

Diseño de un sistema de análisis de datos para Trayectorias Educativas UTVT

Designing a data analysis system for UTVT Educational Paths

MAYA-PÉREZ, P. Norma†*, ORONA-LÓPEZ, Miguel Ángel y BARRON-ADAME, J. Miguel''

Universidad Tecnológica del Valle de Toluca. Carretera, Del Depto del Distrito Federal km 7.5, 52044 Santa María Atarasquillo, Méx.

Universidad Tecnológica del Suroeste de Guanajuato. Carretera Valle de Santiago - Huamimaro Km. 1.2, 20 de Noviembre, 38400 Valle de Santiago, Gto.

ID 1^{er} Autor: P. Norma, Maya-Pérez / ORC ID: 0000-0001-5531-5358, Researcher ID Thomson: S-7924-2018, CVU CONACYT ID: 429523

ID 1^{er} Coautor: Miguel Ángel, Orona-López / ORC ID: 0000-0003-0075-1253, Researcher ID Thomson: S-8754-2018, CVU CONACYT ID: 678742

ID 2^{do} Coautor: J. Miguel, Barron-Adame / ORC ID: 0000-0001-8308-9474, Researcher ID Thomson: S-4649-2018, CVU CONACYT ID: 221435

DOI: 10.35429/JET.2019.9.3.10.23

Recibido: 15 de Julio, 2019; Aceptado 20 de Septiembre, 2019

Resumen

El artículo presenta el primer prototipo del sistema web diseñado para el pre procesamiento de datos con Minería de Datos (DM), que incluye la recopilación de datos de fuentes externas y heterogéneas, preparar los datos importados, a través de un proceso de selección y limpieza, transformación con identificadores de variables y parametrizadas a valores discretos, siendo pre-procesadas a datos útiles, generando una vista minable, requisito previo al modelado con DM y Redes Neuronales Artificiales (RNA). El desarrollo del sistema se basa en la metodología ágil, el Modelo Vista Controlador y herramientas tecnológicas PHP puro, Laravel, Composer, Ajax, JQuery, Json, Bootstrap, HTML5, CSS3 y MySQL; publicando la primera versión en un sitio web y continuando su aplicación en el pre procesamiento en una muestra de estudiantes de Tecnologías de la Información y Comunicación de la Universidad Tecnológica del Valle de Toluca. Generando como resultado de la primera etapa de investigación, una herramienta y un aporte tecnológico, que coadyuve al análisis de trayectorias educativas y seguimiento oportuno a estudiantes con riesgo a deserción o reprobación; en la que se pretende en una etapa posterior, incorporar un modelo de DM y RNA.

Estrés académico, Estudiantes de ingeniería, Desempeño en matemáticas

Abstract

This paper presents the first prototype of the web system designed for data preprocessing with Data Mining (DM), which includes the collection of data from external and heterogeneous sources, preparing imported data, through a selection and cleaning process, transformation with variable identifiers and parameterized to discrete values, being pre-processed to useful data, generating a minable view, prerequisite to modeling with DM and Artificial Neural Networks (ANNs). The development of the system is based on the agile methodology, the Model View Controller and technological tools: pure PHP, Laravel, Composer, Ajax, JQuery, Json, Bootstrap, HTML5, CSS3 and MySQL; publishing the first version on a website and continuing its application in pre-processing in a sample of students of Information and Communication Technologies of the Technological University of Valle de Toluca. Generating as a result of the first stage of research, a tool and a technological contribution, which supports the analysis of educational trajectories and timely monitoring of students at risk of dropping out or failing; in which it is intended at a later stage, to incorporate a DM and RNA model.

Analysis of data, Educational trajectories, Web system

Citación: MAYA-PÉREZ, P. Norma, ORONA-LÓPEZ, Miguel Ángel y BARRON-ADAME, J. Miguel. Diseño de un sistema de análisis de datos para Trayectorias Educativas UTVT. Revista Teoría Educativa. 2019. 3-9: 10-23.

* Correspondencia del Autor (Correo electrónico: norma.maya@utvtol-edu.mx)

† Investigador contribuyendo como primer autor.

Introducción

En las últimas décadas han existido grandes avances en la investigación y desarrollo de software en cinco áreas de las Tecnologías de la Información (TI) como: minería de datos (DM) por sus siglas en inglés Data Mining, big data, cloud computer, inteligencia artificial (IA) e inteligencia de negocios; consideradas como la base para la explotación de la información y pieza fundamental en la inteligencia organizacional y que hacen la diferencia con respecto a los sistemas tradicionales, en busca el descubrimiento del conocimiento denominado KDD por sus siglas en inglés (Knowledge Discovery from Databases) como un proceso no trivial de identificar patrones válidos, novedosos, potencialmente útiles y en última instancia comprensibles a partir de los datos (Formia, 2013); este proceso define una metodología del ciclo de vida de desarrollo de un proyecto de DM que consta de varias etapas, partiendo con un gran volumen de datos iniciales, que son seleccionados a través de un pre procesamiento para obtener una vista minable que se utiliza en la construcción de un modelo de DM o de IA y finalmente generar el conocimiento y patrones procesados, que son interpretados y útiles en la toma de decisiones. Los modelos de DM e IA se han aplicado principalmente en las áreas del conocimiento en la medicina, finanzas, negocios organizacionales, procesos industriales y pocos estudios en educación, representando un desafío tecnológico progresivo en distintas dimensiones de la sociedad.

Derivado de los estudios progresivos en esta área en los últimos diez años, surge la idea de desarrollar el proyecto centrado en el análisis de datos basado en técnicas, métodos y algoritmos de DM para emitir reportes sistematizados de estudiantes con riesgo a deserción o reprobación escolar, en los programas educativos de Técnico Superior Universitario (TSU) de Tecnologías de la Información y comunicación (TIC) de la Universidad Tecnológica del Valle de Toluca (UTVT); esto a través de un sistema web de trayectorias educativas, para el seguimiento de estudiantes vulnerables y aplicar acciones estratégicas escolares oportunas.

El documento describe la problemática, justificación, objetivos del desarrollo del sistema, hipótesis, marco teórico de referencia, metodología aplicada y resultados obtenidos de del primer diseño del sistema web.

Problemática

De un análisis exploratorio sobre el gran número de herramientas de software (SW) de DM e IA que existe actualmente, se identifica, que en su gran mayoría requieren de licenciamiento comercial entre los que se destacan: Matlab, XLMiner que es un complemento de Excel, IBM SPSS Statistics, SAS Enterprise Miner, Rapid Miner, Oracle Data Mining; y solo un número mínimo de herramientas open source (sin costo) como: WEKA, KNIME (Konstanz Information Miner) y el entorno estadístico R; además de verificar que el SW existente, no se adaptan al entorno específico de la investigación de este trabajo, en la que se pretende aplicar métodos y técnicas de DM, sobre el análisis de trayectorias educativas y seguimiento de estudiantes de TSU de TIC de UTVT.

De esta manera se plantea la necesidad de diseñar el sistema web, en una primera aproximación para generar una herramienta tecnológica que coadyuve en este fenómeno representativo, dejando a un lado el uso de herramientas tecnológicas tradicionales que actualmente se utilizan para el análisis de datos y que a veces resultan ser tardías ante determinadas situaciones escolares por parte de los coordinadores o tutores de grupo del programa educativo (PE) de TIC.

El propósito del trabajo es, generar una herramienta tecnológica diseñada estratégicamente, a través de metodologías de ingeniería de software (SW), arquitectura del Modelo Vista Controlador (MVC) y herramientas web, para generar un sistema, que permita aplicar las fases de un proyecto de DM, inicialmente con datos de estudiantes de TIC, de ser posible en todos los PE de UTVT y como proyección futura con Universidad Tecnológica de Suroeste de Guanajuato (UTSOE) e instituciones interesadas.

Justificación

Observando el proceso manual de recopilación y análisis de datos escolares provenientes de otras áreas y en diferentes formatos para dar seguimiento de la trayectoria escolar en el PE de TSU en TIC, además de la duplicidad de tareas realizadas por el personal académico, administrativo y directivo de la UTVT para atender situaciones académicas de estudiantes vulnerables a la deserción escolar, surge la necesidad de diseñar el sistema web que permita automatizar este proceso y coadyuvar en la toma de decisiones oportunas.

Objetivos

Objetivo General

Diseñar el primer prototipo de un sistema web adaptativo al pre procesamiento de datos, tomando una muestra de estudiantes de TSU TIC, que permita gestionar y recopilar los datos de fuentes externas, selección y limpieza, transformación de datos a variables, parametrización de variables y generar una vista minable, para la implementación futura de un modelo DM.

Objetivos Específicos

- Analizar el proceso de desarrollo de un proyecto de DM.
- Analizar los modelos de arquitectura y desarrollo de SW para el diseño del sistema web.
- Diseño del sistema web.
- Desarrollo de las fases de pre procesamiento DM en el sistema web.
- Realizar pruebas de verificación del sistema WEB antes de la publicación.
- Publicación del sistema en su primera versión en un sitio web.
- Implementar el sistema web con datos de muestra seleccionada.
- Analizar los resultados de la implementación del primer prototipo

Hipótesis

El primer prototipo del sistema web, realiza el pre procesamiento DM, generando una vista minable con datos de la muestra representativa de estudiantes TIC-UTVT y sirve de base para el modelo DM de análisis de datos de trayectorias educativas, como proyección futura.

Marco Teórico

Ingeniería de Software

El desarrollo del sistema web se basa en la arquitectura MVC y en el marco de referencia de la Ingeniería de SW, que integra de metodologías, modelos, técnicas y herramientas tecnológicas con el propósito de obtener un producto de SW de calidad que cumpla con la usabilidad, funcionalidad, fiabilidad, eficiencia y capacidad, para el cual está destinado en satisfacer las necesidades del cliente (Pressman, 2006; Kendall & Kendall, 2011).

En tanto que Someerville (2005), separa estas herramientas de apoyo en modelos y metodologías de desarrollo de SW. Define un modelo como una representación simplificada del proceso para el desarrollo de SW, presentada desde una perspectiva específica; y la metodología como un enfoque estructurado para el desarrollo de SW que incluye modelos de sistemas, notaciones, reglas, sugerencias de diseño y guías de procesos.

Un modelo para el desarrollo de SW es una representación abstracta de un proceso. Cada modelo representa un proceso desde una perspectiva particular y así proporciona información parcial sobre el proceso.

Estos modelos generales no son descripciones definitivas de los procesos del SW, más bien son abstracciones de los procesos que se pueden utilizar para el desarrollo del SW.

Puede pensarse en ellos como marcos de trabajo del proceso y que pueden ser adaptados para crear procesos más específicos (Cendejas, 2014). Los modelos más usados en la práctica actual de la ingeniería de SW son:

- Cascada: se adapta perfectamente en proyectos que son rígidos, donde se especifiquen muy bien los requerimientos y se conozca muy bien la herramienta a utilizar. Señalando que rara vez se aplican a proyectos reales no sigue una secuencia lineal; tarda mucho tiempo ya que debe pasar por el proceso de prueba y hasta que el SW no esté completo no se opera (EcuRed, 2016).
- Desarrollo evolutivo (espiral): este modelo reduce los riesgos porque se revisa en cada iteración, antes de que se conviertan en problemas y es ideal para crear productos con diferentes versiones mejoradas y nuevos requerimientos como se hace con el SW actual, ya que este ciclo de vida no es rígido ni estático e incorpora objetivos de calidad (Someerville, 2005). El inconveniente que presenta es que genera mucho tiempo en el desarrollo del sistema, además requiere un alto costo de desarrollo y experiencia en la identificación de riesgos.
- Desarrollo basado en componentes: Se basa en un número significativo de componentes reutilizables y se integran en el sistema, más que en desarrollarlos desde cero (Someerville, 2005). Reduce la cantidad de SW, costos y riesgos, ya que se reutiliza código, permite una entrega más rápida, es el más utilizado actualmente para grandes sistemas de Información (SI) (Pressman, 2006). La única desventaja es que requiere de alto compromiso a los requerimientos y esto da lugar a un sistema que puede no cumplir con las necesidades reales de los usuarios (Pressman, 2006).

Señalando que los modelos descritos, no se excluyen mutuamente, con frecuencia se utilizan en conjunto especialmente para el desarrollo de grandes sistemas e integrando las metodologías denominadas pesadas, entre las más populares son: Rational Unified Process (RUP) y Microsoft Solutions Framework (MSF).

Adicionalmente con el progreso tecnológico global, se tiene que responder a nuevas oportunidades y mercados, condiciones económicas cambiantes y la aparición de productos y servicios de competidores, se requiere un desarrollo rápido y eficiente, ante estas necesidades y que aún las metodologías pesadas no cubren en la totalidad en este tipo de proyectos, se utilizan las metodologías ágiles (Cendejas, 2014).

Las metodologías ágiles de desarrollo de SW, buscan la satisfacción del cliente, la entrega temprana del SW incremental, equipos pequeños de alta motivación, métodos un tanto informales y una simplicidad general del desarrollo; cuyos procesos están diseñados para generar SW útil de forma rápida e interactiva que entrelazan la especificación, el diseño, el desarrollo y las pruebas; permitiendo el éxito de muchos de los sistemas desarrollados en distintas áreas (Cendejas, 2014). Los métodos de desarrollo ágil de SW más utilizados son: SCRUM, Extreme Programming (XP). Adicionalmente existen: Lean Software Development (LSD), Agile Modeling Adaptive Software Development (ASD), Agile Unified Process (AUP), Agile Data Method, Dynamic Systems Development Method (DSDM), Lean Software Development (LSD), y Microsoft Solutions Framework (MSF), etc (EcuRed, 2016).

En la investigación se realizó un análisis comparativo entre modelos tradicionales y metodologías pesadas con respecto a las metodologías ágiles, que se muestra en la Tabla 1, cuyas diferencias no solo se refieren al proceso en sí, sino también al contexto de equipo y organización, que es más favorable a cada uno de estas filosofías de procesos de desarrollo de SW (Cendejas, 2014).

Metodología Ágil	Modelos y Metodologías Tradicionales
Pocos artefactos. El modelo es prescindible, modelos desechables.	Más artefactos. El modelo es esencial, mantenimiento en los modelos.
Pocos roles, más genéricos y flexibles.	Más roles y más específicos.
No existe un contrato tradicional, debe ser bastante flexible.	Existe un contrato prefijado.
El cliente es parte del equipo de desarrollo.	El cliente interactúa con el equipo de desarrollo mediante reuniones.
Orientada a proyectos pequeños, y en el mismo lugar.	Aplicables a proyectos de cualquier tamaño, pero suelen ser especialmente efectivas/usadas en proyectos grandes.
La arquitectura se va definiendo y mejorando a lo largo del proyecto.	Se promueve que la arquitectura se defina tempranamente en el proyecto.
Énfasis en los aspectos humanos: el individuo y el trabajo en equipo.	Énfasis en la definición del proceso: roles, actividades y artefactos.
Basadas en heurísticas provenientes de prácticas de producción de código.	Basadas en normas provenientes de estándares seguidos por el entorno de desarrollo.
Se esperan cambios durante el desarrollo del proyecto.	Se espera que no ocurran cambios de gran impacto durante el desarrollo.

Tabla 1 Análisis comparativo de modelos tradicionales y metodologías ágiles de desarrollo de SW

Fuente: elaboración propia a partir de Cendejas, 2014

La metodología pesada del Proceso Unificado Ágil, AUP por sus siglas en inglés Agile Unified Process, se deriva como una versión simplificada de RUP, considerado en un marco iterativo e incremental, que describe de forma simple y fácil desarrollar aplicaciones de SW usando técnicas ágiles y conceptos que aún se mantienen válidos en RUP, aplicando un desarrollo dirigido de pruebas o TDD por sus siglas en inglés (test driven development), gestión de cambios ágiles y la refactorización de base de datos para mejorar la productividad del desarrollo y entrega de proyectos de pequeña y mediana dimensión.

Flores (2016) menciona que AUP es un proceso altamente formal porque especifica muchas actividades y artefactos involucrados en el desarrollo de un proyecto software, que puede ser adaptado a RUP metodología denominada pesada y éste a su vez ligero y ágil.

Atiende la gestión de riesgos; propone que aquellos elementos con alto riesgo obtengan prioridad en el proceso de desarrollo y sean abordados en etapas tempranas del mismo, donde se crean y mantienen listas identificando los riesgos, especialmente relevante en el desarrollo de prototipos ejecutables durante la base de elaboración del producto, demostrando la validez de la arquitectura para los requisitos clave del producto y que determinan los riesgos técnicos. El proceso de AUP establece un modelo más simple que RUP por lo que reúne en una única disciplina el modelado de negocio, requisitos, análisis y diseño. El resto de disciplinas (Implementación, Pruebas, Despliegue, Gestión de Configuración, Gestión y Entorno) coinciden con las restantes de RUP (Flores, 2016).

Las fases de AUP se presenta en la figura 1 (EcuRed, 2016):

- **Iniciación:** en esta fase se identifica el alcance y dimensión del proyecto, la propuesta de la arquitectura de desarrollo de SW y el presupuesto asignado.
- **Elaboración:** se identifica y valida la arquitectura de desarrollo de SW.
- **Construcción:** se desarrolla el SW en forma incremental con base en las prioridades funcionales de los participantes.
- **Transición:** se realiza la validación e implantación de un Sistema de Información.

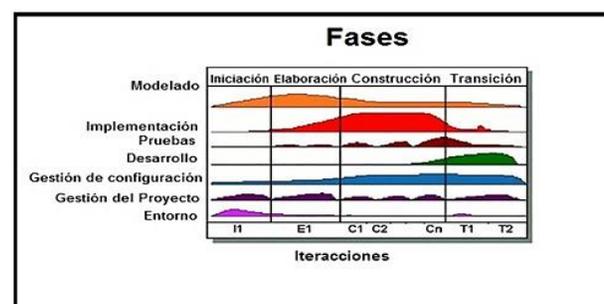


Figura 1 Fases de AUP

Fuente: EcuRed, 2016

En AUP se describen cuatro flujos de trabajo de ingeniería de SW son: modelado, implementación, pruebas y despliegue. Además incluye tres flujos de trabajo: gestión de configuración, entorno y gestión de proyectos que sirven de apoyo en el diseño del sistema (EcuRed, 2016).

Esta metodología trabaja en conjunto con el Lenguaje de Modelado Unificado UML por sus siglas en inglés (Unified Modeling Language) donde se realizan los diagramas que sirven de base para el desarrollo del sistema web (Rational Software, 2011; Uptodown, s.f.).

UML es un lenguaje de modelado visual no es un lenguaje de programación, se ha adoptado como un estándar internacional, que permite visualizar, especificar, construir y documentar artefactos de un desarrollo de SW que puede ser un Sistema de Información o plataforma WEB, etc., donde se captura decisiones y conocimiento sobre los sistemas que se deben construir; permite entender, diseñar, hojear, configurar, mantener, controlar la información sobre los sistemas orientados a objetos e incorporar mejores prácticas en el entorno de desarrollo de SW (Rumbaugh, Jacobson & Booch, 2000; Fernández & García, 2018).

UML incluye trece diagramas para el modelado, previo al diseño y desarrollo del SW, se clasifican en dos categorías: diagramas de estructura y de comportamiento (Quirón, 2005). *Los diagramas de estructura* representan elementos de un sistema o función y reflejan las relaciones estáticas entre ellas.

- Clases: muestra una colección de elementos de modelado declarativo (estáticos), tales como: clases, asociación, generalización, dependencia, realización e interfaz representando una vista estática y gestión del modelo.
- El diagrama de paquetes presenta cómo se organizan los elementos de modelado en paquetes y las dependencias entre ellos, incluyendo importaciones y extensiones de paquetes y componentes, presentando la vista de Implementación de SW.
- Componentes: representa los componentes de una aplicación, sistema o empresa. Los componentes representan la relación, interacción e interfaces públicas vistas en la implementación.

- Despliegue: muestra cómo y dónde se desplegará el sistema. Los equipos físicos y los procesadores se representan como nodos y la construcción interna puede ser representada por nodos o artefactos embebidos. Como los artefactos se ubican en los nodos para modelar el despliegue del sistema, la ubicación es guiada por el uso de las especificaciones de despliegue.
- Objetos: presenta los objetos y sus relaciones respecto al tiempo. Se considera un caso especial de un diagrama de clases o un diagrama de comunicaciones con una vista estática y gestión del modelo.
- Estructuras de composición: clasificador (tal como una clase, un componente o un caso de uso), incluyendo los puntos de interacción de clasificador con otras partes del sistema vista en la implementación.

Los diagramas de comportamiento representan las características de comportamiento de un sistema o proceso de negocio (Quirón, 2005). Los diagramas UML más usuales son: el de actividades, de estados, secuencia, casos de uso, comunicaciones, etc. Existen herramientas de SW con licencia comercial y open source para el diseño del modelado UML, entre las más usuales son: UML Diagrammer, Microsoft Visio, Visible Analyst, Visual Paradigm, Umbrello UML Modeller, Visual UML, Rational Rose, Argo UML, Star UML, TinyUML, Frame UML (Rumbaugh, Jacobson & Booch, 2000). Siendo más utilizadas las open source: Argo UML (Uptodown, s.f.), Dia, Frame UML, entre otras.

Arquitectura MVC para desarrollo de SW

La arquitectura MVC permite definir un patrón de arquitectura de desarrollo de SW, que propone separar el código de los programas por sus diferentes responsabilidades, es utilizada para implementar sistemas que requiere el uso de interfaces de usuario, crear SW más robusto con un ciclo de vida más eficaz y que potencialice la facilidad de mantenimiento, reutilización del código y la separación de conceptos como se muestran en la figura 2 (desarrolloweb, 2019).

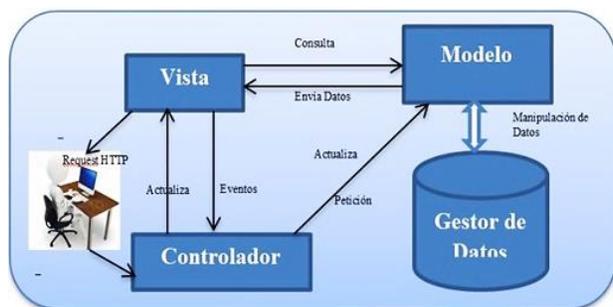


Figura 2 Modelo MVC

Fuente: desarrolloweb, 2019

Los tres componentes de esta arquitectura MVC, descritos en desarrolloweb (2019) son:

- El Modelo: es la capa donde se trabaja con los datos, por tanto contendrá mecanismos para acceder a la información y también para actualizar su estado. Los datos están almacenados en una base de datos (BD), por lo que los modelos deben tener todas las funciones que accederán a las tablas en cuanto a la ejecución de select, update, insert o en general de las transacciones. Además envía a las vistas la información que solicita el usuario y las peticiones de acceso o manipulación de la información llegan a través del controlador.
- El Controlador: sirve de enlace entre la vista y el modelo de acuerdo a las requerimiento de desarrollo de SW, ya que contiene el código necesario para responder a las acciones que solicita la aplicación como por ejemplo, visualizar un elemento, búsquedas de información, realizar una compra, entre otras.
- La Vista contiene el código de la aplicación o sistema de información que presentan la interfaz del usuario, por lo tanto requiere de la información que le envía el modelo para presentar como salida (desarrolloweb, 2019).

Minería de Datos (DM)

El KDD es un proceso no trivial de identificar patrones válidos, novedosos, potencialmente útiles y en última instancia comprensibles a partir de los datos. La diferencia es que busca el descubrimiento del conocimiento sin una hipótesis predefinida o preconcebida con respecto a sistemas tradicionales de explotación de datos basados en la existencia de una hipótesis o modelos previos.

Define una metodología que provee una representación completa del ciclo de vida de un proyecto DM (Formia, 2013). El proceso de KDD se divide en cinco fases principales que se presentan en la figura 3, descritas por Hernández, Ramírez & Ferri (2004) citado por Galán (2015).

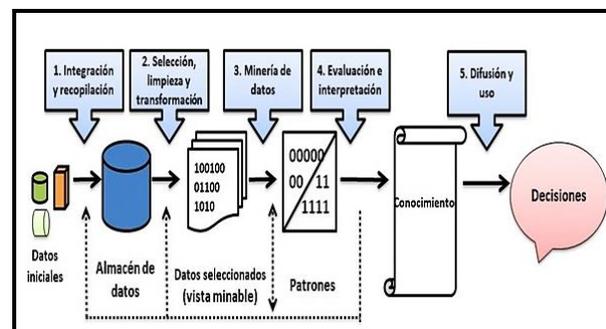


Figura 3 Fases Proceso KDD

Fuente: elaboración propia a partir de Galán, 2015

Se observa en la figura 3, que la minería de datos es la tercera fase de KDD y representa un área de las TI donde integra una variedad de técnicas de análisis de datos y extracción de modelos, capaz de obtener patrones, describir tendencias y regularidades, predecir comportamientos y explorar información computarizada proveniente de fuentes heterogéneas y grandes volúmenes, permitiendo a los individuos y a las organizaciones comprender y modelar de una manera más eficiente y precisa el contexto en que deben actuar y tomar decisiones inteligentes (Galán, 2015). La DM surge como una tecnología que intenta ayudar a comprender el contenido de una base de datos, donde los datos representan la materia prima bruta, convirtiéndose en información cuando el usuario le atribuye algún significado en especial, es decir extraer conocimiento que coadyuve a la toma de decisiones inteligentes (Molina & García, 2012).

Desarrollo del Sistema Web

Esta sección presenta el desarrollo del primer prototipo del sistema Web para el pre procesamiento de datos con Minería de Datos (DM), generando una vista minable, requisito previo al modelado con DM.

Metodología de Investigación

El desarrollo de un sistema web, se inicia con una investigación exploratoria y explicativa como lo menciona Hernández, Fernández & Baptista (2010), con el propósito de identificar las necesidades, requerimientos en la gestión de trayectorias educativas y formas de recopilar y analizar datos de los estudiantes universitarios; así como el entorno de desarrollo guiado con la ingeniería de SW, que garantice la entrega de un sistema web de calidad. En este sentido se establece un plan estratégico para el desarrollo e implementación del sistema web que se presenta en este trabajo. En donde se definen:

- Técnicas de recopilación de datos escolares de fuentes disponibles.
- Selección de la muestra del PE de TIC para el desarrollo del primer prototipo del Sistema.
- AUP como metodología de desarrollo de SW.
- Herramientas de SW a utilizar y equipo de desarrollo del prototipo del sistema web.
- Importación de archivos (.csv, xls, txt) para la recopilación de datos en el sistema.
- Validación de datos útiles para un análisis y seguimiento de trayectorias escolar del estudiante con herramientas externas al sistema, debido a que se desarrollará en una segunda etapa, el generar a través del sistema el análisis y seguimiento de trayectorias de estudiantes de manera automatizada, siendo otro valor agregado de este proyecto.
- Validación del pre procesamiento de datos del primer prototipo diseñado, de la muestra seleccionada.

Metodología de desarrollo de sistema del sistema web

Del análisis exploratorio y comparativo de los modelos tradicionales de desarrollo de SW, las metodologías pesadas RUP y el desarrollo ágil de SW, descritas en la tabla 1, se determina que el diseño y desarrollo del sistema WEB, se realice a través de la arquitectura del MVC y Proceso Unificado Ágil (AUP), debido que es una versión simplificada de RUP, el modelado UML con el SW ArgoUML donde se diseñan los diagramas que sirven de base el desarrollo del sistema web del pre procesamiento de datos con DM (Rational Software, 2011).

Además de considerar la naturaleza y tamaño del proyecto que se presenta en este trabajo, se inicia el diseño del primer prototipo del sistema.

Señalando que AUP tiene las mismas cuatro fases que RUP con la diferencia en que no está tan centrado en la documentación, reteniendo parte de la formalidad y combina técnicas ágiles con el propósito de mejorar la productividad. La gestión resulta sencilla, aun cuando algunos desarrolladores no lo utilizan, debido al gran número de artefactos que se requiere (Torrecilla, 2012). Las fases desarrolladas en el prototipo del sistema web se describen a continuación.

Iniciación

En esta etapa se inicia con las siguientes actividades:

- Se define el alcance del proyecto respecto al desarrollo tecnológico que incluye el diseño de una interfaz del sistema web y la programación del pre procesado DM previo al modelado.
- Se diseña el plan de desarrollo basado en AUP, identificando los recursos tecnológicos Hardware (HW) y SW en el desarrollo e implementación, el tiempo asignado y la capacidad en el manejo de las herramientas tecnológicas que se utilizan.
- Se definen los Riesgos: tiempo asignado, el nivel de conocimiento en las herramientas tecnológicas viables, disponibilidad y acceso a la información escolar para la recopilación e integración de datos.

- Preparación del entorno del desarrollo basado en el análisis de metodologías descritas en la Tabla 1 y patrones de diseño de SW más utilizados, flexibles y dinámicos.

Elaboración

En esta etapa se diseña un plan de desarrollo de proyecto que incluye las siguientes actividades:

- Identificar y validar la arquitectura de desarrollo MVC y metodología ágil AUP (Cendejas, 2014; desarrolloweb, 2019; Gil, 2008).
- Se diseña un plan de desarrollo que indica las principales actividades y el tiempo estimado para dar cumplimiento a los objetivos planteados.
- La ejecución del entorno del proyecto se guía con el plan de desarrollo.
- Se determina que el equipo de desarrollo integrado por profesores de TIC y en la implementación se realiza, con la participación de personal académico de la UTVT y UTSOE.
- Se realiza el modelado UML previo al desarrollo del sistema WEB, utilizando el SW de argoUML (Rational Software, 2011; Rumbaugh, Jacobson & Booch, 2000). D) donde se desarrollaron los diagramas de:
 - Despliegue: consiste en modelar los distintos componentes físicos utilizados en la implementación y la relación entre ellos.
 - Paquetes: muestran los diferentes subsistemas lógicos en que se divide el sistema y las dependencias que existen entre ellos, basado en MVC.
 - Secuencia: muestra la evolución en el tiempo de mensajes que intercambian los diferentes actores y objetos en un determinado escenario que se presentan en el sistema web.
 - Casos de Uso: describe la relación o interacción entre los actores que participan en el procesamiento de DM y el sistema.

- Diseño de la base de datos con MySql y rutinas de control necesarias para la integración de datos previos al pre procesamiento DM en el sistema.

Construcción o Desarrollo

Después del modelado y diseño de la base datos del sistema se realizan las siguientes actividades:

- Selección de herramientas open source para el desarrollo del sistema basado en el MVC (desarrolloweb, 2019), a través de una programación orientada a objetos (POO) en PHP puro, integrando procesos, rutinas y herramientas de otras tecnologías open source como Laravel, Composer, Ajax, Jquery, Json, Bootstrap, HTML5, CSS3, Python y NuSoap (Maya, Herrera y Decena, 2017; Styde, s.f.; Taylor, 2019 & Oracle MySQL, 2019), en un entorno flexible que integra tecnologías de código abierto, que en conjunto permita automatizar el proceso de recopilación de datos (importación o extracción) y pre procesamiento de datos de estudiantes de TSU TIC, previo al modelado DM.

Adicionalmente se utiliza el gestor de base de datos de MySql que auxilia el intercambio de información entre los diferentes procesos requeridos en el sistema, que permite almacenar la información clave y con acceso desde un sitio web publicado en su primera versión (Maya, Herrera y Decena, 2017; Styde, s.f.; Taylor, 2019 & Oracle MySQL, 2019).

- Programación de las interfaces iniciales y los módulos del proceso DM que se integran en el sistema web.

Transición

En esta etapa final se realiza lo siguiente:

- Validación del diseño del sistema web y Configuración del sitio web para la publicación del sistema web.

En este sentido se logra el diseño del primer prototipo del sistema publicado en un sitio web que incluye:

- La integración de datos de fuentes externas, extraídos de otros Sistemas de Información que maneja actualmente la UTVT.

- Realizar el pre procesamiento de DM que incluye selección y limpieza, transformación y clasificación de datos, que se preparan antes de aplicar un modelo DM o de IA: cuyas actividades realizadas son: el diseño, la programación y ejecución de pruebas.

En la ejecución de pruebas de funcionalidad y usabilidad, así como los cambios requeridos se realizaron de manera iterativa, donde permitió revisar el funcionamiento de los procesos del sistema, con la finalidad de conocer la eficiencia y tiempo de respuesta, a través de las siguientes actividades:

- Ejecutar consultas de datos integrados para conocer el tiempo de respuesta.
- Analizar los resultados emitidos por la plataforma WEB para sí saber si son correctos.
- Revisar los procesos después de su ejecución.
- Comprobar el buen funcionamiento con diferentes velocidades de la red de datos y equipos de cómputo.

Una vez terminado estas pruebas, se valida el sistema WEB, continuando con una prueba piloto con datos de una muestra seleccionada del PE de TSU de TIC-UTVT.

Resultados

En esta sección se presentan los resultados del sistema web, en su primer prototipo, orientado al pre procesamiento de DM implementado con datos de una muestra seleccionada de estudiantes de la Universidad en estudio, así como la evaluación de usabilidad, funcionalidad, fiabilidad, eficiencia del sistema, una vez publicada en el sitio web <http://proyectoticsmn.com/cvaati01/public/index.php>.

El prototipo del sistema web, presenta en primera instancia una interfaz de inicio de sesión descrita en la figura 4, donde se solicita la autenticación del usuario para permitir o denegar el acceso como un mecanismo de seguridad para el manejo de la información implementada en la UTVT y una vez validado el nivel de acceso, se mostrará los módulos que le corresponde:

Definiendo roles en cuatro niveles de acceso como: administrador, analista de datos, docente e usuario invitado.



Figura 4 Inicio de Sesión del sistema web

Fuente: elaboración propia

Cuando el usuario ya tiene su acceso se le presentará la interfaz principal descrita en la figura 5, que incluye: Integración, Pre Procesamiento DM, Catálogos, donde se realizan primero las actividades previas al proceso de DM para el modelado y emisión de reportes.



Figura 5 Interfaz Inicial del sistema web

Fuente: elaboración propia

El proceso previo de DM se inicia en el módulo de Integración, en la que se recopila los datos fuentes disponibles, a través de archivos de tipo CSV, TXT y SQL que se generan del Sistema de Automatización de Información Integral de estudiantes de TSU en TIC-UTVT y que son cargados al sistema web, como se muestra en la figura 6.



Figura 6 Interfaz de Integración de Datos al sistema web
Fuente: elaboración propia

Posteriormente se solicita la ruta del archivo a subir en el sistema web y al concluir se envía un mensaje de carga correcta, como se muestra en la figura 7.

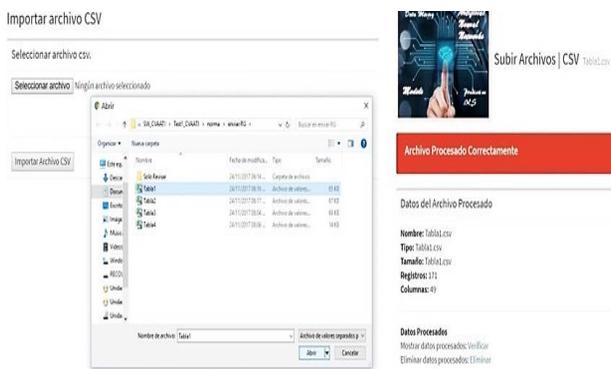


Figura 7 Interfaz de Importación de archivos al sistema web
Fuente: elaboración propia

Continuando con el módulo de pre procesamiento DM descrito en la figura 8.



Figura 8 Interfaz de Pre procesamiento DM en el sistema web
Fuente: elaboración propia

En la figura 8 se observa que se realizan los siguientes procesos previos del modelado con DM.

- Selección y limpieza: aquí se revisa en el archivo de datos importados en el sistema web, que no existan datos nulos o erróneos, en donde se puede modificar o eliminar las variables que no desea que participen en el entrenamiento del modelado DM.
- Transformación de Datos: donde se preparan los datos que representan las variables relevantes seleccionadas para el modelado DM, en la que primero se define en el módulo de catálogos el nombre de las variables como se muestra en la figura 9.



Figura 9 Interfaz de Catálogos definición de variables en el sistema web
Fuente: elaboración propia

Adicionalmente se asignan los valores categóricos discretos de las variables definidas en el módulo de catálogos y parametrización, como se muestra en la figura 10.



Figura 10 Interfaz de Catálogos definición de Parametrización de variables en el sistema web
Fuente: elaboración propia

Una vez definido el nombre y parametrización de variables, se inicia el pre procesamiento de DM, donde se transforman los datos en variables. En esta sección se pueden realizar tres actividades

- *Asignar nueva variable*, en la que se selecciona el id, descripción de la variable en estudio y los aspectos que se desean incluir en la misma, mostrado en la figura 11.



Figura 11 Interfaz de Asignación de nuevas variables en el sistema
 Fuente: elaboración propia

- Interfaz de Mostrar las variables asignadas presentadas en la figura 12, en caso de no estar correctamente seleccionadas las variables, se tiene la opción de eliminar y volver a asignar de forma correcta.

Mostrar Asignación de Variables Pre



Figura 12 Interfaz de Mostrar variables en el sistema web
 Fuente: elaboración propia

- La interfaz de Eliminar Variable que visualiza la figura 13, elimina todas las variables que ya se habían seleccionado, por lo que debe tener especial cuidado es realizar esta acción.

Eliminar Asignación de Variables Pre Procesado DM



Figura 13 Interfaz de Eliminar variables en el sistema web
 Fuente: elaboración propia

- *Parametrización*: aquí se asignan los valores categóricos discretos de las variables definidas en la parametrización de catálogos como se muestra en la figura 14, que serán utilizados en el procesamiento del modelado de DM que se proponga desarrollar. Esta sección incluye dos subprocesos que son: iniciar el proceso de parametrización y mostrar datos parametrizados, como se describe en la figura 14 y 15.



Figura 14 Interfaz de Catálogo de parametrización de variables
 Fuente: elaboración propia

Mostrar Datos Parametrizados csv



Figura 15 Subprocesos de parametrización de variables en el sistema web
 Fuente: elaboración propia

En lo que respecta a la interfaz de modelado solo se exporta el archivo generado con la vista minable que son la fuente de patrones de entrenamiento, para posteriormente realizar el Modelado DM en el SW de Matlab y WEKA sobre trayectorias educativas que impactan significativamente en la deserción estudiantil como parte de la investigación. Dejando como trabajo futuro implementar en su totalidad los procesos de DM y el modelado de trayectorias educativas para los programas educativos de TIC de UTVT y de ser posible de UTSOE.

Agradecimiento

Agradecemos el apoyo recibido de la Universidad Tecnológica del Valle de Toluca (UTVT) y la colaboración de profesores de la Universidad Tecnológica del Suroeste de Guanajuato (UTSOE) que hicieron posible el desarrollo de este proyecto.

Conclusiones y perspectivas futuras

El trabajo presenta una revisión literaria de DM, metodologías de la Ingeniería de SW, la arquitectura de desarrollo en la que se basa el diseño del prototipo del sistema web para el pre procesamiento de datos con DM, donde se argumentó que es una tarea complicada del desarrollador de SW, en seleccionar cual era la mejor herramienta metodológica y tecnológica que se adaptaba al desarrollo en un entorno particular y personalizado para cumplir el objetivo de este trabajo, es decir cuál es el traje a la medida, por lo que después del análisis exploratorio de las herramientas de SW de DM entre las más citadas Matlab, IBM SPSS Statistics, SAS Enterprise Miner, Rapid Miner, Oracle Data Mining, WEKA, KNIME, lenguaje R, etc. se observó que la mayoría tiene un costo de licenciamiento, además del entorno de desarrollo y solo se adapta en forma parcial, algunos procesos de DM de la investigación de este trabajo.

Determinando el desarrollo del sistema en un entorno WEB personalizado y adaptando estrategias tecnológicas y herramientas web, la metodología de desarrollo AUP y el MVC que ofrece como ventaja: separar la lógica del negocio de la lógica del diseño, obtener sistemas de gran calidad en poco tiempo, facilidad en la realización de pruebas unitarias de cada componente desarrollado y prototipos en menor tiempo, reutilización de los y desarrollos altamente escalables.

Concluyendo que el sistema web en su primera versión, se logra obtener resultados satisfactorios en su primera versión publicada para el pre procesamiento de datos con DM en TIC-UTVT y con base en los objetivos planteados, se cumple con la integración de datos, el pre procesamiento generando la vista minable y extracción de datos DM, que son la base para continuar en una segunda etapa del proyecto, donde se pretende integrar el desarrollo del modelo DM del seguimiento de trayectorias educativas, que impactan en gran parte en el fenómeno de deserción y reprobación estudiantil; así como la emisión automatizada de reportes de alertas de estudiantes vulnerables identificados y enviar a los tutores o personal académico asignado, para dar seguimiento oportuno. El sistema web pretende convertir a un escenario dinámico e innovador a través de DM, en la que representa una ventaja competitiva y adaptativa al generar conocimiento y coadyuvar en la toma de decisiones.

Referencias

Cendejas, J. (2014). "Implementación del modelo integral colaborativo (MDSIC) como fuente de innovación para el desarrollo ágil de software en las empresas de la zona centro - occidente en México". (*Tesis Doctoral. Centro Interdisciplinario de Posgrados, Investigación y Consultoría. UPAEP*). México.

Desarrolloweb. (2019). Desarrollo web APACHE CTL Obtenido de: <http://www.desarrolloweb.com/articulos/que-es-mvc.html>.

EcuRed. (2016). Enciclopedia EcuRed Conocimiento con todos y para todos, Obtenido de: <https://www.ecured.cu/>

Fernández, J. & García, I. (2018). RICA O Recuperación e Integración de Contenidos Audiovisuales empleando Ontologías. (*Tesis Magister: Facultad de Informática. Universidad de Murcia*). pp. 18-19. España.

Flores, E. (2011). Ingeniería de Software Mex. TL. Obtenido de <http://ingenieriadesoftware.mex.tl>.

- Formia, S. (2013). La deserción en cursos universitarios. Construcción de modelos sobre datos de la UNRN usando técnicas de Extracción de Conocimiento. (*Tesis Magister. Facultad de Informática. UNLP*). Argentina.
- Galán, V. (2015). “Aplicación de la Metodología CRISP-DM a un proyecto de Minería de Datos en el entorno Universitario”. (*Tesis de grado, Escuela Politécnica Superior. Ingeniería en Informática, Universidad Carlos III de Madrid*). Obtenido de http://e-archivo.uc3m.es/bitstream/handle/10016/22198/PFC_Victor_Galan_Cortina.pdf?sequence=1. Madrid.
- Gil C. (2008). RUP: Metodología en los sistemas y aplicaciones basadas en la Web. Avances investigación en Ingeniería. Centro de Investigación y Facultad de Ingeniería de la Universidad Libre, 5(8).
- Hernández, R., Fernández, C., & Baptista, M. (2010). Metodología de la investigación (5 ed.). México: Mc Graw-Hill.
- Hernández, J., Ramírez, M., & Ferri, C. (2004). Introducción a la Minería de Datos. Pearson.
- Kendall K. & Kendall J. (2011). Análisis y diseño de Sistemas. México: 5ª Edición. Pearson Educación.
- Maya P. N., Herrera C. J. & Decena G.E, (2017), “Aplicando reingeniería al sistema web de evaluación docente”. Ecorfan. Educación Técnica, Perú. 1(2).pp. 31-43, ISSN: ISSN 2523-2460.
- Molina, J., & García, J. (2012). Técnicas de Análisis de Datos, Aplicaciones prácticas utilizando Microsoft Excel y Weka. *Universidad Carlos III de Madrid*. Madrid.
- Oracle MySQL. (2019). Source Databases MySQL, Obtenido de: <https://www.mysql.com/>
- Pressman, R. (2006). *Ingeniería de software. Un enfoque práctico*. España. McGraw.Hill.
- Quirón, A. (2005). *Epidata: Introducción a UML 2.0*. Obtenido de http://www.epidataconsulting.com/tikiwiki/tiki-read_article.php?articleId=15
- Rational Software Development Company. (2011). Rational Unified Process: Best Practices for Software development Teams Recuperado de https://www.ibm.com/developerworks/rational/library/content/03July/1000/1251/1251_bestpractices_TP026B.pdf.
- Rumbaugh, J., Jacobson, I., & Booch, G. (2000). *El Lenguaje Unificado de Modelado UML. Manual de Referencia*. Madrid: Pearson Educacion.
- Styde. (s.f.). Desarrollo Web, Ajax y Laravel usando jQuery, Json. Obtenido de: <https://styde.net>.
- Someerville, I. (2005). *Ingeniería del software* (Séptima ed.). Madrid, España: Pearson.
- Taylor Otwell. (2011-2019). The Laravel Ecosystem. Obtenido de: <https://laravel.com/>.
- Torrecilla, P. (2012). nosolopau.com. Obtenido de: <http://nosolopau.com/2012/06/07/mas-sobre-el-proceso-unificado-agil-fases-y-disciplinas>.
- Uptodown.(s.f.). ArgoUML. Obtenido de: <https://argouml.uptodown.com/windows>