

Factores de elección de licenciaturas en la División Académica de Informática y Sistemas (DAIS) aplicando minería de datos

*Manuel Villanueva*¹
*José Ronzón*²
*Joel Luna*³

Resumen

El objetivo de la investigación permite identificar los factores que influyen en los estudiantes, al momento de seleccionar una de las cuatro Licenciaturas que ofrece la División Académica de Informática y Sistemas (DAIS) de la Universidad Juárez Autónoma de Tabasco, México. Su desarrollo se realizó en seis Subsistemas de Instituciones de Educación Media Superior de carácter público estatal y federal del estado de Tabasco. El enfoque metodológico se fundamenta mediante la aplicación del proceso de KDD “Knowledge Discovery in Databases” conocido como Descubrimiento de Conocimiento en Bases de Datos, el cual es interpretado como “El proceso no trivial de identificar patrones válidos, nuevos, potencialmente útiles y en última instancia comprensible en los datos”, utilizando la técnica de Predicción de Regresión Lineal de la Herramienta Weka. Para la obtención de información y respectivo análisis, se diseñó un cuestionario aplicado a los seis Subsistemas, cuyos resultados informacionales fueron almacenados en una base de datos y se depuró para aplicarle la técnica de minería de datos. Los resultados obtenidos por la Herramienta Weka, muestran los motivos y factores que influyen en la toma de decisiones de los alumnos al momento de la selección de una carrera. Se prevé que los resultados han de ser considerados por las autoridades académicas para la toma de decisiones y definición de estrategias que les permitan promocionar sus licenciaturas.

Palabras clave: factores de elección, minería de datos, técnica de predicción, regresión lineal, weka.

Recibido: 22/03/14 Devuelto para revisión 02/04/14 Aceptado: 30/04/14

¹ Licenciado en Física. Especialistas Ciencias de la Computación. Magister en Cómputo Estadístico. Magister en Ciencias de la Computación. Magister en Ingeniería de Sistemas. Actualmente cursa el Doctorado en Sistemas Computacionales en la Universidad del Sur.

Correo electrónico: manuel.villanueva@ujat.mx

² Licenciado en Economía. Magister en ciencias en Planificación de Empresas y Desarrollo Regional. Profesor Investigador de la División Académica de Informática y Sistemas de la Universidad Juárez Autónoma de Tabasco. Actualmente es el responsable del grupo de investigación: Modernización de Sistemas de Software en la Calidad Organizacional y la Línea de Generación y Aplicación del Conocimiento.

Correo electrónico: jronson@hotmail.com, jose.ronson@ujat.mx

³ Licenciado en Sistemas Computacionales de la Universidad Juárez Autónoma de Tabasco.

Correo electrónico: leoj-alm@hotmail.com

Choice Factors of Academic Degrees in Computer Science and Systems Division of Applying Data Mining

Abstract

The objective of the research makes it possible to identify the factors that influence the students, at the time of selecting one of the four undergraduate degrees offered by the Division of Academic Computing and Systems (DAIS) of the University of Juarez Autonomous Tabasco, Mexico. Its development was carried out in six subsystems of Secondary Education Institutions of a public nature of the state and federal state of Tabasco. The methodological approach basis through the implementation of the KDD process "Knowledge Discovery in Databases" known as Knowledge Discovery in databases, which is interpreted as "the non-trivial process of identifying valid patterns, new, potentially useful, and ultimately understandable in the data", using the technique of linear regression prediction tool "WEKA". For the obtaining of information and analysis, design a questionnaire administered to the six subsystems, informational whose results were stored in a database and debugged to apply the technique of data mining. The results obtained by the "WEKA" tool, show the reasons and factors that influence the decision-making of the students at the time of the selection of a career. It is anticipated that the results have to be considered by the academic authorities for the decision-making and definition of strategies that will enable them to promote their undergraduate degrees.

Key words: Factors of choice, Data mining, Predictive Technical, Linear regression, "WEKA"

Introducción

La técnica de minería de datos en la actualidad se posiciona de forma estratégica y a su vez promueve ventajas en el proceso de la toma de decisiones en determinadas empresas e instituciones.

Asimismo, es una técnica considerada en el ámbito mundial, sobre todo se resalta su importancia en el área educativa, al considerar que a través de esta, se generan y almacenan altos volúmenes de datos, que permiten soportar y determinar en las Instituciones Edu-

cativas, cuáles son las demandas de la sociedad actual Calleja (2010).

De igual forma, se desea conocer los factores que influyen en la elección de una licenciatura por parte del estudiante, así como también determinar si este domina el conocimiento sobre el contexto a enfrentar o cuáles son los aspectos influyentes de este, en momentos cruciales, en caso de requerir llevar a efecto una toma de decisión. En tal sentido, si se considera, que los lapso de tiempo para poder decidir son relativamente cortos, aunado con los diversos cambios durante la transición de la adolescencia,

y madurez intelectual, se prevé la efectividad de toma de decisiones asertivas, al considerar que esta define el rumbo que tomará su vida y conforma el estilo de la misma, González y Álvarez (2008).

De manera muy concreta, en el estado de Tabasco y de acuerdo con el objeto de estudio la División Académica de Informática y Sistemas (DAIS), de la Universidad Juárez Autónoma de Tabasco (UJAT), se publicó el primer estudio formal relacionado con algunas de las razones principales de la elección de la Licenciatura en Informática Administrativa, cuyos datos estadísticos muestran el decreciente número de estudiantes que han ingresado a las carreras de la DAIS, que ofrece la UJAT, según (Villanueva, Zapata, Arceo, y Ronzón, 2004, p. 36)

Tabla 1
Serie histórica de nuevo ingreso

P.E.	2007	2008	2009	2010	2011	2012	2013
LIA	132	178	215	157	126	102	74
LSC	127	205	248	257	214	197	138
LT	0	38	49	36	22	6	21
LTI	0	23	34	22	8	21	21
Tot.	259	444	546	472	370	326	254

Fuente: Gil, (2010), Gil, (2011); Gil, (2012), Piña, (2013) y Piña, (2014)

Los resultados presentados evidencian la evolución de una problemática iniciada desde el año 2009 y en la actualidad prevalece, situación que promovió la conformación de 11 comisiones para responder al incremento de la matrícula en la DAIS.

Una de estas comisiones, específicamente la de Investigación Educativa, logró integrar dos Proyectos de Investigación Institucionales denominados:

- estudio sobre factores de elección de las licenciaturas que ofrece la División Académica de Informática y Sistemas y,
- estudio de demanda potencial en el ámbito educativo medio superior de Tabasco, en el área de tecnologías de la información.

Sin embargo, en la actualidad no se dispone de un estudio integral que determine cuáles son los factores que influyen en la toma de decisiones y selección de los estudiantes correspondientes a la educación media superior, y que a su vez apoye a las autoridades de la DAIS para lograr incrementar el número de alumnos a las carreras ofertadas.

Es por ello, que la presente investigación aplicada, titulada “*Estudio sobre factores de elección de las licenciaturas que ofrece la DAIS. Aplicando Minería de datos. Caso: Método de Regresión, mediante la Técnica de Regresión Lineal*”, busca detectar estos factores, con el fin de contribuir con la superación de la situación expuesta.

Objetivo

El objetivo se fundamenta en la necesidad de identificar los factores que influyen en los alumnos para la elección del nivel medio de educación superior mediante la minería de datos, específicamente hacia las licenciaturas que ofrece la División Académica de Informática y Sistemas (DAIS) de la Universidad Juárez Autónoma de Tabasco, México.

Su desarrollo, se llevó a efecto mediante el análisis de los datos a través de la aplicación del método de regresión, específicamente la técnica de regresión lineal, mediante la Herramienta WEKA.

Metodología

Para el análisis y recolección de datos, se utilizó el enfoque cuantitativo, el cual permite mediante el uso de las técnicas de recolección y el análisis de datos, así como de métodos estadísticos, determinar estudios de casos concretos para llegar a una descripción general o comprobar las hipótesis establecidas previamente, es decir busca la mayor claridad entre los elementos que conforman el problema, buscando que tenga definición, los limita y con exactitud saber dónde inicia el problema según Hernández, Fernández y Baptista, (2007).

Población y muestra

Con base de lo expuesto, para lograr la recopilación de la información relacionada con las preferencias de la elección de una Licenciatura en el Área de Tecnologías de la In-

formación y Comunicaciones, fue considerada como muestra del estudio los alumnos del Centro de Bachillerato Tecnológico Agropecuario (CBTA), Educación Media Superior a Distancia (COBATAB EMSAD), Colegio de Bachilleres de Tabasco (COBATAB), Instituto de Difusión Técnica (IDIFTEC), Colegio Nacional de Educación Profesional Técnica (CONALEP), Centro de Bachillerato Tecnológico Industrial y de Servicios (CBTIS) / Centro de Estudios Tecnológicos Industrial y de Servicios (CETIS) y Centro de Estudios Científicos y Tecnológicos del Estado de Tabasco (CECYTE).

Muestra cuyos resultados permiten que las autoridades de la DAIS conozcan los principales factores de incidencia y se les facilite el definición o dominio de estrategias que soporten sus decisiones de forma adecuada y a su vez se proyecte el incremento de la matrícula estudiantil. Asimismo, la investigación está basada en un proceso que se denomina “fases del proceso de extracción del conocimiento”, el Knowledge Discovery in Database (KDD). Es un proceso interactivo e iterativo.

Iterativo porque al regresar y avanzar para obtener información de calidad del proceso se interaccionan con estos, es decir que las fases correspondientes 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 calidad, a su vez es interactivo por que interviene el usuario con la validación de la información extraída.

De acuerdo Hernández, Ramírez y Ferri (2004) (citado por Brito, Rosete y Acosta, 2008, p.2) el proceso de KDD, se organiza en torno a cinco pasos, que se describen:

- fase de Integración y recopilación de datos.
- fase de selección, limpieza y transformación de datos.
- fase de la Minería de datos.
- Fase de evaluación e interpretación.
- Fase de difusión.

Desde la perspectiva teórica esta investigación considera como base el proceso de descubrimiento de conocimiento las fases descritas por Hernández, Ramírez, y Ferri, (2004), las cuales se fundamentan desde una repetición interactiva de los pasos que se especifican:

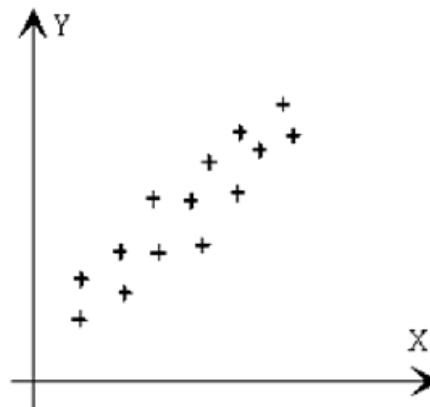
- limpieza de datos: eliminar los datos ruidosos o irrelevantes.
- integración de datos: combinar los datos desde múltiples fuentes.
- selección de datos: recuperar desde la base de datos los datos relevantes para la tarea como objeto de análisis.
- transformación de los datos: transformar los datos o llevarlos a formas apropiadas para extraer el conocimiento.
- Minería de datos: aplicar métodos inteligentes para extraer los patrones de los datos.
- Evaluación de patrones: identificar los patrones verdaderamente interesantes representando conocimiento basado en medidas de interés.

- Representación del conocimiento: utilizar técnicas de representación y visualización para mostrar al usuario el conocimiento extraído.
- Presentación de los conocimientos.

Como herramienta de soporte ha sido aplicada la técnica de regresión lineal. Concebida de esta forma cuando la función es lineal, es decir, requiere la determinación de dos parámetros: la pendiente y la ordenada en el origen de la recta de regresión, $y = ax + b$.

La regresión permite además, determinar el grado de dependencia de las series de valores X e Y, prediciendo el valor estimado que se puede obtener para un valor x que no esté en la distribución.

Gráfica 1
Regresión lineal



Fuente: Morates (1997)

También se hizo uso de la Herramienta Weka (Waikato Environment for Knowledge Analysis). Utilizada para la operatividad de la técnica de "Regresión Lineal Automatizada" concebida como una plataforma de Software para aprendizaje automático y minería de datos escritos en Java y desarrollado en la Universidad de Waikato. Weka, es un software libre distribuido bajo licencia GNU-GPL (Licencia Pública General del Sistema operativo GNU), donde GNU es un sistema Operativo desarrollado bajo UNIX que es considerado un software libre.

Weka, como software programado en Java está orientado a la extracción de conocimientos desde bases de datos con grandes cantidades de información. (Alonso, 2013). Existen otras herramientas similares como *Oracle Data Miner Clementine*, pero el hecho de que Weka sea desarrollado bajo licencia GPL, lo hace una alternativa muy interesante.

Resultados

Los resultados determinan el análisis de los subsistemas según se mencionan, asimismo permiten el análisis y la interrelación de las variables obtenidas en la investigación y la base de datos que se encuentra en el proyecto "Factores de elección de las carreras de la DAIS", con el fin de modelar la información obtenida.

Análisis del subsistema

1. CBTA

Se determinó que los motivos que influyen en la selección de las carreras en los alumnos de este subsistema inciden con: aspectos vocacionales y habilidades personales.

De igual forma, se considera en la escala de *Muy importante* el grado de interés que le dan a los factores que influyen en su elección.

Factores:

- la conveniencia de esa carrera para las personas de género masculino.
- la posibilidad de conseguir empleo.
- el ser accesible, para efectos de costos

Análisis del subsistema

2. CETIS/CBTIS

Los resultados en este subsistema son coincidentes con los obtenidos del subsistema 1 CBTA. Sin embargo, el grado de importancia que le dan a los factores que influyen en su elección es: *Importante*.

Factores:

- el ser accesible por su costo
- el poseer conocimientos o estudios previos.
- las calificaciones obtenidas en las materias afines a la carrera

Análisis del subsistema

3. CECYTE

Se encontró que los motivos que conllevan a la elección de la carrera en los alumnos de este subsistema se basan en el consejo de familiares y amigos. También se detectó el grado de importancia que le ptorgan a los factores que influyen en su elección es: *Muy importante*

Factor:

- la influencia de amigos
- la conveniencia de esa carrera para las personas del mismo género.
- la posibilidad de conseguir empleo

Análisis del subsistema 4. COBATAB

Planteles presenciales.

Se identificó que los motivos que inducen a la elección de carrera en los alumnos de este Subsistema son: por tener vocación y habilidades personales. Y el grado de importancia que le dan a los factores que influyen en su elección es: *Muy importante*

Factor:

- la aceptación de mis padres
- el hecho de ser accesible por su costo
- la conveniencia de esa carrera para las personas de sexo masculino

Planteles a distancia.

Con respecto a este subsistema, se encontró que los motivos que impulsan a la elección de carrera en los alumnos de este plantel son diversas y a su vez el grado de relevancia que le dan a los factores que influyen en su elección es: *Muy importante*

Factor:

- la conveniencia de esa carrera para las personas del mismo sexo.

- la aceptación de mis padres
- el hecho de tener conocimientos o estudios previos

Análisis del subsistema 5. CONALEP

Prevalece en este subsistema, la selección de la carrera en función de la vocación, así como del dominio de habilidades personales, y la escala en su elección es: *Muy importante*

Factor:

- la obtención de buenos ingresos
- el ser accesible por su costo
- la influencia de mis amigos

Análisis del subsistema 6. IDIFTEC

La vocación y el dominio de habilidades personales, aspectos claves de la elección de carrera, y como grado de importancia que le dan a los factores que influyen en su elección es: *Importante*

Factor:

- la posibilidad de conseguir empleo.
- el reconocimiento social que tienen los profesionistas de esta carrera
- la obtención de buenos ingresos

Conclusiones y recomendaciones

La aplicación de la herramienta de KDD y el análisis de las pruebas realizadas mediante el algoritmo de clasificación Regresión Lineal de la

la aplicación de la herramienta de KDD y el análisis de las pruebas realizadas mediante el algoritmo de clasificación Regresión Lineal de la herramienta Weka, se concluye satisfactoriamente, luego de aplicar cada una de las fases o etapas de KDD en donde fue necesario limpiar y dividir la base de datos única para su mejor análisis.

Por lo cual, se determina a continuación los tres principales motivos influyentes en los alumnos de nivel medio superior de carácter público, para elegir una de las cuatro licenciaturas que ofrece la DAIS:

a).-“A los egresados de la carrera les pagan bien” con un 57%, “consejo de familiares y amigos” con un 29 % y “por tener vocación y habilidades personales” con un 14%.

b).-Se determina también, que el grado de importancia que le dan los alumnos a los factores que ellos mismos eligieron para su elección de la carrera son: “Muy importantes” con el 57% e “Importante” con un 43%.

c).- Como conclusión general en este proyecto es que se alcanzaron los objetivos propuestos, ya que se obtuvieron satisfactoriamente los 10 factores de elección que influyen en los alumnos de nivel medio superior al momento de elegir una carrera los cuales son: “La conveniencia de esa carrera para las personas de género afín” con un 21%, “el ser accesible por su costo” con un 17 %, “el poseer conocimientos o estudios previos” con un 13%, “La posibilidad de conseguir empleo” con un 13%, “La

influencia de mis amigos” con un 13%, “El campo de trabajo es amplio y con futuro” con un 8%, “La obtención de buenos ingresos” con un 8%, “la aceptación de mis padres” con un 8%, “las calificaciones obtenidas en las materias afines a la carrera” con un 4%, y, “el reconocimiento social que tienen los profesionistas de esta carrera” con un 4%.

d).- Conocer los motivos y los factores que impulsan a los alumnos del nivel medio superior a la toma de decisión sobre que licenciatura van a elegir, es de gran utilidad para las autoridades universitarias, con el fin afianzar la toma de decisiones correspondientes, así como definir estrategias que permitan dar a conocer a profundidad las licenciaturas que ofrece la DAIS.

Las dificultades encontradas para llegar a estos resultados, se conciben como no relevantes ya que algunos de estas fueron errores de captura de datos y caracteres inválidos encontrados en la base de datos. Con respecto a lo anterior, para el uso de la base de datos se usó una hoja de cálculo de Open Office que al ser de licencia de código abierto, facilitó su utilidad al ser utilizado sobre cualquier sistema operativo, facilitando así la conversión del archivo a formato de .csv antes de realizar el análisis con el software de Weka.

Bibliografía

Alonso C. (2013). Obtenido de Waitako Environment for Knowledge Analysis. Introducción Básica. Departamento de Informática de la universidad de Va-

- lladolid. Recuperado de <http://www.infor.uva.es/~calonso/IAII/Aprendizaje/weka/IntroduccionWeka.pdf>
Recuperado el 22/03/2013.
- Brito, R., Rosete, A. y Acosta, R. (2008) Desarrollo de un proceso de KDD en el ámbito docente: Preparación de los datos. 14 Convención científica de ingeniería y arquitectura.
- Calleja, A. (2010). Minería de Datos con Weka para la predicción del Precio de Automóviles de Segunda Mano. Recuperado de https://riunet.upv.es/mineria-de-datos-con-weka/PFC_DSIC80_AgustínCalleja_2.pdf
Recuperado el 25/02/2013.
- Gil, C. (2011). Segundo informe de Actividades 2009. UJAT. Primera Edición 2010, Villahermosa, Tabasco, México.
- Gil, C. (2011). Tercer Informe de Actividades 2010. UJAT. Primera Edición 2011, Villahermosa, Tabasco, México.
- Gil, C. (2012). Cuarto Informe de Actividades 2011. UJAT. Primera Edición 2012, Villahermosa, Tabasco, México.
- González, R. y Álvarez, I. (2008). *La orientación vocacional en los alumnos de nuevo ingreso*. Universidad Nacional Autónoma de México. México, Distrito Federal.
- Hernández, J., Ramírez, Q. y Ferri R. (2004). *Introducción a la Minería de Datos*. Madrid: Pearson Prentice Hall.
- Hernández, R.; Fernández, C. y Baptista, P. (2007). *Metodología de la Investigación*. México. McGraw-Hill.
- Morates, D. (1997). *Técnicas de regresión lineal mediante WEKA.*, San Francisco. Morgan Robbison.
- Piña, J. (2013). Primer Informe de Actividades 2012. UJAT. Primera Edición 2013, Villahermosa, Tabasco, México.
- Piña, J. (2014). Segundo Informe de Actividades 2013. UJAT. Primera Edición 2014, Villahermosa, Tabasco, México.
- Villanueva, M., Javier, A., Arceo, G., Zapata, T. y Ronzón, J. (2004). *Estudio de Egresados de la Licenciatura en Informática Administrativa COHORTE 1997*. Universidad Juárez Autónoma de Tabasco. División Académica de Informática y Sistemas. Cunduacán, Tabasco: UJAT.