

Identificación de Estilos de Aprendizaje en Alumnos Universitarios de Computación de la Huasteca Hidalguense mediante Técnicas de Minería de Datos

Felipe de Jesús Núñez Cardenas¹, Raúl Hernández Palacios¹, Víctor Tomás Tomás Mariano¹, Ana María Felipe Redondo¹

¹Universidad Autónoma del Estado de Hidalgo - Escuela Superior Huejutla. Corredor Industrial S/N, Parque de Poblamiento, Huejutla de Reyes, Hidalgo. CP. 43000. Tel. 771-7172000, ext. 5880,5881
e-mail: felipe.huejutla@hotmail.com

Resumen--- *Las carreras de Computación o áreas afines son consideradas como las más difíciles dentro de las instituciones de educación superior, sin embargo una forma de lograr mejores resultados en la impartición de la enseñanza en estas áreas consiste en detectar los estilos de aprendizaje de los alumnos con la finalidad de implementar estrategias didácticas acordes a la manera de aprender de ellos. Esta investigación se centra en la aplicación de técnicas de minería de datos para descubrir las combinaciones de estilos de aprendizaje mostradas por estudiantes de educación superior de computación o áreas afines, en la región de la Huasteca Hidalguense. El modelo seguido para determinar los estilos de aprendizaje es el desarrollado por Neil Fleming y Collen Mills llamado VARK [1] con el cual podemos determinar los estilos de aprendizaje de un alumno dentro de las cuatro modalidades que cubre este modelo Visual, Auditivo, Lectura/Escritura y Quinésico. Dado que nuestra hipótesis es que el estilo de aprendizaje dominante en los alumnos es el Quinésico. Los resultados obtenidos indican que existen estilos de aprendizaje dominantes en los diferentes niveles de la duración de la carrera.*

Palabras clave --- *Estilos de Aprendizaje, Minería de Datos, Agrupamiento*

1. ESTADO DEL ARTE

El problema referido en esta investigación corresponde al de describir los estilos de aprendizaje en los alumnos que cursan carreras de computación o afines en la Huasteca Hidalguense, el proyecto consiste en la investigación, elaboración y desarrollo de una propuesta de un modelo

operativo viable para solucionar problemas, requerimientos o necesidades de organizaciones o grupos sociales.

A partir del año 2006 surgen diversos trabajos relacionados con la aplicación de técnicas de minería de datos para describir y predecir estilos de aprendizaje en los alumnos, principalmente en Argentina tenemos "Identificación de Estilos de Aprendizaje dominantes en Alumnos de Informática" realizado en el 2006 [2] y "Minería de Datos para describir estilos de Aprendizaje" en el 2007 [3], realizados por Rossana Constaguta en la Universidad Nacional de Santiago del Estero en Argentina.

En el primero de ellos se utiliza el instrumento creado por Felder y Soloman [4] el cual consiste en cuarenta y cuatro preguntas con dos opciones de respuesta cada una, en las cuales el estudiante debe elegir necesariamente una de ellas, la herramienta utilizada para analizar la información recabada es WEKA [5] y la técnica empleada es Clustering, usando el algoritmo FarthestFirst. Utilizando gráficos de Torta para el Despliegue de los Datos. Los resultados obtenidos fueron los estilos de aprendizaje vinculados *Activo-Sensitivo-Visual-Secuencial* y *Activo-Sensitivo/Intuitivo-Visual-Secuencial*.

En el segundo utilizan el modelo de Felder y Silverman[6] el cual es particularmente aplicable a estudiantes de ingeniería, a diferencia del primer trabajo en este consideraron cinco dimensiones y no cuatro, el resto del proyecto coincidió con su primer trabajo. En los resultados agregan el

esquema de recomendaciones de enseñanza de acuerdo a los tipos de aprendizaje dominante.

En el año 2010 en Coahuila, México, Juan Ramón Olague Sánchez y Sócrates Torres Ovalle desarrollan el proyecto “Aplicación de Técnicas de Minería de Datos y Sistemas de Gestión de Contenidos de Aprendizaje para el Desarrollo de un Sistema Informático de Aprendizaje de Programación de Computadoras” [7], cuyo objetivo principal es que a través de la descripción de estilos de aprendizaje en los alumnos de la carrera de Ciencias Computacionales desarrollar software que facilitara a aprender mejor la asignatura de programación de computadoras.

2. DESARROLLO

2.1 Estilos de Aprendizaje

Para llevar a cabo esta investigación primeramente ha sido necesario conocer los diferentes modelos de estilos de aprendizaje que existen enfocados a estudiantes de ingeniería y ciencias, de tal manera que para este trabajo se centro en tres modelos. En el modelo de Kolb [8], la de Felder y Silverman [6] y el modelo VARK [1]. La teoría de Felder-Silverman [6] se ha utilizado en entender los estilos de aprendizaje de estudiantes en ingeniería y ciencias. Según esta teoría existen cuatro dimensiones dicótomas y en gradiente en el aprendizaje. La primera dimensión es la sensorial/intuitiva, la segunda es como se percibe la información de forma visual o verbal, la tercera es el procesamiento de información en la cual están los estilos secuenciales y globales y la cuarta la dimensión activo y reflexivo. La Teoría de Kolb [8] maneja las cuatro dimensiones de aprendizaje (afectiva, perceptiva, simbólica y comportamental), las cuales son representadas como un cono, cuya base representa los estados más bajos de desarrollo, y el vértice de ese desarrollo representa el hecho de que esas dimensiones se integran al máximo. En esta investigación se determinó que los dos instrumentos anteriores, que se utilizan para recabar información son demasiado extensos y propensos a errores por parte de los alumnos, por lo tanto, en este trabajo se utiliza un tercer modelo llamado VARK.

2.2 Modelo VARK

Neil Fleming y Colleen Mills [1] desarrollaron un instrumento para determinar la preferencia de los alumnos al procesar la información desde el punto de vista sensorial. Los autores no hablan de fortalezas, sino de preferencias sensoriales.

Descripción del modelo:

En el trabajo consideran que las personas reciben información constantemente a través de los sentidos y que el cerebro selecciona parte de esa información e ignora el resto. Las personas seleccionan la información a la que le prestan atención en función de sus intereses, pero también influye cómo se recibe la información. Si, por ejemplo, después de una excursión se le pide a un grupo de alumnos que describan alguno de los lugares que visitaron, probablemente cada uno de ellos hablará de cosas distintas. No puede recordarse todo lo que pasa, sino parte de lo que sucede en el entorno. Algunos se fijan más en la información visual, otros en la auditiva y otros en la que se recibe a través de los demás sentidos, o de la lectura y escritura.

El modelo toma el nombre de **VARK** por las siglas en inglés de las modalidades sensoriales que identificaron: Visual, Auditivo, Lectura y Quinésico.

2.3 Minería de Datos

La minería de datos, es un campo de las ciencias de la computación referido al proceso que intenta descubrir patrones en grandes volúmenes de conjuntos de datos. Utiliza los métodos de la inteligencia artificial, aprendizaje automático, estadística y sistemas de bases de datos. El objetivo general del proceso de minería de datos consiste en extraer información de un conjunto de datos y transformarla en una estructura comprensible para su uso posterior [9].

2.4 Técnica de Agrupamiento

Agrupar datos dentro de un número de clases preestablecidas o no, partiendo de criterios de distancia o similitud, de manera que las clases sean

similares entre sí y distintas con las otras clases. Su utilización ha proporcionado significativos resultados en lo que respecta a los clasificadores o reconocedores de patrones, como en el modelado de sistemas. Este método debido a su naturaleza flexible se puede combinar fácilmente con otro tipo de técnica de minería de datos, dando como resultado un sistema híbrido[10].

Un problema relacionado con el análisis de cluster es la selección de factores en tareas de clasificación, debido a que no todas las variables tienen la misma importancia al momento de agrupar los objetos. Otro problema de gran importancia y que actualmente despierta un gran interés es la fusión de conocimiento, ya que existen múltiples fuentes de información sobre un mismo tema, los cuales no utilizan una categorización homogénea de los objetos. Para poder solucionar estos inconvenientes es necesario fusionar la información al momento de recopilar, comparar o resumir los datos.

2.5 WEKA (Herramienta de Minería de Datos)

Weka (Entorno Waikato para el Análisis del Conocimiento) escrito en java es una conocida suite de software para máquinas de aprendizaje que soporta varias tareas típicas de minería de datos, especialmente pre procesamiento de datos, agrupamiento, clasificación, regresión, visualización y características de selección. Sus técnicas se basan en la hipótesis de que los datos están disponibles en un único archivo plano o relación, donde cada punto marcado es etiquetado por un número fijo de atributos. WEKA proporciona acceso a bases de datos SQL utilizando conectividad de bases de datos Java y puede procesar el resultado devuelto como una consulta de base de datos. Su interfaz de usuario principal es el Explorer, pero la misma funcionalidad puede ser accedida desde la línea de comandos o a través de la interfaz de flujo de conocimientos basada en componentes .

En atención a esta modalidad de investigación, se introdujeron dos fases en el estudio, a fin de cumplir con los requisitos involucrados en este proyecto. En la primera de ellos inicialmente se desarrolló una descripción de los estilos de

aprendizaje de los alumnos. En la segunda fase aún en proceso y atendiendo a los resultados de la descripción obtenida se presentaran propuestas de estilos de enseñanza acordes a los resultados de estilos de aprendizaje predominantes con la finalidad de mejorar los resultados académicos de los alumnos de computación en la región de la Huasteca Hidalguense.

2.6 Metodología

Para la realización de esta investigación se ha elegido la metodología CRISP-DM [11] que se enfoca a la implementación de minería de datos como lo muestra la figura 1. El estándar incluye un modelo y una guía, estructurados en seis fases, algunas de estas fases son bidireccionales, lo que significa que algunas fases permitirán revisar parcial o totalmente las fases anteriores.

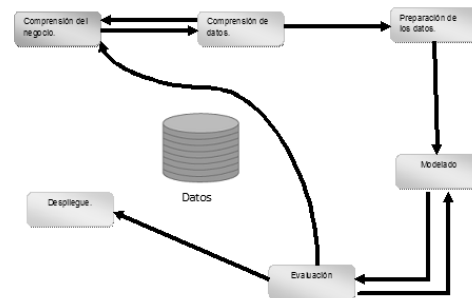


figura 1 Metodología CRISP-DM

Las fases son: a) comprensión del negocio o caso de estudio, b) comprensión de los datos, c) Preparación de los datos, d) Modelado, e) Evaluación y f) Despliegue.

3. RESULTADOS

3.1 Comprensión del Caso de Estudio

Población: Para el caso de esta investigación se consideran todos los alumnos que cursan carreras de Computación o afines dentro de las instituciones Superiores de la Huasteca Hidalguense, tal es el caso de la Ingeniera en Tecnologías de la Información impartida en la Universidad Tecnológica de la Huasteca Hidalguense, la Ingeniería en Sistemas Computacionales y Licenciatura en Informática impartidas en el Instituto Tecnológico de Huejutla y la Licenciatura en Sistemas Computacionales impartida en la

Escuela Superior Huejutla de la Universidad Autónoma del Estado de Hidalgo, todas ellas ubicadas en la ciudad de Huejutla de Reyes Hidalgo. Muestra: Considerando que todas las instituciones se encuentran localizadas en la ciudad de Huejutla de Reyes , Hidalgo, el instrumento se aplica al 80% de la población que equivale a 850 alumnos aproximadamente.

El instrumento para la recolección de la información esta conformado por 16 preguntas con cuatro opciones cada una, las cuales representa cada una de las modalidades, el alumno puede responder una o más opciones, o dejar en blanco si no aplica a sus preferencias, a continuación en la figura 2 se muestra un ejemplo de pregunta del cuestionario VARK.

1. Está ayudando a una persona que desea ir al aeropuerto, al centro de la ciudad o a la estación del ferrocarril. Ud.:
- iría con ella.
 - le diría cómo llegar.
 - le daría las indicaciones por escrito (sin un mapa).
 - le daría un mapa.

figura 2 Estructura de reactivos del Cuestionario VARK

3.2 Compresión de los Datos

Para esta etapa del proyecto se determinaron los criterios para la obtención de los datos de la aplicación del instrumento, dentro de esta etapa se establecen las características de evaluación del instrumento de acuerdo al modelo VARK aplicado. Por cada instrumento aplicado el resultado es un conjunto de cuatro números que representan las incidencias de los cuatro estilos de aprendizaje que se desean describir, a continuación se describe la manera de evaluación de cada cuestionario aplicado.

Por ejemplo, supongamos que el alumno contesta de la pregunta 3 las categorías B y C, entonces los estilos asignados a esa pregunta corresponden el visual y el lectura/escritura respectivamente como muestra la siguiente figura.

Pregunta	Categoría A	Categoría B	Categoría C	Categoría D
3	K	V	R	A

figura 3 Ejemplo de Evaluación del Cuestionario VARK

Para cada una de las preguntas se encuentran asignados de la siguiente manera:

Pregunta	Categoría A	Categoría B	Categoría C	Categoría D
1	K	A	R	V
2	V	A	R	K
3	K	V	R	A
4	K	A	V	R
5	A	V	K	R
6	K	R	V	A
7	K	A	V	R
8	R	K	A	V
9	R	A	K	V
10	K	V	R	A
11	V	R	A	K
12	A	R	V	K
13	K	A	R	V
14	K	R	A	V
15	K	A	R	V
16	V	A	R	K

figura 4 Asignación de Valores al Cuestionario VARK

Para hacer el cálculo de la puntuación de cada alumno se suma los valores de cada pregunta

- Total de V's marcadas =
- Total de A's marcadas =
- Total de R's marcadas =
- Total de K's marcadas =

De tal manera que por cada cuestionario aplicado se tendrán cuatro valores que corresponden a la sumatorias de cada una de las modalidades marcadas por el alumno, por ejemplo (5,6,7,8) que significa que en el cuestionario evaluado se encontró 5 V, 6 A, 7 R y 8K respectivamente. Para el cálculo de todos estos valores se ha desarrollado una macro en el Excel como una forma de automatizar el proceso.

3.3 Preparación de los Datos

Una vez obtenido los resultados de la aplicación del instrumento se acumularon en un archivo de texto para el análisis grupal, posteriormente se utiliza como editor de texto a la aplicación block de notas con la finalidad de darle el formato al archivo .arff como se muestra en la figura 5, esto se realizo con cada uno de los archivos obtenidos por grupo.

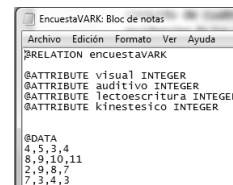


figura 5 Integración de Archivos .arff para WEKA

3.4 Modelado

Para esta investigación se decidió utilizar tareas de minería de datos del tipo descriptivo. En particular,

se aplicó clustering [12] para identificar subgrupos homogéneos dentro de la muestra de estudiantes encuestada. Para ello se utilizó al software WEKA. En particular, para la ejecución se seleccionó la opción Clúster. Además, se optó por el algoritmo kmeans por tratarse de un problema de k centros, Para detectar los estilos de aprendizaje dominantes se decidió mantener en cuatro la cantidad de clústers a generar. Una vez calculados los distintos clústers, ya dentro de la fase de posprocesamiento, se procedió a reemplazar cada uno de los elementos incluidos en los centroides por el código del estilo de aprendizaje relacionado. Luego, se procedió a contar la cantidad de ocurrencias de cada estilo a fin de determinar la combinación presente en los mismos, en la figura 6 se muestra el resultado de la corrida del modelo empleado.

```

--- Run Information ---
Scheme: weka.classifiers.kMeans -1 -M 1000 -J 1000 -L 4 -H 4 -B 1.0 -C 0.5 -O
Instance: Encuesta/ARLK
Instances: 671
Attributes: 4
  Visual
  Auditivo
  Lectura/Escritura
  Quinésico
Test Model: evaluate on training data
--- Model and evaluation on training set ---

XMeans
-----
Requested iterations : 1
Iterations performed : 1
Splits prepared : 4
Splits performed : 0
Cutoff factor : 0.5
Percentage of splits accepted by cutoff factor : 0.0%
Cutoff factor : 0.5

Cluster centers : 4 centers

Cluster 0
6.1970802391970803 7.46985401489854 7.21897810218978 1.8.24087591240875
Cluster 1
2.6806060806080608 3.6787878787878787 4.280909090909091 7.498969696969697
Cluster 2
2.7015708080808082 2.5.848167929292929 4.952079501191830 4.272251308900824
Cluster 3
4.157808370786917 2.72471910110285956 5.61797528089888 3.640449482020247

Distortion: 150.069927
ChiSquare: 47.478917
Clustered instances:
0 137 (20%)
1 165 (25%)
2 191 (28%)
3 178 (27%)

```

figura 6 Corrida del Modelo Kmeans

3.5 Evaluación

En la construcción del modelo Kmeans, de las 850 registros que formaba nuestra vista minable, Weka tomó 671 instancias (78%) para construir el modelo y 179 instancias (21%) para probarlo, con una precisión del 77%. El algoritmo se ejecuto con 4 clusters y con 10 semillas, usando la distancia Euclidiana con 500 iteraciones. De igual manera se utilizaron 4 atributos correspondientes a los estilos de aprendizaje analizados. La evaluación se desarrollo sobre datos de entrenamiento representada en la siguiente tabla:

Attribute / #Cluster	Full Data	0	1	2	3
	671	137	165	191	178
Visual	3.9341	6.197	2.6806	2.7015	4.1573
Auditivo	4.9277	7.4593	3.6787	5.8481	2.7247
Lectura/Escritura	5.5201	7.2189	4.2909	4.9528	5.6179
Quinésico	5.9131	8.2408	7.4989	4.2722	3.6404

3.6 Despliegue

En la figura 7, se presenta la integración de los diferentes agrupamientos, así como su composición de los diferentes estilos de aprendizaje detectados durante la explotación de los resultados de los cuestionarios aplicados a los alumnos.

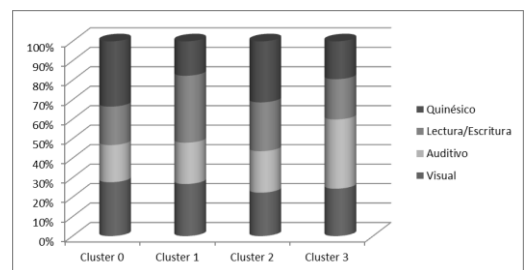


figura 7 Resultados de los Clúster obtenidos

Como puede observarse el estilo de aprendizaje dominante en los estudiantes es el Quinésico, sin embargo, existen algunos multimodales que manejan los alumnos.

4. CONCLUSIONES Y TRABAJO FUTURO

Los resultados obtenidos muestran una tendencia de los alumnos de computación dentro de las tres instituciones de educación Superior analizadas es el estilo de aprendizaje Quinésico como el dominante, sin embargo, se pueden observar varios multimodal o combinaciones de estilos que manejan los alumnos tal son los caso de del mutimodal Visual /Quinésico, o del Auditivo / Visual / Quinésico, y Lectura/Escritura / Quinésico, por lo que se abrirán nuevos proyectos para el desarrollo de material didáctico y guías didácticas de las materias con mas alto índice de reprobación enfocadas a este estilo de aprendizaje. En lo que concierne al uso de la Minería de Datos en este proyecto se aplicarán técnicas de predicción con algoritmos de arboles de decisión y K vecinos mas cercanos, con la finalidad de predecir no solamente estilos de aprendizaje, sino inclusive la permanencia de los alumnos dentro de las instituciones y estimar los índices de deserción y reprobación que pudieran ser reducidos empleando

estrategias desde el programa institucional de Tutorías de cada Institución Educativa.

BIBLIOGRAFÍA

- [1] Fleming, Neil D. (2001). *Teaching and Learning Styles: VARK Strategies*, Editorial Christchurch, Nueva Zelanda
- [2] Costaguta, R., & Gola, M. (2006). *Identificación de estilos de aprendizaje dominantes en estudiantes de informática*. In XV Congreso Argentino de Ciencias de la Computación.
- [3] Durán, E. y R. Costaguta; "Minería de datos para descubrir estilos de aprendizaje". *Revista Iberoamericana de Educación*, ISSN 1681-5653 (en línea), 42(2), 1-10, <http://www.rieoei.org/deloslectores/1674Duran.pdf>. Acceso: 14 de Enero 2013 (2007)
- [4] Felder, R. M., Y Soloman, V. (1984): *Index of Learning Styles*. <http://www.ncsu.edu/felder-public/ILSpage.html> [Consulta: Agosto 2012].
- [5] Software WEKA. Disponible en: <http://www.cs.waikato.ac.nz/ml/weka>
- [6] Felder, R. M. y Silverman, L. K. (1988): "Learning and Teaching Styles in Engineering Education Application". *Engr. Education*, vol. 78 (7), pp. 674-681.
- [7] Sánchez, J. R. O., & Ovalle, S. T. (2010). *Aplicación de técnicas de minería de datos y sistemas de gestión de contenidos de aprendizaje para el desarrollo de un sistema informático de aprendizaje de programación de computadoras*. X Congreso Nacional de Investigación Educativa
- [8] Kolb, D. (1984). *Experiential learning: experience as the source of learning and development*. Englewood Cliffs, NJ: Prentice Hall.
- [9] Han, J., y Kamber, M. (2001): *Data Mining: Concepts and Techniques*. USA, Academic Press.
- [10] Witten, I. y Frank, E. (1999). *Data Mining: Practical machine, learning tools and techniques with Java implementations*, Morgan Kaufmann Publishers, EUA.
- [11] Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). *CRISP-DM 1.0*. CRISP-DM Consortium.
- [12] Hernández Orallo José, R.Q. (2004). *Introducción a la Minería de Datos*. España: Pearson.

CURRICULUM



Felipe de Jesús Núñez Cárdenas es Maestro en Ciencias de la Administración con Especialidad en Informática egresado del Centro de posgrado en Administración e Informática A. C. (CPAI). Profesor Investigador Asociado, incorporado a la Licenciatura en Sistemas Computacionales de la Escuela Superior Huejutla UAEH, en Huejutla de Reyes, Hidalgo, desde el año 2011. Su interés en áreas de investigación incluye:

Ingeniería de Software, Sistemas de Información, Base de Datos, Minería de Datos, últimas publicaciones: Portal Integral de Educación Superior en el XIV Congreso Internacional Sobre educación electrónica, virtual y a distancia. TELEEDU 2007 (Bogotá Colombia), Plataforma Cluster basada en CentOS en el 7º Congreso Universitario en TIC en el ICBI de la UAEH en 2012, Encriptación y comprensión de Archivos en un modelo Cliente- Servidor y Alta Disponibilidad en Cluster bajo CentOS en la vigesimotercera Reunión Internacional de Otoño de IEEE en 2012.



Raúl Hernández Palacios, Maestro en Ingeniería de Computadores y Redes, egresado de la Universidad de Granada, España en 2007. Actualmente como Profesor Investigador y Perfil Promep de la Escuela Superior de Huejutla de la Universidad Autónoma del Estado de Hidalgo, incorporado a la Licenciatura de Sistemas Computacionales de la misma Universidad. En la actualidad realizando estancia de investigación en el Centro de Investigación en Tecnologías de la Información y Comunicación (CITIC-UGR) de la Universidad de Granada. Interés en área de investigación de Sistemas Distribuidos; sistemas de archivos en red; optimización de comunicaciones eficientes en clúster de computadoras con mecanismos como Multicast, Socket Direct Protocol (SDP), Remote Direct Memory Access (RDMA); alta disponibilidad en sistemas clúster a nivel de red, servidor y almacenamiento (iSCSI).



Víctor Tomás Tomás Mariano es Maestro en Ciencias Computacionales egresado de la Universidad Autónoma del Estado de Hidalgo (UAEH), en Pachuca de Soto, Hidalgo, México en el 2007. Profesor investigador titular, con perfil PROMEP y está incorporado a la Licenciatura en Sistemas Computacionales de la Escuela Superior Huejutla UAEH, en Huejutla de Reyes, Hidalgo, desde el año 2008. Su interés en áreas de investigación incluye: planeación de trayectorias, *maze search problem*, laberintos virtuales 2D y 3D, computación inteligente. Entre las publicaciones más recientes están: Algoritmo para la Generación de Laberintos de Conexión Múltiple en 2D. Vigésimosegunda Reunión Internacional de Otoño, ROC&C2011, de IEEE. Propuesta para la generación de laberintos ampliados en 2D. Simposio Iberoamericano Multidisciplinario de Ciencias e Ingeniería SIMCI 2011. Zempoala, Hidalgo.



Ana María Felipe Redondo, Maestra en Comercio Electrónico egresada del Instituto Tecnológico de Estudios Superiores de Monterrey (ITESM), en el Campus Hidalgo, México en el 2005. Profesora por asignatura en la Licenciatura en Administración de la Escuela Superior Huejutla UAEH, en Huejutla de Reyes, Hidalgo, desde el año 2012. Su interés en áreas de investigación incluye: Tecnología

Educativa, Uso de Software Libre orientado a Comercio Electrónico y Plataformas Educativas.