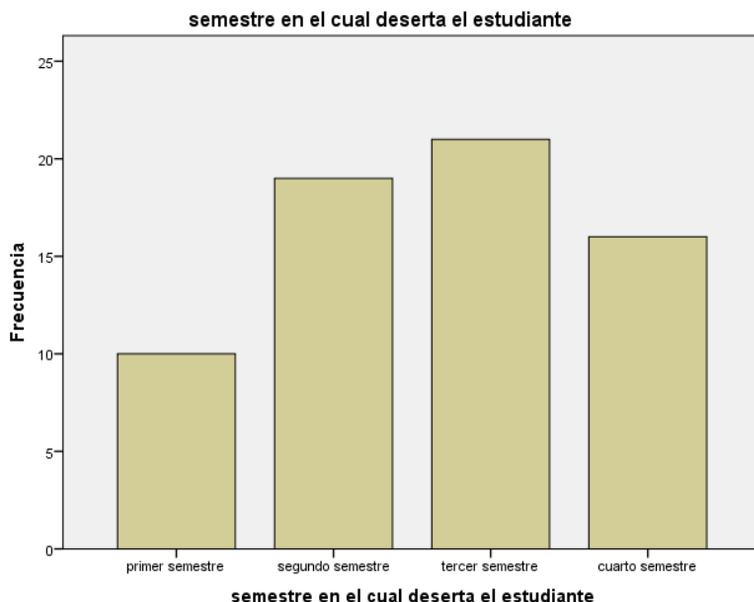


Gráfico 1

Abandono de estudiantes por semestre

Frecuencia de la cantidad de estudiantes que abandonan según semestre académico.

Hasta aquí resultados descriptivos, pero cabe preguntarse cómo se relaciona la retención con otras variables no visibles descriptivamente. Para responder a ello se realizó un análisis de datos que permitió reconocer qué condiciones son las que aumentan la probabilidad de permanencia o abandono.

Método y resultados

La fase de análisis de los datos comprendió la recolección de información de los estudiantes beneficiarios de la BNA para 2016, obtenidos desde diferentes sistemas. La selección de las variables a utilizar para este estudio se basó principalmente atendiendo al factor de disponibilidad, aunque muchas de éstas han sido consideradas determinantes en la bibliografía sobre deserción estudiantil universitaria (Gil, Paredes y Sánchez, 2013; Larroucau et al., 2015). Inicialmente se contó con 27 variables que se detallan a continuación.

Tabla 5
Variables iniciales

Variable	Descripción
Permanencia	Si el estudiante permanece o abandona su carrera inicial
Sexo	Sexo del estudiante
Carrera	Carrera a la cual ingresó el estudiante
Preferencia	Lugar de preferencia de postulación a la carrera
Lugar	Lugar en el cual quedó seleccionado
BEA	El estudiante tiene Beca de Excelencia Académica (BEA)
Puntaje	Puntaje ponderado de ingreso
Región	Región de origen del estudiante
Comuna	Comuna de procedencia del estudiante
Promnota	Promedio de Enseñanza Media
NEM	Puntaje NEM asignado al promedio de notas
Lenguaje	Puntaje en la PSU de lenguaje
Matemáticas	Puntaje en la PSU de matemáticas
Historia_Sco	Puntaje obtenido en la PSU de Historia y Ciencias Sociales
Ciencias	Puntaje obtenido en la PSU de Ciencias
Prommlm	Promedio ponderado entre las pruebas de Matemáticas y de Lenguaje
Vía Ingreso	Vía de ingreso (considera vías de ingreso especial)
Ptje. ranking	Puntaje otorgado al estudiante con base en su posición relativa en cada contexto educativo en el que estuvo
PPN1	Promedio ponderado de notas primer semestre
PPN2	Promedio ponderado de notas segundo semestre
PPN3	Promedio ponderado de notas tercer semestre
PPN4	Promedio ponderado de notas cuarto semestre
Avance	Porcentaje de avance del estudiante en relación con la malla curricular de su carrera inicial
Estado	Estado académico en el cual el estudiante finaliza; considera hasta el cuarto semestre
Semestre	Semestre en el cual deserta el estudiante
Periodo	Año en el cual deserta
Reingreso	Considera a aquellos estudiantes que, habiéndose retirado de su carrera inicial, reingresan a la misma universidad

Descripción de la totalidad de las 27 variables inicialmente contempladas.

Una vez construida la base de datos, se realizó el proceso de selección de las variables de entrada para el modelamiento. Así se determinó qué variables, de todas las obtenidas para la base de datos, presentan una mayor relevancia para este estudio o cuáles aportaban una información redundante o secundaria. Se realizó la selección de variables para utilizar en los análisis de datos cumpliendo dos criterios: disminuir el riesgo de colinealidad entre variables y propender a la disminución de variables que ya hayan sido ampliamente estudiadas según la bibliografía revisada, de manera de obtener resultados novedosos que aportaran nueva información.

Existen distintas técnicas que ayudan a resolver la colinealidad (Méndez-Ramírez, Moreno-Macías, Méndez Gómez-Humarán y Murata, 2014). De entre ellas se escogió el método de conglomerados bietápico, en búsqueda de variables significativas que formaran al menos tres conglomerados. Se considera esta información como novedosa ya que, por una parte, la formación de un único clúster homogeniza el valor de cada variable para el agrupamiento y, por otro, la formación de dos conglomerados resulta redundante frente a la naturaleza dicotómica de la variable dependiente del estudio: *permanencia*. También se consideró el conocimiento teórico para justificar la eliminación de variables. En el primer agrupamiento se formaron dos clústeres con un alto nivel de coherencia interna y de separación entre conglomerados. Las variables más significativas para el agrupamiento resultaron ser *periodo*, *comuna*, *región*, *BEA*, *carrera*, *sexo* y *semestre*. Todas estas variables tienen en común el ser nominales y algunas de ellas también dicotómicas, por lo que es esperable que constituyan dos clústeres distintos. En esta primera instancia el análisis no arrojó información relevante o que no fuera previamente conocida para el agrupamiento; por ejemplo, si se considera que para variables como *sexo* existe abundante evidencia sobre una mayor tendencia de persistencia en la educación superior para estudiantes de sexo femenino (OCDE, 2009; MINEDUC, 2012; Larrocau et al., 2015). De esta manera se procedió a eliminar precisamente estas variables para realizar un nuevo agrupamiento bietápico en el que se produjo un único clúster. Con relación a esto es que se comenzaron a descartar, una a una, determinadas variables, considerando como criterio que el análisis arrojara agrupamientos novedosos. Variables recursivas

de colinealidad resultaron ser los promedios ponderados de notas, porque es esperable que éstos se correlacionen con el abandono de los estudiantes, ya que son precisamente determinantes de reprobación de asignaturas y consecuentemente de una baja académica. Lo mismo aparece conceptualizado en la bibliografía (Tinto, 1975) como abandono involuntario y se excluyó precisamente por no aportar información novedosa para la predicción del abandono. Por los mismos motivos es que otro caso representativo de colinealidad corresponde a la variable *estado*, en la que se precisan, entre otras, las condiciones de “alumno regular” o “eliminado académicamente”, las que naturalmente correlacionan con *permanencia*. Basándose en la medida de silueta de cohesión y separación que establece el nivel de bondad del modelo, se concluyeron los ensayos de modo que las variables que finalmente son consideradas para este estudio son; *permanencia*, *reingreso*, *ptje. ranking*, *NEM*, *avance*, *puntaje*, *matemáticas*, *preferencia* y *lenguaje*. Estas variables fueron las que finalmente produjeron un modelo de tres conglomerados con medida de silueta de cohesión y separación regular, lo que se considera bueno para iniciar el análisis. La cantidad de conglomerados indica que no hay colinealidad, dado que se elimina la tendencia dicotómica en la generación de clústeres. También el mayor número de reglas que se generan en comparación con los modelos previos sirve para entregar evidencia en este punto.

Con la segmentación propuesta en la base de datos se procedió al análisis, considerando los atributos resultantes del proceso de selección de variables de entrada al modelo. Para aplicar dos diferentes modelos de análisis de datos y evaluar el comportamiento de cada uno por separado y en conjunto, se escogieron dos estrategias que corresponden a clúster k-medias y árbol de decisiones. El análisis de conglomerados de k-medias es un análisis de clúster no jerárquico, que corresponde a una división de la población en K grupos fijados o *a priori*. Con este tipo de análisis descriptivo de los datos se puede argumentar qué variables influyen en las diferencias entre los grupos. El primer paso fue estandarizar las variables, buscando que todas ellas tuvieran roles equitativos en el análisis; para este propósito se utilizaron puntajes Z. Con el fin de determinar el número de grupos, se formaron inicialmente dos conglomerados y se identificó

la pertenencia de cada elemento a esos dos grupos; luego se repitió el proceso, formando 3, 4 y 5 y así hasta 9 ($k = 1, 2, \dots, 9$). Finalmente, el número de grupos escogidos para el análisis fue de dos, de modo que el agrupamiento por medio de k-medias entregó una distribución distinta a la obtenida con el análisis bietápico y que se muestra en la tabla 6.

Tabla 6
Número de casos

Válidos	Perdidos	Clúster 1	Clúster 2
228	22	114	114

Considera cantidad de casos válidos y casos perdidos, y de casos para cada clúster.

La tabla 7 muestra los valores finales de los centros de los conglomerados. Estos valores son las medias de cada variable en cada conglomerado final, y los centros de los conglomerados finales reflejan los atributos del caso prototipo para cada conglomerado.

Tabla 7
Centros de clústeres finales

Variable con puntuación Z	Clúster 1	Clúster 2
Porcentaje de avance del estudiante	,16304	-,12152
El estudiante reingresa a la universidad	,14257	-,17008
Si el estudiante permanece o abandona su carrera inicial	-,31215	,04710
Puntaje NEM asignado al promedio de notas	,72065	-,70720
Puntaje en la prueba de lenguaje	,13782	-,07649
Puntaje en la prueba de matemáticas	,41311	-,34867
Preferencia de postulación a la carrera	-,05147	,02087
Puntaje ponderado de ingreso	,74682	-,71773
Ranking	,67525	-,71587

Las distancias entre los centros de los clústeres finales es de 2,657.

En relación con la tabla 7, en un primer análisis tentativo se puede hipotetizar que serán significativas aquellas variables en las cuales las distancias entre un conglomerado y otro son más grandes con relación a su distancia con los centros de clúster final. De este modo, serían poco significativas variables como *lenguaje* y

preferencia. Se consideró necesario hacer más verídica la correlación y, por tanto, se utilizó el método análisis de varianza (ANOVA). En la tabla 8 se puede observar que aquellas variables que tienen una mayor influencia en el proceso de agrupamiento son las variables *NEM* (252,608), *matemáticas* (38,356), *puntaje* (263,274) y *ranking* (241,778). Asimismo, los niveles de significación también indican qué variables aportan más y menos en el agrupamiento; variables no significativas son *avance* (0,030), *lenguaje* (0,99) y *preferencia* (0,582).

Tabla 8
ANOVA

Variable con puntuación Z	Estadísticos F	Significancia
Porcentaje de avance del estudiante	4,798	,030
El estudiante reingresa a la universidad	5,183	,024
Si el estudiante permanece o abandona su carrera inicial	9,098	,003
Puntaje NEM asignado al promedio de notas	252,608	,000
Puntaje en la prueba de lenguaje	2,746	,099
Puntaje en la prueba de matemáticas	38,356	,000
Preferencia de postulación a la carrera	,304	,582
Puntaje ponderado de ingreso	263,274	,000
Ranking	241,778	,000

Análisis de varianza (ANOVA).

Respecto de la tabla ANOVA, se puede concluir que los factores importantes que discriminan entre los dos grupos son: *NEM*, *matemáticas*, *puntaje* y *ranking*. Resulta novedosa la inclusión de la variable *matemáticas*, puesto que esta no aparece dentro de la bibliografía revisada. Se realiza una interpretación de estos resultados en el capítulo de conclusiones y discusión.

Al analizar el clúster se observa la formación de dos grupos que se muestran en la tabla 9. El grupo 1 corresponde a los estudiantes que en su mayoría abandonan, con promedios bajos en las variables que resultaron significativas para el agrupamiento: *NEM*, *matemáticas*, *puntaje* y *ranking*. El grupo 2 corresponde a un grupo de estudiantes que se caracteriza por tener promedios ponderados altos y medios superiores. Se asume, con base en el conocimiento teórico, que *reingreso* es una variable no significativa para el análisis predictivo del riesgo de abandono, ya que es una condición que se produce

con posterioridad a la deserción inicial de los estudiantes. Tinto (1975, 1989) distingue entre *drop out* y *stop out*, y allí el reingreso correspondería al segundo caso. Para este análisis no se repara en esta distinción, no solo por no aportar en la predicción del riesgo de deserción inicial, sino que además por falta de datos, pues este concepto toma en consideración el reingreso al sistema universitario en general y para esta muestra se contó solo con los datos de reingreso a la universidad del caso estudiado en particular.

Tabla 9

Diferencias de promedios en variables significativas para el agrupamiento

Variable	Grupo 1	Grupo 2
NEM	630,49	638,52
Matemáticas	549,66	567,35
Puntaje	598,34	611,23
Ranking	679,47	682,85

Variables seleccionadas según técnica de análisis k-medias.

Posteriormente, se procedió a una inducción de reglas para realizar predicciones a través de un modelo de árbol de decisiones. Con este paso se buscó determinar qué variables tienen mayor capacidad predictiva para el riesgo de deserción. Una clasificación basada en árboles de decisión (CBAD) es un diagrama de flujo con una estructura de árbol, que produce predicciones basándose en combinaciones de valores en los campos de entrada. El algoritmo de árbol de decisiones genera un modelo de análisis de datos mediante la creación de una serie de divisiones en el árbol. Éstas divisiones se representan como nodos, las ramas representan los arcos de conexión entre nodos y las hojas son nodos terminales en los que se refleja la decisión final. El algoritmo agrega un nodo al modelo cada vez que una columna de entrada tiene una correlación significativa con la columna de predicción. La forma en que el algoritmo determina una división varía en función de si predice una columna continua o una columna discreta. El algoritmo de CBAD utiliza la selección de características para guiar la elección de los atributos más útiles. Un árbol representa una segmentación de los datos que se crea mediante la aplicación de una serie de reglas simples (Marquez-Vera, Cano, Romero, Noaman, Mousa Fardoun & Ventura). Cada regla asigna una

observación a un segmento basada en el valor de una entrada, dando como resultado una jerarquía de segmentos dentro de segmentos. Uno de los objetivos de la técnica es obtener un árbol lo más pequeño posible, pues un árbol pequeño consigue explicar mejor los datos y además conduce a menos errores para predecir el resultado de futuros registros, en comparación con un árbol más complejo. Para seleccionar en cada paso el nuevo atributo para discriminar, se sigue como heurística la elección de aquel que produce los subconjuntos de datos más homogéneos (Marquez et al., 2018).

Existen diversos métodos para la inducción de árboles de decisión, cada uno de ellos ofrece diferentes capacidades. Se escogió el método de detección automática de interacción de chi-cuadrado (CHAID), entre otras razones por la naturaleza categórica de la variable dependiente. El método consiste en un rápido algoritmo de árbol estadístico y multidireccional que explora datos de forma rápida y eficaz, creando segmentos y perfiles respecto del resultado deseado y permitiendo la detección automática de interacciones mediante chi-cuadrado. En cada paso, CHAID elige la variable independiente que presenta la interacción más fuerte con la variable dependiente y calcula la categoría de destino predicha para cada nodo del árbol. Si la prueba de chi-cuadrado es significativa, se crea una nueva partición estadísticamente diferente del nodo de origen. Las pruebas de chi-cuadrado también se utilizan para fusionar categorías en nodos únicos, y si los recuentos para categorías no son significativamente diferentes de la variable dependiente se fusionan en un mismo nodo. Esto permitió restringir aún más la posibilidad de colinealidad entre las variables. El nivel de significación por defecto para la división de nodos y la fusión de categorías es igual a 0,05.

Una vez ingresadas las nueve variables seleccionadas para el análisis, el algoritmo de árbol de decisión realizó una poda de variables, de modo que el modelo final para esta técnica incluye tan solo las variables independientes *matemáticas* y *avance*. En concreto, la aplicación de la CBAD indica que son estas dos variables las que explican la deserción estudiantil. No afectan la predicción de deserción estudiantil en la muestra las demás variables que el modelo dejó excluidas.

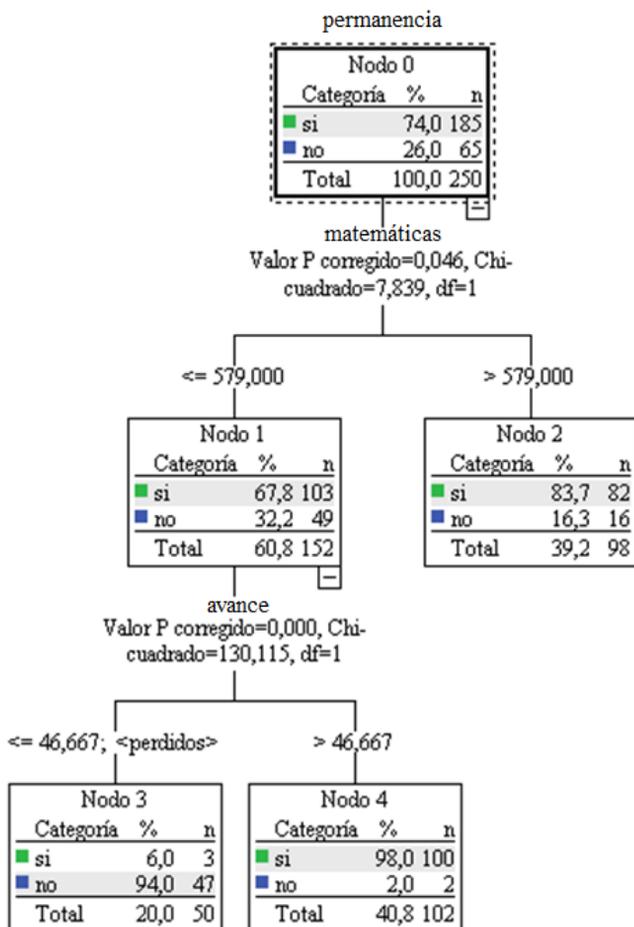


Figura 1
Resultado del análisis de árbol de soluciones

Establece puntajes de corte en las variables *matemáticas* y *avance* para determinar si los estudiantes permanecen o abandonan.

Para probar la calidad del modelo se utilizó el procedimiento de Validación por División Muestral y se dividieron aleatoriamente los datos en dos submuestras. La primera con el 80% de los registros, para determinar la CBAD; la segunda con el 20% restante de registros, para probar la capacidad de predicción del modelo. Con base en la matriz de confusión presentada, la *ratio* de precisión de la predicción alcanzó un 91,6%. La matriz de confusión contiene información acerca de las predicciones realizadas por un método o sistema de

clasificación, comparando para el conjunto de individuos la tabla de aprendizaje o de prueba con la predicción dada, versus la clase a la que estos realmente pertenecen.

La tabla 10 muestra que 182 predicciones de permanencia fueron realizadas correctamente para el 80%, mientras que 3 no. Para el 20%, 47 predicciones de abandono fueron realizadas correctamente, mientras que 18 no. En general, 229 de 250 predicciones fueron correctas para un 91,6% de efectividad en las predicciones; o, lo que es lo mismo, la precisión P, que corresponde a la proporción del número de predicciones que son correctas respecto del total. La precisión positiva, en tanto, es la proporción de casos positivos que fueron identificados correctamente. En el modelo la precisión positiva alcanza un 98,4% para la permanencia; lo mismo puede ser llamado “sensibilidad”. La precisión negativa es la proporción de casos negativos que fueron identificados correctamente. En el modelo esta predicción alcanza un valor de 72,3% para el abandono; lo mismo puede ser llamado “exhaustividad”.

Tabla 10
Matriz de confusión

Observado	Predicción Sí	Predicción No	Predicción correcta
El estudiante permanece	182	3	98,4%
El estudiante abandona	18	47	72,3%
Porcentaje global	80,0%	20,0%	91,6%

Tres son los falsos negativos y 18 los falsos positivos, con 182 predicciones correctas para permanencia y 47 para abandono.

Para medir el nivel de la predicción realizada por una CBAD se usa una tabla de contingencia, que describe los casos acertados y los casos erróneos de la predicción. A esta tabla de contingencia se le llama “matriz de confusión” y en cada columna de la matriz se muestra el número de predicciones de cada clase, mientras que en cada fila se muestran las instancias de la clase real.

La tabla 11 muestra la estimación del riesgo de predicción de la CBAD, o tasa de error, que indica un 8,4% de posibilidades de errar la predicción. Este valor es complementario con el porcentaje global de predicción de 91,6%. Es posible obtener aún mayor precisión para el

error de predicción y, en general, esta estimación se hace utilizando un procedimiento de validación cruzada. En este estudio se utilizó $n=10$ particiones, que es el valor que SPSS propone por defecto, y se obtuvo un error ajustado final del 19,2%.

Tabla 11
Estimación del riesgo

Método	Estimación	Desviación error
Resustitución	,084	,018
Validación cruzada	,192	,025

La validación cruzada ajusta el error de predicción en un 19,2% de error final.

El resultado de la evaluación presenta al modelo como estable y aceptable. Los resultados del análisis del árbol de decisión encontraron que la profundidad de la estructura del árbol es de tres niveles de profundidad, con cinco nodos totales y tres nodos terminales. El primer y segundo nodo se formaron con la variable *matemáticas*. El tercer y cuarto nodo con la variable *avance*. Para la variable *matemáticas* se determinó una significancia de 0,046 con el puntaje de corte 579,000, es decir, el modelo de árbol establece que todos aquellos estudiantes que obtuvieron puntaje inferior a 579 puntos presentan un mayor riesgo de abandono, en comparación con aquellos que obtuvieron puntajes iguales o superiores al corte. Bajo el puntaje de 579 se produce un nodo terminal, es decir, ya no se producen más ramas. Por sobre el puntaje 579 o igual, en cambio, aparece otro nodo con la variable *avance*, es decir, que la variable *avance*, además, se puede determinar cómo significativa para la predicción del abandono, lo que, según la bibliografía revisada y el conocimiento sobre los datos, aporta consistencia al modelo. Para esta última variable se establece 46,667 como porcentaje de corte, es decir el modelo de árbol establece que todos aquellos estudiantes que obtuvieron porcentaje de avance inferior a 46,6% presentan un mayor riesgo de abandono en comparación con aquellos que obtuvieron puntajes iguales o superiores al corte. Finalmente, se considera que la incorporación de la variable *matemáticas* aporta nueva información, lo que además resulta consistente con los resultados encontrados mediante agrupamiento de k-medias.

Por último, se realizó una prueba de entrenamiento teniendo en cuenta las 27 variables iniciales, y se observó que el modelo no cambia sustancialmente en relación con los resultados esperados según la bibliografía revisada y el conocimiento de los datos: éste ingresa como variables independientes significativas: *avance*, *estado*, *PPN1*, *PPN2*, *PPN3* y *PPN4*. Solo cabe considerar que, para las variables *avance* y *estado*, los niveles de significancia son superiores con $P=0,0$. Esto se explica por sí solo desde el conocimiento de los datos, ya que es natural que “abandono” y “permanencia” estén fuertemente vinculados a estas variables, que no son sino otra forma de referirse a la variable *respuesta*.

Conclusiones y discusión

Los resultados son relevantes e invitan a la toma de decisiones basada en evidencia. Los perfiles de deserción estudiantil obtenidos mediante las técnicas utilizadas indican que éstas son capaces de generar modelos consistentes con la realidad observada y con el respaldo teórico. Lo que se buscó fue encontrar propiedades emergentes dentro de los sistema de bases de datos disponibles, que permitieran predecir el abandono más allá de los factores ampliamente revisados en la bibliografía o conocidos por la experiencia, como son el avance curricular o los promedios de notas, los que correlacionan significativamente con permanencia y deserción, pero que, sin embargo, se producen con posterioridad al ingreso de los estudiantes a la educación superior y una vez que ya han enfrentado sus primeras evaluaciones y experiencias académicas. Esto impide generar un SAT que permita focalizar inicialmente a los estudiantes con mayor riesgo de deserción durante el momento más próximo al ingreso de éstos a la universidad, el periodo más crítico, como lo señalan diversas investigaciones (Biggio et al., 2015; Baeza et al., 2016; Donoso-Díaz et al., 2018). Los resultados observados en la muestra estudiada y obtenidos mediante las técnicas de clúster k-medias y árbol de decisiones, sugieren que esta respuesta podría estar en algunas variables de ingreso, tales como *NEM*, *puntaje*, *ranking* y *matemáticas*. Se destaca la aparición de la variable *matemáticas*, la que no aparecía en la bibliografía revisada.

Futuros análisis deberían determinar si estos resultados son válidos para el amplio espectro de la población estudiantil. De verificarse, se sugiere considerar estos resultados para determinar un porcentaje de corte que permita focalizar las distintas intervenciones y esfuerzos que las IES dirijan hacia aquellos estudiantes que, por razón de esta variable, presenten un mayor riesgo de deserción temprana o inicial. También queda pendiente estudiar qué habilidades subyacen a los resultados en la prueba de matemáticas y que hacen que éstos se relacionen con la retención: ¿será la destreza en matemáticas en sí misma la que mejora la retención, o será que la destreza en matemáticas da cuenta de ciertas características del estudiante que aumentan su probabilidad de retención? Responder esto resulta fundamental, puesto que de las distintas respuestas que surjan se puede derivar diferentes consecuencias en la política educativa para aquellos programas que buscan asistir el proceso de transición entre la educación media y la superior, entre ellos la BNA. De considerarse el primer caso, bastaría con enfatizar en la intervención sobre habilidades matemáticas; sin embargo, aunque esto puede resultar evidente para las carreras más vinculadas al área, lo mismo no queda tan claro para las demás carreras, lo que nos lleva a preguntar qué características del estudiante estaría evaluando la prueba. De considerarse el segundo caso, habría que preguntarse qué dominios o habilidades cognitivas u oréticas (no cognitivas) son las que podría estar evaluando la prueba de matemáticas. Una respuesta podría estar en las denominadas “habilidades de preparatividad académica” (Baeza, Rehbein y Antivilo, 2016), asumiendo que éstas son las que generan brecha entre la formación de la educación media y la primera certificación de educación superior, tal como aparece declarado en el Marco Nacional de Cualificaciones (CNED, 2017). El concepto de “preparatividad académica” fue propuesto para definir al conjunto de condiciones de un estudiante para ingresar y tener éxito en una institución de educación superior, lo que implica que sea capaz de aprobar los cursos básicos de manera que, sucesivamente, pueda continuar aprobando los cursos superiores que presentan mayor dificultad (Baeza et al., 2016). Esto, en todo caso, exige otra investigación en la que se pueda evaluar el nivel de pertinencia de la PSU de matemáticas en relación con las habilidades de preparatividad académica, o, en su defecto, el nivel de pertinencia de estas habilidades con la nueva

“Prueba de Transición” que se aplicará durante 2020 y 2021, y, posteriormente, con la nueva prueba de acceso que comenzará a aplicarse desde 2022. Por lo pronto se asume que, de ser verídica esta relación, permitiría identificar las principales debilidades para abordar con los estudiantes en tránsito a la educación superior, de manera que las IES les apoyen eficiente y oportunamente, para que así puedan mejorar sus índices de retención.

En cualquier caso, los resultados obtenidos abren oportunidades para la creación de nuevos modelos usando otras técnicas de clasificación y predicción, u otras más complejas, como pueden ser redes neuronales y regresión logística, las que permitirían un análisis comparativo de los factores que influyen en la deserción estudiantil. Si consideramos la complejidad que puede tener la generación de un SAT, se sugiere una revisión permanente de la influencia de las distintas variables de ingreso en la capacidad predictiva de la deserción que permanezca acorde a la realidad de cada IES, así como también una permanente generación de nuevas variables que arrojen información adicional de la que se contaba para este estudio, como podrían ser los instrumentos que utilizan las IES para caracterizar a sus ingresados a primer año. El desafío pendiente sigue siendo la generación de un SAT propio para cada IES, que entregue especial importancia al desarrollo de planes de acción preventivos y no solamente planes asistencialistas de naturaleza correctiva (Castillo et al., 2019). Si se considera que la bibliografía confirma que las IES con mejores tasas de retención son las que conocen los intereses y características de sus estudiantes, entonces resulta indispensable que la institución genere una exhaustiva caracterización de sus estudiantes según áreas de interés e identifique los recursos de apoyo disponibles (Donoso-Díaz et al., 2018) y los factores motivacionales relacionados con su compromiso con metas académicas y de logro. Por un lado, el seguimiento sistemático de éstos permite orientar la gestión institucional a las necesidades de su propia población de estudiantes y, por otro, generar acciones de intervención que provoquen cambios en los factores de riesgo para así disminuir la deserción estudiantil (Díaz, 2008). Mecanismos de esta naturaleza permiten apoyar a los estudiantes y utilizar efectivamente la información disponible en los registros institucionales (Soto et al., 2017).

Referencias

- Altbach, P. (2016). *Global Perspectives on Higher Education*. Baltimore: The Johns Hopkins University Press.
- Arriaga, J., Burillo, V., Carpeño, A. y Casaravilla, A. (2011). Caracterización de los tipos de abandono. Dividamos el problema y venceremos fácilmente. *I Conferencia Latinoamericana sobre el Abandono en la Enseñanza Superior*, Nicaragua.
- Baeza, M., Rehbein, L. y Antivilo, A. (2016). Desarrollo y validación preliminar de una escala de preparatividad académica para la educación superior. *Formación Universitaria*, 9(4), 63-74, DOI: 10.4067/S0718-50062016000400008
- Basso, P. (2016). Educación Superior en Chile: El fracaso del modelo neoliberal. *Revista Interdisciplinaria de Filosofía y Psicología*, Arica, 11(37), 21-48.
- Bernardo, A., Cervero, A., Esteban, M., Fernández, A., Solano, P. y Agulló, E. (2018). Variables relacionadas con la intención de abandono universitario en el período de transición. *Revista d'Innovació Docent Universitària*, (10), 122-130.
- Biggio, M., Vásquez, S. y García, S. (2015) Deserción en estudiantes de nuevo ingreso a carreras de diseño. El caso de la universidad de Buenos Aires, Argentina. *Revista Electrónica Actualidades Investigativas en Educación*, 15(1), 1-23.
- Blok, L. (2016). *Inequidad en el acceso a la educación superior. Análisis del Programa Propedéutico en la Universidad de Santiago de Chile*. Tesis de Maestría en Estudios Latinoamericanos, Universidad de Leiden.
- Castaño, E., Gallón, S., Gómez, K. y Vásquez, J. (2004). Deserción estudiantil universitaria: Una aplicación de modelos de duración. *Lecturas de Economía*, (60), 39-65.
- Castillo, P. Morales, T. y Miranda, C. (2019). Evaluación de un programa de apoyo psico-social en torno a los conceptos de persistencia y retención universitaria. *Revista Brasileira de Educação*, 24, e240058. DOI: <https://doi.org/10.1590/s1413-24782019240058>
- CNED. (2017). *Marco Nacional de Cualificaciones para la Educación Superior*. Consejo Nacional de Educación, Chile.
- Del Valle, R. (2017). Acceso inclusivo en Chile y resultados académicos a siete años de la instalación del propedéutico UC TEMUCO. *Congresos CLABES*. Recuperado de <https://revistas.utp.ac.pa/index.php/clabes/article/view/1660>

- Departamento de Evaluación, Medición y Registro Educacional. (16 de diciembre de 2020). Santiago, Chile: DEMRE. Recuperado de <https://demre.cl/>
- Díaz, C. (2008). Modelo conceptual para la deserción estudiantil universitaria chilena. *Estudios Pedagógicos*, 34(2), 65-86
- Donoso-Díaz, S., Neira, T. y Donoso, G. (2018). Sistemas de Alerta Temprana para estudiantes en riesgo de abandono de la Educación Superior. *Ensaio: Avaliação e Políticas Públicas em Educação*, 26(100), 944-967. DOI: <https://dx.doi.org/10.1590/s0104-40362018002601494>
- Donoso, S. y Schiefelbein, E. (2007). Análisis de los modelos explicativos de retención de estudiantes en la universidad: una visión desde la desigualdad social. *Estudios Pedagógicos*, 33(1), 7-17.
- Donoso, S., Donoso, G. y Arias, O. (2011). Iniciativas de Retención y Mejora de Desempeño de Estudiantes de Educación Superior. *Calidad en La Educación*, (33), 16-61.
- Espinoza, O. y González, L. (2015). Equidad en el sistema de educación superior de Chile: acceso, permanencia, desempeño y resultados. En A. Bernasconi (Ed.), *La educación superior de Chile. Transformación, desarrollo y crisis* (pp. 517-579). Santiago de Chile: Ediciones UC.
- Espinoza, O., González, L. y Latorre, C. (2009). Un modelo de equidad para la educación superior: análisis de su aplicación al caso chileno. *Revista de la Educación Superior*, México, D.F., 38(150), 97-112.
- Figuroa, L., Maillard, B., Véliz, N., Toledo, S., Andaur, C., Quezada, S. y González, M. (2016). Programas propedéuticos: Equidad, Calidad e Inclusión a la Vida Universitaria. *Congresos CLABES*. Recuperado de <https://revistas.utp.ac.pa/index.php/clabes/article/view/996>
- Gairín-Sallán, J. (2015). *Los sistemas de acceso, normativa de permanencia y estrategias de tutoría y retención de estudiantes de Educación Superior*. Madrid: Wolters Kluwer.
- Gil, F., Paredes, R. y Sánchez, I. (2013). *El ranking de las notas: inclusión con excelencia*. Santiago de Chile: Centro de Políticas Públicas UC, Temas de la Agenda Pública, año 8, n° 60.
- Larroucau, T., Ríos, I. y Mizala, A. (2015). Efecto de la incorporación del ranking de notas en el proceso de admisión a las universidades chilenas. *Pensamiento Educativo*, 52(2), 95-118.
- Marquez-Vera, C., Cano, A., Romero, C., Noaman, A. Y. M., Mousa Fardoun, H & Ventura, S. (2016). Early Dropout Prediction using Data Mining:

- A Case Study with High School Students. *ExpertSystems*, 33, 107-124. DOI: <https://doi.org/10.1111/exsy.12135>
- Marshall, G., San Martín, J., Gil, F., Donoso, S., Roberto, S. y Grau, N. (2017). *Estudio acerca de la Validez Predictiva del Ranking de Notas*. Santiago de Chile: Consejo de Rectores de las Universidades Chilenas (CRUCH).
- Méndez-Ramírez, I., Moreno-Macías, H., Méndez Gómez-Humarán, I. y Murata Ch. (2014). Conglomerados como solución alternativa al problema de la multicolinealidad en modelos lineales. *Ciencias Clínicas*, 15(2), 39-46. DOI: <https://doi.org/10.1016/j.cc.2015.08.002>
- Ministerio de Desarrollo Social, Chile. (2016). *Informe de Desarrollo Social 2016. Encuesta de Caracterización Socioeconómica Nacional – CASEN*. Santiago de Chile.
- Ministerio de Desarrollo Social. (12 de agosto de 2019). *Banco Integrado de Programas Sociales y No Sociales*. BIPS. Santiago, Chile. Recuperado de <https://programassociales.ministeriodesarrollosocial.gob.cl>
- Ministerio de Educación. (2011). *Fortalecimiento Institucional*. Recuperado de http://dfi.mineduc.cl/index2.php?id_portal=59&id_seccion=4207&id_contenido=19911
- Ministerio de Educación. (2012). *Deserción en la educación superior en Chile*. Serie Evidencias. Santiago de Santiago de Chile: MINEDUC.
- OECD. (2009). *Reviews of National Policies for Education Tertiary Education in Chile*. Paris: OECD Publishing. DOI: <https://doi.org/10.1787/9789264054189-es>
- OECD. (2016). *Education at a Glance 2016: OECD Indicators*. Paris: OECD Publishing. DOI: <http://dx.doi.org/10.187/eag-2016-en>
- OECD. (2018). *Education at a Glance 2018: OECD Indicators*. Paris: OECD Publishing. DOI: <http://dx.doi.org/10.1787/eag-2018-en>
- Orellana, M., Moreno, K. y Gil, F. (2015). *Inclusión a la universidad de estudiantes meritorios en situación de vulnerabilidad social*. Santiago de Chile: UNESCO.
- Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. (2017). *Guía para asegurar la inclusión y la equidad en la educación*. Título original: A Guide for ensuring inclusion and equity in education. París: UNESCO.
- Saldana, M. y Barriga, O. (2010). Adaptación del modelo de deserción universitaria de Tinto a la Universidad Católica de la Santísima Concepción, Chile. *Revista de Ciencias Sociales*, (16), 616-628.

- Sánchez, T., Silva, M. y Jouannete, C. (2011). Equidad en el acceso, permanencia, logros y resultados de la educación superior chilena. En M. Jiménez (Ed.), *Propuestas para la Educación Superior Chilena, Foro Aequalis y las Transformaciones Necesarias* (pp. 21-88). Santiago de Chile: Aequalis.
- Servicio de Información de Educación Superior. (2015). *Principales resultados evolución retención de 1er año 2010-2014*. Santiago de Chile: SIES.
- Servicio de Información de Educación Superior (2016). *Informe retención de primer año de pregrado cohortes 2011-2015*. Santiago de Chile: SIES.
- Servicio de Información de Educación Superior. (2017). *Informe retención de primer año de pregrado cohortes 2011-2016*. Santiago de Chile: SIES.
- Soto, V., Díaz, C. y Chiang, M. (2017). Beca de nivelación académica: resultados de un programa de intervención en estudiantes de origen vulnerable en la universidad de concepción. *Educação em Revista*, (33), e160071. DOI: <https://dx.doi.org/10.1590/0102-4698160071>
- Tinto, V. (1975). Dropout from Higher Education: A Theoretical Synthesis of Recent Research. *Journal of Higher Education*, (45), 89-125.
- Tinto, V. (1989). Definir la deserción: una cuestión de perspectiva. *Revista de Educación Superior ANUIES*, México, (71).
- Universidad de Valparaíso. (2017). *Anuario*.
- Zapata, G. y Tejada, I. (2016). *Educación Superior en Chile. Informe Nacional 2011-2016*. Santiago de Chile: Centro Interuniversitario de Desarrollo-UNIVERSIA.

Recibido: 31/03/2020

Aceptado: 16/12/2020