

Sistema inteligente para tutores del bachillerato tecnológico del Estado de Morelos, México.

Smart system for tutors of the technological high school of Morelos State, Mexico.

Rodolfo Villa Noriega* (1).
Centro de Bachillerato Tecnológico Industrial y de Servicios No. 136
rodolfo.villa.noriega@gmail.com.

Francisco Javier Cartujano Escobar (2). Tecnológico Nacional de México / Instituto Tecnológico de Zacatepec,
francisco.ce@zacatepec.tecnm.mx.

Leticia Santa Olalla Ocampo (3). Tecnológico Nacional de México / Instituto Tecnológico de Zacatepec,
leticia.so@zacatepec.tecnm.mx.

Claudia Noguera González (4). Tecnológico Nacional de México / Instituto Tecnológico de Zacatepec,
claudia.ng@zacatepec.tecnm.mx.

*corresponding author.

Artículo recibido en diciembre 21, 2020; aceptado en febrero 16, 2021.

Resumen.

La baja eficiencia terminal y bajo aprovechamiento académico son unos de los principales retos que la Educación Media Superior tiene que afrontar. En respuesta a esta problemática la Secretaría de Educación Pública en conjunto con la Educación Media Superior, ha creado desde el año 2009 el Sistema Nacional de Bachillerato, que en el marco de la Reforma Integral de la Educación Media Superior se ha implementado un programa de acciones tutoriales que atienden esta problemática a nivel operacional. El presente trabajo describe el desarrollo de un sistema inteligente que sirve como herramienta a los tutores en la etapa de análisis de información y toma de decisiones. Para extraer conocimiento de la información generada por la institución educativa y los tutores se utilizaron y evaluaron diferentes técnicas de minería de datos mediante WEKA. Posteriormente se utilizó el conocimiento generado para desarrollar una herramienta que permita predecir si un estudiante tiene altas probabilidades de desertar o tener un bajo aprovechamiento académico. Estas proyecciones son utilizadas por un Sistema Experto construido mediante JESS para sugerir a los tutores las acciones tutoriales a implementar para cada uno de los estudiantes.

Palabras clave: Minería de Datos Educativa, Sistemas Expertos, Sistemas Inteligentes, Tutoría.

Abstract.

Low terminal efficiency and low academic achievement are one of the main challenges that High School Education has to face. In response to this problem, La Secretaría de Educación Pública in conjunction with La Educación Media Superior has created since 2009 El Sistema Nacional de Bachillerato, which within the framework of La Reforma Integral de la Educación Media Superior has implemented a program of actions tutoriales that address this problem at an operational level. This work describes the development of an intelligent system that serves as a tool

for tutors in the information analysis and decision-making stage. To extract knowledge from the information generated by the educational institution and the tutors, different data mining techniques were used and evaluated using WEKA. Subsequently, the knowledge generated was used to develop a tool that allows predicting if a student has a high probability of dropping out or having a low academic achievement. These projections are used by an Expert System built by JESS to suggest to the tutors the tutorial actions to implement for each of the students.

Keywords: Educational Data Mining, Expert Systems, Intelligent Systems, Tutoring.

1. Introducción.

Entre los retos más complejos del sistema educativo nacional se encuentran el abandono escolar, el bajo aprovechamiento académico y bajos índices de eficiencia terminal. En la educación media superior, estos retos, son particularmente atendidos ya que, en muchos casos, truncan el desarrollo y crecimiento personal de los estudiantes generando repercusiones negativas en la sociedad como lo son la desigualdad entre las y los jóvenes, estancamiento de la movilidad social, incremento de la pobreza y marginación, trabajos poco remunerados entre otros, que disminuyen el nivel competitivo del país (Subsecretaría de Educación Media Superior, 2014). Dentro del Sistema Nacional de Bachillerato (en adelante SNB), en el contexto de la Reforma Integral de Educación Media Superior (en adelante RIEMS), la acción tutorial es considerada como un mecanismo de apoyo para la operación y gestión del proceso enseñanza aprendizaje bajo el enfoque por competencias. Se enfoca al acompañamiento de los estudiantes durante su bachillerato con la finalidad de contribuir en su formación integral mediante actividades de orientación, asesoría y apoyo (Martínez et al., 2010).

Desafortunadamente, según lo reportado por (Santizo, 2017), existen diferencias importantes entre el diseño del programa de tutorías y su puesta en práctica, independientemente del subsistema en el que se analice. Algunas de las diferencias son: la tutoría se imparte a todo un grupo reduciendo el tiempo de atención personalizada a los estudiantes que así lo requieren, los tutores no tienen el perfil para detectar y atender las problemáticas de los estudiantes, los tutores no cuentan con una herramienta de apoyo para el análisis de datos y generan reportes en papel con un control estadístico pobre y no hay forma de utilizar esa información para implementar acciones tutoriales adecuadas.

Se ha demostrado que la forma en que se aplica el programa de tutoría si repercute en el aprovechamiento de los estudiantes y en la probabilidad de desertar (Corral, 2014) por lo que es de suma importancia buscar mecanismos que disminuyan los riesgos que puedan afectar su efectividad. Entre los riesgos tenemos tutores mal capacitados, falta de información con la que cuentan los tutores de los estudiantes, pero, sobre todo, la falta de tiempo para atender los problemas de una manera más particular (Santizo, 2017).

Al proporcionar una solución sustentada en las tecnologías de la información que facilite el diagnóstico automático de problemas académicos de los estudiantes, se ayudará a resolver en gran medida la problemática expuesta anteriormente, de tal manera que se beneficia a los estudiantes al recibir una atención tutorial en función de las variables que inciden en su deserción y aprovechamiento. El resultado del presente proyecto será posible replicarlo en planteles inscritos en el SNB como un apoyo a sus programas de tutoría institucional, independientemente del subsistema, lo cual puede aumentar el alcance de su implementación.

2. Métodos.

2.1. Minería de datos educacional.

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 (Hernández, 2016). La minería de datos comparte una gran cantidad de técnicas con el Aprendizaje Automático por lo que resulta difícil realizar una distinción entre ambos conceptos (Berlanga, 2016). 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. La tarea de minería de datos real es el análisis automático o semiautomático de grandes cantidades de datos para extraer patrones interesantes hasta ahora desconocidos, como los grupos de registros de datos (análisis clúster), registros poco usuales (la detección de anomalías) y dependencias. Una aplicación de minería de datos podría identificar varios grupos en los datos que luego pueden ser utilizados para obtener resultados más precisos de predicción por un sistema de soporte de decisiones. Ni la recolección de datos, preparación de datos, ni la interpretación de los resultados y la información son parte de la etapa de minería de datos, pero pertenecen a todo el proceso KDD (Knowledge Discovery in Databases) como pasos adicionales.

La Minería de Datos Educativa (en adelante EDM por sus siglas en inglés de Educational Data Mining) es la aplicación de las técnicas de la Minería de Datos a datos generados en contextos de la educación (Khare et al., 2018). Es un puente donde convergen principalmente dos disciplinas: la educación por un lado y las ciencias de la computación por el otro, donde la Minería de Datos y el Aprendizaje Automático son subáreas de las ciencias de la computación que permiten encontrar nueva información en los datos educativos (Bakhshinategh et al., 2018).

2.1.1 Knowledge Discovery in Databases.

La extracción de conocimiento está principalmente relacionada con el proceso de descubrimiento conocido como *Knowledge Discovery in Databases* (Hernández, 2016), que descubre conocimiento e información potencialmente útil dentro de los datos contenidos en algún repositorio de información. Es un proceso repetitivo que explora volúmenes muy grandes de datos para determinar relaciones y extraer información de calidad que puede usarse para encontrar modelos dentro de los datos (Deepashri et al., 2017).

Las etapas del proceso KDD descritas en la Figura 1, se dividen en 5 fases (Fayyad & Stolorz, 1997) y son:

- **Selección de datos.** En esta etapa se determinan las fuentes de datos y el tipo de información a utilizar.
- **Preprocesamiento.** Esta etapa consiste en la preparación y limpieza de los datos extraídos desde las distintas fuentes de datos en una forma manejable, necesaria para las fases posteriores.
- **Transformación.** Consisten el tratamiento preliminar de los datos, transformación y generación de nuevas variables a partir de las ya existentes con una estructura de datos apropiada.
- **Minería de datos.** Es la fase de modelamiento propiamente en donde métodos inteligentes son aplicados con el objetivo de extraer patrones previamente desconocidos, válidos, nuevos, potencialmente útiles y comprensibles y que están contenidos u “ocultos” en los datos.
- **Interpretación y evaluación.** Se identifican los patrones obtenidos y que son realmente interesantes, basándose en algunas medidas y se realiza una evaluación de los resultados obtenidos.

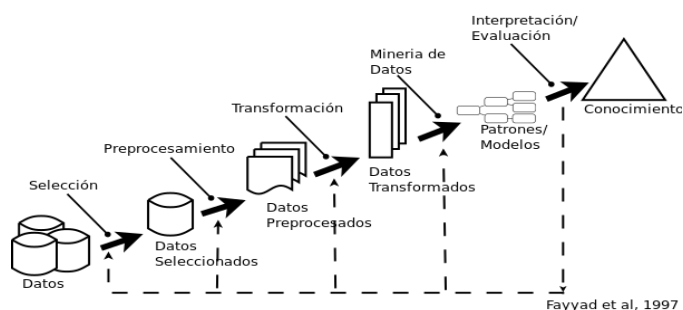


Figura 1. Proceso KDD.

2.1.2 Técnicas de Minería de Datos.

Las técnicas de la minería de datos provienen de la inteligencia artificial y de la estadística. Dichas técnicas no son más que algoritmos más o menos sofisticados que se aplican sobre un conjunto de datos para obtener unos resultados. Las técnicas más representativas según Hernández Cedano (2016) son:

- **Redes neuronales:** Son un paradigma de aprendizaje y procesamiento automático inspirado en la forma en que funciona el sistema nervioso de los animales. Se trata de un sistema de interconexión de neuronas en una red que colabora para producir un estímulo de salida.
- **Regresión lineal:** Es la más utilizada para formar relaciones entre datos. Rápida y eficaz pero insuficiente en espacios multidimensionales donde puedan relacionarse más de 2 variables.
- **Árboles de decisión:** Un árbol de decisión es un modelo de predicción utilizado en el ámbito de la inteligencia artificial. Dada una base de datos se construyen estos diagramas de construcciones lógicas, muy similares a los sistemas de predicción basados en reglas, que sirven para representar y categorizar una serie de condiciones que suceden de forma sucesiva, para la resolución de un problema.
- **Agrupamiento o clustering:** Es un procedimiento de agrupación de una serie de vectores según criterios habitualmente de distancia; se tratará de disponer los vectores de entrada de forma que estén más cercanos aquellos que tengan características comunes.

2.2 Sistemas inteligentes basados en reglas.

Los sistemas basados en reglas constituyen un campo de estudio importante dentro de la Inteligencia Artificial ya que nos ayuda a capturar la experiencia humana en la resolución de problemas, con el fin de alcanzar decisiones consistentes y repetibles. En ellos la representación del conocimiento se identifica por medio de la heurística o formas de proceder de los expertos. Son interesantes, especialmente en aquellos dominios en donde escasean los expertos como por ejemplo medicina, ingeniería, etc. ya que proporcionan un medio eficaz para difundir ampliamente razonamientos escasos y específicos (Ramírez et al., 2016).

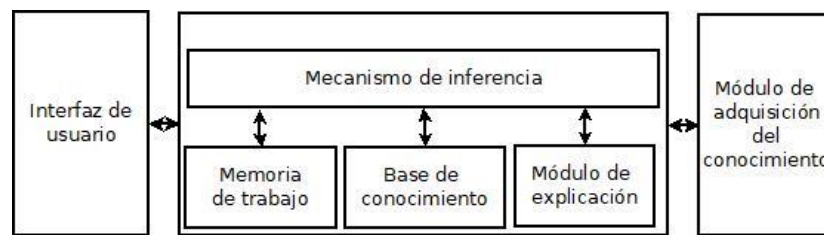


Figura 2. Arquitectura de un sistema experto.

Un sistema basado en reglas según (Ferrer et al., 2015), es un sistema que contiene los elementos mostrados en la Figura 2 y se describen a continuación:

- **Una base de conocimiento (BC):** contiene el conocimiento extraído del experto en un dominio específico. Para realizar la representación del conocimiento se han desarrollado diversas formas como lo son las reglas de producción, marcos, las redes semánticas, entre otras.
- **Mecanismo de inferencia (MI):** selecciona las reglas que se pueden aplicar y las ejecuta, con el fin de obtener alguna conclusión, es decir, realizar un procesamiento o interpretación del conocimiento controlando el proceso de razonamiento. Su objetivo es encontrar conclusiones aplicando el conocimiento almacenado en la base de conocimiento a los datos de entrada.
- **Base de hechos (BH) o memoria de trabajo:** acumula un conjunto de hechos establecidos, que se usan para determinar qué reglas puede aplicar el mecanismo de inferencias.
- **Interfaz de usuario:** es el puente entre el sistema experto y el usuario no experto. Permite al usuario describir al sistema el problema al ingresar información en forma de hechos, por otro lado, permite también al sistema presentar los resultados de forma que el usuario pueda entenderlos.
- **Módulo de adquisición del conocimiento:** permite al experto modificar la base de conocimiento ya sea actualizando el conocimiento ya existente o agregando nuevo conocimiento.
- **Módulo de explicación:** es utilizado para que el sistema experto describa el procedimiento que utilizó para llegar a una conclusión. Es útil cuando se requieren tomar decisiones basados en los resultados del sistema experto.

Existen implementaciones de sistemas expertos que no contienen base de conocimiento, estos son conocidos como *Shells* y permiten desarrollar sistemas expertos constituyendo una herramienta de apoyo en el proceso de creación de bases de conocimiento y facilitar la utilización de los sus componentes al usuario (Prada, 2016).

Jess es un motor de reglas para la plataforma Java. Para usarlo, se especifica la lógica en forma de reglas utilizando uno de dos formatos: el lenguaje de reglas Jess o XML. También proporciona algunos de sus propios datos para que las reglas operen. Las reglas pueden crear nuevos datos, o pueden hacer cualquier cosa que pueda hacer el lenguaje de programación Java.

3. Desarrollo.

3.1. Arquitectura del sistema.

En la Figura 3 se muestra la arquitectura general de la aplicación donde se definen las tecnologías que se utilizaron y los componentes principales del sistema.

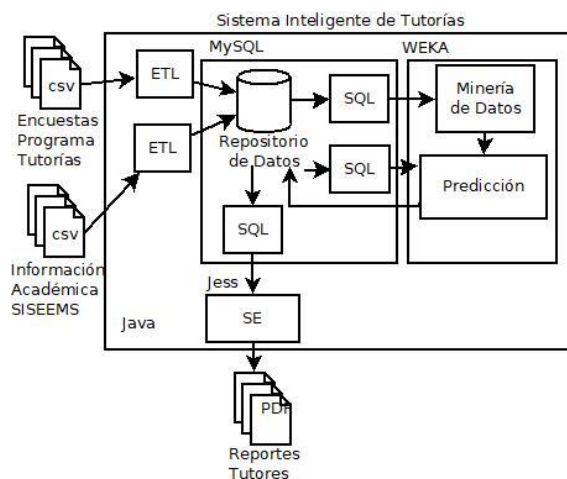


Figura 3. Arquitectura del sistema.

El sistema utiliza una base de datos relacional en MySQL donde se depositan los datos de las fuentes de información. Los datos son cargados por un módulo de la aplicación que permite seleccionar los archivos de datos para procesarlos, limpiarlos, transformarlos (en caso necesario) e insertarlos en la base de datos. Posteriormente se diseñaron consultas SQL que permiten obtener *DataSets* para los procesos de minería de datos y para el sistema experto.

El proceso de minería de datos es realizado con Weka. Este proceso permite obtener un modelo el cual es utilizado para obtener una predicción de cada uno de los estudiantes, dicha predicción también es almacenada en el repositorio de datos.

El sistema experto se desarrolló con el *shell* Jess. Utiliza los datos de los estudiantes, así como la predicción realizada por Weka. Este módulo genera listas de estudiantes para cada una de las acciones tutoriales.

Por último, se desarrolló una interfaz gráfica que permite automatizar cada uno de los procesos descritos anteriormente.

3.2 Minería de datos.

Para la generación del modelo de predicción se utilizó la técnica de redes neuronales siguiendo las fases del proceso KDD, descritas a continuación:

- **Selección de datos.** En esta etapa de seleccionaron las fuentes de información. Para este proyecto se determinaron dos fuentes de información:
 1. Datos de estudiantes que obtiene el programa de tutorías del bachillerato tecnológico. Los tipos de datos incluidos en esta fuente de información son personales, socioeconómicos, de hábitos de estudios, estilos de aprendizaje e información acerca de la convivencia entre los estudiantes. Para este estudio se utilizaron dos muestras de información. la primera con 240 registros con 143 atributos con información personal y socioeconómica del semestre febrero – junio del 2018. En la segunda muestra se obtuvieron 458 registros con 143 atributos del semestre agosto – diciembre del 2018.
 2. Datos académicos. Esta fuente de información incluye calificaciones y asistencias parciales por materia y, datos personales obtenidos del CURP de estudiante. En total se cuentan con 13752 registros producto del proceso de evaluación del semestre enero – junio del 2018 con 25 atributos.
- **Preprocesamiento.** Se desarrollo un proceso ETL que consta con dos componentes:
 1. Una base de datos implementada en MySQL que almacena dos datos preprocesados de las dos fuentes de información.
 2. Una aplicación desarrollada en Java la cual procesa las fuentes de información evalúa los datos obtenidos de las evaluaciones del programa de tutorías y carga la información a la base de datos.
- **Transformación.** Se obtuvo el género y la edad del CURP de los estudiantes para complementar los datos ya existentes. El conjunto de datos presentó el problema de clases desbalanceadas. Para evitar un sobre aprendizaje en el proceso de la construcción del modelo de predicción se balancearon las clases mediante el uso del filtro *Resample* de Weka, que generó una submuestra de los datos de forma aleatoria Villa-Noriega (2019). Por último, se seleccionaron 35 atributos descritos en la Tabla 1, con los cuales se formó el conjunto de datos para la siguiente fase. El total de registros generado en esta fase fue de 4239 registros.

Tabla 1. Atributos del conjunto de datos.

N	Atributo	N	Atributo	N	Atributo
1	edad	13	trabaja	25	resultadoActitudClase
2	genero	14	ingresoFamiliar	26	resultadoCondicionesExternas
3	tipoEscuela	15	tipoVivienda	27	resultadoMetodoPersonal
4	promedioSecundaria	16	viviendaPropia	28	resultadoProgramacionPersonal
5	razonIngreso	17	tieneComputadora	29	resultadoConvivencia
6	padeceEnfermedad	18	tieneInternet	30	resultadoNuevoIngreso
7	vivePadre	19	tiempoTraslado	31	carrera
8	escolaridadPadre	20	gastoTransporte	32	materia
9	viveMadre	21	resultadoAuditivo	33	semestre
10	escolaridadMadre	22	resultadoVisual	34	turno
11	vivePadres	23	resultadoKinestesico	35	tipoAcreditacion
12	sostenEconomico	24	resultadoVAK		

- **Minería de datos.** Para la generación del modelo de predicción se utilizaron redes neuronales mediante el algoritmo *MultiLayerPerceptron* de Weka y una validación cruzada con 10 particiones para verificar su precisión.
- **Interpretación y evaluación.** El algoritmo generó un modelo de predicción con el 97.42% de precisión con un total de 4130 instancias clasificadas correctamente y 109 incorrectas. En la Tabla 2 se muestran los resultados de la validación del modelo donde se detalla el número de instancias que se clasificaron de cada una de las clases.

Tabla 2. Matriz de confusión de la red neuronal.

CLASES	A	NA	NP	Predicción Correcta	Predicción Incorrecta
A	1325	78	10	93.8%	0.7%
NA	19	1392	2	98.5%	0.3%
NP	0	0	1413	100%	0.04%

3.2 Aplicación del modelo predictivo.

Uno de los objetivos del presente trabajo es generar una herramienta para el apoyo a la toma de decisiones por parte del docente tutor. Esta herramienta deber realizar un análisis de estudiantes de manera automatizada, obteniendo una predicción de su deserción. Para lograr el objetivo anteriormente mencionado, es necesario realizar la predicción aplicando el modelo generado en Weka a los datos del estudiante. Posteriormente almacenar dicha predicción en el repositorio de datos para su posterior utilización. Para ello se generó otro módulo en la aplicación Java que permite obtener los datos de los estudiantes, evaluar a cada estudiante con el modelo y finalmente guardar la predicción.

En el Algoritmo 1 podemos observar el código Java donde se aplica el modelo predictivo a cada uno de los registros de los estudiantes. Se obtiene el modelo generado en Weka en un objeto *Classifier* con la finalidad de poder realizar la clasificación de forma masiva. Posteriormente de obtiene el modelo de datos de un archivo *ARFF* con los datos de los estudiantes idéntico al *DataSet* con el que se generó el modelo. Se leen los datos de los estudiantes para clasificarlos con el método *classifyInstance()* de la clase *Classifier*. Por último, se guarda la predicción de cada instancia en la base de datos para su utilización en el futuro.

Algoritmo 1. Aplicación del modelo de predicción en Java

```
Classifier cls = (Classifier) weka.core.SerializationHelper.read("Clasificador.model");
```

```
model = new Instances(  
    new BufferedReader(  
        new FileReader("Archivo.ARFF")));
```

```
for (int i = 0; i < model.numInstances(); i++)  
{  
    Instance inst=model.instance(i);  
    double pred = cls.classifyInstance(inst);  
    data[i][45]=inst.classAttribute().value((int) pred);  
    CalificacionesSiseemsPredictions().save(data);  
}
```

3.3. Implementación de un sistema inteligente.

En el programa de tutorías del bachillerato tecnológico no existe un Plan de Acción Tutorial. La Subsecretaría de Educación Media Superior ha proporcionado documentos base donde se establece que es responsabilidad del docente tutor definir las acciones tutoriales a partir del diagnóstico de los estudiantes y la identificación de las necesidades particulares de cada estudiante. Para definir las acciones tutoriales del sistema se realizaron mesas de trabajo con los docentes tutores. En estas mesas de trabajo los docentes tutores definieron 137 acciones tutoriales.

3.3.1. Reglas del sistema experto en Jess.

Continuando con las mesas de trabajo de los docentes tutores, una vez que definieron las acciones tutoriales, determinaron los valores de las características de los estudiantes para cada una las acciones tutoriales formando

reglas de producción. Estas reglas se transformaron a reglas en el lenguaje CLP que es el que utiliza el *shell* Jess. En la Figura 4 podemos observar un fragmento de la plantilla con las características de los estudiantes y sus posibles valores en lenguaje CLP. Al final del análisis se llegó a un total de 47 slots definidos para utilizarse en el sistema experto.

```
(deftemplate Estudiante
  (slot numeroControl)
  (slot generacion) ;{2016-2019,2015-2018,2017-2020,2018-2021}
  (slot edad) ;numeric
  (slot genero) ;{H,M}
  (slot tipoEscuela) ;{Pública,Privada}
  (slot promedioSecundaria) ;numeric
  (slot razonIngreso) ;{Convicción,Padres,Amigos,Otro}
  (slot periodo) ;{'SEMESTRAL 2 - 2017 ','SEMESTRAL 1 - 2018'}
  (slot padeceEnfermedad) ;{No,Si}
  (slot vivePadre) ;{Si,No}
  (slot escolaridadPadre) ;{Secundaria,'Preparatoria o bachillerato',Primaria,Licenciatura,'Ma
  (slot viveMadre) ;{Si,No}
  (slot escolaridadMadre) ;{'Preparatoria o bachillerato',Primaria,Secundaria,'Sin escolaridad
  (slot vivePadres) ;{Si,No}
  (slot sostenEconomico) ;{Ambos,'Padre o tutor','Madre o tutora',Yo,'Un familiar'}
  (slot trabaja) ;{No,Si}
  (slot ingresoFamiliar) ;{'$2500 - $5000','Menos de $2500','$7500 - $10000','$5000 - $7500','
  (slot tipoVivienda) ;{Casa,Apartamento,'Vivienda rural'}
  (slot viviendaPropia) ;{No,Si}
  (slot tieneComputadora) ;{Si,No}
  (slot tieneInternet) ;{Si,No}
  (slot tiempoTraslado) ;numeric
```

Figura 4. Definición de la plantilla Estudiante en Jess.

Posteriormente se definieron las reglas con el lenguaje CLP en el cual se utilizaron las características de los estudiantes para definir cada una de las acciones tutoriales a realizar. Una de las tareas medulares del docente tutor es realizar una intervención cuando el estudiante presenta deficiencias en su rendimiento académico el cual puede ser de una determinada materia. En la Figura 5 se observa un ejemplo de la definición de una acción tutorial de canalización académica de la asignatura de Álgebra. La regla define las características que los estudiantes deben de cumplir para que el sistema recomiende la asesoría académica como acción tutorial. El estudiante debe de cursar la carrera “Componente básico y propedéutico”, cursar la materia de “Álgebra”, tener “Insuficiente” por lo menos en uno de los resultados de la actitud en clase, condiciones de estudio externas, método de estudio y programación personal, además haber sido clasificado como “Riesgo moderado” o “Riesgo Alto” en el perfil de deserción de nuevo ingreso. Por último, si el estudiante en la evaluación del primer parcial obtuvo menos de ocho o si en los valores obtenidos por los modelos de predicción en la acreditación es posible que repruebe o deserte entonces se canaliza a asesorías académicas de álgebra.

```
(defrule asesoria-algebra
  "Asesoría de Álgebra"
  (Estudiante {carrera == "COMPONENTE BASICO Y PROPEDEUTICO"
    && materia == "ÁLGEBRA"
    && (resultadoActitudClase == "INSUFICIENTE"
      || resultadoCondicionesExternas == "INSUFICIENTE"
      || resultadoMetodoPersonal == "INSUFICIENTE"
      || resultadoProgramacionPersonal == "INSUFICIENTE")
    && (resultadoNuevoIngreso == "Riesgo moderado" || resultadoNuevoIngreso == "Riesgo alto")
    && (call < 8 || (tipoAcreditacionP == "NA" || tipoAcreditacionP == "NP"))
  })
  =>
  (printout t "Canalizar a las asesorias de Álgebra" crlf))
```

Figura 5. Regla para la acción tutorial de canalización académica.

Otro de los objetivos de la tutoría en el bachillerato tecnológico es realizar el acompañamiento del estudiante desde que inicia sus estudios hasta que los finaliza. En la regla de la Figura 6 se define una acción tutorial de canalización a

asesorías académicas de una materia de especialidad. La diferencia entre esta regla y la anterior es que esta última pide que el estudiante pertenezca a la especialidad de “Programación”, y que no se utilizan los resultados del test de perfil de ingreso. Esto es debido a que esta materia pertenece al tercer semestre, por lo que el estudiante pudo haber obtenido un resultado en el perfil de ingreso que no describe al estudiante en el presente.

```
(defrule asesoria-programacion-orientada-eventos
"Asesoría de APLICA LA METODOLOGÍA DE DESARROLLO RÁPIDO DE APLICACIONES CON PROGRAMACIÓN ORIENTADA A EVENTOS"
(Estudiante (carrera == "PROGRAMACIÓN"
  && materia == "APLICA LA METODOLOGÍA DE DESARROLLO RÁPIDO DE APLICACIONES CON PROGRAMACIÓN ORIENTADA A EVENTOS"
  && (resultadoActitudClase == "INSUFICIENTE"
    || resultadoCondicionesExternas == "INSUFICIENTE"
    || resultadoMetodoPersonal == "INSUFICIENTE"
    || resultadoProgramacionPersonal == "INSUFICIENTE")
  && (call < 8 || tipoAcreditacionP == "NA" || tipoAcreditacionP == "NP")
))
=>
(printout t "Canalizar a las asesorías de Programación con RAD Orientada a Eventos" crlf))
```

Figura 6. Regla para la acción tutorial de canalización académica de carrera.

Las reglas y las plantillas se almacenaron en un archivo de texto plano CLP con la finalidad de utilizarlas con el motor de inferencia de JESS.

3.3.2 Implementación del sistema experto en Java.

Para la implementación del sistema experto se utilizó Jess como API. Jess recibe la lista de los estudiantes con cada una de sus características que obtenemos de la información proporcionada por los estudiantes y la información que generó el sistema aplicando los modelos predictivos obtenidos con la ayuda de WEKA. Esta lista es obtenida mediante consultas SQL y contienen los datos de los estudiantes de un grupo en específico en un determinado momento por lo que está definido por el “Grupo”, “Periodo de Evaluación”, “Turno” y “Especialidad”. Estos datos son ingresados mediante un formulario en Java mostrado en la Figura 7.

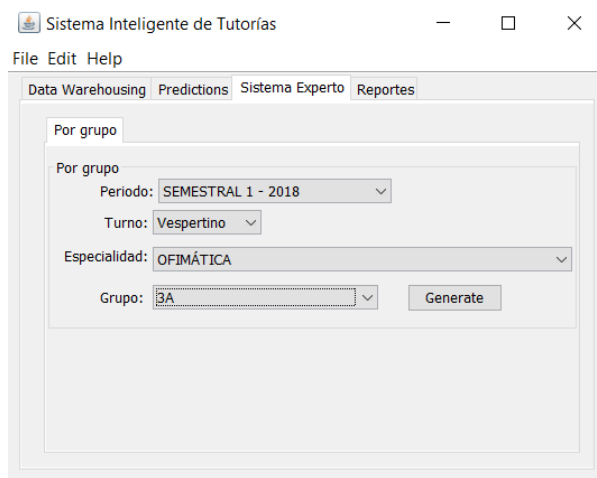


Figura 7. Formulario del sistema experto.

Cada uno de los registros obtenidos en la consulta es llamado instancia. Cada instancia es enviada a Jess como un hecho. Se obtienen las instancias JESS desde la base de datos a partir de los datos anteriormente mencionados, para posteriormente evaluar cada hecho con la ayuda de una clase que contiene el método *afirmarHecho*.

La estructura principal del método *afirmarHecho* se muestra en el Algoritmo 2. Este método crea un hecho y utiliza un objeto *RETE* llamado *engine*, ambos proporcionados por el API de JESS. Mediante el objeto *engine* aceptamos hechos a la memoria de trabajo de JESS y con el método *run()* evaluamos el hecho.

Algoritmo 2. Método *afirmarHecho*

```
public void afirmarHecho(JessInstance instance) throws JessException
{
    Fact fact = new Fact("Estudiante", this.engine);

    fact.setSlotValue("numeroControl", new Value(instance.getNoControl(), RU.STRING));
    fact.setSlotValue("generacion", new Value(instance.getGeneracion(), RU.STRING));
    fact.setSlotValue("edad", new Value(instance.getEdad(), RU.INTEGER));
    ...
    fact.setSlotValue("calP", new Value(instance.getCalP(), RU.INTEGER));
    fact.setSlotValue("tipoAcreditacion", new Value(instance.getTipoAcreditacion(), RU.STRING));
    fact.setSlotValue("tipoAcreditacionP", new Value(instance.getTipoAcreditacionP(), RU.STRING));

    engine.assertFact(fact);
    engine.eval("facts");
    engine.run();
    engine.clear();
}
```

Por último, se utilizó un manejador de eventos de JESS donde especificamos las acciones que se realizarán cada que una regla es aceptada. En el Algoritmo 3 se muestran las acciones que definen si una regla es aceptada. Si es así, obtiene los datos del estudiante para posteriormente enviar la acción tutorial a la base de datos.

Algoritmo 3. Evento de Jess que guarda la acción tutorial

```
if (JessEvent.DEFRULE_FIRED)
{
    Activation actRule = (Activation) je.getObject();
    String nameRule = actRule.getRule().getName();

    JessInstance estudiante = new JessInstance(engine.findFactByTemplateName("MAIN::Estudiante"));
    estudiante.setAccionTutorial(nameRule);
}
```

3.3.3. Reportes.

Como se explicó en la metodología, cuando se acepta una regla del sistema experto, el sistema genera un registro en la base de datos con la acción tutorial de cada uno de los estudiantes. Este conjunto de registros son los que conforman la propuesta de plan de acción tutorial. Para comprobar la generación del plan tutorial se introdujeron hechos al sistema. Estos hechos de prueba pertenecen a los estudiantes de un grupo completo de la asignatura de Álgebra adscritos al componente básico y propedéutico. En la Figura 8 se observa la respuesta de Jess al ingresar los datos de un estudiante de este grupo. Se puede observar que el estudiante con número de control 19317051360414 está adscrito al componente básico y propedéutico, se han detectado problemas en los resultados de los test de habilidades académicas y se obtuvo "NA" (No Acreditado) en la predicción de la materia de Álgebra. En la calificación del primer parcial el estudiante obtuvo una calificación de 7 y en el perfil de nuevo ingreso obtuvo un resultado de "Riesgo Moderado".

f-1 (MAIN::Estudiante	(resultadoVisual 65.0)
(numeroControl "19317051360414")	(resultadoKinestesico 71.66999816894531)
(generacion "2019-2022")	(resultadoVAK "AUDITIVO")
(edad 16)	(resultadoActitudClase "SUFICIENTE")
(genero "H")	(resultadoCondicionesExternas "BIEN")
(tipoEscuela "Pública")	(resultadoMetodoPersonal "INSUFICIENTE")
(promedioSecundaria 8.0)	(resultadoProgramacionPersonal "INSUFICIENTE")
(razonIngreso "Otro")	(resultadoConvivencia "DETECTADO")
(periodo "SEMESTRAL 1 - 2019")	(resultadoNuevolIngreso "RIESGO MODERADO")
(padeceEnfermedad "No")	(carrera "COMPONENTE BASICO Y PROPEDEUTICO")
(vivePadre "No")	(materia "ÁLGEBRA")
(escolaridadPadre "Preparatoria o bachillerato")	(semestre 1)
(viveMadre "Si")	(turno "VESPERTINO")
(escolaridadMadre "Preparatoria o bachillerato")	(cal1 7)
(vivePadres "Si")	(cal1P 0)
(sostenEconomico "Madre o tutora")	(cal2 0)
(trabaja "Si")	(cal2P 0)
(ingresoFamiliar "\$2500 - \$5000")	(cal3 0)
(tipoVivienda "CASA")	(cal3P 0)
(viviendaPropia "Si")	(cal 7)
(tieneComputadora "Si")	(calP 0)
(tieneInternet "Si")	(tipoAcreditacion "A")
(tiempoTraslado 40)	(tipoAcreditacionP "NA")
(gastoTransporte 30.0)	
(resultadoAuditivo 73.33000183105469)	

Figura 8. Datos introducidos en Jess.

Estas características coinciden con las características necesarias para que el sistema recomiende canalizar a este estudiante a asesorías académicas de Álgebra generando un registro en la base de datos de acciones tutoriales. Esta regla es mostrada en la Figura 4.

Una vez que el sistema ha evaluado a los estudiantes, es posible obtener las acciones tutoriales recomendadas mediante un formulario de Java mostrado en la Figura 9. Los datos necesarios para obtener las acciones tutoriales son: el periodo de evaluación, turno, especialidad y grupo.

Figura 9. Formulario para obtener acciones tutoriales.

El resultado es un archivo PDF como el mostrado en la Figura 10. Este reporte contiene una lista de acciones tutoriales con el número de control y nombre completo de los estudiantes con la finalidad de que el docente tutor

realice la implementación de dichas acciones tutoriales. En el último registro del reporte podemos observar el número de control del estudiante descrito en la Figura 8 por lo que el sistema nos sugiere que se canalice a asesorías de Álgebra. Cabe mencionar que en la figura mostrada se han ocultado parte de los datos, debido a que se utilizaron datos reales para el proyecto.

Acciones tutoriales			
CANALIZACIÓN ASESORÍA ACADÉMICA ÁLGEBRA			
NoControl	Apellido Paterno	Apellido Paterno	Apellido Paterno
19317051360684	C	S	ABIE
19317051360434	G		YES
19317051360593	G	B	JUAN
19317051360431	M		ROG
19317051360415	L	H	ELVI
19317051360420	M	F	JESS
19317051360422	M	N	JOSE
19317051360428	M	N	LUIS
19317051360429	M	C	MARI
19317051360425	P	E	KEILI
19317051360706	R	V	BRA
19317051360592	R	V	JOSL
19317051360414	S	A	EDW

Figura 10. Estudiantes canalizados a asesoría académica.

Conclusiones.

En el presente trabajo se realizó un análisis de los datos que el programa de tutorías del bachillerato tecnológico genera. Se encontraron dos fuentes principales de datos; la generada por el programa de tutorías y la generada por el departamento de servicios escolares. Estas fuentes de datos se unieron y se consolidaron en una base de datos mediante un repositorio de datos.

Con la información almacenada y consolidada en la base de datos se generó un modelo de predicción mediante redes neuronales. El modelo generado fue utilizado para realizar un módulo de predicción de deserción de estudiantes. La predicción se almacenó en la base de datos del sistema con la finalidad de utilizar esta nueva información en la construcción de un sistema experto que genera el plan de acción tutorial para cada estudiante.

Se realizaron pruebas en un ambiente real obteniendo los siguientes resultados:

- El sistema experto permite al usuario generar acciones tutoriales a partir de los datos de los estudiantes y las predicciones de la deserción estudiantil.
- El sistema propone acciones tutoriales personalizadas para cada estudiante las cuales son consideradas como un plan de acción tutorial. Este plan es generado en documentos electrónicos para facilitar al tutor su implementación.
- El sistema reduce el tiempo de análisis de datos y propuesta de acciones tutoriales lo que permite implementar acciones tutoriales pertinentes y a tiempo.
- Cada acción tutorial propuesta por el sistema también es almacenada en la base de datos, logrando de esta manera obtener los resultados de cada ejecución en cualquier momento.

- El sistema realiza un análisis de estudiantes cuando sea necesario, inclusive cuando los datos del estudiante cambian a través del tiempo, como lo son las calificaciones parciales registradas por los docentes o actualización de los datos manejados por el programa de tutorías.
- El sistema cumple con los lineamientos de programa de tutorías de los planteles inscritos al Sistema Nacional de Bachillerato por lo que su implementación puede realizarse en más planteles dentro del Estado e inclusive del País.

Los resultados de este trabajo de investigación pueden sentar las bases para realizar trabajo futuro o mejoras como las mencionadas a continuación:

- Existen más técnicas de minería de datos que pueden aplicarse al conjunto de datos. Dichas técnicas permitirán generar nuevos modelos predictivos y descriptivos que pueden aportar más conocimiento del fenómeno del aprovechamiento académico y la deserción estudiantil.
- Es posible el desarrollo de un sistema integral que incorpore el proceso de generación de modelos de predicción al sistema mediante el API de Weka.
- El sistema experto puede ser mejorado agregando nuevas características, como lo son el encadenamiento hacia atrás y el módulo de explicación. Esta mejora permitiría mayor interacción de parte de los tutores por lo que se podrían agregar nuevas reglas o mejorar las ya existentes.
- Por último, debido a la importancia que los datos tienen en la aplicación de técnicas de minería de datos, es recomendable la implementación de un sistema que permita análisis de datos donde se obtenga más información del dominio del problema. En este proyecto se generó un repositorio de datos mediante una base de datos. Este componente puede ser mejorado con una herramienta más adecuada como lo son los Data Warehouse. Esta herramienta permite aplicar operaciones complejas a los datos que después pueden ser utilizados en la aplicación de técnicas de minería de datos.

Agradecimientos.

Especial agradecimiento al Centro de Bachillerato Tecnológico, Industrial y de Servicios No. 136 por permitir utilizar sus datos para la realización del proyecto de investigación reportado en este artículo.

Referencias Bibliográficas.

- Bakhshinategh, B., Zaiane, O. R., ElAtia, S., & Ipperciel, D. (2018).** *Educational data mining applications and tasks: A survey of the last 10 years.* Education and Information Technologies, 23(1), 537–553. <https://doi.org/10.1007/s10639-017-9616-z>.
- Berlanga, A. (2016).** *El camino desde la Inteligencia Artificial al Big Data.* Índice. Revista de estadística y sociedad, 68 (9–11). ISSN: 1697-2325.
- Corral, C. M. (2014).** *El rendimiento académico de los jóvenes en relación a diferentes estilos de acción tutorial de un Centro de Bachillerato Tecnológico* (Tesis de maestría). Tecnológico de Monterrey, México.
- Deepashri, Kamath, A. (2017).** *Survey on Techniques of Data Mining and its Applications.* International Journal of Emerging Research in Management & Technology, 6(2), 198-201. ISSN: 2278-9359.
- Fayyad, U., Stolorz, P. (1997).** *Data mining and KDD: Promise and challenges.* Future Generation Computer Systems, 13(2–3), 99–115. [https://doi.org/10.1016/S0167-739X\(97\)00015-0](https://doi.org/10.1016/S0167-739X(97)00015-0).
- Ferrer, R., Jiménez, K., Argüelles, D. & Montes, A. (2015).** *Sistema experto para la elección del tipo de recuperación en canteras de materiales de construcción.* Revista Cubana de Ciencias Informáticas, 9(3), 33–48.

- Hernández, J. Á. (2016).** *Modelo de minería de datos para identificación de patrones que influyen en el aprovechamiento académico* (Tesis de maestría). Instituto Tecnológico de La Paz, México.
- Khare, K., Lam, H., & Khare, A. (2018).** *Educational Data Mining (EDM): Researching Impact on Online Business Education*. *On the Line: Business Education in the Digital Age*, 37–53. <https://doi.org/10.1007/978-3-319-62776-2>.
- Martínez, A. L., Virgen, I.J., Prieto, L. N., Alcalde, L. A. (2010).** *La Tutoría en el Sistema de Educación Media Superior*. Universidad de Guadalajara. Retrieved from <http://www.sems.udg.mx/sites/default/files/BGC/Tutoria en el SEMS UDG Nov 2010.pdf>.
- Prada, D. (2016).** *Sistema Experto aplicado al análisis de crédito y la gestión de cartera de la empresa Masterkey LTDA* (Tesis de pregrado). Fundación Universitaria Católica Lumen Gentium, Colombia.
- Ramírez, N.V., Azuara, G., Corona, F. J., Ramirez, J. R. & Martínez, H. S. (2016).** *Sistema Experto Determinista Para El Diagnóstico De Enfermedades Eritemato-Escamosas En Prolog*. *Pistas educativas*, México (119), 205–216.
- Santizo, E. (2017).** *La función de la tutoría en educación media superior: Retos entre su diseño e implementación*. Reflexiones, análisis y experiencias sobre la tutoría en la educación media superior, México: Ediciones del Lirio. ISBN 978-607-28-0978-9.
- Subsecretaría de Educación Media Superior. (2014).** *Yo no abandono 9 Manual para ser un mejor tutor en planteles de educación media superior*. (9). Retrieved from https://www.gob.mx/cms/uploads/attachment/file/706/yna_manual_9.pdf.
- Villa-Noriega, R. & Cartujano-Escobar, F. J. (2019).** *Determinación del perfil de deserción en estudiantes de bachillerato mediante técnicas de predicción de minería de datos*. *Congreso internacional en tecnología, innovación y docencia*, 2(2), 201–208. ISSN: 2594-0465.

Información de los autores.



Rodolfo Villa Noriega, Maestro en Ciencias de la Ingeniería e Ingeniero en Sistemas Computacionales, ambos títulos otorgados por el Instituto Tecnológico de Zacatepec. Se ha desempeñado en el sector privado como desarrollador de sistemas. Actualmente docente de las carreras técnicas de Ofimática y Programación en el Centro de Bachillerato, Industrial y de Servicios No. 136, en Jojutla, Morelos.



Francisco Javier Cartujano Escobar, Doctor en Administración con Especialidad en Sistemas de Información egresado del ITESM, Campus Ciudad de México; Maestro en Ciencias de la Computación e Ingeniero en Sistemas Computacionales, ambos títulos otorgados por el ITESM, Campus Cuernavaca. Obtuvo mención honorífica al terminar la carrera de Ingeniería. Se ha desempeñado en el sector privado como gerente de sistemas y como profesor investigador del Departamento de Computación del ITESM, Campus Ciudad de México. Ha pertenecido al SNI del CONACYT. Actualmente catedrático de la carrera de Ingeniería en Sistemas Computacionales y de la Maestría en Ingeniería en el Instituto Tecnológico de Zacatepec.



Leticia Santa Olalla Ocampo graduada como Licenciada en Informática en 1992 por el Instituto Tecnológico de Zacatepec (ITZ), graduada como Maestra en Ciencias de la Computación en 2008 por el Centro Nacional de Investigación y Desarrollo Tecnológico (CENIDET). Actualmente catedrática de la carrera de Ingeniería en Sistemas Computacionales y de la Maestría en Ingeniería en el Instituto Tecnológico de Zacatepec. Área de interés Ingeniería de Software.



Claudia Noguero González, Maestra en Ciencias Computacionales, egresada del Centro nacional de Investigación y Desarrollo Tecnológico (CENIDET) obtuvo el grado en diciembre de 1996, obtuvo su Licenciatura en Informática en el Instituto Tecnológico de Zacatepec en septiembre de 1993. Ha sido miembro del Sistema Estatal de Investigadores. Actualmente catedrática de la carrera de Ingeniería en Sistemas Computacionales en el Tecnológico Nacional de México campus Zacatepec. Áreas de interés todo lo relacionado a bases de datos.