

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



**MODELO DE PREDICCIÓN DEL DESEMPEÑO DE BÚSQUEDA  
DE INFORMACIÓN EN LÍNEA EN ESTUDIANTES DE  
EDUCACIÓN BÁSICA**

**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

2017

## 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).

En el marco del proyecto iFuCo (*Enhancing learning and teaching future competences of online inquiry in multiple domains*), formado por investigadores de Chile y Finlandia, el cual desea investigar y modelar los comportamientos y competencias de investigación en línea de estudiantes de enseñanza básica, se propone la construcción de un modelo de predicción del comportamiento de búsqueda de información en línea en estudiantes de educación básica el cual se vaya perfeccionando a través del registro de datos históricos y que de un feedback en tiempo real.

La investigación será guiada por la metodología KDD con el fin de descubrir patrones en los datos que permitan la creación de un modelo de predicción del comportamiento de búsqueda. Además, para apoyar el proceso de investigación, se desarrollará una plataforma que funcione como extensión de la plataforma NEURONE (*oNlinE inqUiry expeRimentatiON systEm*). La plataforma propuesta alimentará y perfeccionará el modelo de predicción y entregará predicciones en tiempo real. Esta plataforma se guiará bajo la metodología RAD (*Rapid Application Development*) la cual se orienta a un desarrollo iterativo e incremental para la rápida construcción de prototipos de *software*.

# TABLA DE CONTENIDOS

<b>Capítulo 1. Objetivos y alcances de la solución</b>	<b>1</b>
1.1 Objetivo general . . . . .	1
1.2 Objetivos específicos . . . . .	1
<b>Capítulo 2. Descripción del problema</b>	<b>2</b>
2.1 Motivación . . . . .	2
2.2 Revisión de la literatura . . . . .	3
2.3 Definición del problema . . . . .	5
<b>Capítulo 3. Descripción de la solución propuesta</b>	<b>7</b>
3.1 Características de la solución . . . . .	7
3.2 Propósito de la solución . . . . .	7
3.3 Alcances y limitaciones de la solución . . . . .	8
<b>Capítulo 4. Metodología, herramientas y ambiente de desarrollo</b>	<b>9</b>
4.1 Metodología a usar . . . . .	9
4.1.1 Metodología usada en la investigación . . . . .	9
4.1.2 Metodología usada para el desarrollo . . . . .	9
4.2 Herramientas de desarrollo . . . . .	10
4.2.1 Herramientas de <i>hardware</i> . . . . .	10
4.2.2 Herramientas de <i>software</i> . . . . .	10
4.3 Ambiente . . . . .	10
<b>Capítulo 5. Plan de trabajo</b>	<b>11</b>
<b>Referencias bibliográficas</b>	<b>14</b>
<b>Apéndice A. Capítulo Apéndice</b>	<b>15</b>
A.1 Sección del apéndice . . . . .	15
A.1.1 Subseccion del apéndice . . . . .	15
<b>Apéndice B. Another Appendix Chapter</b>	<b>16</b>

## ÍNDICE DE TABLAS

5.1. Plan de trabajo propuesto. . . . .	11
B.1. Ejemplo de una tabla. . . . .	16

## ÍNDICE DE FIGURAS

A.1. A scientific diagram using the <code>pgfplots</code> package by Christian Feuersaenger using the same colors which are also used for the layout . . . . .	15
---	----

# **CAPÍTULO 1. OBJETIVOS Y ALCANCES DE LA SOLUCIÓN**

## **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 básica.

## **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 (oNlinE inqUiry expeRimentatiON systEm, NEURONE desde ahora en adelante).
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. Implementar los modelos de predicción del comportamiento de búsqueda en línea de estudiantes de básica.
6. Implementar y evaluar una plataforma de enseñanza de competencias informacionales, que en base a los datos provistos por NEURONE pueda predecir el desempeño de búsqueda de un estudiante.

## 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 y la diversidad de fuentes de información. 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 (Association *et al.*, 2000). 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 (Tseng, Hwang, Tsai, Tsai *et al.*, 2009).

Actualmente, en Chile la enseñanza de competencias informacionales es cubierta en bibliotecas universitarias y cursos introductorios de mallas universitarias (Marzal & Saurina, 2015). De acuerdo con Urrea y Castro, 2016, 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 se prioriza la reiteración de la 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 Urrea y Castro, 2016. 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 en el contexto del proyecto de investigación “*Enhancing Learning and Teaching Future Competences of Online Inquiry in Multiple Domains*” (iFuCo, desde

ahora en adelante) [24], el cual pretende abordar la temática de la alfabetización informacional en estudiantes de enseñanza básica con el objetivo de estudiar sus patrones de comportamiento y ofrecer modelos curriculares adecuados respecto al tema [3].

### 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 lugar el estudio del comportamiento de estudiantes. En segundo lugar, técnicas de minería de datos y plataformas de aprendizaje de máquina aplicada al contexto educacional.

Cuando se aplica minería de datos en instituciones educativas, la disciplina se conoce como minería de datos educacional (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 MDE 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 (Romero & Ventura, 2010). La disciplina se enfoca en el diseño de modelos para mejorar las experiencias del aprendizaje y la eficiencia organizacional (Pandey & Sharma, 2013). El principal objetivo de la MDE es visto por diferentes investigadores como: i) modelado del estudiante, ii) modelado del dominio, iii) sistema de aprendizaje, iv) construir modelos computacionales, y v) estudiar los efectos de los recursos [7-9].

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 (Bichsel, 2012). En este contexto, University of Georgia desarrolló un modelo para predecir la tasa de graduación y abandono estudiantil en un ambiente en línea (Morris, Wu & Finnegan, 2005). 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 Baepler y Murdoch, 2010.

Baker *et al.*, 2010 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, Pardos y Rosé, 2015 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, 2010 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, 2015 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 y Mahroeian, 2015 consolida las variantes de algoritmos de clustering aplicados al contexto de MDE. 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, 2005 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, 2012 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, 2005 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, 2014 evalúa el rendimiento de los estudiantes, donde selecciona algunos atributos mediante minería de datos, haciendo uso de una red neuronal multicapa perceptrón 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, 2014 aplica los mismos métodos propuestos por Borkar y Rajeswari, 2014.

Abdullah, Malibari y Alkhozai, 2014 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, 2014 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, 2013 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, 2013 evalúa el rendimiento de diferentes algoritmos de clasificación y análisis predictivo, basado en el trabajo de Borkar y Rajeswari, 2013 y propone técnicas de preprocesamiento de datos para lograr mejores resultados.

Tal como se muestra en los antecedentes anteriores, las investigaciones en MDE se realizan mayoritariamente en aprendizaje *online* y en casos puntuales en educación superior, por lo que es limitada la información respecto a educación básica o media, específicamente en la predicción de errores y fracaso escolar (Márquez-Vera, Cano, Romero & Ventura, 2013). Para mayor información de trabajos relacionados con la MDE, consultar los siguientes reviews [30-33].

### 2.3. DEFINICIÓN DEL PROBLEMA

El proceso de búsqueda de información involucra diferentes procesos cognitivos, habilidades, variables de comportamiento y entorno de una persona. En el contexto de la enseñanza de la alfabetización informacional, 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?

**RQ 3 :**

## 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<sup>1</sup> (*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 en particular debe 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 su base de datos, alimentando y perfeccionando el modelo. El ciclo de construcción, evaluación y optimización del modelo se describe en la Figura 3.1, donde el modelo está en una continua optimización.

### 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

---

<sup>1</sup>Traducción libre

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, estos datos tienen su propio contexto y origen que limitan la generalización de 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 [24], 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 (*oNlinE inqUiry expeRimentatiON systEm*), trabajo de memoria de un estudiante de la carrera de Ingeniería de Ejecución en Computación e Informática de la Universidad de Santiago de Chile [25].
4. La solución funciona como un sistema predictor del desempeño de estudiante en la búsqueda de información, sin ofrecer acciones correctivas en caso de bajo desempeño.

## 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 será guiada por la metodología de desarrollo de *software* *Rapid Application Development* (RAD, desde ahora en adelante) [27]. A continuación, se explica el uso de ambas metodologías en el trabajo propuesto. Para mayor información sobre las metodologías a ocupar, ver ANEXO B.

#### 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” [29]. 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

Respecto a la componente de desarrollo de *software*, se recurre a un enfoque 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 [28].

## **4.2. HERRAMIENTAS DE DESARROLLO**

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

### **4.2.1. Herramientas de *hardware***

El desarrollo se llevará a cabo con procesador Intel Core i7 7ma Generación *KabyLake* de 3.6 Ghz, con memoria Ram de 16 GB y 2 TB de disco duro. Además, los despliegues de prueba se realizan sobre un servidor privado virtual (VPS, por sus siglas en inglés) con el sistema operativo GNU/Linux Ubuntu Server alojado en el proveedor DigitalOcean<sup>1</sup>.

### **4.2.2. Herramientas de *software***

En cuanto herramientas *software*, el desarrollo se llevará a cabo en la distribución GNU/Linux Debian en su versión 9.0. El modelo se llevará a cabo en Tensorflow. Para el análisis estadístico se hará uso de R. Además, cada módulo desarrollado estará contenido en contenedores de Docker para facilitar el despliegue en producción del modelo desarrollado. Todo el trabajo realizado, tanto código como documento escrito estará bajo el sistema de control de versiones Git. Finalmente, se hará uso de  $\text{\LaTeX}$  para el documento escrito.

## **4.3. AMBIENTE**

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 sistemas colaborativos.

Después de finalizar cada iteración, la retroalimentación a este proyecto es ofrecida por miembros del equipo de investigación del proyecto iFuCo y el profesor guía, quien además, brinda apoyo en aspectos tecnológicos y metodológicos. Finalmente, el equipo de desarrollo de este trabajo de tesis es unipersonal, con colaboración en fundamentos teóricos de otros tesis y memoristas involucrados en el proyecto.

---

<sup>1</sup><https://www.digitalocean.com/>

## CAPÍTULO 5. PLAN DE TRABAJO

El presente proyecto contempla 612 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, y un horario de trabajo desde las 9:00 hasta las 18:00 hrs 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.

Tabla 5.1: Plan de trabajo propuesto.  
Fuente: Elaboración propia, (2017).

Actividad	Duración (HH)
Actualización del estado del arte centrado en trabajos recientes	1
Exploración, preprocesamiento y transformación de los registros del proceso de búsqueda de información obtenidos por NEURONE	3
Definición de las características para la construcción de modelos de predicción del desempeño de búsqueda de estudiantes	1
Comparación y selección de los algoritmos y/o técnicas de minería de datos para la construcción de modelos	1
Validación de los algoritmos y/o técnicas con datos conocidos	5
Evaluación de los modelos construidos utilizando métricas de desempeño	2



## REFERENCIAS BIBLIOGRÁFICAS

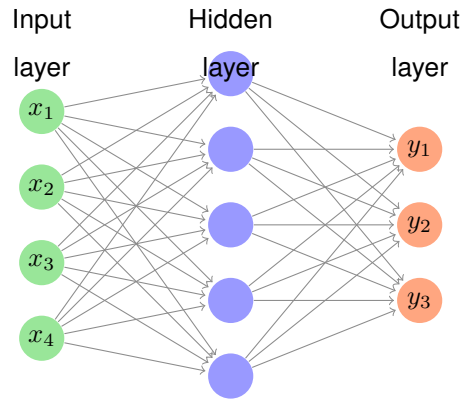
- Abdullah, A., Malibari, A. & Alkhozai, M. (2014). Student's performance prediction system using multi agent data mining technique. *International Journal of Data Mining & Knowledge Management Process*, 4(5), 1. (ver pág. 5).
- Akinola, O., Akinkunmi, B. & Alo, T. (2012). A data mining model for predicting computer programming proficiency of computer science undergraduate students. (ver pág. 4).
- Association, A. L. *et al.* (2000). Information literacy competency standards for higher education. (ver pág. 2).
- Baepler, P. & Murdoch, C. J. (2010). Academic analytics and data mining in higher education. *International Journal for the Scholarship of Teaching and Learning*, 4(2), 17. (ver pág. 3).
- Baker, R. *et al.* (2010). Data mining for education. *International encyclopedia of education*, 7(3), 112-118. (ver pág. 3).
- Bichsel, J. (2012). *Analytics in higher education: Benefits, barriers, progress, and recommendations*. EDUCAUSE Center for Applied Research. (ver pág. 3).
- Borkar, S. & Rajeswari, K. (2013). Predicting student's academic performance using education data mining. *International Journal of Computer Science and Mobile Computing (IJCSMC)*, 2(7), 273-279. (ver pág. 5).
- Borkar, S. & Rajeswari, K. (2014). Attributes selection for predicting student's academic performance using education data mining and artificial neural network. *International Journal of Computer Applications*, 86(10). (ver pág. 4).
- Dutt, A., Aghabozrgi, S., Ismail, M. A. B. & Mahrooieian, H. (2015). Clustering algorithms applied in educational data mining. *International Journal of Information and Electronics Engineering*, 5(2), 112. (ver pág. 4).
- Jayakameswaraiah, M. & Ramakrishna, S. (2014). A study on prediction performance of some data mining algorithms. *International Journal*, 2(10). (ver pág. 4).

- Koedinger, K. R., D'Mello, S., McLaughlin, E. A., Pardos, Z. A. & Rosé, C. P. (2015). Data mining and education. *Wiley Interdisciplinary Reviews: Cognitive Science*, 6(4), 333-353. (ver pág. 4).
- Lahtinen, E., Ala-Mutka, K. & Järvinen, H.-M. (2005). A study of the difficulties of novice programmers. En *Acm Sigcse Bulletin* (Vol. 37, 3, pp. 14-18). ACM. (ver pág. 4).
- Márquez-Vera, C., Cano, A., Romero, C. & Ventura, S. (2013). Predicting student failure at school using genetic programming and different data mining approaches with high dimensional and imbalanced data. *Applied intelligence*, 38(3), 315-330. (ver pág. 5).
- Marzal, M. Á. & Saurina, E. (2015). Diagnóstico del estado de la alfabetización en información (ALFIN) en las universidades chilenas. *Perspectivas em Ciência da Informação*, 20(2), 58-78. (ver pág. 2).
- Merceron, A. & Yacef, K. (2005). Educational data mining: A case study. En *AIED* (pp. 467-474). (ver pág. 4).
- Morris, L. V., Wu, S.-S. & Finnegan, C. L. (2005). Predicting retention in online general education courses. *The American Journal of Distance Education*, 19(1), 23-36. (ver pág. 3).
- Oskouei, R. J. & Askari, M. (2014). Predicting academic performance with applying data mining techniques (generalizing the results of two different case studies). *Computer Engineering and Applications Journal*, 3(2), 79-88. (ver pág. 5).
- Pandey, M. & Sharma, V. K. (2013). A decision tree algorithm pertaining to the student performance analysis and prediction. *International Journal of Computer Applications*, 61(13). (ver pág. 3).
- Romero, C. & Ventura, S. (2010). Educational data mining: a review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 40(6), 601-618. (ver págs. 3, 4).
- Sarala, V. & Krishnaiah, J. (2015). Empirical study Of data mining techniques in education system. *International Journal of Advances in Computer Science and Technology (IJACST)*, 15-21. (ver pág. 4).
- Shazmeen, S. F., Baig, M. M. A. & Pawar, M. R. (2013). Performance evaluation of different data mining classification algorithm and predictive analysis. *IOSR Journal of Computer Engineering*, 10(6), 01-06. (ver pág. 5).

- Tseng, J. C., Hwang, G.-J., Tsai, P.-S., Tsai, C.-C. *et al.* (2009). Meta-analyzer: A web-based learning environment for analyzing student information searching behaviors. *International Journal of Innovative Computing, Information and Control*, 5(3), 567-579. (ver pág. 2).
- Urra, M. C. V. & Castro, S. O. (2016). 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, 20-40. (ver pág. 2).

## APÉNDICE A. CAPÍTULO APÉNDICE

### A.1. SECCIÓN DEL APÉNDICE



**Example Diagram with a Line Break in the Title  
(using the text width option in the title style)**

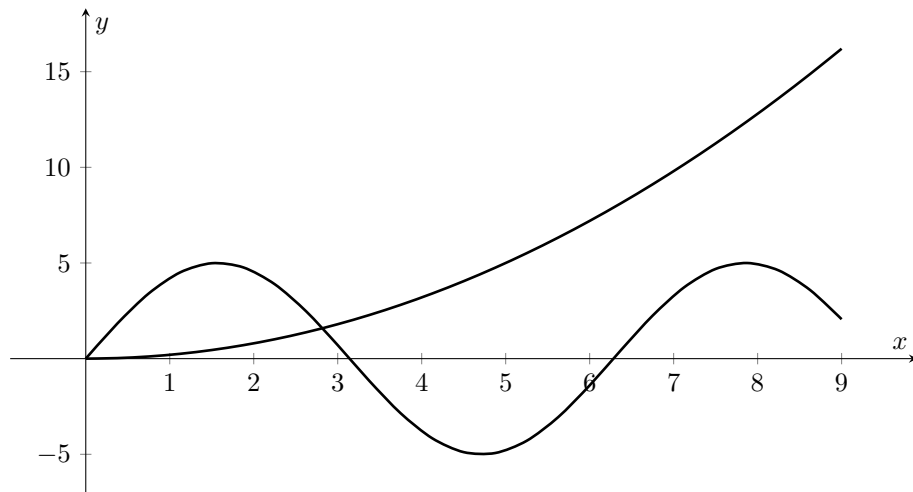


Figura A.1: A scientific diagram using the `pgfplots` package by Christian Feuersaenger using the same colors which are also used for the layout

Fuente: Elaboración propia, (2017).

#### A.1.1. Subseccion del apéndice

## APÉNDICE B. ANOTHER APPENDIX CHAPTER

Como se puede apreciar en la Tabla B.1.

Tabla B.1: Ejemplo de una tabla.  
Fuente: Elaboración propia, (2017).

header1	header2	header3
1	2	3
4	5	6
7	8	9