

Sistema de Categorización y Visualización de Escuelas de Nivel Básico por desempeño académico y estrato socioeconómico

“SCV DAYES”

Trabajo Terminal No. 2021-B048

*Alumnos: Rubio de la Llera Ricardo Emilio, *Zubillaga Trillo Blas*

Directores: M. en C. Hernández Rubio Erika, Dr. Meneses Viveros Amilcar

**email: bzt2799@gmail.com*

Resumen – Desarrollar un sistema que presente el estrato socioeconómico de una zona geográfica y represente a las escuelas de nivel básico por desempeño académico. Se obtendrán los estratos socioeconómicos de las regiones geográficas según los datos del Instituto Nacional de Estadística y Geografía (INEGI) y del censo de población y vivienda 2020. El desempeño de la escuela se obtendrá a partir de los datos de la prueba PLANEA Educación Básica a través de algoritmos de clasificación y agrupamiento de minería de datos. El sistema tendrá una visualización y agrupamiento de minería de datos con una visualización geográfica interactiva e intuitiva. Con el desarrollo de este sistema se busca tener una herramienta que sirva a instituciones públicas o privadas en la toma de decisiones en la identificación y atención de las instituciones educativas que se vean afectadas académicamente por el factores del entorno socioeconómico en el que se encuentran.

Palabras clave – Estrato Socioeconómico, Minería de Datos, Rendimiento Académico, Visualización Geográfica.

1. Introducción

A través de la historia y del desarrollo de la educación pública en México, han existido diferentes épocas en las que las características y la calidad de la educación han crecido y se han visto envueltas en diferentes condiciones de atraso y de innovación para el país, es por ello que, durante el desarrollo de la modernidad y de la tecnología, es importante identificar y atender las necesidades del crecimiento en la educación pública mexicana y especialmente en los niveles básicos, en donde la calidad de la educación se forma hasta una educación superior o universitaria.

Dependiendo de los logros, metas y objetivos previstos alcanzados, se puede determinar la calidad de una escuela, además la importancia de obtener el nivel adecuado de calidad a través de la adquisición de los conocimientos básicos, como ejemplos se puede mencionar el realizar operaciones básicas de matemáticas (suma, resta, multiplicación, división).

Un segundo punto de vista se refiere a considerar la calidad en términos de relevancia. En este sentido los programas educativos de calidad serán aquellos que incluyan contenidos valiosos y útiles: que respondan a los requerimientos necesarios para formar integralmente al alumno, para preparar excelentes profesionales, acordes con las necesidades sociales, o bien que provean de herramientas valiosas para el trabajo o la integración del individuo a la sociedad.

Una tercera perspectiva del concepto de calidad se refiere a recursos y procesos. Un plan de alta calidad será un plan que tenga los recursos necesarios y los utilice de manera eficaz. Por tanto, óptimas plantas físicas, laboratorios, programas de formación docente, buenos sistemas académicos o administrativos, técnicas de enseñanza adecuadas y equipamiento suficiente son necesarios para lograr la calidad.

Es por ello que, a través de diferentes etapas, el gobierno mexicano ha tratado de conocer las necesidades y las condiciones del país a través de diferentes instituciones, por ejemplo, el Instituto Nacional de Estadística y Geografía (INEGI), el cual por casi más de dos décadas ha recabado distinta información estadística y geográfica por lo que, al reconocer la importancia de toda la información recabada por el instituto, resulta de forma interesante hacer uso de esta información en el apoyo a la atención de una problema social como lo es la educación pública y especialmente la educación pública básica.

INEGI realiza el censo de población y vivienda cada 10 años. A partir de estos datos puede hacer estimados de indicadores como el de estratos socioeconómicos de regiones geográficas, el cual se calcula de datos como grado máximo de estudio, ingreso mensual, y la infraestructura con la que cuenta una vivienda, entre otras. INEGI considera 7 estratos socioeconómicos (el más bajo es el estrato socioeconómico 1, donde se carecen de servicios básicos, y el estrato económico más alto es el 7).

Si bien, se han identificado factores pedagógicos que afectan el desempeño escolar dentro del aula, también se debe considerar el contexto donde se encuentra ubicada una escuela. Por ejemplo, como menciona Romero [2] se presenta que hay una fuerte correlación entre el uso del internet y el desempeño a través de la prueba enlace. Sin embargo, si la escuela se encuentra en una zona geográfica con un estrato socioeconómico bajo, es muy probable que no tenga posibilidad de brindar servicios de internet a sus alumnos. De esta forma resulta de particular interés poder identificar y asociar el desempeño de las escuelas al estrato socioeconómico donde se localizan.

En complemento y en la preparación profesional como Ingenieros en Sistemas Computacionales, en el equipo de trabajo, resulta interesante hacer uso de la información geográfica y estadística recabada por el Instituto Nacional de Estadística y Geografía (INEGI) y por otras herramientas que evalúan la calidad del aprendizaje en la sociedad mexicana como la Prueba PLANEA Educación Básica realizada a través del Censo de Escuelas, Maestros y Alumnos de Educación Básica y Especial (CEMABE), haciendo uso y reconociendo la importancia de la explotación de datos para encontrar una posible relación entre nivel académico y condiciones socioeconómicas con el fin de notificar y reconocer las relaciones entre estas que provoquen al déficit académico.

Gracias a las herramientas digitales web que proporciona el Instituto Nacional de Estadística y Geografía (INEGI) para la visualización de datos [4,5], la sociedad en su generalidad o algunas otras entidades pueden hacer uso de ella y obtenerlas en diferentes formatos de lectura, sin embargo dentro de las herramientas que proporciona no incluyen el análisis en la relación de la calidad o del desempeño académico en la educación pública básica, sin embargo, nos ofrece las herramientas geográficas en donde se localizan así como el número de la población que cursa la educación básica, en su complemento, a los resultados de las evaluaciones de la misma forma se publican y pueden ser obtenidos, de esta forma, uno de los primeros pasos es reconocer la forma en que las instituciones nos proporcionan el conjunto de datos.

De esta forma, se pretende que a través de los conocimientos adquiridos a través de la preparación como ingenieros, se procese la información a través de diferentes herramientas y metodologías para la explotación de datos. Por otro lado, la última información que se tiene en relación a los estratos socioeconómicos de INEGI se realizó en 2004 con los datos del censo de población y vivienda 2020. Diversas instituciones públicas y privadas han realizado la computación del estrato socio económico a partir de los datos del censo de población y vivienda 2010, pero no son públicos. En marzo de del 2021, INEGI liberó los datos del censo de población y vivienda 2020. Estos datos aún no están totalmente procesados para obtener indicadores como el del estrato socioeconómico. Por lo que se requiere hacer este procesamiento. Esto es, tener actualizado el valor de los indicadores.

Finalmente a través de la integración de diferentes tecnologías que apoyan la lectura y el análisis de los datos para la explotación de la información se pretende elaborar un sistema que sea un auxiliar para la visualización de las relaciones que pueden existir entre el desempeño académico y las condiciones socioeconómicas asemejándose a las ventajas de las herramientas web que nos proporciona el Instituto Nacional de Estadística y Geografía (INEGI).

En la Tabla 1 se encuentran los productos o sistemas que son similares a nuestra propuesta. Actualmente el conjunto de herramientas web que proporciona el INEGI son las únicas herramientas similares que se han encontrado con características similares, además, indagando a través del repositorio de trabajos terminales y tesis [8] de la Escuela Superior de Cómputo y otras instituciones de educación pertenecientes al Instituto Politécnico Nacional, se encontraron trabajos que hace uso de la minería de datos, pero utilizados hacia otros enfoques, como por ejemplo, la administración, la medicina, sin embargo no se encontró un trabajo similar con un enfoque al desempeño académico. De la misma forma se encontraron trabajos que hacen uso de un sistema de visualización geográfica, sin embargo no presentan las relaciones anteriormente mencionadas.

Además y de acuerdo a los resultados obtenidos por Romero [2], estableció resultados acerca de la calidad educativa de las escuelas de Nivel Básico en México, sin embargo, al hacer uso de un conjunto de información de los resultados de la prueba ENLACE, para los motivos de este trabajo, ya que resultan de poca relevancia ya que estos datos no se encuentran actualizados a comparación de los que serán utilizados de la Prueba PLANEA Educación Básica y además no muestran una relación con el estrato socioeconómico, es por ello que en la siguiente Tabla 1 no son mencionados.

SOFTWARE	CARACTERÍSTICAS	PRECIO EN EL MERCADO
Herramientas Web INEGI	Aplicaciones web que permiten observar la información estadística general y por rubros generalizados a través de elementos gráficos como gráficos de barra, histogramas sin realizar ningún tipo de comparación o relación entre la información que se presenta de forma visual [4].	Gratuito
Clasificación Por nivel socioeconómico de las regiones geográficas de México	Trabajo de fin de máster que presenta desarrollo para el análisis de datos relacionados a las condiciones socioeconómicas de las regiones geográficas de México, así como su procesamiento, establece correlaciones y realiza una clasificación, sin embargo, los resultados concluyentes solo son presentados de una forma analítica, sin ser representados en un sistema de visualización geográfica, ni establece un tipo de relación al desempeño académico de dichas zonas geográficas [9].	Gratuito
Solución Propuesta	Aplicación web que permiten observar la información estadística a través de elementos gráficos como un sistema de visualización geográfica interactiva, realizando un tipo de comparación y relación entre el desempeño académico y el estrato socioeconómico	Gratuito

Tabla 1. Resumen de productos similares.

2. Objetivo

General – Desarrollar un sistema que sirva de apoyo a organismos privados o públicos con el interés que, a través de un sistema de georeferencia permita ver una relación entre el desempeño académico de unidades académicas de Nivel Básico localizadas en las Regiones de la República Mexicana y las variables que determinan el estrato socioeconómico, mostrando los posibles problemas que están afectando directamente el nivel académico a través de la visualización geográfica de la relación entre los factores del estrato socioeconómico y el desempeño académico, el cual se obtendrá con base a la información del Censo de Escuelas, Maestros y Alumnos de Educación Básica y Especial (CEMABE) y el estrato socioeconómico se calculará con los indicadores del Instituto Nacional de Estadística y Geografía (INEGI) a partir de los datos del Censo de Población y Vivienda del 2020.

Específicos

- Separar la información relacionada a la extensión territorial subdividida nombrada como el Área Geoestadística Básica (AGEB), de la información que proporciona el Instituto Nacional de Estadística y Geografía (INEGI).
- Relacionar la ubicación geográfica de las unidades académicas de Nivel Básico en la República Mexicana con la información relacionada a la extensión territorial subdividida (AGEB).
- Separar y relacionar los niveles socioeconómicos determinados por el INEGI con las extensiones territoriales subdivididas (AGEB), de la información que proporciona el Instituto Nacional de Estadística y Geografía (INEGI).
- Categorizar a las unidades académicas de Nivel Básico en la República Mexicana por desempeño académico tomando como criterio los resultados obtenidos a través de la aplicación de la Prueba PLANEA Educación Básica realizada a través del Censo de Escuelas, Maestros y Alumnos de Educación Básica y Especial (CEMABE).
- Analizar el conjunto de algoritmos de clasificación y agrupamiento para aplicar la minería de datos sobre la información separada y categorizada para establecer una correlación entre los niveles socioeconómicos y el desempeño académico.
- Escoger entre el conjunto de algoritmos de clasificación y agrupamiento para aplicar la minería de datos, el más adecuado para establecer una correlación entre los niveles socioeconómicos y el desempeño académico.

3. Justificación

El proyecto que se desarrolla a través de este documento resulta importante atenderlo, ya que este contribuye a una de las principales problemáticas sociales en la República Mexicana, es decir, el déficit académico en la escuelas de Nivel Básico ahora bien no para atender directamente al problema, si no para notificar a través de un sistema de georeferencia las posibles relaciones con el entorno socioeconómico que provoquen esta deficiencia académica, buscando que organizaciones públicas como la Secretaría de Educación Pública (SEP) u otras instituciones educativas públicas o privadas en la República Mexicana puedan atender aquellas posibles causas que generan este comportamiento problemático.

Actualmente existen sistemas o conjuntos de datos contenidos en archivos que facilitan el despliegue y clasificación de información estadística, inclusive hoy en día resulta más fácil encontrar aplicaciones web o de escritorio que facilitan la visualización de cierto tipo de información. Sin embargo, al existir diferentes herramientas que despliegan esta información almacenada, no sucede con la información que relaciona específicamente al nivel académico o el desempeño académico con los niveles socioeconómicos.

Sitios web como los que proporciona el Instituto Nacional de Estadística y Geografía (INEGI) son las principales herramientas para visualizar y extraer información estadística, que de forma general, son fáciles de operar en la visualización de los datos concretos pero sin ningún tipo de relación que se puedan encontrar o calcular entre ellos.

Además, en estas herramientas no se pueden realizar bajo un tipo de lenguaje de consulta estructurado para bases de datos relacionales, lo que impide tener un sistema de consultas generalizado. También la estructura de los datos no permite trabajarlos de una forma sencilla, por ejemplo, se sabe que INEGI tiene una clasificación por estados, municipios y por Área Geoestadística Básica (AGEB), y en la mayoría de los casos los sistemas de información solo manejan los estados y municipios, lo que dificulta establecer las relaciones entre escuelas, estrato socioeconómico de la zona y los servicios con los que cuenta la zona, debido a que es la clasificación de estrato socioeconómico y los servicios se realizan a nivel AGEB. Más aún la información pública de escuelas solamente contiene la infraestructura con la que cuenta la escuela (número de salones, áreas deportivas, baños, laboratorios) y no se encuentra en el INEGI sino en la SEP. Por lo tanto la información proviene de diferentes fuentes de datos y se requiere un trabajo más sofisticado para extraer el conocimiento. Se aprecia que para consultas más sofisticadas es necesario extraer información de diferentes fuentes y aplicar minería de datos.

Es por ello que en la actualidad, las posibilidades que nos brindan las ciencias de la computación y en particular la minería de datos, es un factor que posibilita la creación de un sistema capaz de procesar la cantidad de información que se genera y almacena respecto a las condiciones socioeconómicas, geográficas y de desempeño académico implementando una interfaz gráfica e intuitividad para la visualización específica de la relación del nivel académico o el desempeño académico con el estrato socioeconómico.

4. Productos o Resultados esperados

A lo largo del desarrollo de dos periodos de evaluación se presentarán los siguientes resultados o productos, sobre los cuales se incluye la documentación que describa la utilización y funcionalidad, así como las cuestiones técnicas y de desarrollo relacionadas al sistema propuesto, de esta forma, se enlistan a continuación los puntos.

1. Documentación técnica del sistema.
2. Manual de Usuario.
3. Sistema.

A continuación se muestra la arquitectura del sistema, el cual, contendrá todo lo anteriormente descrito.

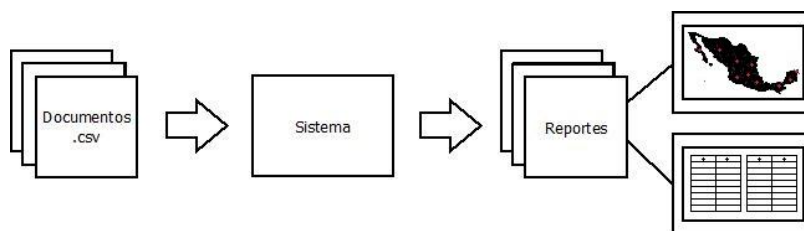


Figura 1. Arquitectura del Sistema.

5. Metodología

La metodología Cross Industry Standard Process for Data Mining (CRISP-DM) proporciona una descripción normalizada del ciclo de vida de un proyecto estándar de análisis de datos, de forma análoga a como se hace en la ingeniería del software con los modelos de ciclo de vida de desarrollo de software [1]. Este modelo cubre las fases de un proyecto, sus tareas respectivas, y las relaciones entre estas tareas. En este nivel de descripción no es posible identificar todas las relaciones; las relaciones podrían existir entre cualquier tarea según los objetivos, el contexto, y el interés del usuario sobre los datos. La secuencia de las fases en esta metodología no es rígida: se permite movimiento hacia adelante y hacia atrás entre diferentes fases.

El resultado de cada fase determina qué fase, o qué tarea particular de una fase, hay que hacer después. Del mismo modo contempla el proceso de análisis de datos como un proyecto profesional, estableciendo así un contexto mucho más rico que influye en la elaboración de los modelos. De esta forma, el ciclo de vida del proyecto de minería de datos consiste en las siguientes seis fases que se muestran y se describen a continuación y de qué forma serán incluidas durante este desarrollo [1,3].

Para este proyecto se utilizará la metodología CRISP-DM la cual está especializada en el desarrollo de aplicaciones y proyectos centrados en la minería de datos. De esta forma se planean lograr hasta las primeras cuatro fases del modelo en la primera entrega del Trabajo Terminal I, las cuales incluyen la recaudación de los requerimientos y los datos georeferenciales acerca de las unidades académicas y las condiciones socioeconómicas delimitadas por las AGEs (Business Understanding), comprensión de la información así como la verificación de la calidad de los datos (Data Understanding), la preparación, análisis, tratamiento y limpieza de la misma (Data Preparation) ya sea a través de herramientas que proporciona lenguajes de programación como Python, el modelado de la interfaz de usuario, así como la realización de un análisis conceptual de los datos para generar modelos analíticos los cuales serán integrados al sistema para el despliegue de los resultados que relacionan los dos conjuntos de datos (Modeling).

Una vez realizados todos estos procesos, se procederá con la evaluación de los modelos (Evaluation) para determinar entre estos, cual resulta ser el adecuado para establecer las relaciones entre el conjunto de datos mencionado, a través de la segunda entrega del Trabajo Terminal se finalizará desarrollando en la sexta fase de la metodología, ya que a través del (Deployment) es requerida de la implementación y el despliegue del sistema, así como sus pruebas con las características indicadas.

6. Cronograma

CRONOGRAMA. Alumno Rubio de la Llera Ricardo Emilio.

TT. No.: 2021-B048

Título del TT: Sistema de Categorización y Visualización de Escuelas de Nivel Básico por desempeño académico y estrato socioeconómico.

[illegible]

TT. No.: 2021-B048

Título del TT: Sistema de Categorización y Visualización de Escuelas de Nivel Básico por desempeño académico y estrato socioeconómico.

[illegible]

7. Referencias

- [1] J. V. Román (08 Agosto 2016). Singular Data & Analytics. [En línea]. Disponible en: <https://www.sngular.com/es/data-science-crisp-dm-metodologia/>. [Último acceso: 13 Septiembre 2021].
- [2] S. D. Romero Garcia, “Minería de datos para determinar la calidad educativa de las escuelas de Nivel Básico en México”, Trabajo fin de Máster, Departamento en computación, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, CDMX, México, 2019.
- [3] D. T. Larose & C. D. Larose, Data Mining and Predictive Analytics. Segunda Edición. Hoboken, New Jersey: Wiley, 2015.
- [4] INEGI (2021). “Portal Principal del Instituto Nacional de Estadística y Geografía”. [En línea]. Disponible en: <https://www.inegi.org.mx/>
- [5] INEGI (2021). “Censo de Población y Vivienda 2020”. [En línea]. Disponible en: <https://www.inegi.org.mx/programas/ccpv/2020/default.html>
- [6] CEMABE INEGI (2021). “Portal Principal del Censo de Escuelas, Maestros y Alumnos de Educación Básica y Especial”. [En línea]. Disponible en: <http://www.censo.sep.gob.mx/>
- [7] V. Galan, “Aplicación de la metodología crisp-dm a un proyecto de minería de datos en el entorno universitario”, Trabajo fin de carrera, Ingeniería en Informática, Universidad Carlos III de Madrid Escuela Politécnica Superior, Madrid, España, 2015.
- [8] DSpace Tesis IPN (2021). “Portal Principal del DSpace Repository del IPN”. [En línea]. Disponible en: <https://tesis.ipn.mx/>
- [9] R. M. Urrutia Hernández, “Clasificación Por nivel socioeconómico de las regiones geográficas de Mexico”, Trabajo fin de Máster, Departamento en computación, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, CDMX, México, 2019.

8. Alumnos y Directores

Rubio de la Llera Ricardo Emilio.- Alumno de la carrera de Ingeniería en Sistemas Computacionales en ESCOM, Especialidad en Sistemas, Boleta: 2014630444, Tel.5527558096, email: rribeiro259@gmail.com

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

Firma: _____

Zubillaga Trillo Blas.- Alumno de la carrera de Ingeniería en Sistemas Computacionales en ESCOM, Especialidad en Sistemas, Boleta: 2015081468, Tel. 5535153146, email: bzt2799@gmail.com

Firma: _____

M. en C. Erika Hernández Rubio.- Maestría en ciencias de la computación CIC-IPN, Licenciatura en ciencias de la informática UPIICSA-IPN , Áreas de Interés: Base de datos, minería de datos y cómputo móvil, Tel. 57296000 EXT 52061, email: ehernandezru@ipn.mx

Firma: _____

Dr. Amilcar Meneses Viveros - Dr. en Ingeniería Eléctrica con opción en computación, por CINVESTAV-IPN. Investigador Cinvestav, del Depto. de Computación, Áreas de interés: Computación Científica, Interacción Humano-Computadora y Computación Sustentable, email: ameneses@cs.cinvestav.mx, (55)574838800 ext 6551.

Firma: _____