

UNIVERSIDAD DE SANTIAGO DE CHILE
FACULTAD DE INGENIERÍA
Departamento de Ingeniería Informática



**MODELO DE PREDICCIÓN DEL COMPORTAMIENTO DE
BÚSQUEDA DE INFORMACIÓN EN LÍNEA EN ESTUDIANTES
DE EDUCACIÓN PRIMARIA**

Propuesta de Tesis

Nombre: Gonzalo Javier Martinez Ramirez

R.U.T.: 18.045.598-1

Año Ingreso: 2010

Teléfono: (+56) 9 96112973

E-mail: gonzalo.martinez@usach.cl

Profesor Patrocinador: Roberto Ignacio González Ibañez

Santiago – Chile

Junio 2017

© **Gonzalo Javier Martinez Ramirez y Roberto Ignacio González-Ibañez, 2017**

- Todos los derechos reservados. Queda prohibida la reproducción total o parcial sin la autorización previa y por escrito.

RESUMEN

Durante la última década, debido a los rápidos avances de las tecnologías de la información y comunicación ha aumentado la cantidad de recursos digitales en Internet, la diversidad de fuentes de información, y además, se ha facilitado el acceso a estos. Asimismo, las búsquedas *web* han pasado a ser parte de las tareas comunes que realizan los estudiantes de los planteles educativos. Considerando la diversidad de fuentes de información y tipos de recursos en línea, resulta necesario desarrollar competencias informacionales durante el proceso de formación en los distintos niveles educativos (primaria, secundaria y universitaria).

Palabras Claves:

TABLA DE CONTENIDOS

Capítulo 1. Objetivos del Proyecto	1
1.1 Objetivo general	1
1.2 Objetivos específicos	1
Capítulo 2. Descripción del Problema	2
2.1 Motivación	2
2.2 Revisión de la literatura	3
2.3 Definición del problema	6
Capítulo 3. Descripción de la Solución Propuesta	7
3.1 Características de la solución	7
3.2 Propósito de la solución	8
3.3 Alcances y limitaciones de la solución	8
Capítulo 4. Metodología, Herramientas, y Ambiente de Desarrollo	9
4.1 Metodología a usar	9
4.1.1 Metodología usada en la investigación	9
4.1.2 Metodología usada para el desarrollo de <i>software</i>	10
4.2 Herramientas de desarrollo	10
4.3 Ambiente de trabajo	10
Capítulo 5. Plan de Trabajo	12
Capítulo . Bibliografía	13
Apéndice A. Capítulo Apéndice	17
Apéndice B. Another Appendix Chapter	18

ÍNDICE DE TABLAS

ÍNDICE DE FIGURAS

4.1. A cow licking its nose. Usage with permission of the photographer Nicole Barth . . .	11
---	----

CAPÍTULO 1. OBJETIVOS DEL PROYECTO

1.1. OBJETIVO GENERAL

Diseñar y evaluar un modelo de predicción en tiempo real del resultado del comportamiento de búsqueda de información en línea de estudiantes de enseñanza primaria.

1.2. OBJETIVOS ESPECÍFICOS

1. Realizar una revisión bibliográfica sobre trabajos recientes relacionados con minería de datos en el contexto educacional.
2. Realizar una exploración, limpieza, pre-procesamiento y transformación de los datos recopilados por la plataforma NEURONE.
3. Definir las características de comportamiento de búsqueda de los estudiantes para la construcción de modelos predicción.
4. Comparar, seleccionar e implementar algoritmos de minería de datos, para la construcción de modelos de predicción.
5. Construcción del modelo de predicción del comportamiento de búsqueda en línea de estudiantes de primaria.
6. Implementar y evaluar una plataforma de aprendizaje de máquina, que en base a la información de NEURONE pueda predecir los resultados del comportamiento de búsqueda de un estudiante.
7. Construir e implementar en la plataforma modelos de clasificación y predicción a partir de los atributos derivados del comportamiento de estudiantes de educación primaria (básica), en el contexto de búsqueda de información en línea.

CAPÍTULO 2. DESCRIPCIÓN DEL PROBLEMA

2.1. MOTIVACIÓN

Durante la última década, debido a los rápidos avances de las tecnologías de la información y comunicación (TICs, desde ahora en adelante) ha aumentado la cantidad de recursos digitales en Internet, la diversidad de fuentes de información, y, además, se ha facilitado el acceso a estos. Asimismo, las búsquedas web han pasado a ser parte de las tareas comunes que realizan los estudiantes de los planteles educativos. En consecuencia, se ha disminuido las visitas a bibliotecas, y el uso de fuentes revisadas y editadas.

La Alfabetización en Información es una disciplina que se define a sí misma en base al desarrollo de destrezas, habilidades y competencias informacionales que permitan ir fortaleciendo el aprendizaje constante y el trabajo colaborativo [1]. Además, favorece la capacidad de buscar, clasificar, y comprender la información, para posteriormente convertirla en conocimiento asimilado y útil. A causa de esto, el estudio, análisis y modelado de las conductas de los estudiantes en ambientes de búsqueda web es esencial para comprender sus niveles de alfabetización informacional [2].

Considerando la diversidad de fuentes de información y tipos de recursos en línea, resulta necesario desarrollar competencias informacionales durante el proceso de formación en los distintos niveles educativos (primaria, secundaria y universitaria). La enseñanza de la alfabetización informacional, se imparte principalmente por bibliotecas universitarias, y en menor medida en la etapa escolar obligatoria [3]. Actualmente, ciertas universidades consideran las competencias informacionales como un requisito de entrada para iniciar estudios universitarios [4].

Actualmente, en Chile la enseñanza de competencias informacionales es enseñada en bibliotecas universitarias y cursos introductorios de mallas universitarias [5]. De acuerdo con [6] los estudiantes universitarios de Chile presentan problemas con las competencias informacionales, ya que no aplican la búsqueda de información de forma crítica. Una de las posibles causas de por qué los estudiantes tienen dificultades con estas competencias es el hecho de que, en los colegios, y en el inicio de su educación, no se prioriza la generación de conocimiento, sino la reiteración de información.

Las consecuencias de no considerar cuándo y por qué se necesita la información, dónde encontrarla, y cómo evaluarla, se ven reflejadas en la evaluación crítica de la información, y en el desempeño de los estudiantes [6]. A causa de esto, existe la necesidad de estudiar el fenómeno de la alfabetización informacional y las competencias de investigación en línea con los objetivos de i) conocer y estudiar los comportamientos de los estudiantes en tareas de búsqueda de información en medios digitales, y ii) obtener modelos para reforzar los niveles de alfabetización informacional.

Esta propuesta de tesis se enmarca dentro en el contexto del proyecto de investigación “*Enhancing Learning and Teaching Future Competences of Online Inquiry in Multiple Domains*” (iFuCo, desde ahora en adelante) [7], el cual pretende abordar la temática de la alfabetización informacional en estudiantes de enseñanza primaria con el objetivo de estudiar sus patrones de comportamiento y ofrecer modelos curriculares adecuados respecto al tema [8].

2.2. REVISIÓN DE LA LITERATURA

En esta sección, se presenta el estado del arte que da soporte a este trabajo, el cual comprende en primer el estudio del comportamiento de estudiantes. En segundo lugar, técnicas de minería de datos y plataformas de aprendizaje de máquina (*Machine Learning*, ML desde ahora en adelante) aplicada al contexto educacional. Finalmente, para una revisión extendida de la literatura, ver ANEXO A. ESTADO DEL ARTE.

La minería de datos (MD, desde ahora en adelante) pretende resolver problemas complejos cuya solución no se puede hallar con técnicas tradicionales como la estadística. Esto se logra al descubrir patrones y predecir tendencias por medio del análisis de datos generados a través de los diferentes sistemas operacionales y transaccionales de una institución y que están almacenados en sus bases de datos [9]. Cuando se aplica minería de datos en instituciones educativas la disciplina se conoce como minería de datos educacionales (MDE, desde ahora en adelante).

La MDE es una disciplina en evolución que usa tecnologías informáticas como son almacenes de datos y herramientas de inteligencia de negocios para descubrir tendencias y patrones sobre datos educacionales. El conocimiento que la MD genera apoya a las autoridades de centros de educación en la toma de decisiones oportunas, y a los profesores para analizar el comportamiento y aprendizaje de sus alumnos [10]. La disciplina se enfoca en el diseño de modelos para mejorar las experiencias del aprendizaje y la eficiencia organizacional [11]. El principal objetivo de la MDE es visto por diferentes investigadores como [10, 12, 13]: i) modelado del estudiante, modelado

del dominio, ii) sistema de aprendizaje, iii) construir modelos computacionales, y iv) estudiar los efectos de los recursos.

Actualmente, la aplicación de MDE, se radica en universidades, tales como Paul Smith's College, la cual utiliza sus datos históricos para mejorar las tasas de retención de alumnos [14]. La University of Georgia desarrolló un modelo para predecir la tasa de graduación y abandono estudiantil en un ambiente en línea [15]. Finalmente, la Purdue University han usado MD para determinar que la evaluación en etapas tempranas y de forma frecuente permite cambiar los hábitos de los estudiantes con calificaciones bajo la media en cursos introductorios. El equipo de investigación ha desarrollado un sistema de alerta académica temprana para saber el desempeño de los estudiantes [16].

Baker *et al.* [17] desarrolló un modelo de predicción usando datos recopilados automáticamente de interacciones entre estudiantes y el software como variables de predicción, y después validando la precisión del modelo al ser generalizado a más estudiantes y contextos. Entonces fueron capaces de estudiar sus avances en el conjunto completo de datos.

Koedinger, D'Mello, McLaughlin *et al.* [18] define que un buen modelo cognitivo de un estudiante debe ser capaz de predecir las diferencias en la dificultad de una tarea, y como el aprendizaje es transferido de tarea en tarea.

Las contribuciones de Romero y Ventura [19] son las relevantes en este campo hasta la fecha. Acercan la minería de datos al contexto educativo y describe los diferentes grupos de usuarios, tipos de entornos escolares y los datos que proporcionan. Luego, exponen las tareas más típicas en el ambiente escolar que pueden resueltas a través de técnicas de minería de datos.

Sarala y Krishnaiah [20] discute las aplicaciones de la minería de datos en instituciones educativas, para extraer la información útil de grandes conjuntos de datos (*datasets*), y proporciona herramientas analíticas para ver y utilizar esta información para tomar decisiones basadas en ejemplos de la vida real.

Dutt, Aghabozrgi, Ismail *et al.* [21] consolida las variantes de algoritmos de clustering aplicados al contexto de EDM. Además, simplifica el diseño los sistemas que aprenden de los datos, utilizando técnicas y algoritmos de minería de datos, tales como, clustering, clasificación y predicción.

Merceron y Yacef [12] establece cómo los algoritmos de minería de datos pueden escoger

información pedagógica importante. El conocimiento obtenido ayuda a mejorar el cómo administrar la clase, como el alumno aprende, y cómo proporcionar un *feedback* a los alumnos.

Akinola, Akinkunmi y Alo [22] aplica técnicas de minería de datos aplicados a estudiar el rendimiento de estudiantes de educación universitaria en cursos de programación. Los resultados demuestran que el conocimiento a priori de física y matemáticas influye de forma positiva en el rendimiento en la programación.

En este mismo contexto, Lahtinen, Ala-Mutka y Järvinen [23] estudia las dificultades de aprender programación, con el objetivo de crear material adecuado para introducir el curso a los estudiantes. De este estudio, se obtuvo las dificultades que sufren los estudiantes al momento de enfrentar tareas de programación.

Borkar y Rajeswari [24] evalúa el rendimiento de los estudiantes, donde selecciona algunos atributos mediante minería de datos, haciendo uso de una red neuronal multicapa perceptron y usando una validación cruzada selecciona las características más influyentes, estableciendo las reglas necesarias para poder detectar las características necesarias para poder predecir el rendimiento de los estudiantes. Jayakameswaraiah y Ramakrishna [25] aplica los mismos métodos propuestos por Borkar y Rajeswari [24].

Abdullah, Malibari y Alkhozai [26] realiza un sistema de predicción del rendimiento de los estudiantes basado en la actividad actual, y mediciones anteriores, clasificando cuales estudiantes rendirán bien, y los que no.

Oskouei y Askari [27] identifica los factores que afectan el rendimiento de los estudiantes en diferentes países, y aplica técnicas de clasificación y predicción para mejorar la precisión de las predicciones de los resultados de los estudiantes. Los resultados muestran que los factores de género, entorno familiar, nivel de educación de los padres, y el estilo de vida, afectan el rendimiento académico de los estudiantes, independiente del país.

En esta misma línea, Borkar y Rajeswari [28] sugiere un método de evaluación del rendimiento de los estudiantes, usando reglas asociativas de minería de datos, estimando el resultado de los estudiantes basado en la asistencia a sus cursos y su avance académico. Shazmeen, Baig y Pawar [29] evalúa el rendimiento de diferentes algoritmos de clasificación y análisis predictivo, basado en el trabajo de Borkar y Rajeswari [28] y propone técnicas de preprocesamiento de datos para lograr mejores resultados.

Tal como se muestra en los antecedentes anteriores, las investigaciones en EDM se realizan mayoritariamente en aprendizaje *online* y en casos puntuales en educación superior, por lo que es limitada la información respecto a educación primaria o secundaria, específicamente en la predicción de errores y fracaso escolar [30]. Para mayor información de trabajos relacionados con la EDM, consultar los siguientes *reviews* [31-33].

2.3. DEFINICIÓN DEL PROBLEMA

El proceso de búsqueda de información involucra diferentes procesos cognitivos, habilidades, variables de comportamiento y entorno de una persona. Por ejemplo, las creencias epistemológicas personales están vinculadas con los métodos de aprendizaje, además de influir en la toma de decisiones.

En el contexto de la enseñanza de la alfabetización en información, las evaluaciones de los cursos se centran principalmente en los resultados de los estudiantes, sin tomar en cuenta el proceso formativo y factores asociados que podrían influir directa o indirectamente sobre los resultados finales y el desempeño de los alumnos.

A partir de lo señalado anteriormente, surgen las siguientes interrogantes (*research questions*, RQ desde ahora en adelante):

- **RQ 1:** ¿De qué manera se puede estimar, durante el proceso de aprendizaje de competencias informacionales, la influencia de diversos factores en el desempeño de los estudiantes?
- **RQ 2:** ¿En qué medida es posible detectar situaciones anormales de conducta, y determinar las causas que llevan a un estudiante fallar durante el proceso de búsqueda de información?

CAPÍTULO 3. DESCRIPCIÓN DE LA SOLUCIÓN PROPUESTA

3.1. CARACTERÍSTICAS DE LA SOLUCIÓN

La solución consiste en una plataforma de aprendizaje de máquina y predicción del rendimiento de los estudiantes en tiempo real, en donde a través de la información obtenida en la plataforma NEURONE se crea un modelo de clasificación y predicción del rendimiento de los estudiantes de enseñanza básica en un curso de alfabetización informacional, específicamente en el tema de investigaciones en línea (*online inquiry*).

Los datos son recopilados y almacenados por NEURONE, estos datos provienen de registros del proceso de buscar información en línea en un sistema cerrado, los cuales son: historial de navegación, consultas realizadas, movimientos del *mouse*, escritura por teclado, número de *clicks* y tiempos de permanencia en páginas web. Además, se conoce con anticipación los documentos y párrafos ideales a seleccionar por parte de los estudiantes.

La información de los estudiantes y el resultado de las evaluaciones definen una buena o mala navegación. La plataforma y el modelo a construir deberá ser capaz de predecir en tiempo real el desempeño de los estudiantes a partir de su comportamiento de búsqueda de información actual.

La plataforma propuesta hará uso de Tensorflow, la cual se conectará con el sistema NEURONE, funcionando como una extensión del mismo, consultando la base de datos de NEURONE. La Figura 3.1, donde un académico hace una consulta a la aplicación, la cual llega a la plataforma propuesta, y el componente Spark Streaming, se encarga de consultar la base de datos de NEURONE y obtener los datos de navegación del usuario en particular, para luego crear y alimentar el modelo de predicción, el cual es almacenado en una base de datos MongoDB. Finalmente, se entrega la predicción del estudiante actual al académico que hizo la consulta.

3.2. PROPÓSITO DE LA SOLUCIÓN

El propósito de la solución consiste en proveer evaluaciones de rendimiento oportunas, que permitan a los docentes aplicar acciones correctivas durante el proceso de formación y desarrollo de competencias informacionales, en cursos de alfabetización informacional.

Con la plataforma propuesta en este trabajo, el académico obtiene una respuesta temprana del comportamiento del estudiante en el proceso de búsqueda de información. Tal como se ve en la Figura 3.2, el estudiante interactúa con el sistema educacional, en este caso NEURONE, y la plataforma propuesta a través de técnicas de minería de datos, informa al académico de los patrones y predicciones del comportamiento del estudiante, con el objetivo de ayudar en la toma de decisiones al académico correspondiente para diseñar y planificar de mejor forma la entrega de contenidos hacia el estudiante.

3.3. ALCANCES Y LIMITACIONES DE LA SOLUCIÓN

Los modelos se construyen a partir de un conjunto de datos específicos, los cuales tienen su propio contexto y origen que limitan los modelos a construir. A continuación, se describen las principales limitaciones y alcances de la solución.

1. El curso de alfabetización informacional y sus respectivos registros de datos, pertenecen al proyecto iFuCo [7], el cual es un trabajo colaborativo entre universidades de Finlandia (University of Tampere, University of Jyväskylä y University of Turku) y de Chile (Universidad de Santiago de Chile y Pontificia Universidad Católica de Chile).
2. Los registros de datos provienen de un estudio enmarcado en un curso de alfabetización en información, aplicado al área de Ciencia y Ciencias Sociales, en ambos países.
3. Los datos son recolectados y almacenados por un sistema externo llamado “NEURONE”, trabajo de memoria de un estudiante de la carrera de Ingeniería de Ejecución en Computación e Informática, de la Universidad de Santiago de Chile [34].
4. La solución funciona como un sistema predictor del comportamiento del resultado de búsqueda de estudiantes, y no ofrece características de un sistema de recomendación.

CAPÍTULO 4. METODOLOGÍA, HERRAMIENTAS, Y AMBIENTE DE DESARROLLO

4.1. METODOLOGÍA A USAR

El presente proyecto presenta una componente de investigación y desarrollo de *software* (I+D). Esto debido a la relación que existe entre ambas componentes, la investigación necesita una herramienta de software de apoyo que permita recibir los datos de NEURONE, alimentar el modelo de predicción y que permita al usuario interactuar con resultados de la predicción a realizar.

La componente de investigación del proyecto será guiada por la metodología Descubrimiento de Conocimiento en Base de Datos (*Knowledge Discovery in Databases*, KDD desde ahora en adelante), mientras que la componente de desarrollo de *software* será guiada por la metodología *Rapid Application Development* (RAD, desde ahora en adelante) [35]. A continuación, se explica el uso de ambas metodologías en el trabajo propuesto. Para mayor información sobre las metodologías a ocupar, ver ANEXO B. DESCRIPCIÓN Y USO DE LAS METODOLOGÍAS.

4.1.1. Metodología usada en la investigación

Respecto a la componente de investigación, esta será guiada bajo la metodología KDD, la cual se define como “un proceso no trivial de identificar patrones en los datos que sean válidos, novedosos, potencialmente útiles y finalmente comprensibles” [36]. En primer lugar, se seleccionan y limpian los datos que se deben extraer para poder realizar el modelado del comportamiento de búsqueda. Luego, se transforman los datos y se realiza minería de datos sobre ellos para buscar los patrones de interés que pueden expresarse como un modelo o que expresen dependencia de los datos. Finalmente, se identifican los patrones realmente interesantes que representan el conocimiento, usando diferentes técnicas, incluyendo análisis estadísticos y lenguajes de consultas para posteriormente interpretar los datos obtenidos.

4.1.2. Metodología usada para el desarrollo de *software*

Respecto a la componente de desarrollo de software, se recurre a un proceso de desarrollo inspirado en la metodología RAD, la cual minimiza la planificación en favor de la creación rápida de prototipos. La planificación se realiza en cada iteración, permitiendo que el software se desarrolle más rápido y se tenga mayor flexibilidad con los requisitos [37].

4.2. HERRAMIENTAS DE DESARROLLO

Las herramientas a utilizar en el trabajo de tesis, se dividen tanto en *hardware* como en *software*.

- *Hardware:*
- *Software:*

4.3. AMBIENTE DE TRABAJO

El ambiente de desarrollo del presente proyecto será tanto en el domicilio particular del candidato a tesista, como también en el Departamento de Ingeniería Informática de la Universidad de Santiago de Chile, específicamente, en el laboratorio de postgrado de dicho departamento.

Después de finalizar cada iteración, el *feedback* a este proyecto es ofrecido por miembros del equipo de investigación del proyecto iFuCo, y el profesor guía, quien, además brinda apoyo en los aspectos tecnológicos y metodológicos. Finalmente, el equipo de desarrollo de este trabajo de título es unipersonal, con colaboración en fundamentos teóricos de otros tesistas involucrados en el proyecto.

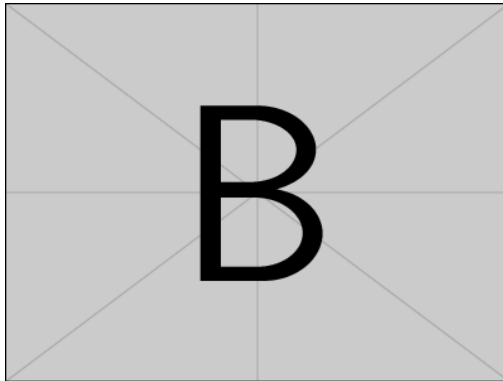


Figura 4.1: A cow licking its nose. Usage with permission of the photographer Nicole Barth
Fuente: Obtenido de www.flickr.com/photos/46311827@N07/14885545396/, (2017).

CAPÍTULO 5. PLAN DE TRABAJO

El presente proyecto contempla 640 horas de trabajo efectivas y se realizará en el transcurso del segundo semestre del año 2017, el cual se inicia el 7 de agosto y termina el 7 de diciembre del presente año contemplando 16 semanas de trabajo. Se dispone como día de trabajo todos los días hábiles de la semana, desde las 9:00 hrs. A.M hasta las 18:00 hrs. P.M considerando una hora de descanso.

El plan de trabajo propuesto se muestra en la Tabla 5.1, dada las metodologías empleadas las actividades se realizan de forma secuencial. Cabe destacar que el alumno candidato a tesista a la fecha de entrega del presente informe, ya ha avanzado el estado del arte e investigación de tecnologías.

BIBLIOGRAFÍA

- [1] A. L. Association *et al.*, "Information literacy competency standards for higher education", 2000 (ver pág. 2).
- [2] J. C. Tseng, G.-J. Hwang, P.-S. Tsai, C.-C. Tsai *et al.*, "Meta-analyzer: A Web-based learning environment for analyzing student information searching behaviors", *International Journal of Innovative Computing, Information and Control*, vol. 5, n.º 3, págs. 567-579, 2009 (ver pág. 2).
- [3] S. A. Weiner, "Who teaches information literacy competencies? Report of a study of faculty", *College Teaching*, vol. 62, n.º 1, págs. 5-12, 2014 (ver pág. 2).
- [4] J. K. Smith, L. M. Given, H. Julien, D. Ouellette y K. DeLong, "Information literacy proficiency: Assessing the gap in high school student's readiness for undergraduate academic work", *Library & Information Science Research*, vol. 35, n.º 2, págs. 88-96, 2013 (ver pág. 2).
- [5] M. Á. Marzal y E. Saurina, "Diagnóstico del estado de la alfabetización en Información (ALFIN) en las Universidades Chilenas", *Perspectivas em Ciência da Informação*, vol. 20, n.º 2, págs. 58-78, 2015 (ver pág. 2).
- [6] M. C. V. Urrea y S. O. Castro, "Alfabetización en información: Estudio de su impacto en estudiantes de último año del pregrado de las facultades de educación y ciencias naturales y exactas en la Universidad de Playa Ancha de Ciencias de la Educación", 2016 (ver págs. 2, 3).
- [7] CONICYT. (). Listado de proyectos seleccionados concurso AKA-CONICYT Edu2 2015, dirección: <http://www.conicyt.cl/pci/files/2015/05/AKA-EDU-2-proyectos-adjudicados.pdf> (ver págs. 3, 8).
- [8] A. of Finland. (). Funding decision - Enhancing learning and teaching future competences of online inquiry in multiple domains (iFuco), dirección: http://webfocus.aka.fi/ibi_apps/WFServlet?IBIF_ex=x_HakKuvaus&HAKNR01=%20294163&UILANG=en&IBIAPP_app=aka_ext (ver pág. 3).
- [9] I. H. Witten, E. Frank, M. A. Hall y C. J. Pal, *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2016 (ver pág. 3).
- [10] C. Romero y S. Ventura, "Educational Data Mining: A Review of the state of the art", *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 40, n.º 6, págs. 601-618, 2010 (ver pág. 3).

- [11] M. Pandey y V. K. Sharma, "A Decision tree algorithm pertaining to the student performance analysis and prediction", *International Journal of Computer Applications*, vol. 61, n.º 13, 2013 (ver [pág. 3](#)).
- [12] A. Merceron y K. Yacef, "Educational Data Mining: A Case study", en *AIED*, 2005, págs. 467-474 (ver [págs. 3, 4](#)).
- [13] J. Kumar, "A Comprehensive study of educational Data Mining", *International Journal of Electrical Electronics & Computer Science Engineering Special Issue-TeLMISR*, págs. 2348-2273, 2015 (ver [pág. 3](#)).
- [14] J. Bichsel, *Analytics in higher education: Benefits, barriers, progress, and recommendations*. EDUCAUSE Center for Applied Research, 2012 (ver [pág. 4](#)).
- [15] L. V. Morris, S.-S. Wu y C. L. Finnegan, "Predicting retention in online general education courses", *The American Journal of Distance Education*, vol. 19, n.º 1, págs. 23-36, 2005 (ver [pág. 4](#)).
- [16] P. Baepler y C. J. Murdoch, "Academic analytics and Data Mining in higher education", *International Journal for the Scholarship of Teaching and Learning*, vol. 4, n.º 2, [pág. 17](#), 2010 (ver [pág. 4](#)).
- [17] R. Baker *et al.*, "Data Mining for education", *International encyclopedia of education*, vol. 7, n.º 3, págs. 112-118, 2010 (ver [pág. 4](#)).
- [18] K. R. Koedinger, S. D'Mello, E. A. McLaughlin, Z. A. Pardos y C. P. Rosé, "Data Mining and education", *Wiley Interdisciplinary Reviews: Cognitive Science*, vol. 6, n.º 4, págs. 333-353, 2015 (ver [pág. 4](#)).
- [19] C. Romero y S. Ventura, "Data Mining in education", *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 3, n.º 1, págs. 12-27, 2013 (ver [pág. 4](#)).
- [20] V. Sarala y J. Krishnaiah, "Empirical study Of Data Mining techniques in education system", *International Journal of Advances in Computer Science and Technology (IJACST)*, págs. 15-21, 2015 (ver [pág. 4](#)).
- [21] A. Dutt, S. Aghabozrgi, M. A. B. Ismail y H. Mahrooian, "Clustering algorithms applied in educational Data Mining", *International Journal of Information and Electronics Engineering*, vol. 5, n.º 2, [pág. 112](#), 2015 (ver [pág. 4](#)).
- [22] O. Akinola, B. Akinkunmi y T. Alo, "A Data Mining model for predicting computer programming proficiency of computer science undergraduate students", 2012 (ver [pág. 5](#)).
- [23] E. Lahtinen, K. Ala-Mutka y H.-M. Järvinen, "A Study of the difficulties of novice programmers", en *Acm Sigcse Bulletin*, ACM, vol. 37, 2005, págs. 14-18 (ver [pág. 5](#)).

- [24] S. Borkar y K. Rajeswari, "Attributes selection for predicting students' academic performance using education Data Mining and artificial neural network", *International Journal of Computer Applications*, vol. 86, n.º 10, 2014 (ver [pág. 5](#)).
- [25] M. Jayakameswaraiah y S. Ramakrishna, "A Study on prediction performance of some Data Mining algorithms", *International Journal*, vol. 2, n.º 10, 2014 (ver [pág. 5](#)).
- [26] A. Abdullah, A. Malibari y M. Alkhozai, "Student's performance prediction system using multi agent Data Mining technique", *International Journal of Data Mining & Knowledge Management Process*, vol. 4, n.º 5, [pág. 1](#), 2014 (ver [pág. 5](#)).
- [27] R. J. Oskouei y M. Askari, "Predicting academic performance with applying Data Mining techniques (Generalizing the results of two different case studies)", *Computer Engineering and Applications Journal*, vol. 3, n.º 2, [págs. 79-88](#), 2014 (ver [pág. 5](#)).
- [28] S. Borkar y K. Rajeswari, "Predicting students academic performance using education Data Mining", *International Journal of Computer Science and Mobile Computing (IJCSMC)*, vol. 2, n.º 7, [págs. 273-279](#), 2013 (ver [pág. 5](#)).
- [29] S. F. Shazmeen, M. M. A. Baig y M. R. Pawar, "Performance evaluation of different Data Mining classification algorithm and predictive analysis", *IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN*, vol. 2278, 2013 (ver [pág. 5](#)).
- [30] C. Márquez-Vera, A. Cano, C. Romero y S. Ventura, "Predicting student failure at school using genetic programming and different Data Mining approaches with high dimensional and imbalanced data", *Applied intelligence*, vol. 38, n.º 3, [págs. 315-330](#), 2013 (ver [pág. 6](#)).
- [31] A. M. Shahiri, W. Husain *et al.*, "A Review on predicting student's performance using Data Mining techniques", *Procedia Computer Science*, vol. 72, [págs. 414-422](#), 2015 (ver [pág. 6](#)).
- [32] K. Sukhija, M. Jindal y N. Aggarwal, "The recent state of educational Data Mining: A Survey and future visions", en *MOOCs, Innovation and Technology in Education (MITE), 2015 IEEE 3rd International Conference on*, IEEE, 2015, [págs. 354-359](#) (ver [pág. 6](#)).
- [33] M. Anoopkumar y A. M. Z. Rahman, "A Review on Data Mining techniques and factors used in educational Data Mining to predict student amelioration", en *Data Mining and Advanced Computing (SAPIENCE), International Conference on*, IEEE, 2016, [págs. 122-133](#) (ver [pág. 6](#)).
- [34] D. Gacitúa, "NEURONE: Sistema de apoyo para la evaluación de competencias de investigación en línea para estudiantes de enseñanza básica", Memoria de pregrado, Universidad de Santiago de Chile, 2016 (ver [pág. 8](#)).
- [35] J. Martin, *Rapid application development*. Macmillan Publishing Co., Inc., 1991 (ver [pág. 9](#)).

- [36] U. Fayyad, G. Piatetsky-Shapiro y P. Smyth, "From Data Mining to knowledge discovery in databases", *AI magazine*, vol. 17, n.º 3, pág. 37, 1996 (ver [pág. 9](#)).
- [37] S. McConnell, *Rapid development: taming wild software schedules*. Pearson Education, 1996 (ver [pág. 10](#)).

APÉNDICE A. CAPÍTULO APÉNDICE

APÉNDICE B. ANOTHER APPENDIX CHAPTER