

Análisis del rendimiento educativo en México respecto a los mejores y peores países de la OCDE

Trabajo Terminal No. 2021-B060

*Alumno: *Morgado Jiménez Arturo*

**e-mail: arturomorgadol117@gmail.com*

Resumen – En este trabajo terminal se busca detectar factores causantes de dos problemas que enfrentan los estudiantes de nivel medio superior, a saber: la deserción académica y el bajo rendimiento escolar. La búsqueda se realizará analizando los datos del rendimiento educativo que mide el Programa para la Evaluación Internacional de Alumnos de la OCDE (PISA, por sus siglas en inglés) a nivel internacional. Esta base de datos está formada por diversos criterios, entre ellos la evaluación de las habilidades matemáticas, científicas y la comprensión lectora. Adicionalmente, se realiza un censo que hace énfasis sobre las condiciones materiales, emocionales y sociales de los estudiantes y las escuelas que son partícipes en la prueba.

El estudio se complementa con datos obtenidos del Instituto Nacional de Estadística y Geografía (INEGI) de México acerca de la tasa de abandono escolar a nivel nacional, el cambio en las matrículas de los distintos niveles educativos y la inversión del estado en educación. El desarrollo del trabajo consiste en aplicar una metodología de análisis y minería de datos para detectar a los países con las mejores y peores evaluaciones, identificando los factores con mayor influencia en las variables de interés.

Los resultados esperados se presentarán en la forma de indicadores clave de desempeño (KPI) mediante técnicas visuales con distintos tipos de gráficas.

Palabras clave – Análisis de datos, Efectos de la pandemia SARS-COV-2, Prueba de PISA, Toma de decisiones

1. Introducción

La cuarta revolución industrial es una etapa en la evolución de la sociedad que está transformando los bienes y servicios que satisfacen las necesidades humanas y a los ecosistemas. Los efectos se han manifestado en diferentes ámbitos, razón por la que se han acuñado términos como Industria 4.0, Agricultura 4.0, Gobierno 4.0, Educación 4.0, entre otros.

Esta revolución industrial se caracteriza por el aprovechamiento de datos e información, además de la automatización de procesos en las actividades humanas. En muchos casos la información es un activo clave utilizado para tomar decisiones adoptando un modelo centrado en los datos «data-centric». En este modelo los datos se separan de las aplicaciones y las plataformas tecnológicas para que los datos se compartan y aprovechen en la organización que los produce [1].

En el contexto de la cuarta revolución industrial, la Educación 4.0 también considera el aprovechamiento de los recursos de procesamiento y manejo de información como transformadores de los procesos de enseñanza-aprendizaje, en diferentes niveles educativos [3]. En México el aprendizaje es un derecho constitucional que se hace efectivo a través del servicio público atribuido al Estado mediante la ley [2] y donde el IPN se reconoce como una institución pública líder en tecnología.

Si bien la Educación 4.0 está relacionada con el aprovechamiento de las tecnologías como habilitadoras y transformadoras de los procesos de aprendizaje, la incorporación de herramientas tecnológicas no es suficiente para mejorar los procesos educativos, puesto que uno de los tres pilares de la Educación 4.0 [3] es la personalización del aprendizaje; refiriéndose a la satisfacción de las demandas de aprendizaje individual dentro y fuera del aula.

Por lo tanto, también es importante considerar las condiciones materiales, psicológicas, emocionales y sociales de los estudiantes, puesto que están identificadas como factores influyentes en los problemas de deserción escolar y el bajo rendimiento académico [6]. De modo que es útil tomar en cuenta los diferentes modelos pedagógicos que se han desarrollado a través del tiempo para la realización de este trabajo.

En este trabajo se recabarán datos de la prueba PISA por parte de la Organización para la Cooperación y el Desarrollo Económico (OCDE), la cual es aplicada de manera internacional cada tres años desde el año 2000. Esta prueba evalúa el rendimiento y hábitos de estudio en estudiantes entre 15 y 16 años de edad. Adicionalmente se complementará el estudio con datos del Banco Mundial y en particular de los estudiantes de México con datos proporcionados por el INEGI con la finalidad de detectar los factores que influyen en el rendimiento de los estudiantes, así como los factores que México comparte tanto con los cinco mejores resultados de la prueba pisa 2018 y 2015, como con los cinco últimos lugares en el ranking de la prueba en ambos periodos.

México es un país que ha tenido muchas transformaciones desde su independencia, en las últimas décadas como consecuencia de la globalización, la brecha en el desarrollo tecnológico-económico se ha incrementado, aunque también han surgido oportunidades que requieren personas calificadas. La desigualdad social, considerándola principalmente a partir de la distribución de ingresos, constituye el indicador más aceptado para medir esos efectos, así como los impactos que ésta representa en las posibilidades de desarrollo.

El proceso enseñanza–aprendizaje necesario para la formación de personas calificadas enfrenta en México, como en otros países, dos problemas: el primero es la deserción escolar, el segundo es el bajo rendimiento académico. La deserción escolar ocurre cuando una persona abandona sus estudios completamente, con esto, la persona en general estará sin oportunidades para acceder a las oportunidades propias del proceso de globalización. El segundo problema tiene como consecuencia la formación de profesionistas con bajo rendimiento que tienen dificultades para desempeñar su actividad profesional en forma adecuada.

En este trabajo se aplicarán técnicas de minería de datos para encontrar los factores que son causantes de la problemática en los aspectos académicos. Entre los factores que influyen en este problema están: los recursos financieros, el interés de los estudiantes por alcanzar estudios superiores y el entorno social. En muchas ocasiones, algunos estudiantes no abandonan sus estudios, pero presentan calificaciones deficientes reflejo de comprender parcialmente los temas curriculares. Estos problemas tienen consecuencias laborales cuando la persona enfrenta dificultades para encontrar empleo bien remunerado o en el ejercicio deficiente de su profesión [6]. Por lo tanto, las consideraciones y prospectos sobre la educación de una entidad como el Instituto Nacional para la Evaluación de la Educación (INEE) de México, son afectados directamente resultando en una sociedad con mayor desigualdad social, menor productividad, etc. Entonces una de las causas de la problemática presentada, como lo es el sistema educativo mexicano, puede verse reflejada en los resultados de la prueba PISA donde México ha ocupado un lugar por debajo de la media entre los países de la OCDE.

Los rezagos en materia educativa de un país impiden la construcción de una sociedad más próspera. Si bien el sistema educativo mexicano se ha expandido a lo largo de su historia, padece de inequidad y baja calidad. Según el INEE, se espera que la educación produzca resultados sociales como una mayor participación ciudadana, mayor productividad, mejores empleos y salarios. También el INEE considera que el derecho a la educación combate la inequidad y permite la construcción de sociedades progresivamente más justas, participativas, democráticas, saludables y tecnológicas.

Proyectos similares que se han llevado a cabo:

Nombre	Características
The Analysis of Mathematics Academic Burden for Primary School Students Based on PISA Data Analysis [9]	Esta investigación toma en cuenta únicamente los resultados de la prueba PISA de 2018 y muestra que la carga académica de los estudiantes chinos de nivel primaria es pesada, lo que provoca ansiedad de aprendizaje en los estudiantes. Con este trabajo de investigación, el departamento de gobernanza de datos correspondiente propone una serie de soluciones para asegurar el desarrollo de la salud física y mental de los estudiantes de primaria de China.
International Comparative Study on PISA Mathematics Achievement Test Based on Cognitive Diagnostic Models [10]	El trabajo considera la prueba PISA del 2018 y sus previas aplicaciones a partir del año 2000 y está apoyado por el Consejo de Becas de China y el Fondo Juvenil de Planificación de Filosofía y Ciencias Sociales de Guizhou. Lo que se pretende con los resultados de este trabajo de investigación es construir un modelo cognitivo que se ajuste a las necesidades de los estudiantes de la educación básica de china, haciendo énfasis en la alfabetización matemática.
La desigualdad de los resultados educativos en Latinoamérica: un análisis desde PISA [11]	Se realiza un análisis sobre la definición de calidad educativa, de la desigualdad en educación y de cómo ésta ha sido estudiada en los reportes de evaluación educativa. Con base en las pruebas de Matemáticas y Lectura de PISA 2015, se realiza un análisis de los resultados focalizándose en países latinoamericanos y en la discusión sobre la dispersión y forma de la distribución de los puntajes. Este trabajo propone al cálculo del coeficiente de variación y del coeficiente de asimetría como alternativas que permiten brindar información acerca de la calidad de los sistemas educativos desde la equidad

Tabla 1. Resumen de proyectos similares.

Estos trabajos realizaron un análisis de datos de la prueba PISA sobre los mismos periodos que nos compete analizar, sin embargo, cada trabajo tiene un objetivo en particular el cual corresponde al país o región de origen de los realizadores. Durante la elaboración de este protocolo, no se encontró ningún otro proyecto de investigación o trabajo terminal que comparta o cuente con un objetivo similar al propuesto.

2. Objetivo

Analizar los conjuntos de datos «datasets» de la prueba internacional de rendimiento académico PISA conjuntamente con datos del INEGI para la detección de patrones de rendimiento. En la prueba PISA también se utilizarán datos de las condiciones materiales, emocionales y sociales de los estudiantes y escuelas que fueron partícipes de la prueba PISA en los años 2018 y 2015, en caso que estén disponibles también los datos de 2022.

Objetivos particulares:

1. Identificar las fuentes de datos que se utilizarán en este trabajo terminal.
2. Aplicar análisis exploratorio de datos, junto con la limpieza correspondiente en caso que se requiera.
3. Aplicar la metodología de minería y análisis de datos, utilizando los resultados de la prueba correspondiente al año 2022 en la que se espera se refleje el impacto de la pandemia SARS-COV-2 sobre el rendimiento académico a nivel internacional con la finalidad de determinar las consecuencias del cambio de paradigma en el proceso enseñanza–aprendizaje presencial a la modalidad remota soportada por redes de computadoras.
4. Realizar un procesamiento y análisis de datos numéricos y textuales.
5. Elaborar presentaciones visualizaciones de los resultados que se obtengan como parte del estudio.

3. Justificación

Conforme a los datos proporcionados por el INEGI sobre la tasa de abandono escolar por entidad federativa según nivel educativo 2000/2001 a 2020/2021 [4], México ha disminuido su tasa de abandono escolar en lo que respecta a los niveles educativos que conforman la educación básica. No obstante, esta tendencia hacia abajo en la tasa de abandono escolar no es el caso en la educación superior a nivel nacional.

El Sistema de Educación Superior de México se ha caracterizado, en las últimas décadas, por un importante crecimiento tanto en matrícula, como en instituciones. De acuerdo con datos de la Secretaría de Educación Pública, en el ciclo escolar 2019-2020, el Sistema Educativo Nacional registró una matrícula total en Educación Superior de 4 millones de alumnos inscritos, en más de 42 mil Programas de Licenciatura y Posgrado, ofrecidos en 4 mil Instituciones de Educación Superior. [3]

Tomando de ejemplo a nuestra casa de estudios, de acuerdo con los resultados publicados por QS World University Ranking, en su edición 2021, el Instituto Politécnico Nacional se ubicó en el rango de las posiciones 701 a 750. En The World University Ranking, el IPN participó por primera vez en la edición 2018; posicionándose en el rango de lugares del 1001 en adelante, de un listado de 1 mil 526 instituciones. El Academic Ranking of World Universities (ARWU) parte de un listado clasificado de más de 2 mil universidades de las cuales publica las mejores mil. Desde su primera participación en 2017, el IPN ha ocupado posiciones entre los lugares 501 al 600. Finalmente, en el University Ranking by Academic Performance (URAP), que se enfoca en valorar la calidad académica de la Educación Superior en el mundo, recopilando datos de 3 mil instituciones, en un esfuerzo por clasificarlas de acuerdo con su rendimiento, el Politécnico ocupó la posición 570 en el ciclo 2020-2021. [3]

Considerando los bajos resultados obtenidos en los rankings internacionales a través de los años, el incremento en las matrículas de las instituciones de educación superior y la transición hacia la educación a distancia de emergencia debido a la pandemia SARS-COV-2 puesto que la solución adoptada asume que tanto estudiantes como docentes disponen del equipamiento y de la conectividad requeridos. [5] Es innegable llegar a la conclusión de que la calidad educativa dentro del IPN ha decaído a lo largo de los últimos 6 años.

Partiendo de este escenario planteado y tomando en cuenta el panorama de la cuarta revolución industrial, los desafíos del siglo XXI exigen la innovación en los modelos educativos, para alcanzar la excelencia académica. [3]

Algunos estudios caracterizan el conocimiento por las diferentes habilidades que se adquieren durante el proceso enseñanza-aprendizaje, entre estas: memorización de conceptos, análisis de textos y síntesis de informes. [6] La OCDE presenta la prueba PISA, la cual tiene por objeto evaluar hasta qué punto los alumnos cercanos al final de la educación obligatoria (nivel medio superior) han adquirido algunos de los conocimientos y habilidades necesarios para la participación plena en la sociedad del saber. El Marco analítico y de evaluación de la prueba PISA presenta definiciones y descripciones más detalladas de los aspectos académicos evaluados los cuales son: La competencia lectora, la alfabetización matemática y la alfabetización científica. A través

de esta prueba, se estudian igualmente una gama de resultados educativos entre los que se encuentran: la motivación de los alumnos por aprender, la concepción que éstos tienen sobre sí mismos y sus estrategias de aprendizaje, así como una serie de criterios sobre los docentes y la administración de las instituciones donde fue aplicada la prueba. Los resultados de esta prueba internacional la cual es aplicada cada 3 años, son públicos y están disponibles en el sitio oficial de la OCDE. Dentro de estos resultados, se encuentran los datasets correspondientes a las respuestas de los cuestionarios de la prueba en cada periodo desde el 2009. Si bien, esta prueba es aplicada a estudiantes de nivel medio superior, podemos aplicar estrategias de análisis de datos sobre los datasets de la prueba de PISA con el propósito de identificar las problemáticas que han resultado en el incremento de la tasa de abandono escolar en la educación de nivel superior. Finalmente utilizando los resultados del trabajo de análisis de datos, se desplegarán una serie de indicadores clave de rendimiento presentados con un reporte dirigido al departamento de gobernanza de datos correspondiente, con el cual esta podrá elaborar o en su defecto encontrar prácticas educativas que integren herramientas de las Tecnologías de la Información y la Comunicación en conjunto con estrategias pedagógicas apoyadas en un modelo de gestión de conocimiento, para propiciar un aprendizaje que se ajuste a las modalidades presenciales y en línea con el propósito de mejorar el rendimiento académico.

4. Resultados Esperados

En la figura 1 se muestra un diagrama del funcionamiento general del producto esperado, siendo utilizado por el departamento de gobernanza de datos para la toma decisiones que puede servir de apoyo en el futuro para elaborar estrategias que puedan incrementar la calidad educativa en la institución, así como fomentar un ambiente educativo que priorice la innovación y proactividad en los estudiantes.

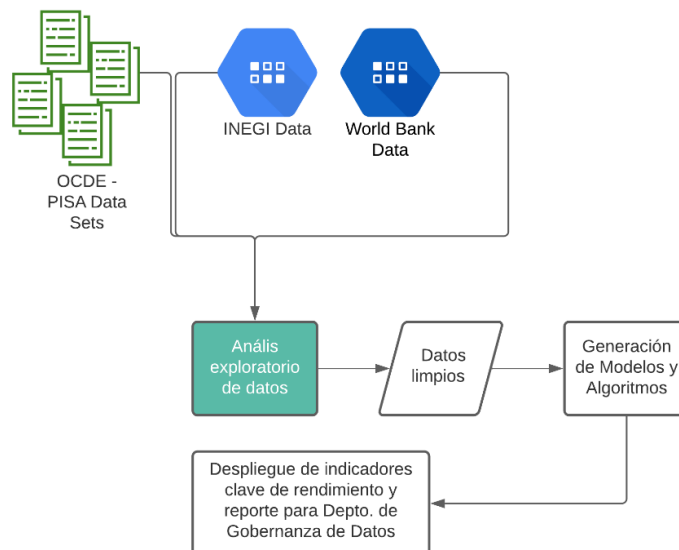


Figura 1 Diagrama de bloques del resultado esperado

5. Metodología

El término Descubrimiento de Conocimiento en Bases de Datos o KDD por sus siglas en inglés, fue acuñado en 1989 para referirse al amplio proceso de encontrar conocimiento en datos y para enfatizar la aplicación de “alto nivel” de métodos particulares de Minería de Datos. KDD es el proceso de usar métodos de minería de datos para extraer lo que se considera conocimiento de acuerdo con la especificación de medidas y umbrales, usando una base de datos junto con cualquier preprocesamiento, submuestreo y transformación requeridos de la misma. Este proceso considera cinco etapas: Selección, Preprocesamiento, Transformación, Minería de datos y la Interpretación / Evaluación. [8]

En resumen, dentro de estas etapas el proceso parte de crear un conjunto de datos objetivo, realizar una limpieza y preprocesamiento de los datos de destino para obtener datos consistentes, transformar los datos utilizando métodos de transformación o reducción de dimensionalidad para entonces comenzar la búsqueda de patrones de interés en una forma de representación particular y finalmente interpretar y evaluar los patrones extraídos.

Bajo este proceso, se han construido un conjunto de metodologías para proyectos de minería de datos. Se ha considerado las dos metodologías vigentes más usadas según la comunidad de KDnuggets (Data Mining Community's Top Resource). [7] Estas metodologías son: SEMMA y CRISP-DM y ambas pueden verse como una implementación del proceso KDD, sin embargo, se

ha llegado a la conclusión de que la metodología CRISP-DM es la más completa. [8] Por lo que se ha decidido realizar este trabajo siguiendo dicha metodología.

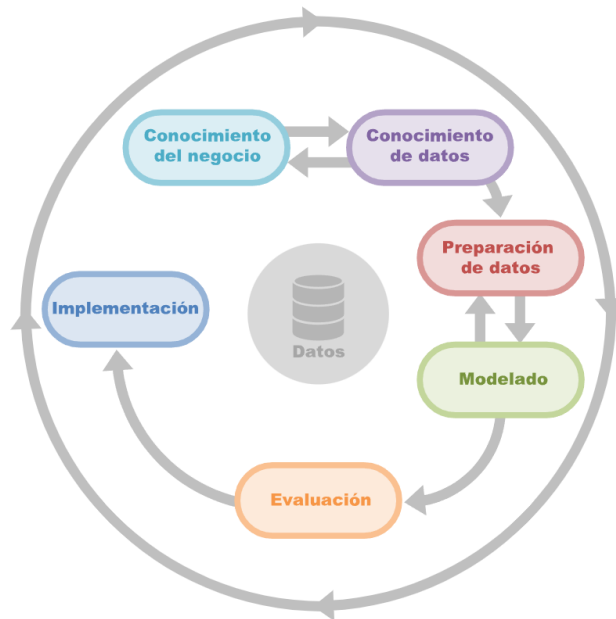


Figura 2 Diagrama de la metodología CRISP-DM

A continuación, se listan los pasos de la metodología CRISP-DM como se planean aplicar en este trabajo.

- **Comprensión/Entendimiento del negocio**
 - Comprensión del proceso enseñanza-aprendizaje.
 - Documentarse de los distintos métodos educativos y taxonomías del conocimiento, así como de los indicadores de rendimiento académico como criterios para la evaluación de la calidad.
- **Comprensión de los datos**
 - Realizar la búsqueda e identificación de las fuentes de datos a utilizar, enfocándonos en los datasets proporcionados por la OCDE de los resultados de las aplicaciones de la prueba PISA.
 - Realizar el análisis exploratorio de datos de los datasets obtenidos.
- **Preparación de los datos**
 - Identificación y distribución de datos faltantes y la identificación de datos atípicos y erróneos.
 - Normalización y estandarización de datos.
 - Imputación de datos.
 - Evaluar la calidad de datos.
- **Modelado**
 - Aplicación de diferentes técnicas de minería de datos como patrones frecuentes, árboles de decisión, agrupaciones y clasificaciones.
- **Evaluación**
 - Se usarán métricas como precisión, sensibilidad, exactitud, F1-Score, dependiendo del resultado de la minería.
- **Despliegue/Implementación**
 - Elaboración de un tablero de control en el cual se muestren los indicadores clave de rendimiento mediante tablas y elementos gráficos.

6. Cronograma

Nombre del alumno(a): Morgado Jiménez Arturo

TT No.: _____

Título del TT: Análisis del rendimiento educativo con base en datos de la prueba PISA.

[illegible]

7. Referencias

- [1] J. R. Yebeles y M. Zorrilla, "Data Governance Framework for Industry 4.0", IEEE Latin America Transactions, mayo de 2021.
- [2] C. Quintero Jaramillo, "El derecho humano a la educación y el servicio público de educación superior en México", Biblioteca Jurídica Virtual del Instituto de Investigaciones Jurídicas de la UNAM, Ciudad de México, México, 2017.
- [3] Dirección general del Instituto Politécnico Nacional. "Documento rector de planeación, Actualización del Programa de Desarrollo Institucional 2019 – 2024", Instituto Politécnico Nacional, Ciudad de México, México. 23 de abril de 2021.
- [4] INEGI, "Tasa de abandono escolar por entidad federativa según nivel educativo, ciclos escolares seleccionados de 2000/2001 a 2020/2021", <https://inegi.org.mx/app/tabulados/interactivos/?pxq=9171df60-8e9e-4417-932e-9b80593216ee> (accedido el 29 de octubre de 2021).
- [5] M.B. Vilizzio, H. Camacho, D.C. Alarcón, etc, *La educación superior en Iberoamérica en tiempos de pandemia Impacto y respuestas docentes*. Madrid - España: Fundación Carolina, 2021.
- [6] C. López López, JM. Olivares Ceja, CC. Vasquez Rincón y JA. Arguelles Cruz, "Buscando las mejores prácticas académicas integrando u-learning y pedagogía" en IEEE ROC&C, Acapulco, Gro, México, 8 marzo 2019.
- [7] J.M. Moine, A.S. Haedo, S. Gordillo, " Estudio comparativo de metodologías para minería de datos", Grupo de investigación en Minería de Datos, UTN Rosario, Facultad de Ciencias Exactas, UBA, Facultad de Informática, UNLP, 2011.
- [8] A. Azevedo, M.F Santos, " KDD, SEMMA and CRISP-DM: A parallel overview", en IADIS European Conference Data Mining 2008, Praga, República Checa, 2008.
- [9] W. Li, " The Analysis of Mathematics Academic Burden for Primary School Students Based on PISA Data Analysis", Frontiers in Psychology, vol. 12, art 229, 2021.
- [10] W. Xiaopeng, W. Rongxiu, C. Hua-Hua, K. Qiping, and Z. Yi, " International Comparative Study on PISA Mathematics Achievement Test Based on Cognitive Diagnostic Models", Frontiers in Psychology, vol. 11, art 2230, 2020.
- [11] L. Medina Gual, "La desigualdad de los resultados educativos en Latinoamérica: un análisis desde PISA." Revista Latinoamericana de Estudios Educativos, Vol. XLVIII, núm.2, pp.45-70, diciembre 2018.

8. Alumno y directores

Arturo Morgado Jiménez.- Alumno de la carrera de Ingeniería en Sistemas Computacionales en ESCOM, Especialidad Sistemas, Boleta: 2015030904, Tel. 5586761099 , email: arturomorgado117@hotmail.com.

Firma:_____

Jesús Manuel Olivares Ceja.- actualmente labora como profesor–investigador de tiempo completo en el CIC–IPN. Tiene el grado de Doctor en Ciencias de la Computación egresado del CIC–IPN en 2002. En 1996 concluyó la Maestría en Ciencias de la Computación en el CINVESTAV, México. En 1991 obtuvo el título de Licenciado en Ciencias de la Computación en la UPIICSA–IPN. Actualmente desarrolla las líneas de investigación de Ciencia de Datos, Modelación del conocimiento, Reconocimiento de patrones para visión por computadora e Informática educativa. Número de empleado 25164, celular 55 1498 1020, email: jolivaresc@ipn.mx

Firma:_____

Nidia Asunción Cortez Duarte. - Maestra en Ciencias en Computación CINVESTAV-IPN 2009, Ing. en Sistemas Computacionales ESCOM-IPN 2006, Profesora Titular en ESCOM Depto. de Ingeniería en Sistemas Computacionales. Áreas de interés: criptografía, seguridad de información, hardware reconfigurable, aritmética computacional, diseño digital. Teléfono: 57-29-6000 ext. 52032, ncortezd@ipn.mx

Firma:_____

CARÁCTER: Confidencial

FUNDAMENTO LEGAL: Artículo 11 Fracc. V y Artículos 108, 113 y 117 de la Ley Federal de Transparencia y Acceso a la Información Pública.

PARTES CONFIDENCIALES: Número de boleta y teléfono.