

DOI: <https://doi.org/10.56712/latam.v4i1.237>

## Análisis de Datos Académicos

### Academic Data Analysis

**Rosa de Guadalupe Cano Anguiano**

Depto. De Sistemas y Computación, Tecnológico Nacional de México/I.T. Colima.  
rcano@colima.tecnm.mx  
<https://orcid.org/0000-0001-7787-9058>  
Villa de Álvarez, México

**Alejandra Yunuen Zuñiga Melchor**

Estudiante de Ingeniería en Sistemas Computacionales, Depto. De Sistemas y Computación, Tecnológico Nacional de México/I.T. Colima.  
y.zunigam@hotmail.com  
Villa de Álvarez, México

**María Elena Martínez Durán**

Depto. De Sistemas y Computación, Tecnológico Nacional de México/I.T. Colima.  
mmartinez@colima.tecnm.mx  
<https://orcid.org/0000-0003-1615-5130>  
Villa de Álvarez, México

**Ariel Lira Obando**

Depto. De Sistemas y Computación, Tecnológico Nacional de México/I.T. Colima.  
alira@colima.tecnm.mx  
<https://orcid.org/0000-0002-6703-3697>  
Villa de Álvarez, México

**Lindalva Ponce Ibarra**

Depto. De Sistemas y Computación, Tecnológico Nacional de México/I.T. Colima.  
lindalva.ponce@colima.tecnm.mx  
<https://orcid.org/0000-0002-7495-5200>  
Villa de Álvarez, México

Artículo recibido: día 30 de noviembre de 2022. Aceptado para publicación: 18 de enero de 2023.  
Conflictos de Interés: Ninguno que declarar.

Todo el contenido de LATAM Revista Latinoamericana de Ciencias Sociales y Humanidades, publicados en este sitio está disponibles bajo Licencia Creative Commons . 

Como citar: Cano Anguiano , R. de G., Zuñiga Melchor, A. Y., Martínez Durán , M. E., Lira Obando, A., & Ponce Ibarra , L. (2023). Análisis de Datos Académicos. *LATAM Revista Latinoamericana de Ciencias Sociales y Humanidades* 4(1), 190–200. <https://doi.org/10.56712/latam.v4i1.237>

## Resumen

En el presente artículo se presenta el diseño de una plataforma web responsiva para el análisis de datos, la motivación para el desarrollo de este proyecto es poder aprovechar los datos que son almacenados por el Departamento de Sistemas y Computación del Instituto Tecnológico Nacional de México Campus Colima de todos los semestres en diferentes formatos que terminan en un folder dentro de un archivero. Durante las pruebas, esta aplicación dio resultados favorables para el procesamiento de datos y creación de consultas dinámicas, permitiendo que en un futuro se puedan agregar nuevos módulos a la plataforma dependiendo de las necesidades del momento, entre ellas, la aplicación de minería de datos para la búsqueda de patrones.

*Palabras clave:* análisis de datos académicos, educación, tecnológico de Colima, consultas dinámicas, ETL

## Abstract

This article presents the design of a responsive web platform for data analysis, the motivation for the development of this project is to take advantage of the data that is stored in the Systems and Computing Department of the National Technological Institute of Mexico Campus Colima all semesters in different formats and ending in a folder within an archivist. During the tests, this application gave favorable results for data processing and creation of dynamic queries, allowing new modules to be added to the platform in the future depending on the needs of the moment, including the application of data mining for The search for patterns

*Keywords:* academic data analysis, education, tecnológico de Colima, dynamic consultations, ETL

## **INTRODUCCIÓN**

El primer paso para extraer información es recolectar los datos. Para proceder al análisis se tienen que tomar en cuenta los KPI (indicadores) que mejor satisfagan a los objetivos establecidos en un principio; este análisis permite examinar los datos en crudo para intentar llegar a una solución con base a la información resultante.

La minería de datos tiene como finalidad prevenir a los dirigentes de las empresas o instituciones sobre situaciones interesantes, anomalías o peligros que no se pueden detectar a simple vista (Martínez Luna, 2011). La minería de datos es un área multidisciplinar, por lo que no es de extrañar que se aplique al área de la enseñanza.

La minería de datos aplicada a la educación, es una disciplina emergente que busca desarrollar nuevos métodos para explorar la información que se genera dentro de los ambientes educativos con el fin de entender la forma en que los estudiantes aprenden (Rosado Gómez & Verjel Ibáñez, 2017), y así, tomar las decisiones que mejor se acoplen en la búsqueda de mejorar el proceso educativo.

Entre los análisis que se pueden llevar a cabo en dichas instituciones están el predecir el desempeño de los estudiantes, crear modelos de estudiantes, comparar el desempeño de los profesores, encontrar patrones en la deserción escolar, entre otros. Es por ello que en las instituciones educativas se llevan a cabo procesos ya sean manuales o automáticos para el concentrado de datos y posterior procesamiento.

Uno de los problemas más recurrentes, no solo en las instituciones educativas sino en cualquier empresa u organización, es la forma en que los datos son almacenados, pues es muy común que se guarden en diferentes gestores, con diferente formato, o que los datos guardados difieran entre semestres.

El caso particular que se analizó durante el desarrollo de este proyecto fue el concentrado de información que el departamento de Sistemas y Computación del Instituto Tecnológico Nacional de México Campus Colima lleva a cabo todos los semestres en hojas de cálculo, las cuales se van llenando de manera manual a lo largo de los seis meses con las entregas de evaluación que los docentes realizan al jefe del departamento.

Dado que el vaciado de información es completamente manual, existe la posibilidad de que ocurran los famosos "errores de dedo" que por falta de tiempo no son corregidos la mayor parte de las veces, también cabe señalar que las gráficas que son generadas en los archivos se hacen mediante referencia de datos, por lo que, con un solo cambio de dato o movimiento afectará la visualización del gráfico o la integridad de la información resultante.

Como una posible solución le fue propuesto al Departamento de Sistemas y Computación el diseño de un formato para las fuentes de datos y Data Warehouse, además del desarrollo de una plataforma web responsiva que permita la lectura de dichas fuentes y el análisis de los datos recabados, dando así oportunidad a los jefes de docencia de realizar consultas dinámicas.

## **MATERIAL Y MÉTODOS**

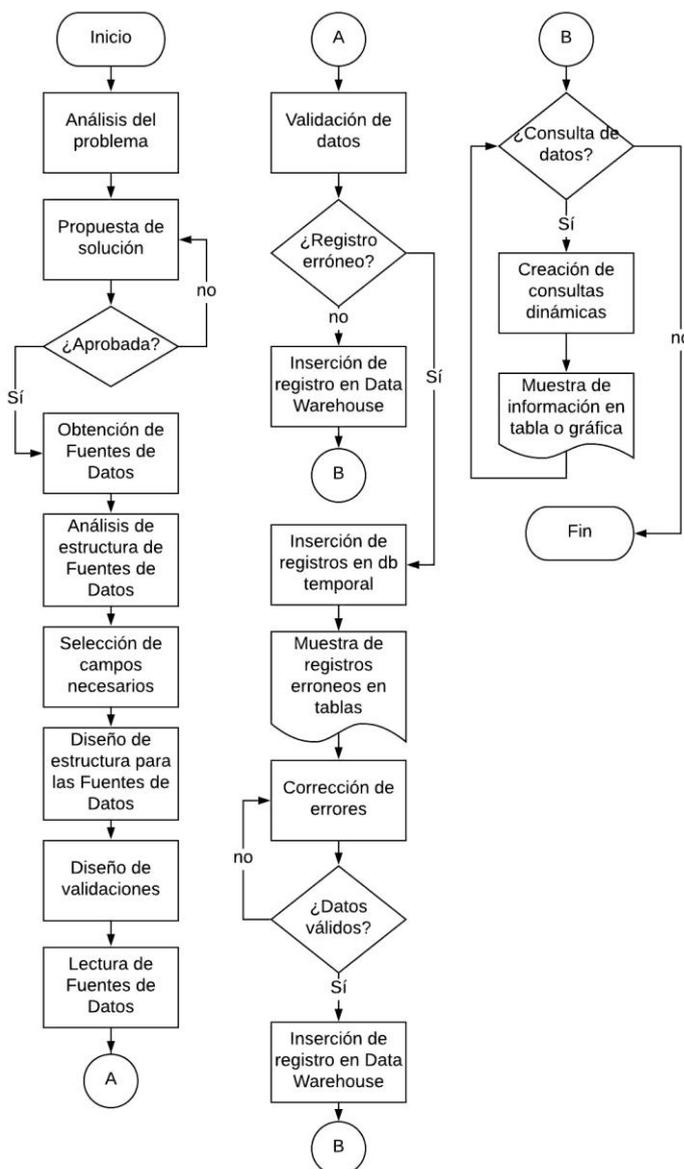
Para el desarrollo del proyecto se consultó una amplia variedad de herramientas, aplicaciones y librerías, entre las cuales se seleccionaron aquellas de código abierto, que cuentan con licencia gratuita o que son utilizadas por el Departamento de Sistemas y Computación. Como fuente de datos se utilizan hojas de cálculo de Excel, en la cual, la primera hoja es la que contendrá la información que será extraída por la plataforma.

La plataforma web responsiva se desarrolló con HTML5, CSS, Bootstrap y las librerías DataTable y Chart para la parte de frontend, para el backend se usó PHP versión 7.3.6 con el framework CodeIgniter versión 3.1.10 que trabaja con una arquitectura Modelo-Vista-Controlador (MVC). Como Sistema Gestor de Base de Datos se utilizó MySQL 2.4.39. También fue necesario el uso de la librería PHPEXcel para la lectura de los archivos.

El enfoque de este proyecto es el análisis de datos, por lo que después de realizar una investigación sobre las metodologías más implementadas para esta tarea, KDD fue seleccionada como la que más se ajustaba a las necesidades planteadas en un principio. De acuerdo a Nigro, Xodo, Corti, Terren, KDD cuenta con 7 fases principales: entendimiento del dominio, selección del conjunto de datos, limpieza y preprocesamiento, eliminar redundancias, elegir la herramienta adecuada de minería de datos, interpretar los resultados y la consolidación del conocimiento (Nigro, Xodo, Corti, & Terren, 2004). Ver figura 1.

**Figura 1**

*Diagrama de flujo de las etapas del proyecto basándose en las fases de la metodología KDD*



Debido a que la plataforma manejará información histórica para la realización de consultas dinámicas, se llevó a cabo la elaboración de una base de datos denominada Data Warehouse, en la cual se almacenarán los datos limpios resultantes del proceso ETL. Para el diseño de esta estructura se contemplaron los diferentes formatos existentes en los cuales se almacenaban los datos del departamento. Cabe señalar que al comienzo del proyecto surgieron algunas dificultades, entre las cuales destaca que no fue posible acceder a los datos almacenados en la base de datos de la institución, por lo cual la plataforma se pensó para la lectura de archivos tipo .XLSX.

Después de diseñar la estructura del Data Warehouse, fue necesario elaborar una para los archivos donde se almacena la información, pues como se acaba de mencionar, durante los años 2012 – 2018 que fueron los datos proporcionados por el departamento, los formatos implementados eran diversos, algunos con más campos que otros, manejaban diferente tipo de dato para el mismo campo o contenían varias versiones del mismo archivo.

Una vez las estructuras estuvieron listas, fue el turno de elaborar las validaciones para los campos a utilizar con ayuda de expresiones regulares; estas validaciones fueron pensadas en orden de respetar el contenido que se deseaba guardar, entre algunas de las que se hicieron se encuentran: revisar si el campo contiene solo letras, para el caso de los índices de aprobación y reprobación almacenados, estos debían cumplir con ser una cantidad mayor o igual a cero pero menor a cien cada uno, y entre los dos no sumar más de cien.

Para la lectura del archivo, con ayuda de PHPExcel, se almacenó cada campo a evaluar en una variable, de esta forma se podían aplicar cuantas validaciones fueran necesarias para asegurar la calidad de la información. En algunos campos se hizo la comparación con catálogos elaborados de manera manual con información extraída de los documentos proporcionados para su análisis y de las retículas que se encuentran en la página oficial del Instituto Tecnológico Nacional de México Campus Colima en la sección de oferta académica; estos catálogos son: docentes, materias, aulas y carreras. Ver figura 2

**Figura 2**

*Oferta académica del Instituto Tecnológico Nacional de México Campus Colima*



La realización de la lectura de archivos se hace dentro de la plataforma una vez que el usuario se ha identificado con sus credenciales correspondientes, después, se procede a evaluar cada campo para ver si cumple con la estructura propuesta, de hacerlo, el registro se inserta al Data Warehouse, de lo contrario, se almacena en una base de datos temporal para su próxima corrección.

Los registros erróneos serán mostrados en tablas de acuerdo a su clasificación, permitiendo una agradable navegación entre las mismas para el usuario. Los datos solo serán vaciados al Data Warehouse si se cumplen con las reglas establecidas en un principio.

Ya que se han corregido y cargado los datos, es turno de realizar las consultas dinámicas, para acceder a esta sección no es necesario acabar de editar un documento en ese momento, simplemente con que se tengan registros previos en la base de datos se podrán mostrar las consultas ya sea en forma de gráfica o tabla; para llegar a ese apartado hay que hacer clic en el botón "Continuar editando" que se había mencionado anteriormente y después buscar la opción en el menú lateral.

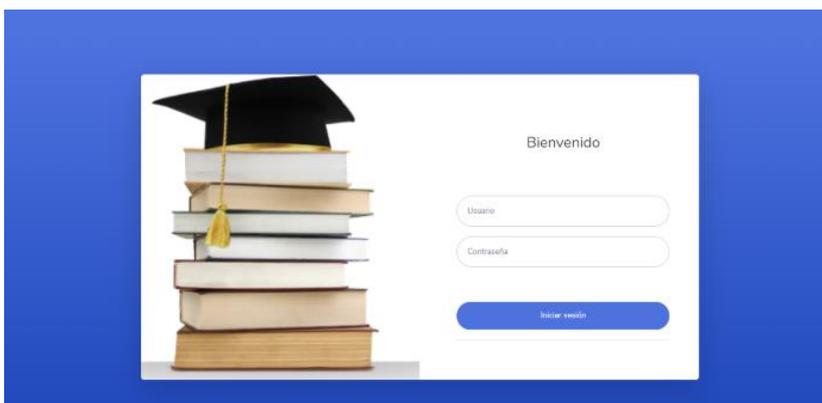
### **RESULTADOS**

Como se explicó al inicio del artículo, la plataforma fue diseñada a partir de estructuras proporcionadas por el Departamento de Sistemas y Computación, por lo que la información extraída por las consultas dinámicas solo arrojará información referente al mismo. Esta plataforma puede considerarse como una primera versión, pues de dársele continuidad en un futuro, es posible que pueda aplicarse a otras carreras dentro del mismo tecnológico, además de agregar los módulos para la minería de datos, pues debido a problemas al comienzo del proyecto el tiempo para desarrollo se vio reducido, dando solo oportunidad a la implementación de consultas dinámicas como primera parte.

Como se mencionó, el usuario solo puede ingresar una vez coloque sus credenciales de manera correcta en el inicio de sesión de la plataforma. Figura 3.

### **Figura 3**

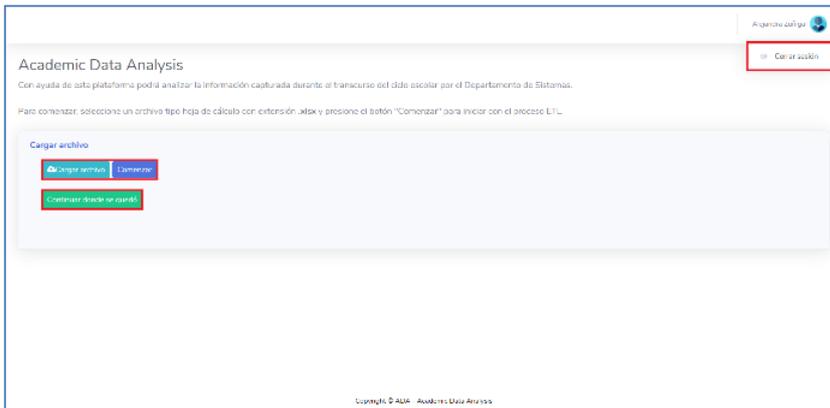
*Inicio de sesión en la plataforma*



En la figura 4 se puede ver la pantalla inicial que se le muestra al usuario, aquí tiene tres opciones, cargar un archivo nuevo, continuar donde se quedó o cerrar sesión, el cual se ubica en la parte superior derecha, al dar clic en el nombre de usuario.

**Figura 4**

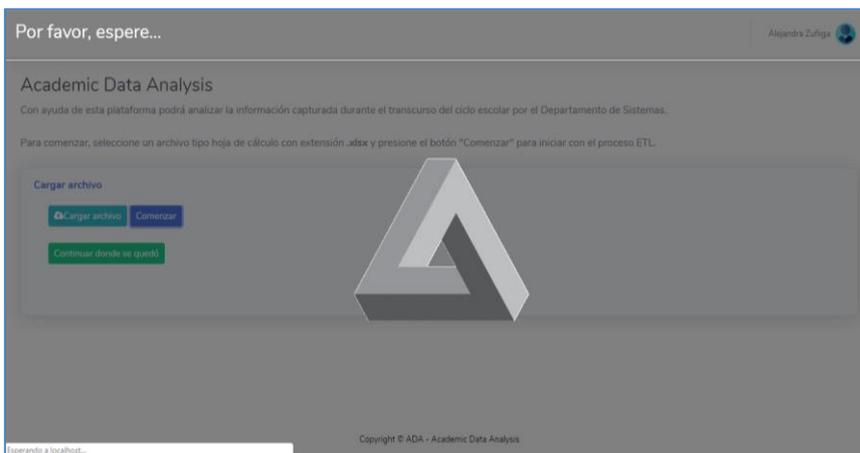
*Pantalla de inicio en la plataforma*



Una vez que se selecciona alguna de las dos opciones que tiene que ver con procesar datos, aparece una animación durante el proceso de carga que bloquea los botones para que el usuario no interrumpa el proceso. Ver figura 5.

**Figura 5**

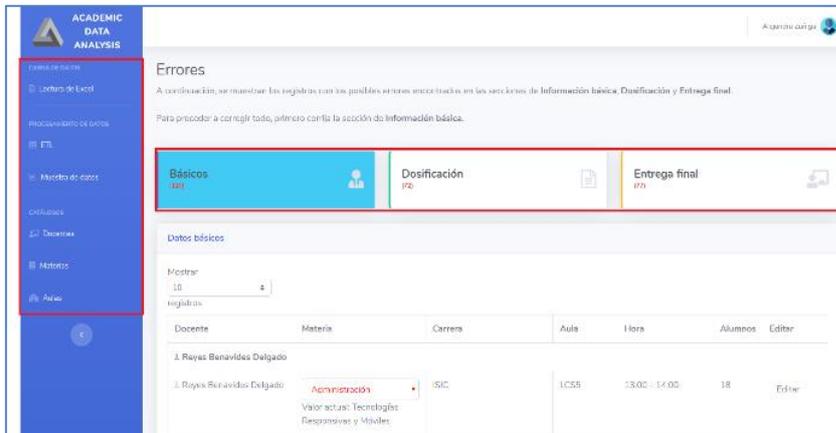
*Pantalla de carga de la plataforma*



En la figura 6 se muestra la vista en la cual puede comenzar a editar información, como se puede ver, en la parte lateral izquierda se encuentra un menú con opciones, en las cuales puede ir a la vista de carga de archivos, corregir errores en el ETL, realizar consultas dinámicas en la muestra de datos y agregar elementos a los catálogos, esto porque, al ser agregados de manera manual y no provenir directamente de la información oficial de la base de datos del tecnológico puede tener registros faltantes. Por otro lado, en los tres grandes bloques señalados se observan los nombres de las tres agrupaciones de datos para una mejor visualización de los errores.

**Figura 6**

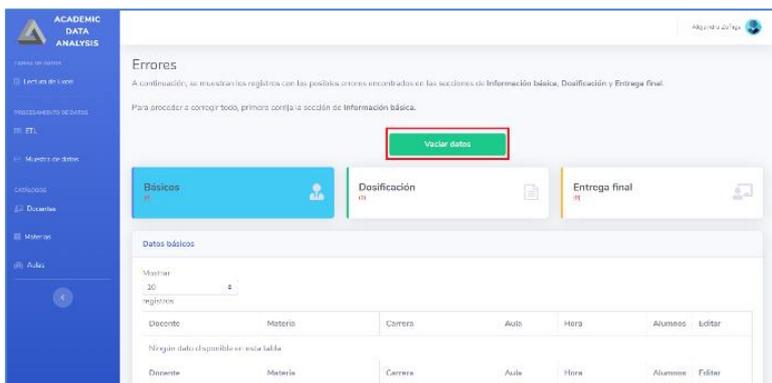
*Tablas para la corrección de errores*



Una vez que se han corregido todos los datos, los contadores de cada tabla indicarán cero, por lo que aparecerá un botón para continuar con el siguiente paso. Ver figura 7.

**Figura 7**

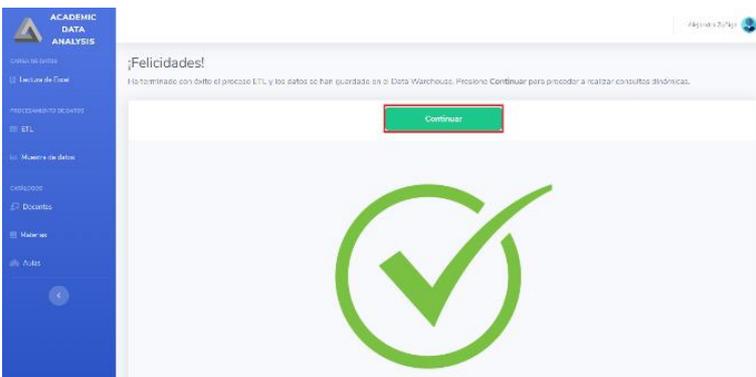
*Errores corregidos*



Ya que se cargaron los datos al Data Warehouse, aparece una vista indicando que puede continuar a las consultas dinámicas. Figura 8.

**Figura 8**

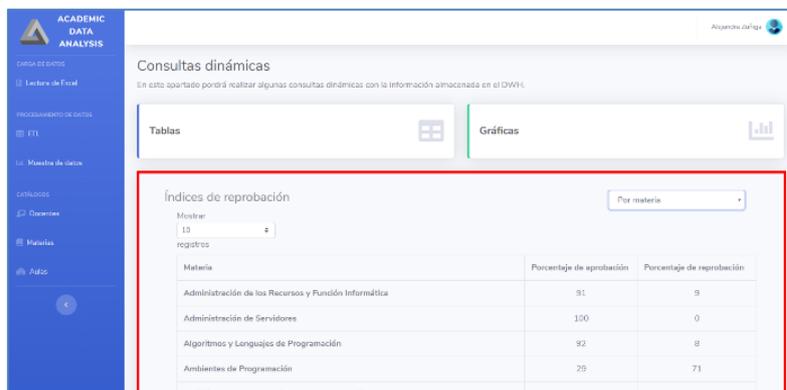
*Datos insertados en el Data Warehouse*



Como se mencionó anteriormente, los resultados pueden ser consultados de dos formas, de manera gráfica o en forma de tabla. A continuación, en las figuras 9 y 10 se muestran las dos formas.

**Figura 9**

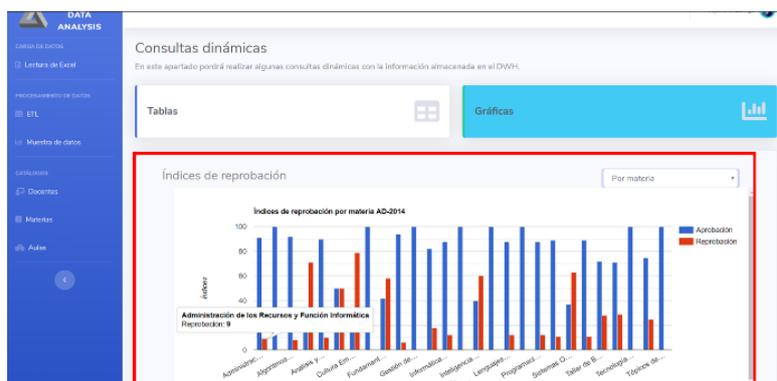
*Muestra de datos en una tabla*



Los datos presentados en las figuras 9 y 10 son de la misma consulta, pero presentado de las dos formas.

**Figura 10**

*Muestra de datos en forma gráfica*



**CONCLUSIONES**

Para el desarrollo de este tipo de plataformas de análisis de datos, se requiere que la institución proporcione vistas con la información de los catálogos, de este modo, se le da un poco más de fiabilidad a los datos al comparar directamente con la fuente.

Como ya se indicó anteriormente, esta es la primera versión de una plataforma que pretende ayudar a los jefes de docencia del Departamento de Sistemas y Computación, de tener buenos resultados en su implementación, podría extenderse a otros departamentos dentro del tecnológico, pero tomando en cuenta que almacenen la información con el formato establecido para el óptimo funcionamiento de la plataforma.

En esta versión para la consulta de datos se manejaron las listas dinámicas, pues el tiempo no ajustó para implementar módulos de minería de datos, y de este modo encontrar patrones de comportamiento en la deserción escolar de los alumnos, como se pretendía hacer.

La selección de implementar dos formas en que se pueda consultar la información surgió porque hay usuarios a los que se les facilitará la visualización a través de gráficos y a otros a través de números plasmados en una tabla.

## REFERENCIAS

Coursera. (15 de Julio de 2019). Obtenido de Programa especializado Análisis de Datos para la toma de decisiones. : <https://es.coursera.org/specializations/analisis-datos>

Huapaya, C. R., Lizarralde, F. A., Arona, G. M., & Massa, S. M. (2012). Minería de Datos Educacional en Ambientes Virtuales de Aprendizaje. XIV Workshop de Investigadores en Ciencias de la Computación, 996-1000.

Martínez Luna, G. L. (2011). Minería de Datos: Cómo hallar una aguja en un pajar. Ciencia - Academia Mexicana de Ciencias, 18-28.

Nigro, H. O., Xodo, D., Corti, G., & Terren, D. (2004). KDD (Knowledge Discovery in Databases): Un proceso centrado en el usuario. VI Workshop de Investigadores en Ciencias de la Computación, 53-58.

Rosado Gómez, A. A., & Verjel Ibáñez, E. A. (2017). Aplicación de la minería de datos en la educación en línea. Revista colombiana de tecnologías de avanzada, 92-98.

UNICEF. (23 de Julio de 2019). Obtenido de Sinopsis: Métodos de Recolección y Análisis de Datos en la Evaluación de Impacto.: [https://www.unicef-irc.org/publications/pdf/brief\\_10\\_data\\_collection\\_analysis\\_spa.pdf](https://www.unicef-irc.org/publications/pdf/brief_10_data_collection_analysis_spa.pdf)