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

MODELO PREDICTIVO DEL DESEMPEÑO DE BÚSQUEDA DE INFORMACIÓN EN LÍNEA EN ESTUDIANTES DE EDUCACIÓN BÁSICA

Gonzalo Javier Martinez Ramirez

Profesor guía: Roberto Ignacio González Ibáñez

Tesis de grado presentado en conformidad a los requisitos para obtener el grado de Magíster en Ingeniería Informática.

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*.

Palabras Claves: Alfabetización informacional, competencias de investigación en línea, comportamiento de estudiantes, minería de datos, modelos de clasificación.

ABSTRACT

Today

Keywords:

TABLA DE CONTENIDOS

Tabla de Contenidos

ÍNDICE DE TABLAS

ÍNDICE DE FIGURAS

CAPÍTULO 1. INTRODUCCIÓN

1.1. ANTECEDENTES Y MOTIVACIÓN

La alfabetización informacional (conocida en inglés como *information literacy*) es definida como "el grupo de habilidades en las que se requiere reconocer cuándo la información es necesaria y tener la habilidad de encontrar, evaluar y usar efectivamente dicha información necesaria"[1] (Association *et al.*, 2000, p. 2). Durante la última década, debido a los rápidos avances de las tecnologías de la información y comunicación (TICs) ha aumentado la cantidad de recursos digitales en Internet y además se ha facilitado el acceso a ellos. Estos avances han provocado una brecha entre el ser humano y la habilidad de reconocer cuando la información es necesaria para satisfacer su necesidad de búsqueda, la cual se puede asociar principalmente a dos razones: En primer lugar, las competencias de alfabetización informacional no son enseñadas ni reforzadas a temprana edad. Segundo, las búsquedas *web* han pasado a ser parte de las tareas comunes que realizan los estudiantes, disminuyendo las visitas a bibliotecas y el uso de fuentes revisadas.

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 (básica, media y universitaria). La enseñanza de la alfabetización informacional se imparte principalmente por bibliotecas universitarias, y en menor medida en la etapa escolar obligatoria (Weiner, 2014). En Chile, la enseñanza de competencias informacionales es cubierta en bibliotecas universitarias y cursos introductorios de mallas universitarias (Marzal & Saurina, 2015). De acuerdo con Urra 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 (Urra & Castro, 2016).

A través de encuestas a estudiantes universitarios, Head (2013, p. 475) establece que al momento de realizar investigaciones el 84 % de los estudiantes universitarios utiliza como fuente primaria de búsqueda Wikipedia[2] y un 87 % consulta a sus amigos, sin verificar la veracidad de la información que obtienen. Como consecuencia, los estudiantes al no ser instruidos en parafrasear, resumir o citar fuentes revisadas, caen al plagio de forma premeditada o no intencionada.

[1] Traducción libre.
[2] https://es.wikipedia.org/

A partir de los argumentos anteriormente expuestos, respecto a la enseñanza de competencias de alfabetización informacional, se puede ver que no ha sido completamente satisfecha y la brecha entre los usuarios e alfabetización informacional permanece abierta.

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*"[3] (iFuCo, desde ahora en adelante), 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 (Sormunen *et al.*, 2017).

1.2. DESCRIPCIÓN DEL PROBLEMA

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 búsqueda de los alumnos.

En el contexto del proyecto de investigación iFuCo, el cual pretende realizar un análisis cuantitativo y cualitativo de la alfabetización informacional y las competencias de búsqueda en línea en estudiantes de enseñanza básica[4] en los países de Chile y Finlandia, 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 búsqueda de la información de los estudiantes?

RQ 2 ¿En qué medida es posible detectar situaciones anormales de conducta, y determinar las causas que llevan a un estudiante a fallar durante el proceso de búsqueda de información?

RQ 3 ¿De qué manera se puede implementar un módulo de clasificación y predicción del desempeño de los estudiantes en la búsqueda de información en herramientas de apoyo de la alfabetización informacional para proporcionar una retro evaluación oportuna a estudiantes y docentes?

[3] https://www.researchgate.net/project/Enhancing-learning-and-teaching-for-future-competences-of-online-inquiry-in-multiple-domains-iFuCo

[4] En otros países es conocido como enseñanza primaria

1.3. SOLUCIÓN PROPUESTA

1.3.1. Características de la solución

La solución consiste en incorporar un módulo en NEURONE (González-Ibáñez, Gacitua, Sormunen & Kiili, 2017) que clasifique y prediga de forma continua el desempeño de búsqueda de los estudiantes de enseñanza básica en un curso de alfabetización informacional, específicamente en el tema de investigaciones en línea[5] (*online inquiry*).

Los datos son recopilados y almacenados por NEURONE, estos datos provienen de registros del proceso de búsqueda de 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.

El módulo propuesto hará uso de Apache Spark[6], el cual es un *framework* de código abierto para el procesamiento de datos masivos, el cual incluye librerías de minería de datos y aprendizaje de máquina. Este módulo 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 ilustra en la Figura 1.1, donde a través de los datos históricos obtenidos de NEURONE construye el modelo, lo evalúa y lo optimiza en un proceso continuo, entregando como resultado la clasificación del desempeño de búsqueda y prediciendo de forma continua a partir del comportamiento actual de búsqueda de información del estudiante.

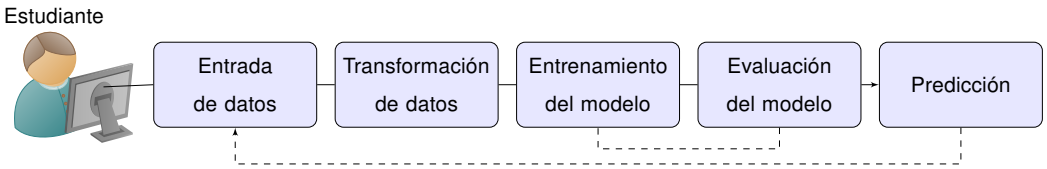


Figura 1.1: Ciclo de construcción y perfecionamiento del modelo

Fuente: Elaboración propia, (2017)

1.3.2. Propósito de la solución

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

[5] Traducción libre.
[6] https://spark.apache.org/

Con el módulo propuesto en este trabajo, el docente obtiene una estimación temprana del desempeño del estudiante en el proceso de búsqueda de información, de tal forma que él pueda guiar al estudiante en el proceso. Tal como ilustra la Figura 1.2, 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 docente de los patrones y predicciones del desempeño de búsqueda del estudiante con el objetivo de ayudar en la toma de decisiones al docente correspondiente para diseñar y planificar de mejor forma la entrega de contenidos hacia el estudiante.

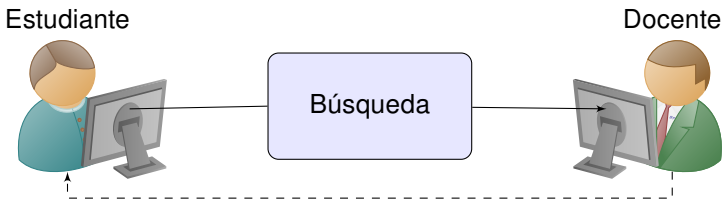


Figura 1.2: Proceso de búsqueda de información de un estudiante

Fuente: Elaboración propia, (2017)

1.4. OBJETIVOS Y ALCANCES DE LA SOLUCIÓN

1.4.1. Objetivo general

Diseñar y evaluar un modelo predictivo del desempeño de búsqueda de información en línea de estudiantes de enseñanza básica.

1.4.2. 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 (acrónimo de oNlinE inqUiry expeRimentatiON systEm).
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, la cual en base a los datos provistos por NEURONE prediga el desempeño de búsqueda de un estudiante.

1.4.3. Alcances

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, 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.

4. La solución funciona como un sistema predictor del desempeño del estudiante en la búsqueda de información, sin ofrecer acciones correctivas en caso de bajo desempeño.

1.5. METODOLOGÍA Y HERRAMIENTAS UTILIZADAS

1.5.1. 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 realizada.

La componente de investigación del proyecto será guiada por la metodología Descubrimiento de Conocimiento en Base de Datos (conocido como KDD, las iniciales de *Knowledge Discovery in Databases*) (Fayyad, Piatetsky-Shapiro & Smyth, 1996), mientras que la componente de desarrollo

será guiada por la metodología de desarrollo de *software* Desarrollo de Rápido de Aplicaciones (conocido como RAD, las iniciales de *Rapid Application Development*) (Martin, 1991). A continuación, se explica el uso de ambas metodologías en el trabajo propuesto.

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" (Fayyad *et al.*, 1996, p. 5). 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 para posteriormente interpretar los datos obtenidos.

Metodología usada para el desarrollo

Respecto a la componente de desarrollo de *software*, se toma en cuenta las condiciones bajo las cuales se desarrolla el proyecto, las cuales se expresan a continuación:

- El sistema es de rápido desarrollo.
- El sistema es de tamaño pequeño.
- Es un proyecto cuyos requerimientos están sujetos a cambios.
- Inicialmente no existe un número total de requerimientos especificado. Estos se irán desarrollando de forma creciente durante el avance del proyecto.
- El desarrollador no cuenta con un conocimiento profundo de la arquitectura y todas las herramientas de desarrollo, por lo tanto, se requiere un tiempo de investigación y aprendizaje.
- Se requiere documentar los aspectos fundamentales de la arquitectura, una vez que se tenga un producto estable. Esta documentación permitirá la continuidad del proyecto.
- Se requiere de varias entregas funcionales, para medir el progreso del proyecto y verificar que se cumplan los objetivos propuestos.

Dado los antecedentes mencionados anteriormente, se determina que el proyecto presenta características que se ajustan bien a un modelo de desarrollo evolutivo enfocado a la generación de prototipos. A partir de esto, se recurre a un enfoque de desarrollo inspirado en la metodología RAD, metodología de desarrollo rápido que 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 una mayor flexibilidad con los requisitos (McConnell, 1996).

1.5.2. Herramientas de desarrollo

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

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[7].

Software

En cuanto herramientas *software*, el desarrollo se llevará a cabo en la distribución GNU/Linux Debian[8] en su versión 9.0. El modelo se llevará a cabo en Spark ML[9]. 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 LaTeX para el documento escrito.

- Anaconda Python 3.5
- Apache Kafka 0.10.2.0

[7] https://www.digitalocean.com/
[8] https://www.debian.org/
[9] https://spark.apache.org/

- Apache Spark 2.1.0
- Flask 0.12.2 [10]
- MongoDB
- Python 3.5

1.6. ORGANIZACIÓN DEL DOCUMENTO

El resto del documento se estructura de la siguiente forma.

Capítulo 2 Se estipulan los conceptos teóricos que se deben definir para tener una base consensuada respecto de los distintos conceptos que se tratan en este documento. En el mismo capítulo se aborda el estado del arte donde se hace una revisión bibliográfica de los últimos avances en el área.

[10] http://flask.pocoo.org/

CAPÍTULO 2. MARCO TEÓRICO

Este capítulo tiene como objetivo entregar las bases teóricas, conceptuales y empíricas que soportan cada desarrollo de esta investigación. En primer lugar, se presenta el marco conceptual donde se entregan las definiciones y conceptos necesarios para abordar esta investigación. En segundo lugar, se presenta el estado del arte relacionado con el tema.

2.1. MARCO CONCEPTUAL

En esta sección se presentan conceptos y bases teóricas respecto a la temática que conduce el desarrollo de este trabajo, el cual tiene relación con el uso de interfaces no tradicionales, específicamente con una interfaz operada con el cuerpo. Además, se indaga sobre ciertas definiciones para establecer lo que se pretende medir en este estudio, lo que involucra la experiencia de usuario y métricas de rendimiento en la realización de tareas. Finalmente, se proponen ciertas características fundamentales respecto de este tipo de proyectos relacionados con diseños experimentales con usuarios.

Human Computer Information Retrieval (HCIR) o Recuperación de Información Humano Computador[1], es el estudio de los métodos que integran la inteligencia humana y la búsqueda algorítmica para ayudar a la gente a mejorar la búsqueda, exploración y aprendizaje de información (Marchionini, 2007). Dentro de esta área interactúan otras disciplinas, como la recuperación de información, llamada en inglés *Information Retrieval* (IR), la que está enfocada principalmente en proveer a los usuarios de fácil acceso a la información de su interés trabajando con la representación, almacenamiento, organización y acceso a objetos de información como documentos, páginas *web*, catálogos en línea y objetos multimedia (Baeza-Yates Ribeiro-Neto, 2011) y la búsqueda de información, llamada en inglés *Information Seeking* (IS) que se entiende como un proceso más orientado al usuario y abierto que IR. En IS, no se sabe si existe una respuesta a la consulta del usuario, por lo que el proceso de búsqueda puede proporcionar el aprendizaje necesario para satisfacer su necesidad de información (Baeza-Yates Ribeiro-Neto, 2011).

[1] Traducción libre.

2.1.1. Búsqueda de información

2.1.2. Rendimiento

En el contexto de la recuperación de información se definen *precision* y *recall* en función de un conjunto de documentos relevantes y un conjunto de documentos recuperados (Powers, 2011), las cuales se definen a continuación.

Precision Métrica que mide la razón de documentos relevantes recuperados con respecto al total de documentos recuperados. Es un valor continuo entre 0 y 1, mientras más cercano a 1, mayor fue su precisión al encontrar los documentos relevantes.

$$Precision = \frac{\left[\{documentos\ relevantes\} \cap \{documentos\ recuperados\}\right]}{\{documentos\ recuperados\}} \tag{2.1}$$

Recall Métrica que mide la razón de documentos relevantes recuperados con respecto al total de documentos relevantes. Es un valor continuo entre 0 y 1, mientras más cercano a 1, mayor fue la recuperación de documentos en base al total del universo disponible.

$$Recall = \frac{\left[\{documentos\ relevantes\} \cap \{documentos\ recuperados\}\right]}{\{documentos\ relevantes\}} \tag{2.2}$$

F1 Métrica que considera a *precision* y *recall* en un promedio ponderado. Es un valor continuo entre 0 y 1, en que un valor cercano a uno permite identificar a los estudiantes con una recuperación de documentos proporcional a su precisión, respecto a la relevancia de estos.

$$F1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \tag{2.3}$$

Coverage Effectiveness Razón entre la cobertura útil, páginas visitadas sobre 30 segundos, y el total de documentos visitados por un estudiante. Es un valor continuo entre 0 a 1, mientras más cercano a 1, mejor fue la efectividad de la cobertura respecto el universo de documentos, respecto al tiempo de permanencia.

$$CE = \frac{UsfCover}{TotalCover} \tag{2.4}$$

Query Effectiveness Razón existente entre la efectividad de la cobertura (fórmula anterior) y el total de consultas realizadas por un estudiante. Esta proporción da indicios del desempeño

del estudiante en torno a la calidad de las consultas efectuadas, en base a la cantidad y eficacia. Sus valores también están en un intervalo continuo entre 0 a 1.

$$QE = \frac{CE}{countQ} \tag{2.5}$$

Search Score Calificación de los estudiantes que se expresa en una escala continua de 0 a 5 puntos. Es una razón entre la cobertura relevante y el total de páginas marcadas activas al final de la tarea.

$$Score = \frac{BMRelv}{ActBM} * 5 \tag{2.6}$$

2.1.3. Alfabetización informacional

La alfabetización informacional (conocida en ingles *information literacy*) es definida como "el grupo de habilidades en las que se requiere reconocer cuándo la información es necesaria y tener la habilidad de encontrar, evaluar y usar efectivamente dicha información necesaria" [2] (Association *et al.*, 2000, p. 2). Es un campo que cubre varias áreas, entre las que se destaca la alfabetización digital, las habilidades de uso de bibliotecas, la ética informacional, la lectura crítica, el pensamiento crítico, los derechos de autor, la seguridad y privacidad, entre otras. A través del estudio de estas áreas como factores que influyen a la alfabetización informacional se puede obtener una visión clara de cómo los estudiantes llevan a cabo sus tareas de obtención y selección de información.

2.1.4. Competencias de investigación en línea

Se definen las competencias de investigación (*inquiry skills* en inglés) como "las habilidades para explorar preguntas, para poder reunir, interpretar y sintetizar diferentes tipos de información y datos, además de desarrollar y compartir una explicación para responder preguntas dadas" [3] (Council *et al.*, 2000, p. 13). En base a este concepto nacen las competencias de investigación en línea (conocidas en inglés como *online inquiry skills*), que son una instancia específica de las competencias de investigación, pero aplicada sobre información disponible en línea (Quintana, Zhang & Krajcik, 2005).

Las competencias de investigación en línea involucran una serie de actividades cognitivas, como generar una pregunta de investigación, buscar información relevante en colecciones digitales, evaluar y seleccionar la información encontrada, e integrar coherentemente la información seleccionada para responder la pregunta original (Eisenberg & Berkowitz, 1990).

[2] Traducción libre.
[3] Traducción libre.

2.1.5. Minería de datos educacional

La minería de datos utiliza una combinación de bases de conocimientos explícita, conocimientos analíticos complejos y conocimiento de campo para descubrir las tendencias y los patrones ocultos, estas tendencias y patrones forman la base de los modelos predictivos que permiten a los analistas realizar nuevas observaciones de los datos existentes (Luan, 2002). La gran cantidad de información generada hoy en día por los estudiantes permite que la minería de datos obtenga datos relevantes y, a través de métodos estadísticos y otras herramientas, relacione la información para conocer si el proceso de enseñanza aprendizaje ha dado resultados positivos.

Mining (2012, p. 9) define la minería de datos educacional (MDE, desde ahora en adelante) como "la teoría que desarrolla métodos, aplica técnicas estadísticas y de aprendizaje automático para analizar los datos recogidos durante el proceso de la enseñanza y aprendizaje" [4]. Actualmente, los usos más generales que se le están dando a la MDE básicamente se enfocan en mejorar la estructura del conocimiento y determinar el apoyo pedagógico al estudiante.

2.1.6. Técnicas de minería de datos

Support Vector Machine

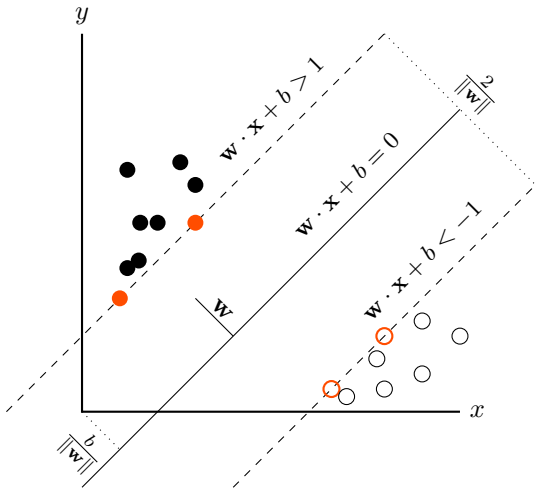


Figura 2.1: SVM

Fuente: Elaboración propia, (2017)

[4] Traducción libre.

Perceptrón multicapa

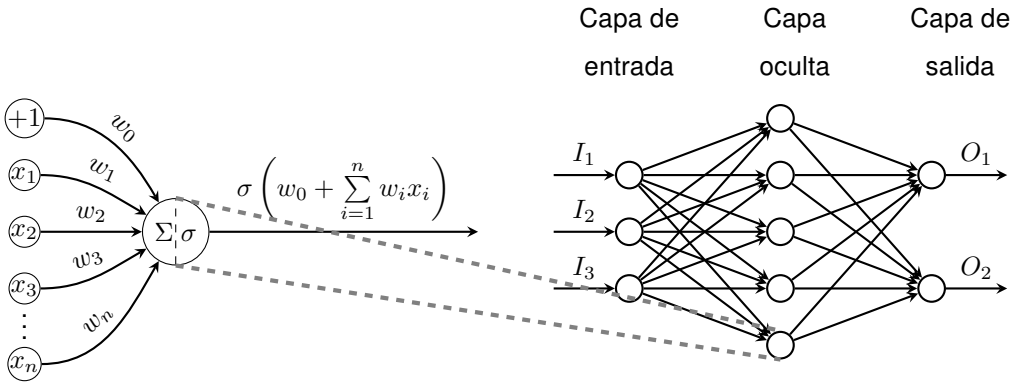


Figura 2.2: Perceptrón multicapa

Fuente: Elaboración propia, (2017)

2.2. ESTADO DEL ARTE

El propósito de esta sección es presentar los últimos trabajos realizados en la línea de investigación que se ha planteado en la sección y capítulo anterior. La búsqueda de estos trabajos considera: la exploración de trabajos con estudios que midan experiencia de usuario y/o medidas de rendimiento en la utilización de interfaces no-tradicionales operadas con el cuerpo para la realización de actividades en distintos contextos.

2.3. MARCO DE INVESTIGACIÓN

En esta sección se propone y formula el marco de investigación que guía este estudio. Para esto se desarrollan las preguntas de investigación que motivan la realización del estudio. Además se presentan las hipótesis que se desean someter a prueba en base a los resultados producto de la realización del estudio que en este documento se propone.

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 búsqueda de la información de los estudiantes?

RQ 2 ¿En qué medida es posible detectar situaciones anormales de conducta, y determinar las causas que llevan a un estudiante a fallar durante el proceso de búsqueda de información?

RQ 3 ¿De qué manera se puede implementar un módulo de clasificación y predicción del desempeño de los estudiantes en la búsqueda de información en herramientas de apoyo de la

alfabetización informacional para proporcionar una retro evaluación oportuna a estudiantes y docentes?

2.4. RESUMEN

En el presente capítulo

Como marco conceptual se definen conceptos importantes que son utilizados en el estudio. Primero la experiencia de usuario dado que implica emociones y actitudes de una persona con respecto al producto o servicio que se esté utilizando suele medirse con instrumentos que requieren una participación directa del usuario como encuestas. Con respecto al rendimiento en el área de recuperación de información se utilizan métricas para evaluarlo a partir fórmulas que usan la cantidad de documentos clasificados y documentos correctamente clasificados.

En el estado del arte se realiza una revisión de distintos estudios de usuario en los que se trabaja con representaciones alternativas de documentos, donde se observaron distintos resultados. Por ejemplo, Nguyen y Zhang (2006) encontraron que, al utilizar una representación visual de documentos, la clasificación de documentos relevantes mejoraba o Tilsner (2009) quien observó un rendimiento mayor con menores tiempos de respuesta utilizando una interfaz visual. Por otro lado, Sebrechts, Cugini, y Laskowski (1999) obtuvieron un rendimiento similar contrastando una interfaz visual con una tradicional. Por otro lado, se ve que en el mercado ya existen algunas herramientas de visualización como Zakta o oSkope las que pueden resultar beneficiosas si se usan para buscar ciertos tipos de información.

Finalmente, en el marco de investigación se plantean preguntas de investigación relacionadas con cómo las personas perciben su interacción con resultados de búsqueda de información a través de representaciones visuales y si es posible mejorar aspectos como la experiencia de usuario el rendimiento utilizando este tipo de representación. Lo anterior lleva a las dos hipótesis con las que se trabaja en este proyecto que de forma resumida son:

CAPÍTULO 3. METODOLOGÍA

El presente capitulo tiene por objetivo exponer los aspectos metodológicos implicados en el desarrollo de la componente de investigación de este trabajo, que es llevado a cabo mediante la metodología Descubrimiento de Conocimiento en Base de Datos (desde ahora en adelante RAD, las iniciales de *Knowledge Discovery in Databases*). En primer lugar, se detalla la metodología KDD. Posteriormente, se presentan los datos utilizados para luego contextualizar la metodología para este trabajo.

3.1. DESCUBRIMIENTO DE CONOCIMIENTO EN BASE DE DATOS (KDD)

CAPÍTULO 4. DESARROLLO DE *SOFTWARE*

En el presente capítulo tiene por objetivo presentar el desarrollo de *software* de apoyo a la experimentación llevado a cabo mediante la metodología de desarrollo de *software* Desarrollo Rápido de Aplicaciones (desde ahora en adelante RAD, las iniciales de *Rapid Application Development*). En primer lugar, se presenta la metodología de desarrollo de *software* donde se define brevemente en que se basa la construcción de la herramienta. En segundo lugar, se hace una definición conceptual de los aspectos básicos que debe cumplir el *software* desarrollado. Luego, se presenta el desarrollo de *software* desde el punto de vista de los prototipos construidos a razón de la metodología ocupada. Finalmente, se describe la arquitectura construida para la herramienta.

4.1. DESARROLLO RÁPIDO DE APLICACIONES (RAD)

La metodología utilizada en el desarrollo de *software* es la metodología RAD (*Rapid Application Development* o Desarrollo Rápido de Aplicaciones), la cual es una metodología de desarrollo que 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 una mayor flexibilidad con los requisitos (Martin, 1991).

Utilizando RAD la planificación del desarrollo de *software* se intercala con la construcción del *software* en sí. La falta de una amplia pre-planificación general, permite que el *software* sea implementado expeditamente y hace que sea más fácil cambiar los requisitos. Cada iteración de RAD, se compone de cuatro etapas (Martin, 1991), las cuales se explican a continuación:

1. **Etapa de definición conceptual**: También conocida como “Planeación de requerimientos”, es la fase donde se definen las funciones del negocio y los alcances de la solución.

2. **Etapa de diseño funcional**: También conocida como “Diseño del usuario” es la fase donde se modela el sistema y sus procesos. Suelen utilizar herramientas CASE para realizar el modelado mencionado.

3. **Etapa de desarrollo**: También conocida como etapa de “Construcción”, es cuando se ejecuta el trabajo planificado en las etapas anteriores y hace el desarrollo propio del sistema.

4. **Etapa de despliegue**: También conocida como “Implementación” es cuando el prototipo es liberado y se entrega para la evaluación por parte del cliente.

4.2. DEFINICIÓN CONCEPTUAL

4.2.1. Requerimientos de *software*

Requisitos funcionales

RF1

RF2

RF3

RF4

RF5

Requisitos no funcionales

A continuación se presenta una lista de requerimientos no funcionales que la aplicación debe cumplir.

RNF1

RNF2

RNF3

RNF4

RNF5

4.3. DESARROLLO

Tabla 4.1: Asociación entre tareas que componen el desarrollo y los prototipos realizados

Tareas/Prototipo	P1	P2	P3	P4
Chapter 1: Introduction	✓	✓	✓	✓
Chapter 2: Background and experimental set-up		✓	✓	✓
Chapter 3: Evaluating and comparing outlier-selection algorithms		✓		
Chapter 4: Stochastic Outlier Selection			✓	
Chapter 5: Meta-features for one-class data sets				✓
Chapter 6: Meta-learning for one-class classifiers				✓
Chapter 7: Conclusions	✓	✓	✓	✓

Fuente: Elaboración propia, (2017)

4.3.1. Prototipo 1

4.3.2. Prototipo 2

4.3.3. Prototipo 3

4.3.4. Prototipo 4

CAPÍTULO 5. RESULTADOS Y ANÁLISIS

CAPÍTULO 6. CONCLUSIONES

En este capítulo se presentan las conclusiones del trabajo realizado y los resultados obtenidos durante este proyecto. En primer lugar, se concluye respecto del grado de cumplimento de los objetivos. En segundo lugar, se concluye respecto de los resultados obtenidos en la evaluación. Después, se revisan las implicancias y alcances que tiene el trabajo realizado. Posteriormente, se listan potenciales trabajos futuros que podrían realizarse a partir de esta investigación. Finalmente, se realizan observaciones y apreciaciones finales del investigador respecto al proyecto.

6.1. OBJETIVOS

En el capítulo introductorio de este trabajo se presentó el objetivo general, el cual plantea el trasfondo de la investigación, este es descompuesto en objetivos específicos que definen la serie de aspectos a satisfacer para alcanzar la consecución del objetivo general de esta investigación. En esta sección se presentan las conclusiones respectivas a los objetivos planteados que guían el desarrollo de este proyecto. En primer lugar, se retoman los objetivos específicos para concluir sobre los procesos realizados para abordar cada uno de estos. En segundo lugar, se retoma el objetivo general del proyecto sintetizando de forma integral los pasos necesarios para alcanzarlo.

6.1.1. Objetivos específicos

A continuación, se listan las conclusiones respectivas a cada uno de los objetivos específicos planteados con el propósito de alcanzar el objetivo general ideado para este proyecto.

- Objetivo 1
- Objetivo 2
- Objetivo 3
- Objetivo 4
- Objetivo 5

6.1.2. Objetivo general

A continuación, se presenta el objetivo general que guía las pautas de este proyecto, el cual es:

En función de este objetivo, complementado con la motivación y el levantamiento del estado del arte realizado en este proyecto, se definen tres preguntas de investigación.

RQ 1

RQ 2

RQ 3

REFERENCIAS BIBLIOGRÁFICAS

Association, A. L. *et al.* (2000). Information literacy competency standards for higher education. (citado en páginas 1, 11).

Council, N. R. *et al.* (2000). *Inquiry and the national science education standards: A guide for teaching and learning*. National Academies Press. (citado en página 11).

Eisenberg, M. B. & Berkowitz, R. E. (1990). *Information problem solving: The big six skills approach to library & information skills instruction.* ERIC. (citado en página 11).

Fayyad, U., Piatetsky-Shapiro, G. & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI magazine*, *17*(3), 37. (citado en páginas 5, 6).

González-Ibáñez, R., Gacitua, D., Sormunen, E. & Kiili, C. (2017). NEURONE: oNlinE inqUiRy experimentatiON systEm. En T. be included in Proceedings of the 80th Annual Meeting of the Association for Information Science & T. (2017) (Eds.). (citado en página 3).

Head, A. J. (2013). Project Information Literacy: What can be learned about the information-seeking behavior of today's college students? (citado en página 1).

Luan, J. (2002). Data mining and its applications in higher education. *New directions for institutional research*, *2002*(113), 17-36. (citado en página 12).

Martin, J. (1991). *Rapid application development*. Macmillan Publishing Co., Inc. (citado en página 6).

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. (citado en página 1).

McConnell, S. (1996). *Rapid development: Taming wild software schedules*. Pearson Education. (citado en página 7).

Mining, T. E. D. (2012). Enhancing teaching and learning through educational data mining and learning analytics: An issue brief. En *Proceedings of conference on advanced technology for education*. (citado en página 12).

Quintana, C., Zhang, M. & Krajcik, J. (2005). A framework for supporting metacognitive aspects of online inquiry through software-based scaffolding. *Educational Psychologist*, *40*(4), 235-244. (citado en página 11).

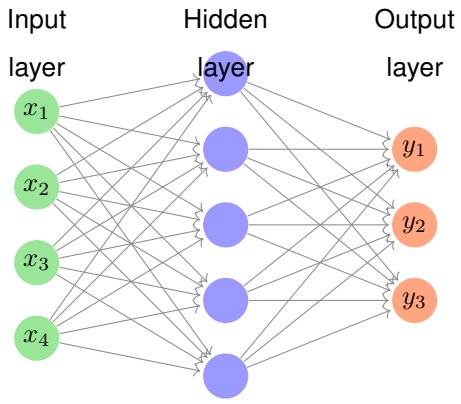
Sormunen, E., González-Ibáñez, R., Kiili, C., Leppänen, P., Mikkilä-Erdmann, M., Erdmann, N. & Escobar-Macaya, M. (2017). A Performance-based Test for Assessing Students' Online Inquiry Competences in Schools. En E. C. in Information Literacy (ECIL) (Ed.). (citado en página 2).

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. (citado en página 1).

Weiner, S. A. (2014). Who teaches information literacy competencies? Report of a study of faculty. *College Teaching*, *62*(1), 5-12. (citado en página 1).

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`)

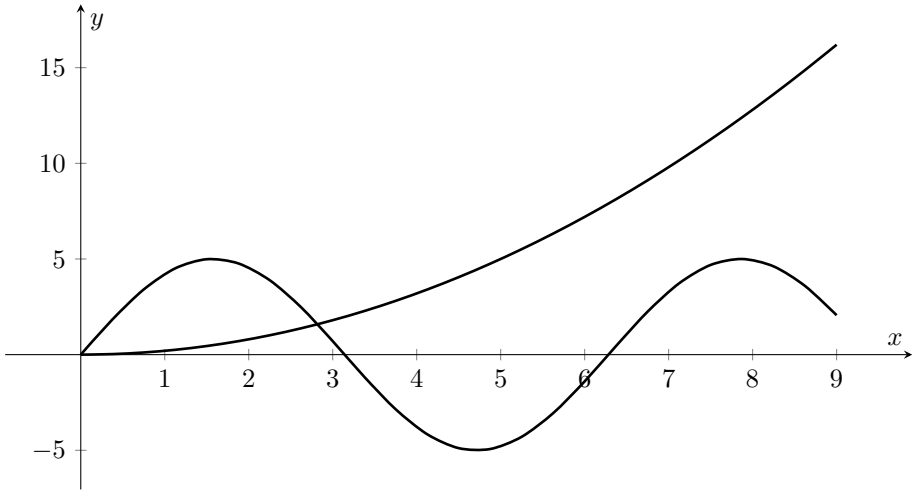


Figura A.1: A scientific diagram using the `pgfplots` package by Christian Feuersaenger *using the same colors which are also used for the layout*
Fuente: Elaboración propia, (2017)

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

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

Fuente: Elaboración propia, (2017)

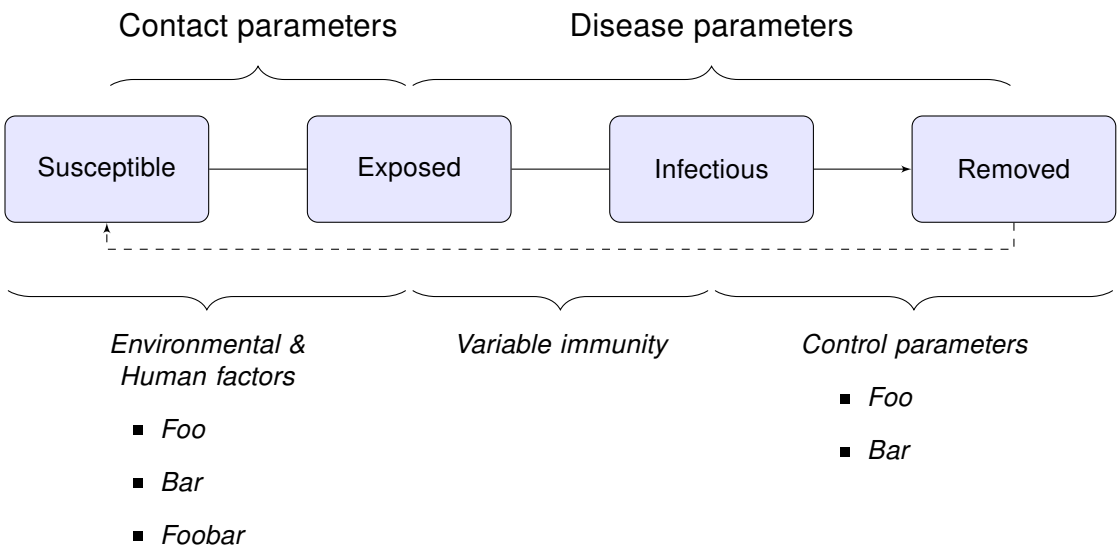
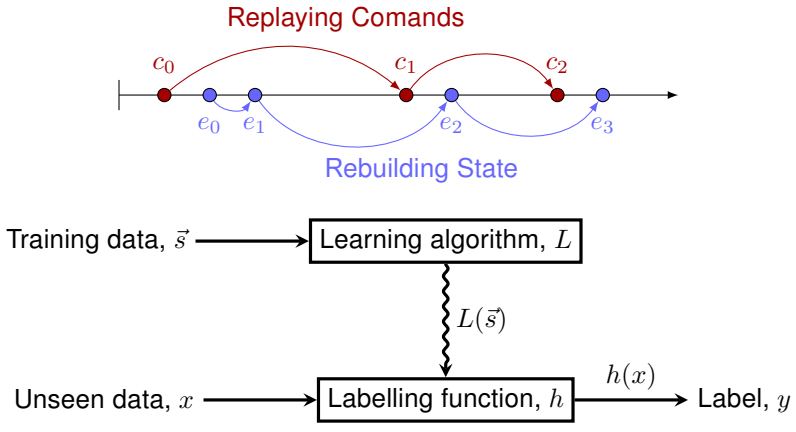
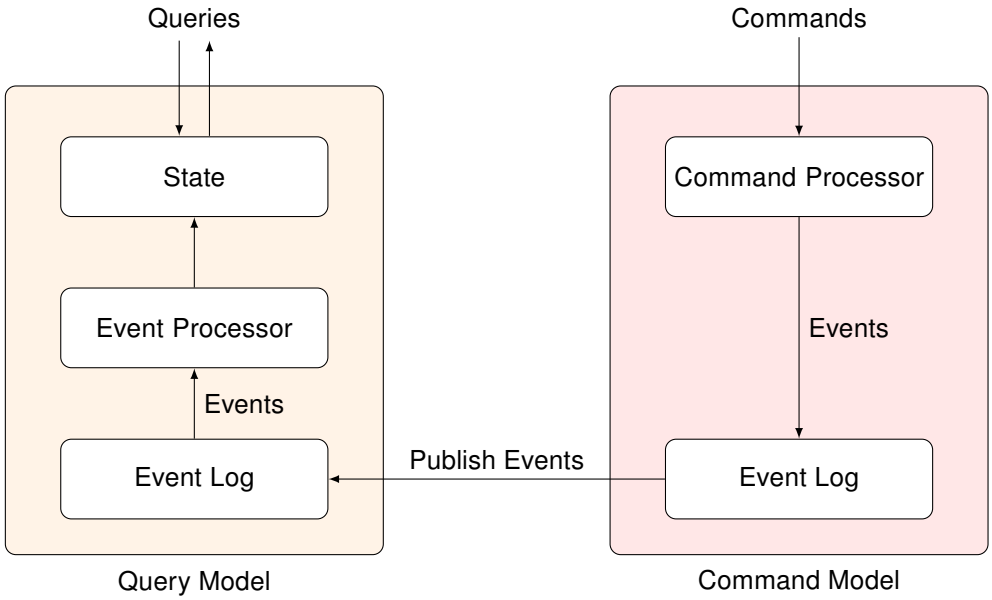
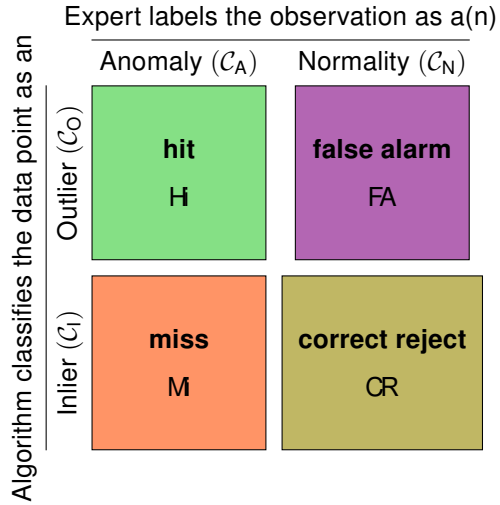
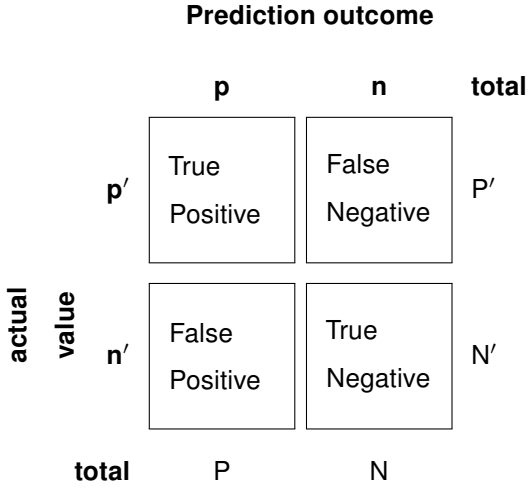


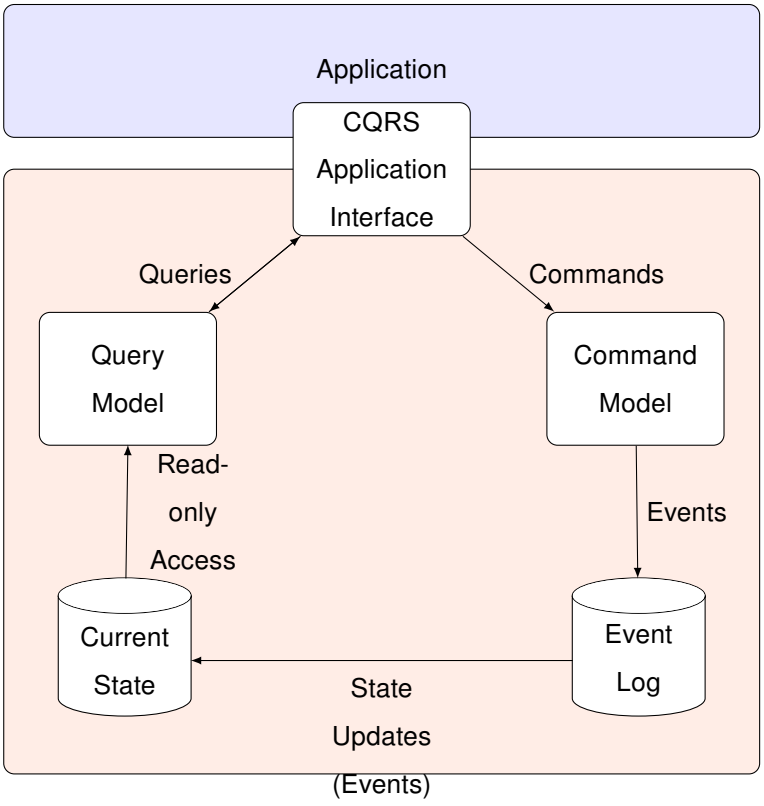
Figura B.1: Flowchart of fundamental disease transmission mechanisms

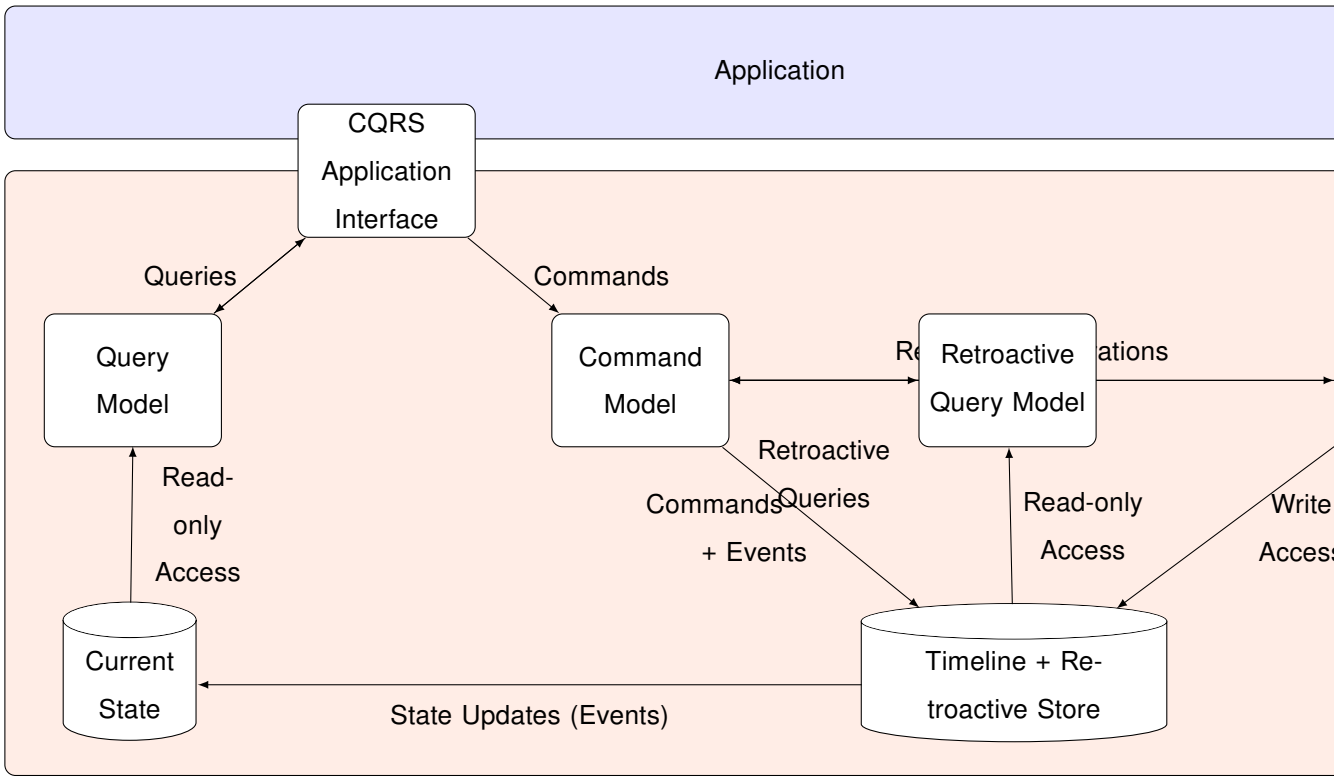



Queries
Commands
State
Command Processor
Event Processor
Events
Events
Publish Events
Event Log
Event Log
Query Model
Command Model


Expert labels the observation as a(n)
Anomaly ($\mathcal{C}_A$)
Normality ($\mathcal{C}_N$)
Algorithm classifies the data point as an
Outlier ($\mathcal{C}_O$)
Inlier ($\mathcal{C}_I$)
hit
H
false alarm
FA
miss
M
correct reject
CR


Prediction outcome
p
n
total
p′
True Positive
False Negative
P′
actual value
n′
False Positive
True Negative
N′
total
P
N


Application
CQRS
Application
Interface
Queries
Commands
Query
Model
Command
Model
Read-
only
Access
Events
Current
State
Event
Log
State
Updates
(Events)


Application
CQRS
Application
Interface
Queries
Commands
Query
Model
Command
Model
Retroactive
Query Model
Retroactive
Commands
+ Events
Queries
Read-only
Access
Read-
only
Access
Current
State
Timeline + Re-
troactive Store
State Updates (Events)
eventually consistent
always consistent

	ml				kdd	
Feature	knn	ddpw	dds	vddlof	loci	
Estimate local density	✓				✓	✓
Estimate global density		✓				
Domain based			✓			