

Educación Digital y la Minería de Datos Educativos



Autores

MSC, Lic. y Prof. Daniel Guillermo Cavaller Riva, Cdor. Cristian Darío Ortega Yubro,
Lic. Emiliano Andrés Dueñas, Lic. Héctor Nicolás Sosa.

<mailto:{daniel.cavaller; cristian.ortega; hector.sosa; emiliano.duenas}@fce.uncu.edu.ar>
<http://fce.uncu.edu.ar>

Resumen: La presente investigación luego de más de diez años de producción de Moodle para el soporte de cátedras, tiene como principal objetivo la conectividad automática entre las bases de datos SIU-Guaraní (la mandatoria), SIU-Kolla y Econet. Esto permitirá entre otras cosas: contar con datos educativos certeros en bases de datos confiables, mejorar la eficiencia administrativa en la carga de información al realizarse por única vez y luego ser compartida, depurar las bases de datos automáticamente al formular una máquina (pieza de software) para autoaprender, incorporando Inteligencia Artificial a través de una machine learning. Además, esto facilitará los aprendizajes didácticos. Esta herramienta estará a disposición de la Secretaría Académica, el Servicio de Apoyo Pedagógico y Orientación al Estudiante (SAPOE), Direcciones de Carrera y de los Profesores Editores para la toma de decisiones y el planteo de estrategias a nivel EIS (Sistema de Información Ejecutiva). Para ello se cuenta con las opciones analíticas de Moodle y las que se pueden desarrollar con Python, con datos educativos en línea, a través de la metodología CRISP-DM.

Palabras claves: Minería de Datos Educativos. Análisis del Aprendizaje. Aprendizaje Automático. Aprendizaje Profundo. Aprendizaje Supervisado. Aprendizaje no Supervisado. Sistema de gestión educativa. Moodle.

DIGITAL EDUCATION AND EDUCATIONAL DATA MINING

Resumen: The main objective of this research, after more than ten years of Moodle production for the support of chairs, is automatic connectivity between the SIU-Guaraní (mandatory), SIU-Kolla and Econet databases. This will allow, among other things: having accurate educational data in reliable databases, improving administrative efficiency in loading information when it is done once and then being shared, automatically debugging databases when formulating a machine (piece of software) to self-learn, incorporating Artificial Intelligence through machine learning. Furthermore, this will facilitate didactic learning. This tool will be available to the Academic Secretariat, the Pedagogical Support Service and Student Orientation (SAPOE), Career Directorates and the Editing Professors for decision-making and the formulation of strategies at the EIS level (Executive Information System). For this, there are Moodle analytical options and those that can be developed with Python, with online educational data, through the CRISP-DM methodology.

Palabras claves: Educational Data Mining. Analysis of Learning. Machine Learning. Deep learning. Supervised Learning. Unsupervised learning. Educational management system. Moodle.

Google Colab: <https://github.com/Moodle>

1. Introducción

El análisis de las bases de datos del sistema de gestión de aprendizaje Moodle ([«Moodle - Open-source learning platform | Moodle.org», s. f.](#)) a través de la minería de datos educativos (EDM) y del análisis del aprendizaje (AA), ofrece una mejora a todo el proceso educativo que se implemente tanto de las ofertas de grado como las de posgrado. Usando las diferentes metodologías y algoritmos para la modelización supervisada a través del aprendizaje automático en una primera etapa y aprendizaje profundo posteriormente, se pueden optimizar los procesos de enseñanza y los procesos del aprendizaje. El análisis del aprendizaje comprende la exploración, medición, recopilación e informe de datos sobre los estudiantes, los profesores y sus contextos, con el fin de comprender y optimizar el aprendizaje y los entornos en los que ocurre ([Monllao Olive, Huynh, Reynolds, Dougiamas, & Wiese, 2018](#)).

A su vez, la minería de datos educativos es una disciplina emergente, preocupada por el desarrollo de métodos para explorar los datos que provienen de sistemas de gestión de aprendizaje y que utilizan esos métodos para comprender mejor a los estudiantes y la configuración en la que aprenden. Este entendimiento hoy día representa una oportunidad de mejora inigualable la que impactará en evidenciar el gran esfuerzo realizado por parte de los profesores, con su gran aporte, y la participación de los estudiantes que son los principales actores en las actividades desarrolladas y a desarrollar del sistema de gestión de aprendizaje Moodle. Los procedimientos estándar para el desarrollo de la analítica de las bases de datos insumen un tiempo valioso. Ese tiempo puede optimizarse notablemente utilizando herramientas analíticas de Inteligencia Artificial a través del desarrollo de modelos de aprendizaje supervisado, tomando como insumo la totalidad de los datos de Moodle, a través de las herramientas integradas y otras herramientas. Los datos abiertos, públicos, personales y los informatizados de Moodle como los datos no tradicionales están expuestos a anomalías, y estas anomalías responden a distintas causales ([Álvarez, Romero, Cadenas, Coronado, & Rodríguez, 2016](#)). Vislumbrando dicho problema se puede realizar un cambio cultural en la institución. Las anomalías en los datos educativos coexisten en las múltiples bases de datos, sean estos datos educativos de origen interno como los de origen externo. Los datos educativos conceptualizados como no tradicionales proceden de diversas fuentes y orígenes, por ejemplo, como los que se detallan:

- El sitio web de diferentes sistemas de gestión de aprendizaje sobre plataformas Moodle,
- Otros sitios y portales de la Administración Pública Nacional, Provincial o Municipal, de Empresas y Organizaciones no Gubernamentales vinculados a la plataforma Moodle,
- El sistema SIU-Kolla («SIU-Kolla - SIU», s. f.), el sistema SIU-Guaraní («SIU-Guaraní - SIU», s. f.),
- Las redes sociales,
- Los correos electrónicos del ecosistema educativo,
- Las imágenes,
- Los datos geográficos, y otras fuentes de datos educativos.

Si las bases de datos educativos y los datos educativos no tradicionales que se tienen a disposición contienen información acerca de todas las actividades de cada uno de los estudiantes, el análisis de ciertas anomalías en los datos educativos, podría dar lugar al descubrimiento de patrones que responden ciertamente a una causa, pudiendo evidenciar

estas causas ciertos defectos y hechos de la comunidad de estudiantes, en tanto se concibe el cambio necesario de enseñanza y aprendizaje, en la carrera Contador Público de la Facultad de Ciencias Económicas y demás carreras, impactando en un cambio en la cultura institucional de la institución educativa, a través de las diferentes cátedras contenidas en los planes de estudio. Los problemas en las bases de datos educativos se corresponden a estrategias de conversión de cultura institucional, como también adecuación de los planes de estudio referenciados a las competencias requeridas de los estudiantes, y las anomalías en los datos educativos se relacionan con los aspectos de no innovar en los cambios de la enseñanza y el aprendizaje, y ese es uno de los tantos problemas a resolver. Esos problemas son independientes de los de E-Learning y del nivel de desarrollo de la región o país, y de cada una de sus provincias, ya que en gran parte tienen que ver con el grado de nivel de cultura Institucional que se posea, la cual se ve reflejada con la cantidad de ruido que posean las bases de datos educativas que manejan. El propósito de la presente investigación es el desarrollo teórico del análisis causal de las anomalías de los datos educativos, teniendo en cuenta el gran volumen de datos educativos que existen, especialmente en las administraciones de ECONET y otras plataformas, demostrando que los procesos metodológicos de la Inteligencia Artificial a través de metodologías de Minería de Datos Educativos, Aprendizaje Automático y el Aprendizaje Profundo, permitan mejorar y/o crear nuevos tipos de prácticas profesionales a través de la definición y fortalecimiento de las competencias requeridas de los estudiantes, optimizando así los aportes de los profesores, entendiendo que esas competencias de los estudiantes no pueden ser estáticas, sino dinámicas.

2. Área de estudio

Se define en el marco de las Teorías Didácticas e Ingeniería Didáctica, trabajando sobre datos de ECONET - Moodle complementados con los Informes de Google Analytics, sumando a ello: el análisis de datos, aplicando Aprendizaje Automático con lenguaje Python en Google Colab. Moodle es la plataforma que utiliza la Facultad de Ciencias Económicas, de la Universidad Nacional de Cuyo que ofrece la posibilidad de acceso a todo el contenido propuesto aprendizaje en virtud de los planes de estudio, para todos los estudiantes de esa casa de estudios, con la alternativa de crear cursos y el envío de tareas entre otras actividades. Particularmente en nuestra provincia Moodle se está aplicando en distintos niveles de educación, desde las escuelas secundarias hasta la universidad. Moodle comprende un conjunto de funciones en dos grupos: actividades y recursos.

- Los recursos son todos aquellos materiales educativos que están digitalizados y que se adjuntan a la plataforma. Ejemplos de recursos son aquellos documentos PDF, imágenes, documentos de texto, presentaciones de PowerPoint, etc.
- Las actividades son partes integradas de la plataforma Moodle que permiten la interacción de los profesores con sus estudiantes. Ejemplos de las actividades son los foros, las lecciones, las encuestas, los cuestionarios, las tareas, y otras. Todos los datos de los estudiantes y de los profesores se almacenan en las múltiples bases de datos de Moodle.

Si bien, la situación actual de público conocimiento que se está atravesando ha dejado de manifiesto el gran trabajo que ha sido desarrollado en la plataforma Moodle, esa situación puntual está generando el clima perfecto para transitar un oportunidad de mejora inigualable y esbozar metodologías y procesos que no se estarían llevando actualmente a cabo como lo es el análisis de las diferentes tablas de la base de datos de Moodle y su adecuación a

través de las competencias necesarias de los estudiantes a lo largo del desarrollo de su plan de estudios académicos. Las herramientas de minería de datos podrían utilizarse para descubrir patrones del conjunto de datos. El problema ocurre cuando en los datos se encuentran anomalías sin detectar su causa, o cuando se está en presencia de ruido en los datos educativos, lo que complicaría aún más el proceso de extracción de conocimiento en virtud de la metodología CRISP-DM. Hay que tener presente que los registros volcados a Moodle contienen datos educativos de cada una de las acciones que el estudiante y los profesores realizan. Por tal motivo recobra vital importancia la corrección de los datos educativos imperfectos a través del análisis de datos educativos. El análisis de datos es el proceso de examinar, limpiar, transformar y modelar grandes conjuntos de datos con el objetivo de descubrir patrones ocultos, correlaciones desconocidas y obtener conocimiento. Colab es una herramienta que permite ejecutar y programar Python desde cualquier dispositivo que se esté utilizando, con la posibilidad de utilizar los recursos computacionales que Google pone a disposición, y el atractivo de la integralidad con TensorFlow. De acuerdo a la presentación que hace Colab, la aplicación puede facilitar el trabajo ya sea de un alumno, un científico de datos o un investigador de inteligencia artificial y de esa forma desarrollar modelos de aprendizaje ([«Google Colaboratory», s. f.](#)).

3. Metodología

La múltiples bases de datos de Moodle está compuesta por diferentes tablas, las cuales algunas de ellos tendrían campos vacíos naturalmente, mientras que otras contendrían muchos datos. Debe tenerse cuidado y omitir de los análisis de datos las bases de datos que contienen datos personales tomando para ello los recaudos previstos en la ley de protección de datos personales, protegiendo así la privacidad de los profesores y de los estudiantes, a través de la aplicación de un proceso de despersonalización de los datos. Se puede identificar las tablas necesarias que contienen los datos para el desarrollo de la analítica de los datos. Algunos ejemplos de tablas, son las que se detallan en Tabla N° 1.

Identificadas las tablas necesarias para el desarrollo del análisis de los datos educativos ([«Reportes del sitio - MoodleDocs», s. f.](#)), se pueden desarrollar subtablas para contribuir al proceso de creación de consultas complejas que podrían seguramente requerir recursos computacionales para ejecutarse y demandar una gran cantidad de tiempo ([Velinov, s. f.](#)).

Por tal motivo, e identificadas o creadas las tablas necesarias que contienen los datos educativos requeridos, se podrían vincular de tal manera las competencias requeridas de los estudiantes y su alcance a través de las distintas materias contenidas en los planes de estudio, generando vinculaciones entre ellas, con el objetivo de facilitar el análisis predictivo en los patrones a descubrir de los estudiantes a través de los datos volcados en Moodle y tomar acciones preventivas que se requieran, en forma automatizada, para la totalidad de los estudiantes.

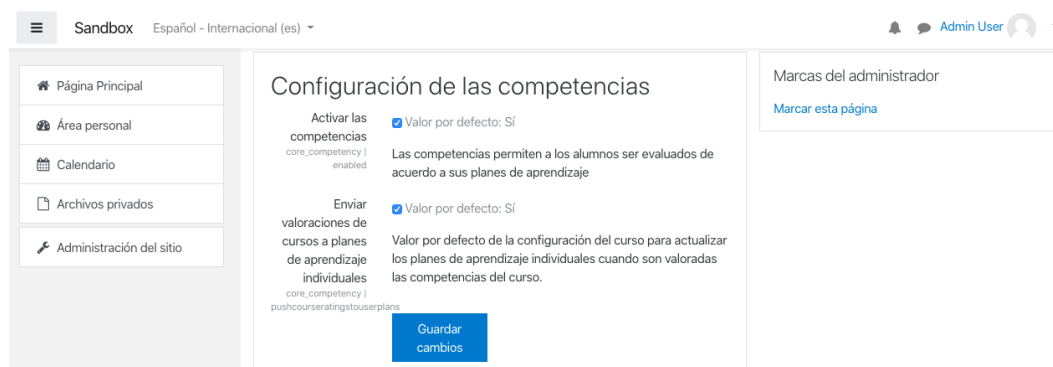
Tabla N° 1 - Ejemplo de tablas de Moodle a considerar.

Nombre de la Tabla	Descripción de la Tabla
mdl_assign	Asignación que un creador de curso ha incluido en un curso.
mdl_course	Entradas para cada curso creado en Moodle. Contiene aproximadamente

	33 campos, incluidos fecha de creación y el idioma.
mdl_lesson	Para cada actividad Lección creada en cualquier curso, se ingresa un registro en esta tabla.
mdl_user	Cada usuario de Moodle tiene una entrada en esta tabla.
mdl_logstore_standard_log	Casi todas las interacciones de los visitantes se registran en esta tabla.
mdl_role_assignment	El rol del usuario en Moodle.

Esa es la razón por la cual debería dedicarse un esfuerzo adicional y construir un óptimo conjunto de datos educativos velando en el concepto de calidad del dato, y poder recopilar información útil y extraer conocimiento. Si bien se vislumbra una actividad ardua y de mucho trabajo, los resultados serían muy satisfactorios, como por ejemplo la posibilidad de automatizar el seguimiento de la totalidad de los estudiantes de la Facultad de Ciencias Económicas y predecir su rendimiento en otras materias, en virtud de la construcción en línea de sus competencias, derivando las predicciones realizadas en acciones automatizadas como las que desarrolla el SAPOE, resultando así un seguimiento basado en inteligencia artificial de cada uno de los estudiantes, y de las necesidades que ellos manifiestan hacia sus profesores [\(Raga & Raga, 2017\)](#).

Imágen N° 1 - Configuración de las competencias de los estudiantes



Por tal motivo ya se tiene en cuenta que si bien el punto de partida en la investigación es el análisis de los datos educativos imperfectos, el impacto será en el sistema de gestión de aprendizaje (LMS), siendo el uso de herramientas de análisis de datos educativos inevitable para alcanzar un objetivo superador que sería:

- Localizar los inconvenientes en los procesos de enseñanza e intentar solucionarlos y ajustarlos.

Hay varios informes, bloques y otros complementos de Moodle que proporcionan análisis de aprendizaje y pueden llevar a lo planteado en el párrafo anterior. Algunos de ellos son los que se conocen como:

- El registro de actividad,
- Finalización de la actividad,
- Participación en los distintos cursos,
- Descripciones generales del curso,
- Estado de finalización del curso,
- Lista de eventos,

- Registros en vivo, etc.

Los mencionados permitirían encontrar correlaciones y recopilar información más que valiosa del conjunto de datos (Singh, 2015). Por defecto, la API de Moodle sólo proporciona capacidades básicas de informes. Sus herramientas de informes le permite a los usuarios, principalmente a los profesores, acceder a los registros de las distintas actividades del curso, para agrupar los datos por cada uno de los estudiantes o por actividad desarrollada, y de esa forma generar gráficos analíticos. Pero las herramientas de análisis para extraer conocimiento no están disponibles en el mismo sistema de gestión de aprendizaje Moodle. No obstante existen una amplia gama de complementos compatibles con Moodle pueden instalarse por separado. Algunas herramientas de análisis típicas para Moodle basadas son:

- X-Ray Learning Analytics.
- Intelliboard.
- Analytics Graphs.
- Analytics local plugin.
- Google Analytics.

Los algoritmos como: las reglas de asociación, predicción, agrupamiento y análisis secuencial de patrones pueden utilizarse para la minería de datos educativos (EDM) («Moodle Research Library - Moodle Research Library», s. f.). Pero existen muchos tipos de técnicas o métodos populares de la Ciencia de Datos aplicables a los objetivos buscados, utilizando los datos de las bases de datos del sistema de gestión de aprendizaje Moodle como insumos para la modelización en el marco de un proceso de extracción de conocimiento, y la literatura acorde recomienda cinco categorías como las más adecuadas a tal fin, ellas son:

- Predicción.
- Agrupamiento.
- Descubrimiento con modelos.
- Minería de datos educativos.
- Extracción de conocimiento para juicio humano.

Ahora bien, Moodle tiene integrado herramientas analíticas de aprendizaje automático, las cuales deben ser habilitadas, tal como se constata en el sitio <https://moodle.org/demo/> accediendo a la demostración de SandBox ingresando al mismo con perfil de administrador. Los procesadores para analítica y predicciones pueden ser en PHP o Python, y se puede definir un intervalo de tiempo para el análisis deseado en virtud al modelo de aprendizaje esbozado. Los modelos analíticos establecidos son los siguientes:

- Cursos en riesgo de no comenzar.
- Cursos en riesgo de abandonar.
- Estudiantes que no han accedido recientemente al curso.
- Estudiantes que no han accedido al curso.
- Próximas actividades pendientes.

Tabla Nº 2 - Técnicas de minería de datos educativos (EDM)

Técnicas	Concepto	Propósito
Predicción	La predicción desarrolla un modelo para estimar el valor de la variable que se desconoce. Hay tres tipos de predicción: regresión, clasificación y estimación de densidad.	Predicir si un estudiante aprobará un curso o no. Predicir si un estudiante necesitará un refuerzo en alguna competencia o no.

Agrupamiento	El agrupamiento es una técnica de aprendizaje no supervisada que tiene como objetivo descubrir similitudes entre los datos y agrupar datos similares de acuerdo con sus características.	Agrupar estudiantes similares basado en su comportamiento de aprendizaje. Descubrir patrones de aprendizaje de los estudiantes de primer año.
Descubrimiento con modelos	En el proceso de descubrimiento con modelos, se desarrolla los modelos mediante técnicas de agrupamiento, predicción, y otras técnicas, usando ese modelo para análisis.	Descubrir la relación entre diferentes comportamiento de aprendizaje de los alumnos de un determinado curso. Analizar las características de cada estudiante.
Minería de datos educativos	La minería de datos educativos descubre relaciones entre variables en un conjunto de datos educativos. Hay cuatro tipos de minería: minería de reglas de asociación, minería de correlación, minería de datos causales y minería de patrones secuenciales.	Descubrir asociaciones de las materias contenidas en los distintos planes de estudio Descubrir el comportamiento en el aprendizaje.
Extracción de conocimiento para juicio humano	Los seres humanos a veces hacen inferencias sobre los datos como suplemento para el método de minería de datos educativos	Etiquetar datos para mejorar modelo predictivo. Descubrir patrones de aprendizaje de los estudiantes

Pero lo más destacable es la opción de Moodle que permite crear los modelos analíticos necesarios y además permite importar otros modelos. La mayoría de los modelos analíticos de aprendizaje no están habilitados de forma predeterminada. La habilitación de los modelos para su uso debe realizarse después de considerar los objetivos institucionales que los modelos deben respaldar, los cuales podrían establecerse siguiendo la metodología CRISP-DM. Al seleccionar o crear un modelo analítico, debe considerarse que la API de Moodle no viene activa, sino que debe habilitarse y configurarse, esa configuración no es del tipo específico y estándar, sino que debe tener un análisis previo de los actores que interactúan en el sistema de gestión educativo Moodle (profesores, autoridades, técnicos etc) y según ese análisis se determinan los objetivos que se configurarán, lo que impactarán en el desarrollo de la inteligencia artificial. Posibles puntos a considerar:

- Predicción.
- Resultado.
- Evaluación del resultado.
- Notificación del resultado.
- Que se puede hacer con ese resultado.

La herramienta está ideada a los efectos de que cada usuario configure el mismo como antes se indicó, pero vienen tres tipos de posibles configuraciones que ofrece moodle y está en nosotros analizar si nos pueden ser de utilidad.

Imágen N° 2 - Moodle machine learning backend



Imágen N° 3 - Modelos analíticos de Moodle



Una de las metodologías analíticas que puede desarrollarse es la del procesamiento del lenguaje natural (NLP) sobre las encuestas de incidentes críticos efectuadas a los estudiantes. La misma podría realizarse sobre un archivo .txt el cual debería ser importado a Google Colab. Previamente, el archivo de texto requiere un trabajo adicional, en el cual debe eliminarse encabezados, pie de página, y todo elemento que altere el análisis. De esa forma el archivo queda lo más depurado posible. Una vez que está depurado, se procede a subirlo a nuestro entorno de trabajo de Google Colab. Hay que tener en cuenta que cuando se sube un archivo con el menú Archivos -> Subir, ese archivo permanece en el entorno de trabajo mientras se esté trabajando. Al cerrar sesión, ese archivo se borra. El código necesario para cargar las librerías, es el siguiente:

```
import collections
import multiprocessing as mp
import re
```

El código siguiente, se utiliza para cargar el texto en el entorno de trabajo:

```
def process(line):
    print(line)
with open('/content/Encuesta.txt') as f:
    data = f.readlines()
    for line in data:
        process(line)
```

Y el resultado se visualiza en la siguiente imagen:

Imágen N° 4 - Procesamiento del Lenguaje Natural - Encuestas de incidentes críticos



Para depurar y lograr una bolsa de palabras, debe definirse una variable y eliminar caracteres innecesario.

Esta herramienta también puede aplicarse para el análisis de las sensaciones de los estudiantes a través de la captura de datos educativos resultantes de las actividades de encuestas. Previamente, para poder diseñar el vocabulario, se debe cambiar el tipo de datos de las palabras pasando de string a lista. El contenido del vocabulario contiene cada palabra y la cantidad de repeticiones de cada una de ellas.

Las bolsas de palabras se desarrollan para orientarlas a modelos de inteligencia artificial como los sistemas de recomendaciones, los modelos de tópicos y la clasificación de las palabras. De esa forma se puede constatar las palabras más utilizadas en las encuestas y construir un modelo que permita analizar las sensaciones de los estudiantes que participaron. En el caso bajo análisis se encuentra suficiente evidencia que demuestran la necesidad de los estudiantes de cursar. Téngase presente que el aislamiento social por motivos del COVID 19 produce en los estudiantes incertidumbre acerca de sus procesos de aprendizaje, más aún cuando estos estudiantes son estudiantes de primer año, que no están familiarizados con el uso de la plataforma, como lo sería un conjunto de estudiantes de segundo, tercer y cuarto año.

El módulo de encuesta predefinida (traducción del inglés Survey) es una actividad de curso que proporciona un conjunto de instrumentos verificados, incluyendo COLLES (Constructivist On-Line Learning Environment Survey) y ATTLS (Attitudes to Thinking and Learning Survey), que se han mostrado útiles para evaluar y estimular el aprendizaje en contextos de aprendizaje en línea. Los profesores pueden usarlas para recopilar datos de sus estudiantes que les ayuden a aprender tanto sobre su clase como sobre su propia enseñanza.

Los resultados que arroje podrían evidenciar los niveles de satisfacción de los estudiantes por cátedra. Ese nivel de satisfacción puede contrastarse con el análisis de las competencias adquiridas, y lograr interesantes conclusiones.

Imágen N° 5 - Tamaño y contenido del vocabulario

```

Encuesta de Incidentes Críticos.ipynb
+ Código + Texto
[51] print("Tamaño de Vocabulario: {}".format(len(vect.vocabulary_)))
print("Contenido del Vocabulario: {}".format(vect.vocabulary_))
print("Features: {}".format(vect.get_feature_names()))

Tamaño de Vocabulario: 990
Contenido del Vocabulario: {'encuesta': 325, 'de': 249, 'incidentes': 484, 'criticos': 230, 'preguntas': 720, 'todas': 920, 'las': 522, 's
Features: ['10', '13', '17', '19', '1er', 'abel', 'abiertos', 'abril', 'acceder', 'accion', 'acciones', 'acerca', 'aclaraciones', 'aclarar

[52] bolsadepalabras = vect.transform(lista)
print("Bolsa de palabras: {}".format(bolsadepalabras))

Bolsa de palabras: (0, 0) 1
(0, 1) 1
(0, 2) 1
(0, 3) 1
(0, 4) 1
(0, 5) 5
(0, 6) 1
(0, 7) 6
(0, 8) 1
(0, 9) 6
    
```

Imagen Nº 6 - Bolsa de palabras y Matriz de documentos - términos

```

Encuesta de Incidentes Críticos.ipynb
+ Código + Texto
[ ] print("Bolsa de palabras: {}".format(repr(bolsadepalabras)))

Bolsa de palabras: <1x990 sparse matrix of type '<class 'numpy.int64''>
with 990 stored elements in Compressed Sparse Row format>

[ ] print("Matriz documentos - términos:\n{}".format(bolsadepalabras.toarray()))

Matriz documentos - términos:
[[ 1 1 1 1 1 5 1 6 1 6 2 1 1 4 1 1 5 17
 1 1 2 1 5 1 30 1 5 5 5 20 1 5 5 10 4 4
 4 6 2 1 1 1 10 10 1 5 1 5 5 5 5 15 5 6
 2 2 1 15 1 1 1 4 5 2 1 1 2 1 1 1 1 1
 1 1 7 1 1 1 2 1 1 1 5 5 6 1 1 2 1 2
 2 5 1 2 5 1 5 5 2 5 5 20 5 5 5 5 5 6
 5 5 5 5 1 1 2 4 2 1 1 5 4 1 10 5 2 1
 2 5 1 5 1 2 5 4 1 1 1 5 5 1 1 1 8 2
 1 4 1 3 39 74 1 5 1 1 1 7 4 17 1 1 1 16
 2 1 1 1 1 3 1 1 7 1 22 1 1 1 1 5 2
    
```

La presente codificación en Python se basa en los modelos desarrollados por el [Dr. Marcelo Luis Errecalde](#), la cual se puede verificar en: https://github.com/merrecalde/rio_2020 en el cual se desarrolla conceptos de la Minería de Textos para el análisis automático de datos, en el presente caso, datos educativos. La encuesta COLLES (Constructivist On-Line Learning Environment Survey) tiene 24 frases agrupadas en seis escalas, cada una de las cuales nos ayuda a tratar un asunto clave acerca de la calidad del entorno de aprendizaje en-línea ([«Configuraciones de encuesta - MoodleDocs», s. f.](#)), por ejemplo:

- Relevancia: ¿qué tan relevante es el aprendizaje en-línea para las prácticas profesionales de cada uno de los estudiantes?
- Reflexión: el aprendizaje en-línea ¿estimula el pensamiento crítico de los estudiantes?
- Interactividad: indica en qué medida los estudiantes se comprometen en-línea en diálogo educativo enriquecido.
- Soporte del Tutor: pone en evidencia qué tan bien habilitan los tutores el que los estudiantes participen en el aprendizaje en-línea
- El Soporte por Pares es sensible y promueve el soporte proporcionado en-línea por cada uno de los compañeros de los estudiantes?
- Interpretación: los estudiantes y tutores entienden las comunicaciones en-línea de los unos a los otros.

Subyacente a la visión dinámica del aprendizaje está una teoría del conocimiento: constructivismo social, que retrata al estudiante como un conceptualizador activo dentro de un ambiente de aprendizaje socialmente interactivo. El constructivismo social es una epistemología, o manera de conocer, en la cual los estudiantes colaboran reflectivamente para co-construir nueva comprensión, especialmente en el contexto de investigación mutua fundamentada en sus experiencias personales ([«Configuraciones de encuesta -](#)

[MoodleDocs», s. f.](#)). La teoría de las “vías del conocimiento” (ways of knowing), originalmente del campo de investigación en género (Belenky et al., 1986) nos proporciona una herramienta de encuesta para examinar la calidad del discurso dentro de un entorno colaborativo. La Encuesta de Actitudes hacia el Pensamiento y el Aprendizaje (ATTLS) es un instrumento desarrollado por Galotti et al. (1999) para medir qué tanto una persona es un conocedor conectado (connected knower = CK) o un conocedor separado (separate knower = SK). Las personas con cocientes CK altos tienden a encontrar más disfrutable el aprendizaje, y son más frecuentemente cooperativas, simáticas y más dispuestas a construir sobre las ideas de otros, mientras que aquellos con cocientes SK más altos tienden a tomar una postura más crítica y argumentativa hacia el aprendizaje. Los estudios han demostrado que estos dos estilos de aprendizaje son independientes entre sí (Galotti et al., 1999; Galotti et al., 2001). Adicionalmente, son solamente un reflejo de las actitudes hacia el aprendizaje, no son capacidades de aprendizaje ni poder intelectual ([«Configuraciones de encuesta - MoodleDocs», s. f.](#)).

Las ante mencionadas son parte de la retroalimentación del sistema de Enseñanza-Aprendizaje, que devuelve tanto social como cualitativamente lo que subyace en el grupo control de educación a distancia. Las mismas se auto procesan con una máquina en colab, de forma que cada vez que se cuente con nuevos datos, la proyección se hace más exacta, en un Modelo de aprendizaje auto reforzado. En este caso son más de 160 entradas vectoriales. La base de construcción de actividades y recursos de moodle se ve reforzada en las teorías de Guy Brousseau, Michele Artigue & Regine Duady e Ives Chevallard (s.f. 2011).

Otro análisis del ecosistema de los datos educativos es el que se relaciona con la aplicación del entorno de trabajo de Google Colab, a través de su lenguaje Python, retroalimentado los modelos de aprendizaje con los datos del sistema de gestión educativo Moodle. Múltiples son los análisis que pueden desarrollarse, pero la potencialidad de la herramienta permite en el desarrollo de análisis estadísticos, la opción de crear modelos predictivos en virtud de esos datos, y graficarlos.

Un ejemplo de ello sería con los datos de ingresos a la plataforma por parte de los estudiantes. Se puede definir el calendario lectivo del año 2018 como serie de datos de entrenamiento, y el calendario lectivo del año 2019 como serie de testeo, para proyectar sobre el calendario lectivo del año 2020 que todavía no transcurre la predicción de participación de los estudiantes, y luego esa predicción contrastar con el efecto COVID-19, para lograr ajustar la predicción.

Los datos son ingresos diarios, se establece que desde el último dato cierto, se haga un análisis predictivo para 120 días más, es decir casi 4 meses después del 26 de abril. El gráfico de la Imágen Nº 9 muestra en puntos negros los datos ciertos, y en línea azul el comportamiento predictivo esperado para los 4 meses posteriores al 26 de abril.

Lo expuesto hasta aquí sería lo correspondiente al desarrollo de inteligencia artificial con modelos supervisados. (Por el ajuste posterior y corrección de la predictibilidad).

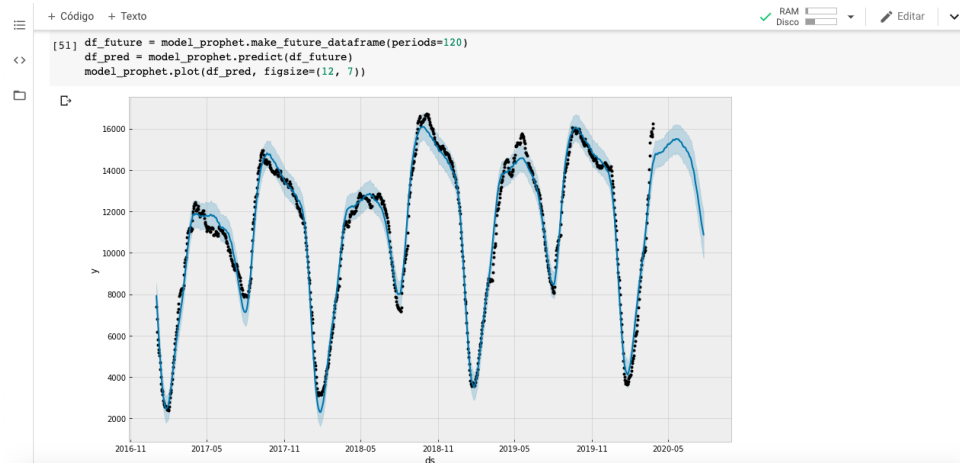
[Imágen Nº 7 - Actividades en Moodle Entorno Google Colab Lenguaje Python](#)



Imagen N° 8 - Recursos en Moodle Entorno Google Colab Lenguaje Python - Estadística Descriptiva



Imagen N° 9 - Serie de tiempo predictiva ingresos a la plataforma Moodle 2017 - 2020



Con la depuración de los datos a través de la librería pandas, los cálculos básicos de estadística no presentan dificultad alguna.

La codificación completa, por un tema de simplicidad y espacio, puede consultarse en el repositorio de Github, con la opción de clonar los archivos necesarios a tal efecto. En enlace es el siguiente: <https://github.com/cristiandarioortegayubro/Moodle>

4. Resultados obtenidos

Se infiere que el tiempo de estadía promedio de un estudiante en sistema de gestión de aprendizaje Moodle se de sólo 5 a 10 minutos por sesión, lo que corresponde a distintos causales:

- Relevancia,
- Baja del contenido innecesario,
- La inexistencia o bajo desarrollo de actividades específicas,
- Resúmenes de la materia en formato papel (en la fotocopidora).

Algunas cuestiones que podrían parametrizarse con el objeto de conseguir mayor calidad en los datos educativos del sistema de gestión de aprendizaje Moodle son como las que se detallan a continuación:

La Base de Datos de ECONET/Moodle no tiene conexión electrónica con el SIU-Guaraní y menos con el SIU Kolla ([«SIU-Kolla - SIU», s. f.](#)); es aquí donde se hace énfasis sobre el punto de anomalías; en principio la base de datos del sistema de gestión de aprendizaje denominado ECONeT/Moodle es una base de datos mala, en la cual el campo registro de cada estudiante no se rellena desde los últimos 3 años.

Mientras que para los estudiantes de los últimos años, es decir de cuarto y quinto año de las distintas carreras si aparece este dato primordial. ¿Por qué se considera a ese dato primordial? Se lo considera así debido a que la Facultad de Ciencias Económicas tiene por obligación entregar el número de registro y ello sucede cuando se completan todos los trámites del alumno frente a la Oficina de Alumnos.

¿Cómo proceder en este caso puntual?, sólo se puede hacer con el documento único o pasaporte; por lo que se presenta otro problema, el campo para el dato del documento único está definido para datos de tipo carácter y debería ser para datos de tipo numérico, pero los pasaportes tiene datos alfanumérico (números y letras) y esta característica limitaría en principio la analítica de datos de los estudiantes cargados en la base de datos del SIU-Guaraní ([«SIU-Guaraní/versión 3.16.0/interfaces/moodle - SIU», s. f.](#)).

Se dice en principio porque en realidad demandaría mayor actividad de preprocesamiento de los datos, de la etapa limpieza y transformación de datos que pertenece a un proceso de extracción de conocimiento, siendo este proceso el utilizado en la minería de datos educativos.

5. Conclusiones y Proyecciones

5.1. Conclusiones

La inclusión de la Ciencia de Datos para extraer conocimiento de los datos de las múltiples bases de datos de la plataforma educativa Moodle en virtud de los procesos educativos, permite esbozar la generación de modelos de aprendizaje con el único objetivo de detectar los patrones de las actividades de los estudiantes y de los profesores y usarlos para mejorar el aprendizaje y fortalecer las competencias individuales de cada uno de ellos.

Por tal motivo la aplicación de diferentes herramientas de análisis recopiladas de los sistemas de gestión de aprendizaje, se puede predecir el desempeño futuro de cada

estudiantes, clasificándolos y detectando aquel grupo de estudiantes que no es probable que tengan éxito, lo que permitiría modificar los estilos y métodos de enseñanza de los profesores, de una forma más eficaz que por ejemplo, a través de cuestionarios efectuados a los alumnos con respecto de los profesores y de la cátedra. Además de la posibilidad de construir una matriz de estudiantes en riesgo, especificados por cada orientación profesional, el análisis del aprendizaje también puede ayudar a los profesores a descubrir si sus estudiantes comprenden el material de los cursos que imparten, de la forma que lo imparten, si los contenidos son los necesarios, o hacer una predictibilidad de los contenidos requeridos, a través de la captura de los datos necesarios para ello. De esa manera, con el análisis del aprendizaje y los algoritmos adecuados, se obtienen las instrucciones relativas al cómo los profesores y los estudiantes deberían proceder con sus actividades, y de esa manera obtener los mejores resultados posibles en el proceso de enseñanza y el sistema educativo. Es oportuno por ello destacar que, no importa a esta altura cómo se recolectan los datos, sino que los mismos sean de calidad, la que se obtiene en las primeras fases de la metodología CRISP-DM, con lo cual es de vital importancia de poner énfasis en aquellos datos imperfectos de las múltiples bases de datos de la plataforma educativa Moodle y el análisis de su causal. Ahora bien, la posibilidad de encontrar mayores datos imperfectos sobre la cantidad de datos educativos de la plataforma Moodle aumenta ya que su utilización se ha visto incrementada a causa del COVID 19. En la imagen siguiente puede visualizarse el incremento a partir del mes de marzo.

5.2. Proyecciones

Migrar de un aprendizaje supervisado a uno no supervisado, con automatización de notificaciones y acciones tanto para estudiantes como para profesores y secretaría académica.

Aunque en principio la investigación parece que incluye solo a dos grupos principales, el de los estudiantes y el de los profesores como partes interesadas en la minería de datos educativos (EDM), en realidad hay otros grupos que deberían estar involucrados en la minería de datos educativos, porque esos grupos tienen otra mirada con sus propios objetivos y razones para participar de un proceso de minería de datos educativos, todos válidos. Esos interesados serían:

- Desarrolladores de cursos y contenidos.
- Investigadores educativos.
- Administradores de Redes.
- Administradores de Moodle.
- Secretaría Académica.
- SAPOE.

Referencias Bibliográficas

<https://colab.research.google.com/notebooks/welcome.ipynb>

Álvarez, K., Romero, B., Cadenas, J., Coronado, D., & Rodríguez, R. (2016).

Arquitectura para la Gestión de Datos Imperfectos en la Era de Big Data. *Revista Venezolana de Computación*, 3(2), 47-56.

Artigue, Michele & Duady, Regine; "Ingeniería Didáctica en la Educación Matemática", Iberoamerican Grupo Editorial ISBN; 978-970-625-119-0

<https://www.todostuslibros.com/autor/artigue-michele-douady-regine>

Brousseau, Guy (2007) *Iniciación al estudio de la teoría de situaciones didácticas* Zorzal ISBN 978-987-599-035-7

Chevallard, Yves, Licenciado en Matemáticas e investigador de la Université d'Aix-Marseille II. Ha sido Director de l'IREM de Aix-Marseille II; desde esa fecha se desempeña como Catedrático Universidad IUFM d'Aix-Marseille; "La Transposición Didáctica. Del Saber Sabio al Saber Enseñado", edit AIQUE, ISBN: 978-950-701-380-5, EAN: 9789507013805, ES, 31-01-2011

https://docs.moodle.org/all/es/Configuraciones_de_encuesta.

PhD Errecalde, Marcelo Luis, Full Professor and Chief research at: Laboratory of Research and Development in Computational Intelligence (LIDIC). Chief Department of Computer Science Universidad de San Luis, "PLN en redes neuronales, machine learning", RIO 2020; FCeyN RIOIV, Córdoba.

Galotti, Kathleen M., et al, A New Way of Assessing Ways of Knowing: The Attitudes Toward Thinking and Learning Survey (ATTLS), May 1999;

<https://link.springer.com/article/10.1023/A:1018860702422>

Álvarez, K., Romero, B., Cadenas, J., Coronado, D., & Rodríguez, R. (2016).

Arquitectura para la Gestión de Datos Imperfectos en la Era de Big Data. *Revista Venezolana de Computación*, 3(2), 47-56.

Configuraciones de encuesta—MoodleDocs. (s. f.). Recuperado 24 de abril de 2020, de https://docs.moodle.org/all/es/Configuraciones_de_encuesta

Google Colaboratory. (s. f.). Recuperado 18 de enero de 2020, de

<https://colab.research.google.com/notebooks/welcome.ipynb>

Monllao Olive, D., Huynh, D. Q., Reynolds, M., Dougiamas, M., & Wiese, D.

(2018). A Supervised Learning framework for Learning Management Systems.

Proceedings of the First International Conference on Data Science, E-Learning and Information Systems. Recuperado de <https://dl.acm.org/citation.cfm?id=3279996>

Moodle Research Library—Moodle Research Library. (s. f.). Recuperado 18 de abril de 2020, de <https://research.moodle.org/>

Moodle—Open-source learning platform | Moodle.org. (s. f.). Recuperado 18 de abril de 2020, de <https://moodle.org/>

Raga, R. C. J., & Raga, J. D. (2017). Monitoring Class Activity and Predicting Student Performance Using Moodle Action Log Data. *International Journal of Computing Sciences Research*, 1(3), 1-16.

Reportes del sitio—MoodleDocs. (s. f.). Recuperado 18 de abril de 2020, de https://docs.moodle.org/all/es/Reportes_del_sitio

Singh, J. (2015, octubre 26). Learning Analytics tools available in Moodle #MoodleResearch #MoodleWorld. Recuperado 18 de abril de 2020, de LMSPulse website: <https://www.lmspulse.com/2015/learning-analytics-tools-available-in-moodle-moodleresearch-moodleworld/>

SIU-Guarani/version3.16.0/interfaces/moodle—SIU. (s. f.). Recuperado 18 de abril de 2020, de <http://documentacion.siu.edu.ar/wiki/SIU-Guarani/version3.16.0/interfaces/moodle>

SIU-Kolla—SIU. (s. f.). Recuperado 21 de abril de 2020, de

<http://documentacion.siu.edu.ar/wiki/SIU-Kolla>

Velinov, A. (s. f.). *Framework for Big Data Analytics of Moodle Data Using Hadoop in the Cloud*. Recuperado de

https://www.academia.edu/37414016/Framework_for_Big_Data_Analytics_of_Moodle_Data_Using_Hadoop_in_the_Cloud