

FACTORES DE ELECCIÓN DE LICENCIATURAS QUE OFRECE LA DIVISIÓN ACADÉMICA DE INFORMÁTICA Y SISTEMAS USANDO MINERÍA DE DATOS. CASO: ÁRBOL DE DECISIÓN, TÉCNICA AD-TREE

López Ramírez Cristina¹
Morcillo Presenta Freddy Alberto²
De los Santos Torres, Guillermo³
Gordillo De La Cruz, Darwin⁴

RESUMEN

En los últimos periodos escolares, se ha visto disminuida la matrícula de algunas de las Licenciaturas ofertadas por la División Académica de Informática y Sistemas (DAIS), notándose una tendencia de su comportamiento a la baja, específicamente en las Licenciaturas de Informática Administrativa, Tecnologías de la Información y Telemática no siendo así para la Licenciatura en Sistemas Computacionales cuyo comportamiento ha sido estable, la investigación determinó los factores o variables que influyen en el interés de los alumnos del Nivel Medio Superior en el estado de Tabasco, para elegir alguna de las Licenciaturas ofertadas en la División. Para la elaboración del estudio se utilizó la metodología de las fases del proceso de Minería de Datos aplicándose la Técnica AD-tree, utilizando el software de Minería de Datos Weka, donde se realizaron diferentes pruebas y se obtuvo como resultado un árbol de decisión.

Palabras claves: Árbol de decisión, Minería de Datos, AD-tree, Factores de Elección, Estudiantes..

SELECTION FACTORS STUDY OF DEGREES OFFERED BY DIVISION ACADÉMICA DE INFORMÁTICA Y SISTEMAS, APPLYING DATA MINING. CASE: DECISION TREE USING THE TECHNIQUE AD-TREE

ABSTRACT

In the latter school periods has been decreased enrollment of some of the degree courses offered by the Academic Division of Informatics and Systems (División Académica de Informática y Sistemas. DAIS) noticing a downward trend in of his behavior, specifically in degrees of Administrative Computing, Information Technology and Telematics being not so for the degree course in Computer Systems whose behavior has been stable, the investigation determined the factors or variables that influence the interest of students in the Middle top-level in the State of Tabasco, to choose some of the degrees offered in the Academic Division of Informatics and Systems; for the elaboration of study used the methodology phases of the process of data mining applying to the AD-Tree technique, using the Weka data mining software, where tests were done different and resulted in a decision tree. .

Keywords: Decision tree, Data mining, AD tree, Choice factors, Students

¹ Universidad Juárez Autónoma de Tabasco-División Académica de Informática y Sistemas (UJAT-DAIS) cristyna2001@hotmail.com

² Universidad Juárez Autónoma de Tabasco-División de Informática y Sistemas (UJAT-DAIS) fmorcillo64@hotmail.com

³ Universidad Juárez Autónoma de Tabasco - División Académica de Informática y Sistemas (UJAT-DAIS)
guillermodelossantos@hotmail.com

Universidad Juárez Autónoma de Tabasco-División Académica de Informática y Sistemas (UJAT-DAIS)
dargoz_88@hotmail.com

1. Introducción

El documento presenta como finalidad determinar cuáles son los factores o variables que influyen en el interés de los alumnos del nivel medio superior en el estado de Tabasco, para elegir alguna de las diferentes licenciaturas de las que oferta la División Académica de Informática y Sistemas (DAIS) de la Universidad Juárez Autónoma de Tabasco (UJAT). Por lo que se necesita saber la manera en como los estudiantes del nivel medio superior eligen su carrera, ya que su decisión incide en el aumento o disminución de la matrícula escolar en las instituciones de nivel superior, tal es el caso de la UJAT-DAIS en la cual se ha visto disminuida la matrícula de nuevo ingreso en los últimos años.

Por lo anterior se realizó un estudio mediante encuestas a través de la aplicación de cuestionarios de preguntas cerradas a los estudiantes del nivel medio superior próximos a egresar, la captura de la información obtenida se integró en una base de datos.

Para la elaboración del estudio se utilizó la metodología de las Fases del Proceso de la minería de datos, mismas que se aplicaron a la técnica Ad-tree, utilizando el software de minería de datos Weka donde se realizaron diferentes pruebas y se obtuvo como resultado un árbol de decisión.

2. Definición del problema

En los últimos periodos escolares se ha visto disminuida la matrícula en las Licenciaturas ofertadas por la DAIS, situación muy notoria por la tendencia de baja de matrícula de nuevo ingreso de las Licenciaturas en: Informática Administrativa, Tecnologías de la Información y Telemática. Actualmente la Licenciatura en Sistemas Computacionales no se encuentra en esta situación, se ha mantenido a lo largo de estos años con un crecimiento mínimo.

En el año 2000, en la DAIS, la matrícula total era de 1159 alumnos inscritos en las diferentes licenciaturas que ofertan, posteriormente en los siguientes años la matrícula total iba en ascenso hasta llegar al 2004 con un total de 2192 alumnos en ese año. Pero a partir del año 2005 ha venido en descenso el número de matriculados, donde en el año 2010 se contaba con un total de 1710 alumnos inscritos, por lo que fue importante determinar las estrategias a seguir para hacer frente al problema del descenso de matrícula, dado que no se tiene una investigación que permita saber las causas que lo provocan.

De acuerdo a lo anterior, surgió la inquietud de realizar un análisis descriptivo que proporcionó más información útil sobre la elección de una Licenciatura en el área de Tecnologías de Información, formulándose los siguientes objetivos:

2.1 **Objetivo General**

Determinar los factores o variables que influyen en la elección de una Licenciatura en el Área de Tecnologías de la Información y Comunicaciones por parte de los alumnos del nivel medio superior de carácter estatal mediante el proceso de minería de datos.

2.2 **Objetivos Específicos**

- Determinar las preferencias en la elección de una licenciatura en el Área de Tecnologías de la Información y Comunicaciones por parte de los alumnos de los diferentes planteles de educación media superior del estado.
- Analizar e Interrelacionar las variables obtenidas en la investigación y la base de datos que se encuentra en el proyecto “Factores de elección de las Carrera de la DAIS”, y así modelar esta información.
- Aplicar la Minería de Datos, haciendo uso de la técnica Ad-Tree.
- Obtener información precisa y exacta de las preferencias de la elección de una licenciatura en el Área de Tecnologías de la Información y Comunicaciones por parte de los alumnos de los diferentes planteles educativos de nivel medio superior en el estado (CBTA, COBATAB, IDIFTEC, CONALEP y CBTIS).

La investigación será importante dado que proporcionará indicadores sobre la elección de la Licenciatura en el Área de Tecnologías de la Información y Comunicaciones por parte de los alumnos del nivel medio superior, y en base a los resultados obtenidos, se podrán tomar las decisiones adecuadas y diseñar las estrategias que permitan incrementar la demanda real de la matrícula de la DAIS que se mantiene alerta ante las nuevas necesidades sociales.

3. **Aspectos metodológicos**

Para el desarrollo de la investigación se utilizó el enfoque cuantitativo, el cual constituye un mayor nivel de integración y precisión de los datos durante todo el proceso de la investigación.

La metodología que a continuación se presenta, está basada en el método de clasificación proveniente del aprendizaje automático conocido en inglés como Alternating Decision Tree (ADTree).

Los ADTree contienen nodos *splitter* y nodos de predicción, puede ser visto como una consistencia de una raíz nodo de dirección y cuatro unidades de tres nodos cada uno. Cada unidad es una regla de decisión y está compuesta por un nodo *splitter* y dos nodos de predicción que son sus hijos. Las estructuras de datos y el

algoritmo son una generalización de los árboles de decisión. El ADTree fue introducido por Yoav Freund y Llew Mason en 1999.

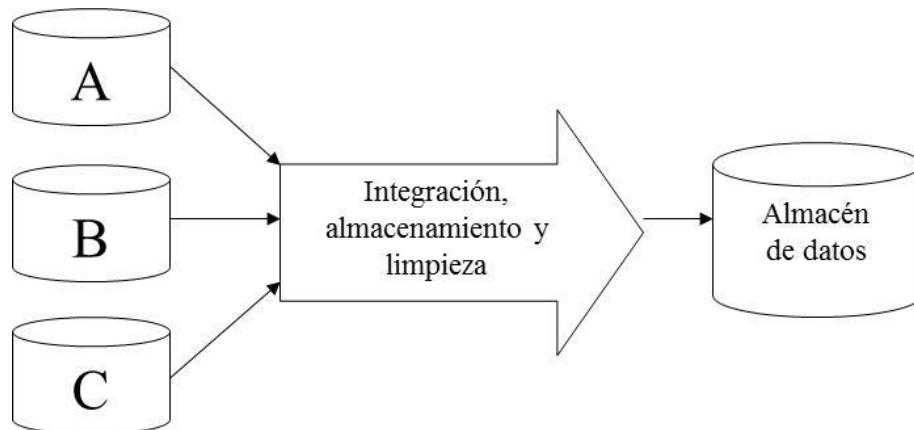
Básicamente, el algoritmo de un ADTree es una estrategia *top-down*; es decir, cada paso de aumento es seleccionado y adiciona una nueva regla o su equivalente a una nueva unidad estable de un nodo *splitter* y dos nodos de predicción. Para el proceso de esta investigación se manipuló el software WEKA (*Waikato Environment for Knowledge Analysis*) el cual es una recopilación de algoritmos de aprendizaje por computador o ML (del inglés *Machine Learning*) para ejecutar tareas de Minería de Datos. Este software facilita comprender y representar de forma segura los resultados mediante tablas, gráficos de calidad y poder compartir los resultados mediante algoritmos de minería de datos, se encuentra en “*Manual de WEKA 2011*”.

Este proyecto está fundamentado en un proceso que se denomina “fases del proceso de extracción del conocimiento”, el KDD es un proceso interactivo e iterativo, porque al regresar y avanzar para obtener información de alta calidad del proceso se interaccionan con los procesos, es decir que las fases del proceso pueden ser en cualquier momento interrumpidas para volver a alguno de sus pasos anteriores, siendo este proceso de iteración muchas veces necesario para poder lograr un descubrimiento de alta calidad, a su vez es interactivo por que interviene el usuario con la validación de la información. (Hernández et. al. (2004)).

Las fases del proceso de KDD constan de cinco pasos:

- Fase de integración y recopilación de datos.
- Fase de selección, limpieza y transformación de datos.
- Fase de evaluación e interpretación.
- Fase de difusión.

Ilustración 1. Integración y recopilación de datos



Las primeras fases del KDD determinan que las fases sucesivas sean capaces de extraer conocimiento válido y útil a partir de la información original. Generalmente, la información que se quiere investigar sobre un cierto dominio de la organización se encuentra: en bases de datos y otras fuentes muy diversas tanto internas como externas muchas de estas fuentes son las que se utilizan para el trabajo transaccional.

Recogida de Información Externa:

Aparte de información interna de la organización, los almacenes de datos pueden recoger información externa:

- Demografías (censo), páginas amarillas, psicografías (perfiles por zonas), uso de Internet, información de otras organizaciones.
- Datos compartidos en una industria o área de negocio, organizaciones y colegios profesionales, catálogos, etc.
- Datos resumidos de áreas geográficas, distribución de la competencia, evolución de la economía, información de calendarios y climatológicas, programaciones televisivas-deportivas, catástrofes.
- Bases de datos externas compradas a otras compañías.
- *OLTP (On Line Transaction Processing)* = suficientes para cubrir necesidades diarias de organización. Ejm. Nómina, facturación, control inventario, etc.
- Insuficientes para funciones más complejas
- Análisis, planificación y predicción → toma decisiones a largo plazo
- Limpieza (*data cleansing*) y criba (selección) de datos:

Se debe eliminar el mayor número posible de datos erróneos o inconsistentes (limpieza) e irrelevantes (criba).

Métodos estadísticos casi exclusivamente.

- Resúmenes e histogramas (detección de datos anómalos).
- Selección de datos (muestreo, ya sea verticalmente, eliminando atributos, denominado “selección de características”, u horizontalmente, eliminando tuplas, denominado “muestreo”).
- Redefinición de atributos (agrupación o separación).

La selección y la limpieza pueden acompañarse de “transformación” de atributos (numeración, discretización).

- El resultado es un conjunto de filas y columnas denominado: VISTA MINABLE.
- La vista minable integra datos de diferentes fuentes, los limpia, selecciona y transforma, y los tipa, con el fin de prepararlos para la modelización.

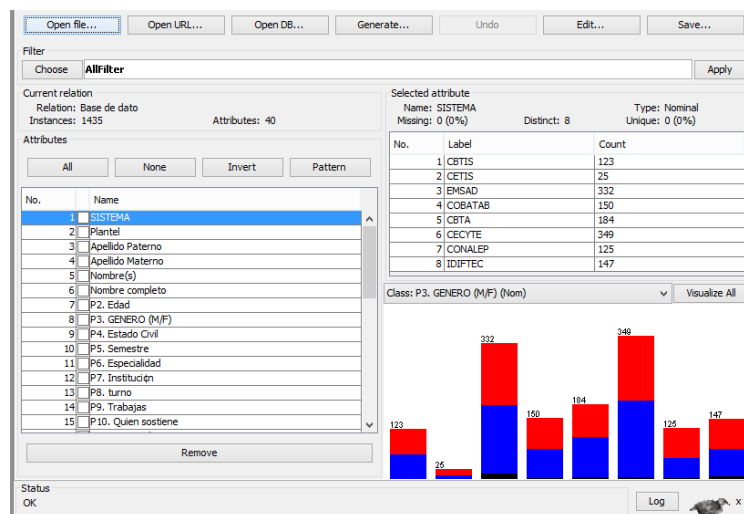
De acuerdo con Hernández et al. (2004), la tecnología de base de datos ha respondido a este reto con una nueva arquitectura surgida recientemente: almacenes de datos (*datawarehouse*).

El almacenamiento de datos, se refiere a la tendencia comercial popular de la recolección y limpieza de datos transaccional para que estén disponibles para el análisis en línea y soporte de decisión de acuerdo a Fayyad et. al (1996). Britos et al. (2005), define: “un *datawarehouse* es un conjunto de datos integrados orientados a una materia, que varían con el tiempo y que nos son transitorios, los cuales soportan el proceso de toma de decisiones de una administración”, esta definición se refiere a que los datos son relacionados a las transacciones generadas por la empresa.

4. Pruebas y resultados

En primer lugar, *Cargar la base de datos*: Para cargar la base de datos se utilizará el botón OPEN FILE de la interfaz Explorer (pestaña *Preprocess*), se seleccionará el directorio data y dentro de él, el fichero Base de datos.csv, esta base de datos contiene un cierto número de registros de alumnos de diferentes planteles de nivel medio superior del Estado de Tabasco; con el objetivo de determinar los factores o variables que influyen en la elección de una Licenciatura en el Área de Tecnologías de la Información y Comunicaciones. El resultado de abrir la base de datos, será una pantalla como la que se muestra en la tabla 1.

Tabla 1. Interfaz Explorer con archivo Base de datos.csv

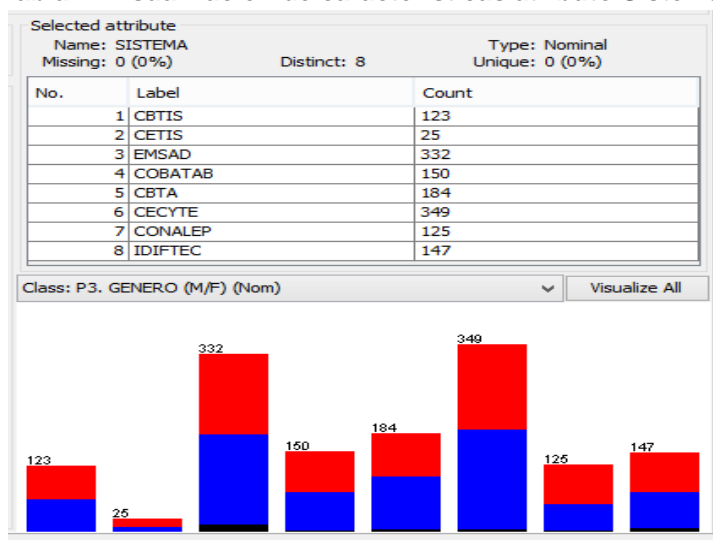


Fuente: Elaborada por los autores

El modelo muestra 40 atributos, haciendo clic sobre cada uno de los atributos, se muestra información sobre el mismo en la parte derecha de la ventana. En el caso de atributos discretos se indica el número de instancias que toman cada uno de los valores posibles; y en el caso de atributos reales se muestran los valores máximo, mínimo, medio y la desviación estándar. Así mismo, se puede mostrar un gráfico en el que las distintas clases se representan con colores distintos, en función de los valores del atributo elegido.

A continuación se describirán algunos de los atributos de acuerdo a lo estipulado para weka. *Sistema*: El tipo de dato es Nominal, tiene valores estándar de los nombres de cada plantel que son 8, donde CBTIS tiene 123 alumnos, CETIS 23 alumnos, así progresivamente hasta llegar al IDIFTEX que cuenta con 147 alumnos a egresar. Las barras representan la cantidad de alumnos de cada plantel y el género a la que pertenecen (F/M); es decir, el color rojo representa a los hombres (M) y el azul a las mujeres (F). Tabla 2.

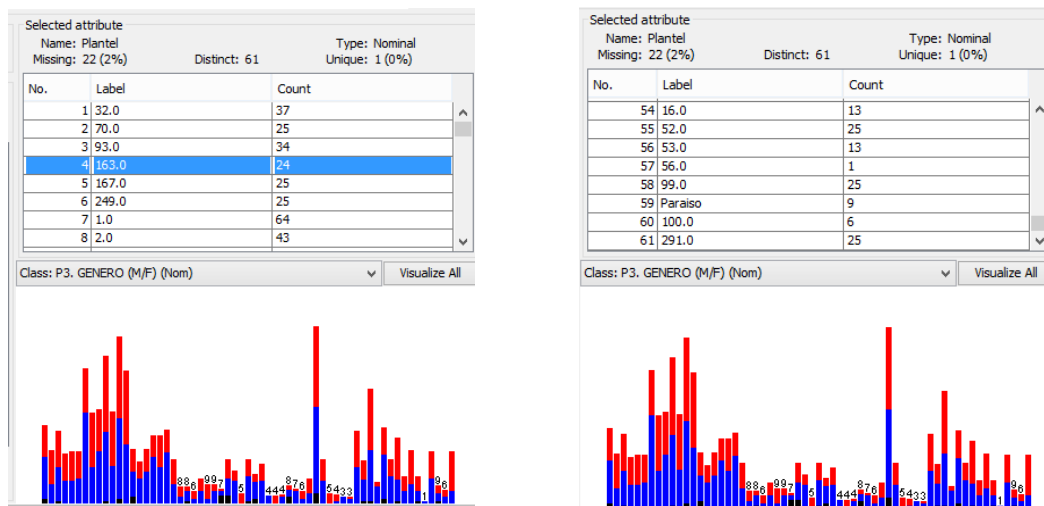
Tabla 2. Visualización de características atributo Sistema.



Fuente: Elaborada por los autores

- a) *Plantel*: El tipo de dato también es Nominal, la cual tiene valores estándar del número del plantel, donde el plantel No. 32 tiene 37 alumnos, el plantel 70 tiene 25 alumnos, así progresivamente hasta llegar al plantel No. 291 que cuenta con 25 alumnos. Las barras significan la cantidad de alumnos de cada institución que en total son 61 planteles, el género al que pertenecen (F/M); el color rojo representa a los hombres (M) y el azul a las mujeres (F). Tabla 3.

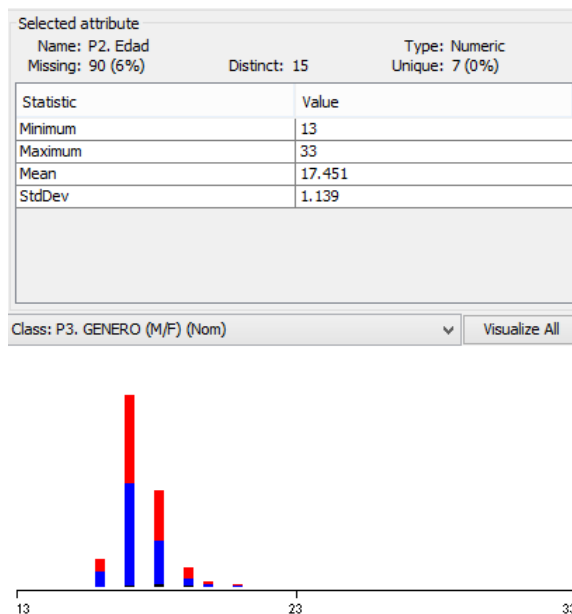
Tabla 3 Visualización de características atributo Plantel.



Fuente: Elaborada por los autores

b) *Edad*: Su tipo de dato es Numérico, tiene un valor mínimo de 13 años y un valor máximo de 33 años de edad, tiene una media de 17.451 y una desviación estándar de 1.139. Si se observa en la Ilustración 3, se tiene a cierta cantidad de alumnos (rojo) y alumnas (azul) que están entre el rango de 13 a 23 años de edad; es decir, tienen un promedio de edad de 16 a 21 años y 0 alumnos(as) entre el valor de 23 y 33 años de edad. Tabla 4.

Tabla 4. Visualización de características atributo Edad.



Fuente: Elaborada por los autores

- c) *Género*: Donde el tipo de dato es Nominal, la cual representa a las mujeres y hombres, con valores estándar de F y M, donde 708 personas son mujeres y 690 son hombres. Tabla 5.

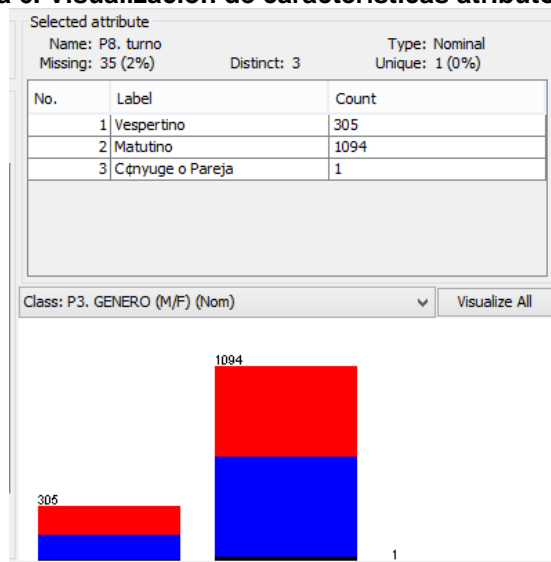
Tabla 5. Visualización de características atributo Género.



Fuente: Elaborada por los autores

- d) *Turno*: Donde el tipo de dato es Nominal, presenta el turno al que pertenecen, podemos apreciar que tenemos a 305 alumnos en el turno Vespertino y 1094 en el matutino. Representados por mujeres (azul) y hombres (rojo). Tabla 6.

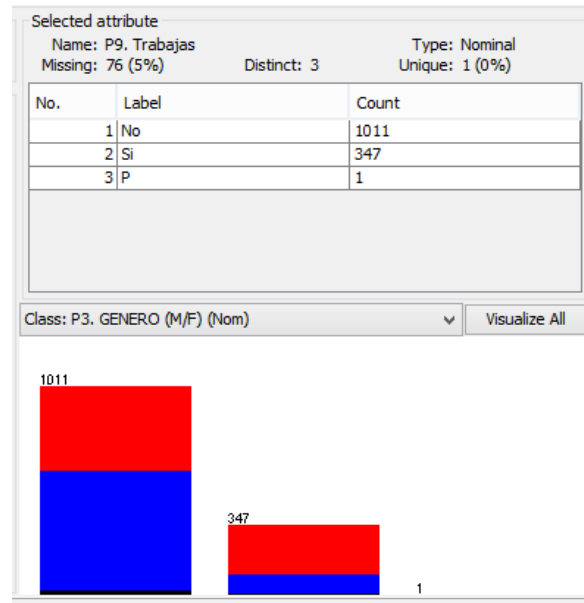
Tabla 6. Visualización de características atributo Turno.



Fuente: Elaborada por los autores

- e) *Trabaja*: Dato del tipo Nominal, y presenta valores estándar de Si y No; se puede ver que tenemos a 1011 alumnos que no trabajan, 347 si lo hacen, tanto hombres (rojo) como mujeres (azul). Tabla 7.

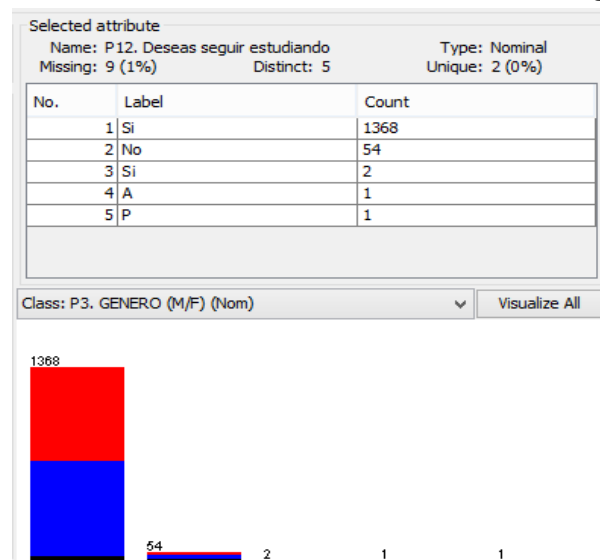
Tabla 7. Visualización de características atributo Trabaja.



Fuente: Elaborada por los autores

- f) *Deseas seguir estudiando*: El tipo de dato es Nominal, tiene valores estándar de Si y No, donde 1368 alumnos desea seguir estudiando y 54 no lo desean. Tabla 8.

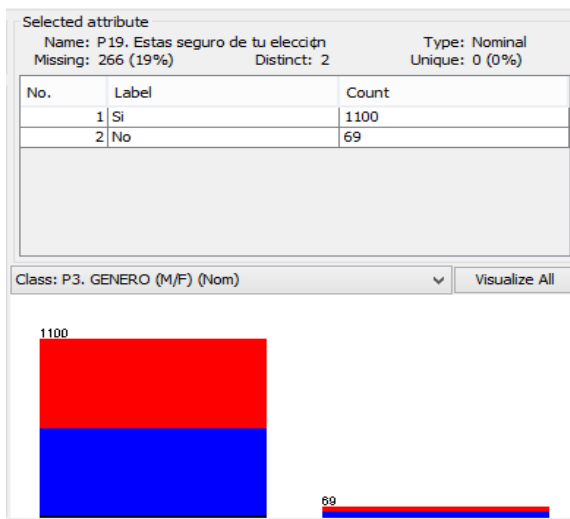
Tabla 8. Visualización de características atributo deseas seguir estudiando.



Fuente: Elaborada por los autores

- g) *Están seguro de su elección*: Dato Nominal, contiene valores estándar de Si y No, donde 1100 alumnos están seguros de su elección y 69 no lo están tanto hombres (rojo) como mujeres (azul). Tabla 9.

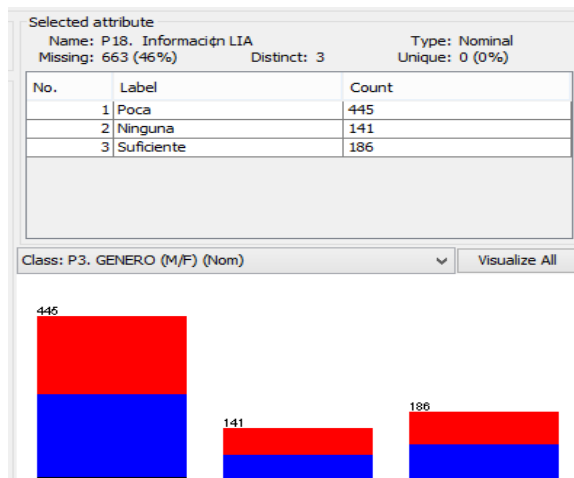
Tabla 9. Visualización de características atributo Están seguro de su elección.



Fuente: Elaborada por los autores

- h) *Información LIA*: Dato Nominal, contiene valores estándar de Poca, Ninguna y Suficiente, donde 445 alumnos tienen poca información sobre la carrera, 141 ninguna y 186 tienen suficiente información, incluyendo hombres (rojo) como mujeres (azul). Tabla 10.

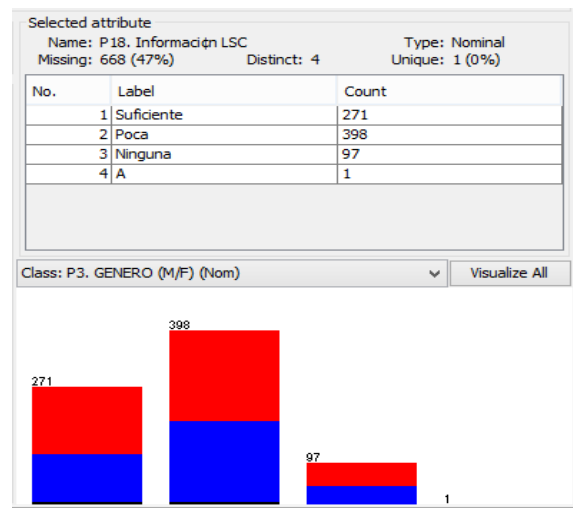
Tabla 10. Visualización de características atributo Información LIA.



Fuente: Elaborada por los autores

- i) *Información LSC*: Dato Nominal, contiene valores estándar de Suficiente, Poca y Ninguna, donde 271 alumnos tienen información suficiente sobre la carrera, 398 poca y 97 ninguna información, incluyendo hombres (rojo) como mujeres (azul). Tabla 11.

Tabla 11. Visualización de características atributo Información LSC.



Fuente: Elaborada por los autores

- j) *Información LTI*: Dato Nominal, contiene valores estándar de Poca, Ninguna y Suficiente, donde 310 alumnos tienen poca información sobre la carrera, 162 ninguna y 148 tienen suficiente información, incluyendo hombres (rojo) y mujeres (azul). Tabla 12.
- k) *Información LT*: Dato Nominal, contiene valores estándar de Suficiente, Poca y Ninguna, donde 42 alumnos tienen información suficiente sobre la carrera, 171 poca y 327 ninguna información, incluyendo hombres (rojo) como mujeres (azul). Tabla 12.

Tabla 12. Visualización de características atributo Información LTI y LT.



Fuente: Elaborada por los autores

4.1 Run Información.

Tipo de clasificador utilizado la base de datos sobre la que se trabaja, el número de instancias (1435), el número de atributos (40) y su nombre y el tipo de test (*cross validation*).

Tabla 13. Ventana Run Information del árbol ADTree.

```
=== Run information ===  
  
Scheme:weka.classifiers.trees.ADTree -B 10 -E -3  
Relation:      Base de dato  
Instances:    1435  
Attributes:   40  
             SISTEMA  
             Plantel  
             Apellido Paterno  
             Apellido Materno  
             Nombre(s)  
             Nombre completo  
             P2. Edad  
             P3. GENERO (M/F)  
             P4. Estado Civil  
             P5. Semestre  
             P6. Especialidad  
             P7. Institucion  
             P8. turno  
             P9. Trabajas  
             P10. Quien sostiene  
             P11. Correo electronico  
             P12. Deseas seguir estudiando  
             P13. A que nivel  
             P14. prueba de Conocimiento  
             P15. Que licenciatura  
             P16. Motivos por los que elegiste  
             P17. Donde te gustaria estudiar  
             P18. Informacion LIA  
             P18. Informacion LSC  
             P18. Informacion LTI  
             P18. Informacion LT  
             P19. Estas seguro de tu eleccion  
             P20.1  
             P20.2  
             P20.3  
             P20.4  
             P20.5  
             P20.6  
             P20.7  
             P20.8  
             P20.9  
             P20.10  
             P20.11  
             P20.12  
             P20.13  
Test mode:10-fold cross-validation
```

Fuente: Elaborada por los autores

4.2 Classifier Model

Seguidamente se muestra el árbol que se ha generado y el número de instancias que clasifica cada nodo. Podemos apreciar que tenemos el resultado del total de nodos ejecutados (31), el número de vaticinador de nodos (21) y posteriormente, el tiempo que tomó construir el modelo (1.67 segundos): Tabla 14.

Tabla 14. Ventana Classifier model del árbol ADTree

```
=== Classifier model (full training set) ===  
  
Alternating decision tree:  
  
: -0.013  
| (1)P9. Trabajas = No: -0.161  
| (1)P9. Trabajas != No: 0.494  
| | (7)P7. Institución = IDIFTEC: 0.897  
| | (7)P7. Institución != IDIFTEC: -0.091  
| (2)P15. Que licenciatura = 3.0: 0.425  
| (2)P15. Que licenciatura != 3.0: -0.085  
| | (5)P16.Motivos por los que elegiste = 3.0: 0.578  
| | | (8)P6. Especialidad = Tec. en Inform tica: -0.952  
| | | (8)P6. Especialidad != Tec. en Inform tica: 0.317  
| | (5)P16.Motivos por los que elegiste != 3.0: -0.029  
| (3)P14. urea de Conocimiento = 1.0: 0.367  
| (3)P14. urea de Conocimiento != 1.0: -0.081  
| | (6)P14. urea de Conocimiento = 3.0: -0.208  
| | (6)P14. urea de Conocimiento != 3.0: 0.083  
| | | (10)P6. Especialidad = Admen.: -0.731  
| | | (10)P6. Especialidad != Admen.: 0.037  
| | (9)Apellido Paterno = Martinez: 0.814  
| | (9)Apellido Paterno != Martinez: -0.019  
| (4)P16.Motivos por los que elegiste = 2.0: 0.412  
| (4)P16.Motivos por los que elegiste != 2.0: -0.052  
  
Legend: -ve = F, +ve = M  
Tree size (total number of nodes): 31  
Leaves (number of predictor nodes): 21  
  
Time taken to build model: 1.67 seconds
```

Fuente: Elaborada por los autores

4.3 Stratified Cross-Validation and Confusion Matrix.

Por último se muestran los resultados del test. Indican la capacidad de clasificación esperable para el árbol y la matriz de confusión. En los resultados observamos que tenemos 65.2361% de certeza y 34.7639% de errores. Tabla 15.

Un dato importante que debemos tener en cuenta son los resultados que nos arroja la matriz de confusión: (Tabla 15). Al analizar la matriz de confusión vemos claramente que los valores con respecto a la variable son favorables.

La matriz se interpreta de la siguiente manera: muestra en las diagonales qué tanto porcentaje de error o de confiabilidad se puede tener en la diagonal; entre mayor

sea el resultado de su respectiva letra mayor es el grado de confiabilidad: en este caso vemos que $484 > 262$ y $428 > 224$ por lo que los resultados obtenidos son confiables. Los resultados nos presentaron un 65.2361% de certeza y una matriz de confusión con 484 puntos buenos contra solo 262 errores para "a", 428 puntos buenos para "b", con solo 224 errores.

Tabla 15. Ventana Stratified cross-validation del árbol ADTree and Confusion Matrix.

```

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      912          65.2361 %
Incorrectly Classified Instances    486          34.7639 %
Kappa statistic                    0.3041
Mean absolute error                 0.4576
Root mean squared error             0.4726
Relative absolute error             91.5383 %
Root relative squared error         94.5307 %
Total Number of Instances          1398
Ignored Class Unknown Instances     37

=== Detailed Accuracy By Class ===

```

Porcentaje de aciertos desglosados por clase.	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.684	0.38	0.649	0.684	0.666	0.69	F
	0.62	0.316	0.656	0.62	0.638	0.687	M
Weighted Avg.	0.652	0.348	0.653	0.652	0.652	0.688	

```

=== Confusion Matrix ===

 a  b  <-- classified as
484 224 |  a = F
262 428 |  b = M

```

Fuente: Elaborada por los autores

5. Conclusiones

La hipótesis obtenida durante el desarrollo de la técnica ADTree mediante WEKA, indica que la mayoría de los alumnos de Nivel Medio de Educación Superior, tienen una gran preferencia hacia la Universidad Juárez Autónoma de Tabasco (UJAT); es decir, indican que la UJAT tiene una enorme demanda potencial conforme a las preferencias de los estudiantes en el Estado de Tabasco para estudiar una Licenciatura.

No obstante, esto no quiere señalar que los educandos prefieran estudiar una Licenciatura en el Área de Tecnologías de la Información y Comunicaciones en la División Académica de Informática y Sistemas (DAIS), ya que prefieren estudiar una Licenciatura totalmente diferente a nuestra División. Sin embargo, la Licenciatura de mayor demanda por parte de algunos escolares es la de Informática Administrativa;

por otra parte las gráficas de los atributos revelaron que los alumnos tienen mayor información sobre la Licenciatura de Sistemas Computacionales, seguida de Informática Administrativa, posteriormente Tecnologías de la Información y por último la Licenciatura en Telemática.

Al tomar la decisión final, los alumnos prefieren una licenciatura por tener competitividad, inspiración o experiencias; también lo hacen por el prestigio de la carrera, en especial en el área de Tecnologías de la Información.

6. Bibliografía

D. García. "Manual de WEKA" 2011

López, N. & Montejó, L. (2012) "Análisis del Desempeño Profesional de los Egresados de la Licenciatura en Sistemas Computacionales de la Cohorte (2008)". Publicada por la Universidad Juárez Autónoma de Tabasco. Villahermosa, Tabasco, México

Villanueva, M.; Ronzón, J.; Moheno, G. & Zapata, T. (2004) "Estudio de Egresados de la Licenciatura en Informática Administrativa (1997)". Publicado por la Universidad Juárez Autónoma de Tabasco. Villahermosa, Tabasco, México.