



**UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MÉXICO**

---

---

**UNIDAD ACADÉMICA PROFESIONAL TIANGUISTENCO**

**IDENTIFICACIÓN DE TRAYECTORIAS ACADÉMICAS  
COMUNES PARA DETERMINAR EL RIESGO DE  
DESERCIÓN ESCOLAR MEDIANTE MINERÍA DE  
PATRONES SECUENCIALES**

**TESIS**

PARA OBTENER EL TÍTULO DE  
INGENIERO EN SOFTWARE

QUE PRESENTA  
**CRISTIAN JOE HERNÁNDEZ LÓPEZ**

DIRECTOR:

Dr. René Arnulfo García Hernández

TIANGUISTENCO, MÉX. AGOSTO 2021

## *Resumen*

La deserción escolar es un problema académico que se presenta en todos los niveles educativos, siendo que la presente investigación se enfoca en el nivel superior, en concreto en la Unidad Académica Profesional Tianguistenco (**UAPT**), de la Universidad Autónoma del Estado de México (**UAEM**).

Dentro de la UAPT se han realizado dos investigaciones sobre deserción escolar: En el primer trabajo: "*Descubrimiento de factores que inciden en la eficiencia terminal de estudiantes de educación superior con árboles de decisión*", García [García, 2012] se dedicó a descubrir aquellos factores que inciden en la eficiencia terminal de estudiantes de la UAPT; una de sus conclusiones importantes fue que el número de créditos tomados y el número de unidades de aprendizaje tomadas hasta el sexto semestre influyen bastante para saber si un estudiante permanece o deserta [García, 2012]. En cuanto al segundo trabajo: "*Identificación de factores que influyen en la deserción escolar de estudiantes universitarios usando algoritmos genéticos y K-vecinos más cercanos*" [Osorio, 2013], este trabajo se enfocó en la identificación de factores que influyen en la deserción escolar. Para ello, Osorio [Osorio, 2013]; obtiene como conclusión que las trayectorias académicas si influyen en la deserción o permanencia de un estudiante [Osorio, 2013].

De acuerdo a los trabajos realizados en la UAPT se concluyó que el estudio de las trayectorias académicas de los estudiantes hasta cierto punto es determinante para saber si el estudiante deserta o no, por lo que es necesario extraer el conocimiento de los historiales académicos de los estudiantes.

El plan de estudios de cada una de las 4 carreras que ofrece la UAPT es flexible en cuanto a la toma de unidades de aprendizaje por parte del estudiante [UAEM-OE, 2016], por lo que, los estudiantes pueden cursar las unidades de aprendizaje en el orden que se les oferta, pueden saltarse unidades de aprendizaje o postergarlas para tomarlas después, pueden cursar

unidades de aprendizaje antes del semestre en que se ofertan; lo que genera un sin número de trayectorias posibles generadas una por cada estudiante. Dado el anterior inconveniente es difícil para la persona interesada saber cuáles trayectorias propician un caso de deserción o un caso de egreso, ya que las trayectorias en sí presentan una similitud debido a que los estudiantes toman la misma carrera.

En la presente investigación se propuso utilizar minería de patrones secuenciales para identificar las trayectorias académicas comunes de estudiantes que desertan o que egresan. Para ello se utilizó el algoritmo Generalización de Patrones Secuenciales (GSP, por sus siglas en inglés) más un método de transformación de patrones secuenciales a reglas de asociación aplicado a bases de datos creadas con historiales académicos pertenecientes a estudiantes de la UAPT.

Los resultados obtenidos pueden ser entendibles por la persona interesada, por lo que la información obtenida se convierte en conocimiento útil, para la toma de decisiones tanto para la institución educativa como para el estudiante, lo que permitiría disminuir el índice de deserción académica.

<i>Contenido</i>	<i>Página</i>
Agradecimientos.....	iii
Dedicatoria.....	iv
Resumen.....	v
<b>1. INTRODUCCIÓN.....</b>	<b>1</b>
1.1 Planteamiento del problema.....	14
1.2 Objetivo General.....	15
1.3 Objetivos Particulares.....	15
1.4 Hipótesis.....	16
1.5 Justificación.....	16
1.6 Organización de la tesis.....	17
<b>2. MARCO TEÓRICO.....</b>	<b>19</b>
2.1 Minería de Datos.....	21
2.2 Metodología KDD.....	21
2.3 Minería de Patrones Secuenciales.....	23
2.3.1 Algoritmos de minería de patrones secuenciales.....	25
2.3.2 Fases de la minería de patrones secuenciales.....	27
2.3.3 Generalización de Patrones Secuenciales (GSP).....	27
2.4 Resumen del capítulo.....	30
<b>3. ESTADO DEL ARTE.....</b>	<b>31</b>
3.1 Las trayectorias académicas: dimensiones personales de una trayectoria estudiantil. Testimonio de un actor.....	31
3.2 Los estudios de trayectorias académicas y profesionales, contribuciones al estado del arte.....	35
3.3 Origen social de los estudiantes y trayectorias estudiantiles en la Universidad Veracruzana.....	37
3.4 Trayectorias académicas generacionales, constitución y diversificación del oficio académico en el caso de los bioquímicos de la Facultad de Medicina.....	42
3.5 Trayectorias académicas y construcciones subjetivas de estudiantes y ex estudiantes virtuales.....	44
3.6 Síntesis de los trabajos con el análisis de trayectorias académicas.....	46
3.7 Descubrimiento de factores que inciden en la eficiencia terminal de estudiantes de educación superior con árboles de decisión.....	48

3.8 Identificación de factores que influyen en la deserción escolar de estudiantes universitarios usando algoritmos genéticos y K vecinos más cercanos.....	50
3.9 Reglas que describen la deserción y permanencia en los estudiantes de la UAP Tianguistenco de la UAEM. ....	51
3.10 Síntesis de los trabajos relacionados al análisis de deserción escolar en la UAP Tianguistenco.....	53
3.11 Análisis de patrones secuenciales en una base de datos de estudiantes. ....	54
3.12 Minería secuencial: análisis de algoritmos y patrones.....	57
3.13 Síntesis de los trabajos relacionados a la aplicación de minería de datos secuenciales a estudiantes .....	58
3.14 Resumen del capítulo .....	59
<b>4. PROPUESTA DE SOLUCIÓN .....</b>	<b>61</b>
4.1 Propuesta .....	61
4.2 Método propuesto.....	62
4.3 Metodología.....	65
<b>5. EXPERIMENTACIÓN.....</b>	<b>67</b>
5.1 Caso de Estudio .....	67
5.2 Fase 1: Integración y recopilación.....	70
5.3 Fase 2: Selección y limpieza .....	73
5.3.1 Construcción de la base de datos .....	74
5.4 Fase 3: Minería de datos.....	77
5.4.1 Método de transformación.....	81
5.5 Fase 4: Evaluación e Interpretación.....	82
5.5.1 Experimentación.....	82
Normalización 0. Prueba general, base de datos completa .....	82
Normalización 1. Prueba general con síntesis de nota, base de datos dividida por semestre y condición académica .....	83
Normalización 2. Prueba de asignación de periodos, base de datos con división de semestres .....	86
Normalización 3. Síntesis recopilatoria de resultados anteriores, base de datos con agrupación resultados .....	88
5.5.2 Evaluación .....	92
5.5.3 Listado de reglas con mayor confianza .....	94
5.5.4 Resultados Evaluación.....	124

5.6 Fase 5: Difusión y uso .....	133
<b>6. CONCLUSIONES Y TRABAJO FUTURO .....</b>	<b>134</b>
6.1 Conclusiones .....	134
6.2 Trabajo Futuro .....	135
<b>REFERENCIAS BIBLIOGRÁFICAS .....</b>	<b>136</b>
<b>ANEXOS .....</b>	<b>143</b>

## Índice de ilustraciones

Ilustración 1. Diagrama de trayectorias académicas comunes. ....	12
Ilustración 2. Diagrama de trayectorias académicas comunes. ....	20
Ilustración 3. Diagrama de KDD.....	22
Ilustración 4. Clasificación de algoritmos SPM.....	26
Ilustración 5. Funcionamiento del algoritmo GSP. ....	30
Ilustración 6. Diagrama de método propuesto.....	62
Ilustración 7. Esquema detallado de construcción de las Bases de Datos.....	63
Ilustración 8. Historial académico de prueba, tomado aleatoriamente.....	72
Ilustración 9. Datos de depuración del historial académico.....	74
Ilustración 10. Trabajo fase 3.....	77
Ilustración 11. Funcionamiento del algoritmo GSP.....	78
Ilustración 12. Formato de Base de Datos.....	80
Ilustración 13. Gráfica de resultados obtenidos.....	91
Ilustración 14. Reglas de culminación tamaño 2.....	96
Ilustración 15. Reglas de baja tamaño 2.....	98
Ilustración 16. Reglas de culminación tamaño 3.....	100
Ilustración 17. Reglas de baja tamaño 3.....	102
Ilustración 18. Reglas de culminación tamaño 4.....	104
Ilustración 19. Reglas de baja tamaño 4.....	106
Ilustración 20. Reglas culminación tamaño 5.....	108
Ilustración 21. Reglas de baja tamaño 5.....	110
Ilustración 22. Reglas de culminación tamaño 6.....	113
Ilustración 23. Reglas de baja tamaño 6.....	115
Ilustración 24. Reglas de culminación tamaño 7.....	118
Ilustración 25. Reglas de baja tamaño 7.....	120
Ilustración 26. Reglas de baja tamaño 8.....	122

Ilustración 27. Regla de baja tamaño 9. .... 124



## Índice de tablas

Tabla 1. Grupos de estudiantes por capital familiar.....	39
Tabla 2. Trayectorias agrupadas por tipo de estudiante.....	41
Tabla 3. Situación escolar del estudiante de acuerdo a sus créditos aprobados.....	42
Tabla 4. Cuadro sinóptico de trabajos relacionados con el estudio de trayectorias académicas.....	47
Tabla 5. Datos de García Lambert.....	49
Tabla 6. Cuadro sinóptico de trabajos relacionados con el estudio de trayectorias académicas en la UAPT de la UAEM.....	53
Tabla 7. Cuadro sinóptico de trabajos relacionados con la minería de datos secuenciales.....	58
Tabla 8. Estadística de la UAP Tianguistenco.....	68
Tabla 9. Índices de deserción UAEM.....	68
Tabla 10. Composición del fragmento del historial.....	75
Tabla 11. Elementos de la Base de Datos.....	79
Tabla 12. Registro total de reglas obtenidas.....	89
Tabla 13. Registro de reglas depuradas.....	89
Tabla 14. Número de reglas separadas.....	90
Tabla 15. Reglas de culminación tamaño 2.....	94
Tabla 16. Reglas de baja tamaño 2.....	96
Tabla 17. Reglas de culminación tamaño 3.....	98
Tabla 18. Reglas de baja tamaño 3.....	100
Tabla 19. Reglas de culminación tamaño 4.....	102
Tabla 20. Reglas de baja tamaño 4.....	104
Tabla 21. Reglas de culminación tamaño 5.....	106
Tabla 22. Reglas de baja tamaño 5.....	108
Tabla 23. Reglas de culminación tamaño 6.....	110
Tabla 24. Reglas de baja tamaño 6.....	113
Tabla 25. Reglas de culminación tamaño 7.....	115
Tabla 26. Reglas de baja tamaño 7.....	118

Tabla 27. Reglas de baja tamaño 8.....	120
Tabla 28. Reglas de baja tamaño 9.....	123
Tabla 29. Sumatorias reglas obtenidas.....	125
Tabla 30. Sumatorias confianzas de reglas.....	125
Tabla 31. Sumatorias aplicando fórmula.....	126
Tabla 32. Sumatorias de reglas.....	127
Tabla 33. Sumatorias de confianza de reglas.....	127
Tabla 34. Sumatorias de reglas.....	128
Tabla 35. Sumatoria de reglas.....	128
Tabla 36. Sumatorias de confianza de reglas.....	129
Tabla 37. Sumatorias de total de reglas.....	129
Tabla 38. Sumatorias de reglas.....	130
Tabla 39. Sumatorias de confianzas de reglas.....	131
Tabla 40. Sumatorias de total de reglas.....	131
Tabla 41. Evaluación y clasificación de historiales al azar.....	132



# 1. Introducción

---

El aprendizaje es una actividad humana primaria que es una forma de adquirir el conocimiento, lo aumenta y mejora su capacidad mediante el estudio o la experiencia [REA, 2015], [Osorio, 2013]. Aquellas personas que se dedican a aprender son llamados aprendices, alumnos o estudiantes.

La educación se concibe como un proceso de aprender a aprender, se orienta no solo a la transmisión de conocimientos e información, sino también a desarrollar la capacidad de producirlos y utilizarlos. Además, la educación es un derecho y bien público indispensable para el desarrollo personal, social y familiar. [SSP, 2011], [Goicovic, 2002]. Por lo que la educación se considera como el principal instrumento para la búsqueda del desarrollo social en forma integral para superar las brechas existentes. La educación dentro de las organizaciones educativas es fundamental para trazar la trayectoria deseada. Además, la educación permite responder eficientemente a las transformaciones sociales: globalización, sociedad, información y conocimiento [Vargas, 2016]. Un aspecto importante en la sociedad es la educación de sus individuos (niños,

adolescentes y jóvenes) que son el futuro de la sociedad misma, ya que su impacto se manifiesta en el desarrollo y crecimiento de la sociedad. Por lo que los jóvenes mejor preparados contribuyen a un mejor desarrollo, siendo que los estudiantes necesitan terminar su formación académica para poder cumplir con las expectativas impuestas en ellos pudiendo realizar las transformaciones sociales que se van presentando.

De acuerdo a la *Secretaría de Educación Pública (SEP)* la función social básica de la educación en México es: *“Ampliar las oportunidades educativas, para reducir desigualdades entre grupos sociales, cerrar brechas e impulsar equidad, al dotar a los alumnos de competencias y conocimientos necesarios, que funcionan como base y estructura para construir una trayectoria individual, comunitaria, productiva e integral”* [SEP-SEMS, 2016].

El proceso educativo como tal es un conjunto de pasos a seguir para aprender y obtener conocimiento unilateralmente, este proceso educativo ya ha sido planeado, probado, validado por diversas instituciones al transcurrir de años, debido a que, con la globalización y las nuevas tecnologías, el proceso educativo se va adaptando a las nuevas necesidades [Rodríguez, 2014]. Dicho proceso para los estudiantes es el mismo en muchos casos, dependiendo de las condiciones y características de las instituciones, países, etc. Por lo que el proceso educativo no es concluido por la totalidad de estudiantes que lo inician.

El papel del sistema educativo universitario es el de contribuir a la difusión de la economía moderna del conocimiento en todo el mundo, ayudando a que se disponga de una fuerza de trabajo más eficiente, permitiendo a sus países mejorar su economía a nivel mundial y haciendo que estos asuman un compromiso de contribuir a ese desarrollo económico – social [Gutiérrez, 2014]. Para que los estudiantes cumplan con el papel educativo universitario es necesario que terminen su formación académica y que no la dejen inconclusa o abandonada.

Dentro de la educación se encuentra un elemento llamado rendimiento escolar, que es el efecto de la relación de los factores multicausales que inciden en el resultado académico de un estudiante, en el que interactúan aspectos sociodemográficos, psicosociales, pedagógicos, institucionales y socioeconómicos [Espinosa, 2014]. Por su parte, Edel [Edel, 2003] define el rendimiento escolar como “*un nivel de conocimientos mostrados en un área o materia comparado con la norma de edad y nivel académico*”. Entendiendo que el rendimiento escolar para cada estudiante es diferente debido a sus capacidades y limitantes, por lo que institucionalmente se trata de obtener un buen rendimiento o regular rendimiento en sus estudiantes.

Sin embargo, hablando de educación es necesario hablar del problema de la interrupción del proceso educativo conocido como *deserción escolar*, que es un problema que enfrenta nuestro país (México) y que está presente en todos los niveles educativos, ya que por ello se truncan los estudios y el aprendizaje [Osorio, 2013], [Edel, 2003]. Es importante mencionar que la problemática de la deserción escolar es llamada de distintas formas tales como: deserción escolar [Osorio, 2013], [García, 2012], deserción, bajo rendimiento y rezago; retraso escolar, abandono escolar, etcétera. [Tinto, 1989], [Abril, 2008], [Díaz, 2008], [Zermeño, 2009], [Fierros, 2014], [SSP, 2011], [Goicovic, 2002], [Espíndola, 2002], [Valero, 2009]; por lo que el problema es abordado desde diferentes puntos de vista o de diferentes unidades de aprendizaje. La deserción escolar se define como el abandono escolar antes de terminar su ciclo educativo o de dejar de asistir de manera prolongada o definitiva por parte del estudiante<sup>1</sup> o la escuela, esta se puede dar de manera temprana o durante el transcurso de ciclo escolar [Dzay, 2012], [Fierros, 2014], [Zermeño, 2009], [SSP, 2011], [Goicovic, 2002].

Muchos estudios sobre deserción escolar se enfocan en los resultados de los estudiantes egresados, los estudian desde diversos enfoques tanto psicológicos, socioeconómicos, académicos, ya que es un problema multidisciplinario [Zermeño, 2009].

---

<sup>1</sup> Se utiliza el término “estudiante” en esta investigación, ya que los trabajos del estado del arte y de antecedentes usan los términos: alumno o estudiante.

Además de ser un problema multifactorial como aquellos presentados por Edel para el rendimiento escolar [Edel, 2003], los presentados por Espíndola para la reducción de deserción o los que la condicionan [Espíndola, 2002], o factores influyentes, condicionantes [García, 2012], [Osorio, 2013], los factores motivacionales [Hassanbeigi, 2010].

Otras investigaciones estudian solo parte del problema de deserción como las conductas de riesgo en los adolescentes y el desempeño académico [Palacios, 2007], riesgos motivacionales [Hassanbeigi, 2010].

Osorio menciona la clasificación de los factores que influyen en la deserción escolar, estos son [Osorio, 2013]:

- Factores fisiológicos: entre estos factores se encuentran los cambios hormonales, las deficiencias en los órganos de los sentidos, la desnutrición, problemas de peso y salud.
- Factores pedagógicos: estos factores se relacionan con la calidad de la enseñanza, como el número de estudiantes por maestro, los métodos y el material didáctico usado por este, la motivación de sus estudiantes, el tiempo de preparación de sus clases por parte del maestro.
- Factores psicológicos: en estos se tienen los desórdenes en las funciones psicológicas básicas, como, los problemas de percepción, memoria y conceptualización.
- Factores sociológicos: son aquellos que incluyen características familiares como las socioeconómicas, la posición económica familiar, el grado de estudios y ocupación de los padres, la calidad del ambiente en el que vive el estudiante.

Las causas generales que propician la deserción escolar se identifican gracias a diversos factores encontrados en distintos estudios, los cuales de forma general los agrupan en: personal, económico, escolar, familiar, matrimonio y unión, término o grado alcanzado de estudios y el factor cultural [SSP, 2011], [Zermeño, 2009].

La deserción escolar acarrea distintas consecuencias tanto para el individuo como para la sociedad en la que se desenvuelven. Estas consecuencias de forma general son las siguientes [Zermeño, 2009]: vagancia, vandalismo, drogadicción, frustración, desempleo y subempleo. Las cuales son consecuencias económicas, sociales, personales y familiares. Así mismo, la deserción genera diversos costos que pueden ser [SSP, 2011], [Zermeño, 2009]:

- Sociales: la fuerza de trabajo es menos competente, más difícil de calificar, su impacto está en el crecimiento económico.
- Privados: la reproducción de las desigualdades sociales, la pobreza y la desintegración social; su impacto está en la población infantil y adolescente.

Los trabajos se enfocan en cómo aparece la deserción mediante la combinación de diversos factores, las causas de estos factores, así como las consecuencias que acarrea, llevando la investigación a la parte personal, del entorno social y académico, incluyendo el punto de vista de los tutores, maestros y demás académicos.

En México la deserción escolar es un problema que se presenta en todos los niveles de educación (Pre escolar, Primaria, Secundaria, Medio Superior y Superior) ya que trae consigo diversas consecuencias; ya que es un problema que caracteriza a la mayoría de las instituciones mexicanas de educación superior [Valero, 2009]. Por lo anterior, Zermeño menciona la deserción escolar como un problema multifactorial y multidisciplinario, que se estudia desde tres ciencias y perspectivas diferentes: Psicología, Sociología y Pedagogía [Zermeño, 2009].

*La Asociación Nacional de Universidades e Instituciones de Educación Superior (ANUIES), desde su fundación en 1950, ha participado en la formulación de programas, planes y políticas nacionales, así como en la creación de organismos orientados al desarrollo de la educación*

superior mexicana, siendo que ésta presenta estudios realizados sobre la deserción escolar, dentro de estos estudios se menciona el índice de eficiencia terminal [ANUIES, 2015].

La eficiencia terminal es un indicador que determina la cantidad de estudiantes que terminan sus estudios en el tiempo ideal asignado por el plan de estudios, es decir, que determina la calidad del rendimiento de una institución, ya que es un indicador cuantitativo y estadístico de logros establecido por un establecimiento escolar usado para su evaluación del funcionamiento y rendimiento del mismo [García, 2012]. De acuerdo a la investigación de García [García, 2012], el índice de eficiencia terminal en México es del 39% según la ANUIES, lo que indica que el restante 61% de estudiantes acaban sus estudios posteriormente al tiempo ideal asignado o no terminan sus estudios.

De acuerdo a las estadísticas presentadas por la SEP en el período escolar 2014 – 2015, la cantidad de estudiantes inscritos en la cual se presenta la educación superior se distribuye de la siguiente manera [SEP-DGPPEE, 2016]: Profesional Asociado o Técnico Superior Universitario tuvo el 4.5% de estudiantes, la Normal 3.5% de estudiantes, la Licenciatura 85.3% de estudiantes y el Posgrado 6.7% de estudiantes.

La estadística presentada de deserción en México es: del 100% de estudiantes en el curso escolar 2007 – 2008 solo el 7.11% desertaron; del 7.11%, se distribuye de la siguiente forma: los desertores son el 30% que se dedican a solamente a estudiar mientras que 70% restante estudian y trabajan al mismo tiempo; en el periodo escolar 2014 – 2015 en el Estado de México hay un índice de abandono escolar de 9.2% [SEP-DGPPEE, 2016], [Zermeño, 2009], [GobBC, 2015], [SSP, 2011].

En la información recabada en la *Universidad Autónoma de Estado de México (UAEM)*, presentada en la agenda estadística 2015, se obtuvieron los siguientes resultados [UAEM-AE, 2016]:



- En toda la UAEM (facultades, centros universitarios, unidades académicas) el índice de deserción escolar es de 7.1%.
- Particularizando, en la Unidad Académica Profesional Tianguistenco (**UAPT**) el índice de deserción escolar fue de 11.2%.

La UAPT es una unidad académica perteneciente a la UAEM que oferta 4 carreras de nivel superior: Ingeniería en Plásticos, Ingeniería en Producción Industrial, Ingeniería en Software y Licenciatura en Seguridad Ciudadana [UAPT, 2015].

La siguiente información sobre la deserción escolar presente en la UAPT se obtuvo de la agenda estadística de la UAEM: 2013, 2014 y 2015 [UAEM-AE, 2016]:

- En el periodo escolar 2012 – 2013 la deserción fue de 13.3%
- En el periodo escolar 2013 – 2014 la deserción fue de 12.4%
- En el periodo escolar 2014 – 2015 la deserción fue de 11.2%.

La estadística de deserción escolar en la UAPT por carrera es la siguiente [UAEM-OE, 2016]:

- Ing. en Plásticos: 13.6%
- Ing. en Producción Industrial: 10.1%
- Ing. en Software: 16.6%
- Lic. en Seguridad Ciudadana: 8.7%

Con los datos estadísticos presentados se puede observar que la deserción en la UAPT está disminuyendo, sin embargo, los estudiantes siguen desertando por diversas razones ya presentadas anteriormente, se observa que la deserción existe en todos los niveles educativos

por lo que se realizan diversas acciones para contrarrestarla, como lo son los estudios psicológicos y sociales, la aplicación de programas de tutorías académicas, los apoyos de becas, entre otros.

La función de un tutor es apoyar al estudiante a través de su formación académica, conocer su trayectoria académica y ayudar a formar al estudiante en aspectos de [Méndez, 2006]:

- La orientación escolar, vocacional y profesional.
- El mejoramiento de las relaciones interpersonales en cada grupo-clase y con otros miembros de la comunidad educativa.
- Facilitar las acciones de un determinado equipo docente con el grupo de estudiantes a ellos confiado.

La tutoría académica universitaria es el proceso de acompañamiento orientado a la ayuda personalizada en la formación universitaria desde un enfoque integral, que necesita de estructura y recursos; cuyos beneficios son para estudiantes y profesorado. Lo que permite el desarrollo de enfoques de enseñanza centrados en el estudiante, una mayor individualización de la formación universitaria, la búsqueda de modelos formativos integrales, la mejora de las tasas de abandono, el desarrollo de competencias transversales, la orientación para la carrera, entre otros aspectos [López, 2017].

Dentro de la UAPT se han realizado dos investigaciones sobre deserción escolar: En el primer trabajo: "*Descubrimiento de factores que inciden en la eficiencia terminal de estudiantes de educación superior con árboles de decisión*", García [García, 2012] se dedicó a descubrir aquellos factores que inciden en la eficiencia terminal de estudiantes de la UAPT. Para ello se utilizaron cuestionarios realizados a los estudiantes que se encontraban cursando sus respectivas carreras, utilizando árboles de decisión para encontrar esos factores. Una de sus conclusiones importantes fue que el número de créditos tomados y el número de unidades de aprendizaje tomadas hasta el sexto semestre influyen bastante para saber si un estudiante

permanece o deserta [García, 2012]. En cuanto al segundo trabajo: "*Identificación de factores que influyen en la deserción escolar de estudiantes universitarios usando algoritmos genéticos y K-vecinos más cercanos*" [Osorio, 2013], el trabajo se enfocó en la identificación de factores que influyen en la deserción escolar. Para ello, Osorio [Osorio, 2013] utilizó una base de datos de estudiantes con información incluida de control escolar de la UAPT y evaluó sus datos con algoritmos genéticos y el algoritmo K – vecinos más cercanos. De este trabajo se obtiene como conclusión que las trayectorias académicas si influyen en la deserción o permanencia de un estudiante [Osorio, 2013].

De acuerdo a los trabajos realizados en la UAPT se concluye que el estudio de las trayectorias académicas de los estudiantes hasta cierto punto es determinante para saber si el estudiante deserta o no, por lo que es necesario extraer el conocimiento de los historiales académicos de los estudiantes.

Cabe resaltar que el plan de estudios de cada una de las 4 carreras que ofrece la UAPT es flexible en cuanto a la toma de unidades de aprendizaje por parte del estudiante [UAEM-OE, 2016], por lo que, los estudiantes pueden cursar las unidades de aprendizaje en el orden que se les oferta, pueden saltarse unidades de aprendizaje o postergarlas para tomarlas después, pueden cursar unidades de aprendizaje antes del semestre en que se ofertan. Es decir, los estudiantes pueden adelantar unidades de aprendizaje; lo que genera un sin número de trayectorias posibles generadas una por cada estudiante. Dado el anterior inconveniente es difícil para la persona interesada (estudiante, tutor, directivo, profesor) saber cuáles trayectorias propician un caso de deserción o en caso contrario un caso de egreso, ya que las trayectorias en sí presentan una similitud debido a que los estudiantes toman la misma carrera (una de las 4 ofertadas).

Normalmente, el tutor y los académicos no conocen de manera específica las trayectorias que llevan a que un estudiante deserte o egrese. Sin embargo, los tutores académicos conocen por experiencia propia que hay ciertas unidades de aprendizaje que son "filtro" para los estudiantes por su alto índice de reprobación. Del mismo modo, los estudiantes van

conociendo cuáles son aquellas unidades de aprendizaje que son más complicadas o con mayor índice de reprobación y deserción de las mismas unidades de aprendizaje. La capacidad que tiene cada profesor como tutor para ponerse del lado del estudiante y comprender su proceso de aprendizaje contribuye a que el estudiante lleve de manera ordenada su formación universitaria, sin importar si el estudiante va retrasando unidades de aprendizaje, con el propósito de concluir con su carrera.

Una **Trayectoria** es aquella línea que se va escribiendo en el espacio como consecuencia del movimiento de un punto por el mismo, definido por la RAE como “Curso que a lo largo del tiempo sigue el comportamiento o el ser de una persona, de un grupo social o de una institución” [Nícuesta, 2015], [REA, 2015]. Por lo que una trayectoria académica es descrita como una línea temporal a lo largo de la formación académica del estudiante, es decir, es un recorrido que sigue el estudiante en un tiempo determinado respecto al desempeño escolar: aprobación y/o reprobación, promedios, etc., a lo largo de sus ciclos escolares [ENCCH, 2018].

Es necesario mencionar que la **historia académica** del estudiante está contenida en un documento que contiene la información académica del estudiante a lo largo de su carrera. Esta información contiene: datos de la escuela, datos del estudiante, carrera, inicio de carrera, unidades de aprendizaje tomadas, periodos en los que curso las unidades de aprendizaje, calificaciones de cada materia, situación académica, promedio del semestre y general, entre otros detalles. Dentro de la UAEM a este documento se le denomina Trayectoria académica. Sin embargo, para esta investigación el término de “**trayectoria académica**” será usado para describir la trayectoria que sigue el estudiante cuando deserta o egresa; a su vez el término “**historial académico**” será usado para referirse al documento que contiene la historia académica del estudiante.

La UAEM mantiene las trayectorias académicas en un registro para cada estudiante, el cual es llamado historial académico. Es necesario mencionar que un historial académico contiene la información académica del estudiante a lo largo de su carrera como: datos de la escuela, datos del estudiante, carrera, inicio de carrera, unidades de aprendizaje tomadas, periodos

en los que tomó las unidades de aprendizaje, calificaciones de cada materia, situación académica, promedio del semestre y general, entre otros detalles.

La información obtenida sobre el desempeño académico en general, incluyendo la deserción escolar se encuentra en repositorios de datos, agendas y reportes estadísticos, informes escolares, historiales académicos, etc. Por lo tanto, es indispensable buscar si hay información en las trayectorias de los historiales académicos que permitan predecir si un estudiante pudiera desertar o no, por lo que se propone hacer uso de herramientas y técnicas computacionales para este fin.

Una trayectoria académica común es una trayectoria que se presenta frecuentemente en el historial académico de los estudiantes. Las trayectorias pueden ser completas, semicompletas, parciales o por fragmentos.

Para esta investigación se trata una trayectoria académica común de la siguiente manera:

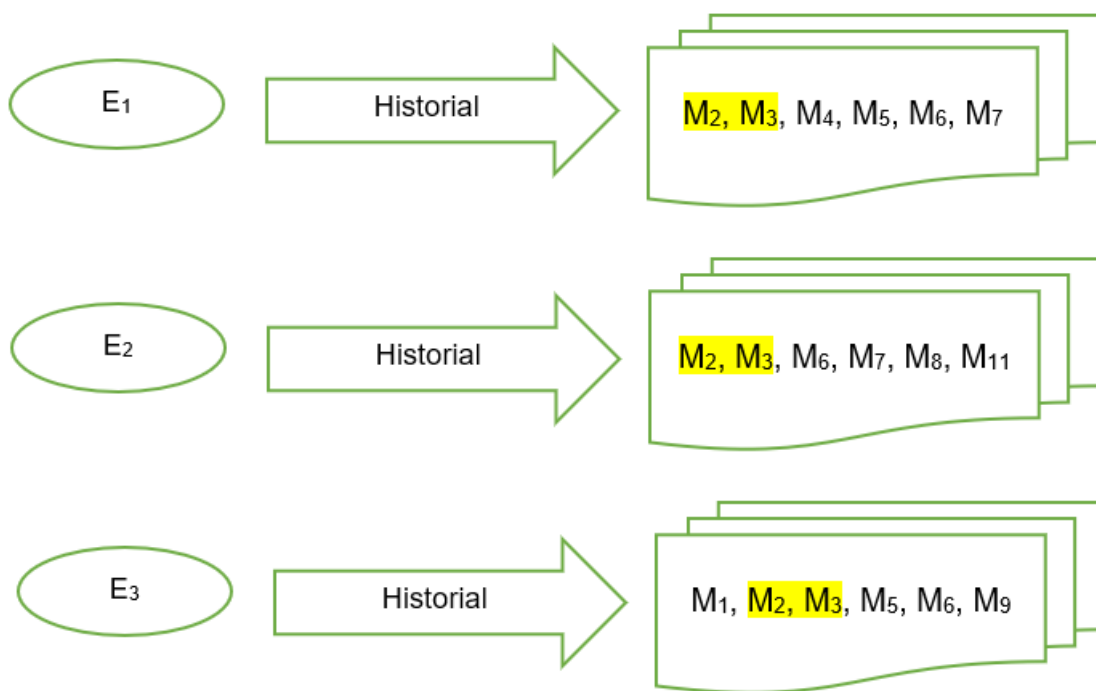
Dado un conjunto de estudiantes, donde cada estudiante tiene una trayectoria, que es un conjunto de unidades de aprendizaje tomadas a lo largo de su carrera distribuidas por la temporalidad de cada semestre.

$$TAC (E_1, E_2, E_3 \dots E_N) = \{ S_1(M_1, M_2, M_3 \dots M_R), S_2(M_1, M_2, M_3 \dots M_S), \dots S_N(M_1, M_2, M_3 \dots M_T) \}$$

Donde:

- TAC: Significa Trayectorias Académicas Comunes.
- E: Es un estudiante con una trayectoria académica.
- S: Representa el semestre en que el estudiante curso las unidades de aprendizaje.
- M: Corresponde a una materia o unidad de aprendizaje.

El siguiente es un diagrama que muestra cómo se presentan las trayectorias académicas comunes:



**Ilustración 1.** Diagrama de trayectorias académicas comunes.

Se observa que la trayectoria de los 3 estudiantes es: M<sub>2</sub> y M<sub>3</sub>, ya que se presenta en 3 casos, una segunda trayectoria es M<sub>5</sub> y M<sub>6</sub>, otra más M<sub>6</sub> y M<sub>7</sub>. Además, se observa que la cantidad de trayectorias aumenta con la cantidad de historiales obtenidos. El término **común** puede considerarse como un parámetro de frecuencia, por lo que se puede presentar de la manera subrayada como se presenta en la ilustración.

Las instituciones educativas tanto de nivel superior, nivel medio superior, nivel secundario y nivel elemental, toman sus decisiones para combatir el problema de la deserción escolar basadas en la información obtenida de informes escolares, investigaciones estadísticas, investigaciones

sobre factores de la deserción, agendas estadísticas. Por lo que, normalmente las instituciones académicas no cuentan con un sistema que les otorgue el conocimiento sobre las unidades de aprendizaje que afectan directamente la deserción, es decir, no cuentan con información sobre las trayectorias académicas de los estudiantes otorgada por un sistema, por lo que la información recabada no puede ayudar a la toma de decisiones basadas en las trayectorias académicas.

La información de cada estudiante se encuentra almacenada en historiales académicos que es recabada por cada generación, la cual es cada vez mayor, por lo que está es almacenada en bases de datos de estos historiales; cada unidad académica, centro universitario o plantel de la UAEM tiene este tipo de registros. Por lo que para poder obtener información que sea de utilidad para la toma de decisiones se hace uso de diversas técnicas o herramientas computacionales encontradas en la Inteligencia Artificial (IA), para ello se utiliza una rama de la minería de datos [García, 2012].

De acuerdo a los datos contenidos en los historiales, el modelo de datos constituido es predictivo ya que se estiman valores futuros o desconocidos de las variables de interés [Hernández, 2004], es decir, que en este trabajo se busca información en las trayectorias de los historiales académicos que permitan predecir si un estudiante puede desertar o no.

Dentro de la IA se encuentra una subrama conocida como Minería de Patrones Secuenciales [Hernández, 2004], que es la técnica que se adapta de mejor manera a los datos almacenados en los historiales académicos, ya que esta se define como la extracción de patrones frecuentes relacionados con el tiempo u otro tipo de secuencia [Hernández, 2004].

Dado que el análisis de patrones secuenciales se encarga de encontrar relaciones entre ocurrencias de eventos secuenciales, es decir, de encontrar si existe un orden específico en estas ocurrencias [Campagni, 2015]. La naturaleza temporal de los datos brinda un mejor

entendimiento de las tendencias o patrones en el tiempo con el fin de encontrar relaciones entre eventos.

### *1.1 Planteamiento del problema*

En muchos estudios sobre deserción escolar no se han estudiado las trayectorias académicas de los estudiantes como base para esas investigaciones. En los estudios realizados en la Unidad Académica Profesional Tianguistenco (UAPT) García [García, 2012] se enfocó en descubrir aquellos factores que inciden en la eficiencia terminal de estudiantes de la UAPT, Osorio [Osorio, 2013] se enfocó en la identificación de factores que influyen en la deserción escolar. Sin embargo, en los trabajos de la UAP concluyen y coinciden en que un factor escolar se presenta en la trayectoria académica del estudiante. Además, el plan de estudios de la UAPT es flexible, lo que da lugar a que existan distintas trayectorias académicas de los estudiantes, siendo estas de difícil análisis, por lo que las trayectorias académicas pueden ser comunes para aquellos estudiantes que desertan o egresan. Conociendo que en la toma de decisiones no se utiliza sistema alguno que ayude a conocer en qué parte de la carrera están desertando, es necesario obtener este conocimiento.

De acuerdo a lo presentado en el capítulo de introducción se formula la siguiente pregunta de investigación en la cual se plantea el problema encontrado:

¿Cuáles son las trayectorias académicas comunes que distinguen a los estudiantes que egresan de los que desertan, a partir de los historiales académicos de la UAPT de la UAEM, para ayudar en la detección de la deserción?



## *1.2 Objetivo General*

Encontrar las trayectorias académicas comunes de los estudiantes que permanecen estudiando, los egresados y desertores en la UAPT perteneciente a la UAEM mediante la minería de patrones secuenciales.

## *1.3 Objetivos Particulares*

- Analizar estudios de deserción escolar y los estudios de minería de datos secuenciales.
- Obtener los datos de todos los estudiantes egresados, los desertores y los que siguen estudiando de la UAPT de los periodos 2008 – 2013.
- Crear las de bases de datos generales y específicas.
- Reutilizar y programar el algoritmo de minería de datos secuencial GSP (Generalización de Patrones Secuenciales).
- Desarrollar el método computacional de identificación.
- Realizar experimentos, documentarlos y analizarlos.
- Evaluar y analizar las reglas obtenidas.
- Conocer si las reglas obtenidas satisfacen la pregunta de investigación
- Agregar conclusiones y trabajo futuro.

## *1.4 Hipótesis*

Si se aplica minería de datos secuenciales usando un algoritmo que encuentre trayectorias (secuencias) frecuentes en un conjunto de datos, a una base de datos de trayectorias académicas comunes, será posible identificar las trayectorias comunes que influyen en si un estudiante deserta o egresa de su carrera.

## *1.5 Justificación*

La información de cada estudiante se encuentra almacenada en historiales académicos que es recabada por cada generación, la cual es cada vez mayor, por lo que esta es almacenada en bases de datos de estos historiales, cabe comentar que cada Unidad Académica, Centro Universitario o Facultad de la UAEM tiene este tipo de registros. Por lo que, para poder obtener información que sea de utilidad para la toma de decisiones se hace uso de diversas técnicas o herramientas computacionales encontradas en la inteligencia artificial (IA), para ello se utiliza una rama de esta llamada minería de datos [García, 2012]. De acuerdo a los datos contenidos en los historiales, el modelo de datos constituido es predictivo ya que se estiman valores futuros o desconocidos de las variables de interés [Hernández, 2004], en este trabajo se busca información en las trayectorias de los historiales académicos que permitan predecir si un estudiante puede desertar o no.

## *1.6 Organización de la tesis*

### **Capítulo 1 Introducción**

En el capítulo se hablará sobre la educación, específicamente sobre el problema de la deserción escolar, así como lo son las causas y consecuencias de ésta, incluyendo datos estadísticos sobre deserción en el país, en la Universidad Autónoma del Estado de México y en la Unidad Académica Profesional Tianguistenco.

Además de mostrar qué se ha investigado en relación a la deserción dentro de la Unidad Académica Profesional Tianguistenco mencionando la importancia de estudiar las trayectorias académicas de los estudiantes por lo que se da a conocer el problema a resolver en la investigación, la justificación, la hipótesis y los objetivos.

### **Capítulo 2 Marco teórico**

En este capítulo se dan los conceptos necesarios para entender el problema computacionalmente, tanto para entender el estado del arte y para entender la posible solución al problema.

### **Capítulo 3 Estado del Arte**

En este capítulo se describen detalladamente los trabajos relacionados a la deserción escolar, el análisis de los trabajos sobre la deserción en la Unidad Académica Profesional Tianguistenco y los trabajos relacionados a la posible solución.

## **Capítulo 4 Propuesta de solución**

En el capítulo se da a conocer la propuesta de solución al problema planteado mediante el caso de estudio descrito y se da a conocer cómo se va a trabajar la metodología fase por fase.

Incluye la metodología de trabajo para la investigación, vista a grandes rasgos.

## **Capítulo 5 Experimentación**

En el capítulo se prueba la hipótesis, al describir cómo se trabajó cada fase de la metodología, al realizar las pruebas o experimentos y al analizar los resultados obtenidos.

## **Capítulo 6 Conclusiones y trabajo futuro.**

En el capítulo se presentan las conclusiones generales de la investigación, así como los inconvenientes que se presentaron, además se presenta el trabajo futuro.



## 2. Marco Teórico

---

En este capítulo se presentan los conceptos básicos para entender el estado del arte, relacionado al problema.

Retomando la información sobre que es una trayectoria académica: una trayectoria académica común es una trayectoria se presenta en una gran cantidad de estudiantes. Las trayectorias pueden ser completas, semicompletas, parciales o por fragmentos.

Para esta investigación se trata una trayectoria académica común de la siguiente manera: Dado un conjunto de estudiantes, donde cada estudiante tiene una trayectoria, que es un conjunto de unidades de aprendizaje tomadas a lo largo de su carrera distribuidas por la temporalidad de cada semestre.

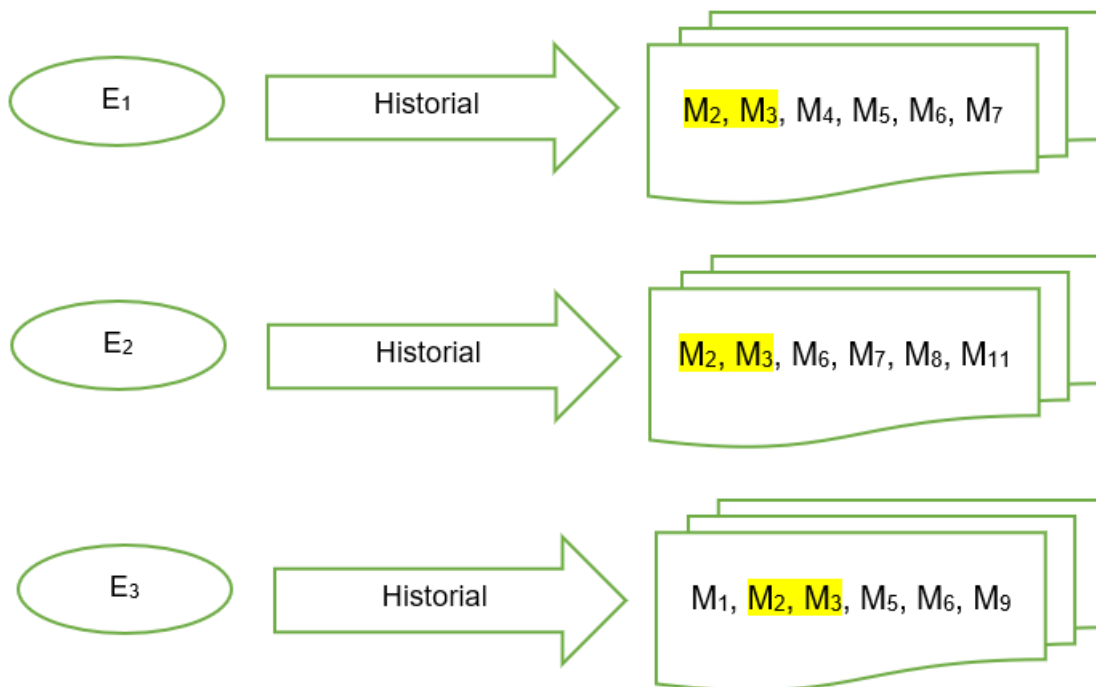
$$TAC (E_1, E_2, E_3 \dots E_N) = \{ S_1(M_1, M_2, M_3 \dots M_R), S_2(M_1, M_2, M_3 \dots M_S), \dots S_N(M_1, M_2, M_3 \dots M_T) \}$$

Donde:

- TAC: Significa Trayectorias Académicas Comunes.
- E: Es un estudiante con una trayectoria académica.
- S: Representa el semestre en que el estudiante curso las unidades de aprendizaje.

- M: Corresponde a una materia o unidad de aprendizaje.

El siguiente es un diagrama que muestra cómo se presentan las trayectorias académicas comunes:



**Ilustración 2.** Diagrama de trayectorias académicas comunes.

Se observa que la trayectoria de los 3 estudiantes es: M2 y M3, ya que se presenta en 3 casos, una segunda trayectoria es M5 y M6, otra más M6 y M7. Además, se observa que la cantidad de trayectorias aumenta con la cantidad de historiales obtenidos. El término **común** puede considerarse como un parámetro de frecuencia, por lo que se puede presentar de la manera subrayada como se presenta en la ilustración.

Para poder obtener la información dentro de los historiales académicos y así poder extraer el conocimiento implícito, se utiliza la Minería de Datos, que es una sub rama de la Inteligencia Artificial.

## *2.1 Minería de Datos*

De acuerdo a García [García, 2012], la minería de datos consiste en la extracción de información útil que se encuentra de manera implícita en los datos; dicha información era previamente desconocida y podrá ser usada para algún proceso posterior. Dentro de la minería de datos se encuentra una metodología conocida como la Metodología de **Descubrimiento de Conocimiento en Bases de Datos (en inglés KDD, Knowledge Discovery in Databases)** [García, 2012].

La minería de datos tiene tres componentes principales Clustering o Clasificación, Reglas de Asociación y Análisis de Secuencias [Pande, 1997].

## *2.2 Metodología KDD*

El **KDD**, de acuerdo con García [García, 2012], es una metodología de la minería de datos que contiene técnicas que permiten extraer conocimiento previamente desconocido a partir de grandes cantidades de datos, para después ocuparlo en algún proceso posterior.

Para la investigación y el desarrollo del método se utilizó esta metodología.

Por su parte Fayyad [Fayyad, 1996] dice que es un proceso general de descubrimiento de conocimiento útil a partir de los datos, ya que identifica información válida, novedosa, potencialmente útil además de patrones comprensibles en los datos.

La metodología de KDD consta de varias etapas iterativas, estas son [Hernández, 2004]:

1. **Integración y recopilación:** se determinan las fuentes de información que pueden ser útiles y dónde conseguirlas.
2. **Selección y limpieza:** se eliminan o corrigen los datos incorrectos y se decide la estrategia a seguir con los datos incompletos.
3. **Minería de Datos:** se decide la tarea a realizar y el método se va a utilizar.
4. **Evaluación e interpretación:** se evalúan los patrones y se analizan por los expertos, si es necesario se vuelve a iterar.
5. **Difusión y Uso:** se hace uso del nuevo conocimiento y se hace partícipe de este a los posibles usuarios.

Este proceso se muestra más detallado en el siguiente diagrama en el cual se pueden observar más subtareas intermedias entre cada fase de KDD [Hernández, 2004]:



**Ilustración 3.** Diagrama de KDD.

Las principales fases de KDD se muestran en la figura anterior, las fases pueden ser reiterativas dependiendo de los resultados obtenidos, dando las vueltas necesarias para obtener los resultados deseados.



## 2.3 Minería de Patrones Secuenciales

Una subrama de la minería de datos es llamada minería de patrones secuenciales (sus siglas en inglés SPM, Sequential Pattern Mining):

- Se define como la extracción de patrones frecuentes relacionados con el tiempo u otro tipo de secuencia [Hernández, 2004].
- SPM es el proceso de extraer ciertos patrones secuenciales cuya base soporte supere un umbral de soporte mínimo predefinido [Slimani, 2013].

Existen varios campos donde se aplica la minería de patrones secuenciales tales como [Motegaonkar, 2014], [Hernández, 2004], [Slimani, 2013]:

- El encontrar información valiosa en desastres naturales
- Análisis del registro de ventas
- Estrategias de marketing
- Secuencias de compras
- Tratamiento médico
- Secuencias de ADN
- Etc.

De acuerdo a Pande [Pande, 1997] la minería de patrones secuenciales tiene aplicaciones como:

- **La industria minorista:** incluyendo correo adjunto, ventas complementarias y satisfacción del cliente.
- **Dominios científicos y empresariales:** muchos de sus resultados se enfocan en investigaciones científicas o en investigaciones de negocios.

- **Dominio médico:** Investigación de enfermedades, identificación de síntomas / enfermedades que preceden a ciertas enfermedades.

En la Minería de Patrones Secuenciales, una secuencia de datos se considera frecuente si ésta se encuentra en al menos un cierto número de registros a esto se le conoce como **umbral mínimo de frecuencia** [García, 2007].

El **soporte de la regla** es el porcentaje de transacciones en un conjunto que contienen la secuencia, además, determina cuánta frecuencia de la regla es aplicable en la transacción del conjunto [Liu, 2011].

La **confianza** determina la predictibilidad de la regla. Se le conoce como umbral de confianza a la frecuencia con que aparece una secuencia en un evento dentro de un registro de datos, dicho umbral hace mención a un mínimo número de veces que aparece [Liu, 2011].

El **GAP** es una restricción de separación máxima entre los eventos dentro de los cuales están las secuencias frecuentes [García, 2007].

Los conceptos mencionados son ocupados como parámetros dentro de los algoritmos de la minería de patrones secuenciales.

Una **regla de asociación** es una regla que implica ciertas relaciones de asociación entre un conjunto de objetos en una base de datos, en este proceso se descubre un conjunto de reglas de asociación con múltiples niveles de abstracción a partir del conjunto o conjuntos relevantes de datos en una base de datos. Las reglas de asociación se utilizan para descubrir hechos que ocurren en común dentro de un determinado conjunto de datos [Pande, 1997].

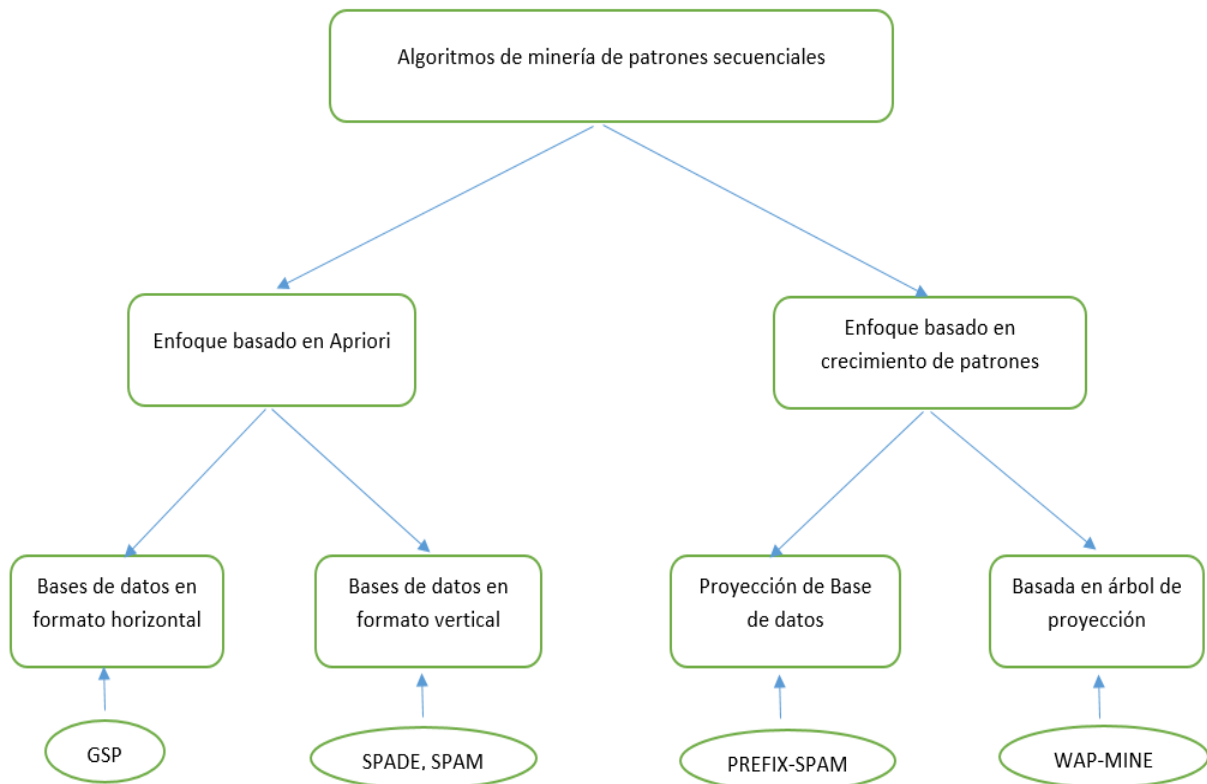
### 2.3.1 Algoritmos de minería de patrones secuenciales

En la minería de patrones secuenciales se encuentran diversos algoritmos con lo que se ha estudiado la misma. De acuerdo a las investigaciones hechas en minería de patrones secuenciales se han ido desarrollando diversos algoritmos, los cuales son [Motegaonkar, 2014], [Agrawal, 2015], [Srikant, 2015]:

- Algoritmos basados en Apriori:
  - *Apriori All*
  - *Apriori Some*
  - *GSP (Generalized Sequential Patterns)*
  - *SPIRIT*
  - *SPADE*
  - *SPAM (Sequential Pattern Algorithm Mining)*
  - *CloSpan (Closed Sequential Pattern Mining)*
  - *CMDS (Closed Multidimensional Pattern Mining)*
  
- Algoritmos de crecimiento de patrones (*pattern-growth*) de minería de secuencias:
  - *Prefix Span (Prefix Projected Sequential Pattern Mining)*
  - *FREESPAN*
  - *WAP-MINE*

Los algoritmos basados en *a priori* indican los patrones secuenciales que se encuentran dentro de la base de datos, en cambio los algoritmos de crecimiento de patrones indican los patrones secuenciales que están dentro de la base de datos más los posibles patrones que se pueden formar a partir de los datos.

Para complementar, la siguiente ilustración muestra la clasificación de los algoritmos de minería de patrones secuenciales [Motegaonkar, 2014]:



**Ilustración 4.** Clasificación de algoritmos SPM.

Como se puede observar los algoritmos de minería de patrones secuenciales se clasifican en 2 ramas: algoritmos basados en el enfoque de apriori y los algoritmos basados en el crecimiento de patrones; de los cuales, los basados en a priori se dividen en 2: Formación horizontal y formación vertical; en cambio los de crecimiento se dividen en proyección y de árbol de proyección [Motegaonkar, 2014].

### 2.3.2 Fases de la minería de patrones secuenciales

El problema de la extracción de patrones secuenciales tiene las siguientes fases [Pande, 1997], [Agrawal, 2015]:

- **Fase de organización u ordenamiento (Sort Phase):** Este paso convierte implícitamente la base de datos de transacciones original en una base de datos de secuencias.
- **Fase de Litemset (Litemset Phase):** En esta fase se encuentra el conjunto de todas las Litemset L. Además de encontrar simultáneamente el conjunto de todas las secuencias de tamaño 1 grandes.
- **Fase de Transformación (Transformation Phase):** En esta fase se determina repetidamente los conjuntos de secuencias largas que contienen en una secuencia de cliente. En una secuencia de cliente transformada, cada transacción es reemplazada por el conjunto de todos los itemsets contenidos en la transacción. Por lo que la secuencia de cliente se representa por una lista de conjuntos de itemsets.
- **Fase de Secuencias (Sequence Phase):** En esta fase se utiliza el conjunto de litemsets para encontrar las secuencias deseadas.
- **Fase Maximal o fase de secuencia maximal (Maximal Phase):** En esta fase se encuentran las secuencias maximales a partir del conjunto de secuencias largas o grandes. En esta fase se utiliza el algoritmo que se usará para encontrar las secuencias maximales, donde cada secuencia Maximal representa un patrón secuencial.

Litemset: es un conjunto de datos con un soporte mínimo.

Itemset: conjunto de elementos que comparten alguna(s) características en común.

### 2.3.3 Generalización de Patrones Secuenciales (GSP)

En la etapa 3 de KDD correspondiente a la minería de datos, se utilizó el algoritmo de minería de patrones secuenciales conocido como Generalización de Patrones Secuenciales (Generalized Sequential Patterns o GSP) [Srikant, 2015].

El algoritmo GSP es un algoritmo de minería de patrones secuenciales que descubre como son los patrones secuenciales dentro de una base de datos secuenciada por un lapso de tiempo determinado.

La evaluación empírica de GPS muestra escalas lineales con el número de datos secuenciales, y que tiene las mejores propiedades de subido escalado con respecto al número de datos secuenciales de transacciones por cada dato secuencial y el número de elementos por transacción. La evaluación empírica del GSP es 20 veces más rápido que el algoritmo A priori All [Srikant, 2015].

A continuación, se muestra el pseudocódigo del algoritmo GSP presentado en el artículo [Liu, 2011]:

### **Algoritmo GSP(S)**

```
C1 init – pass(S); // El primer paso sobre S
F1 { < {f} > | f ∈ C1, f.count / n ≤ minsup }; // n es el número de secuencias en S
Para (k=2; Fk – 1 ∈ ; k++) hacer //paso subsecuente sobre S
Ck candidato –gen-SPM(Fk-1);
Para cada subsecuencia de datos s ∈ S hacer //un escaneo de los datos
Para cada candidato c ∈ Ck hacer
Si c está contenido en s entonces
c.count++; //incrementa el contador de soporte
Fin para
Fin para
Fk { c ∈ Ck | c.count / n ≤ minsup}
Fin para
Refrescar F Uk Fk;
```

Donde:

S = Base de datos de secuencias

C1 = Guarda los candidatos de la generación evaluada

F1 = itemset evaluado

Fk = es el conjunto de k secuencias frecuentes

W = es un temset formado por items

Ck = guarda el conjunto de secuencias candidatas por generación

C = secuencia candidata

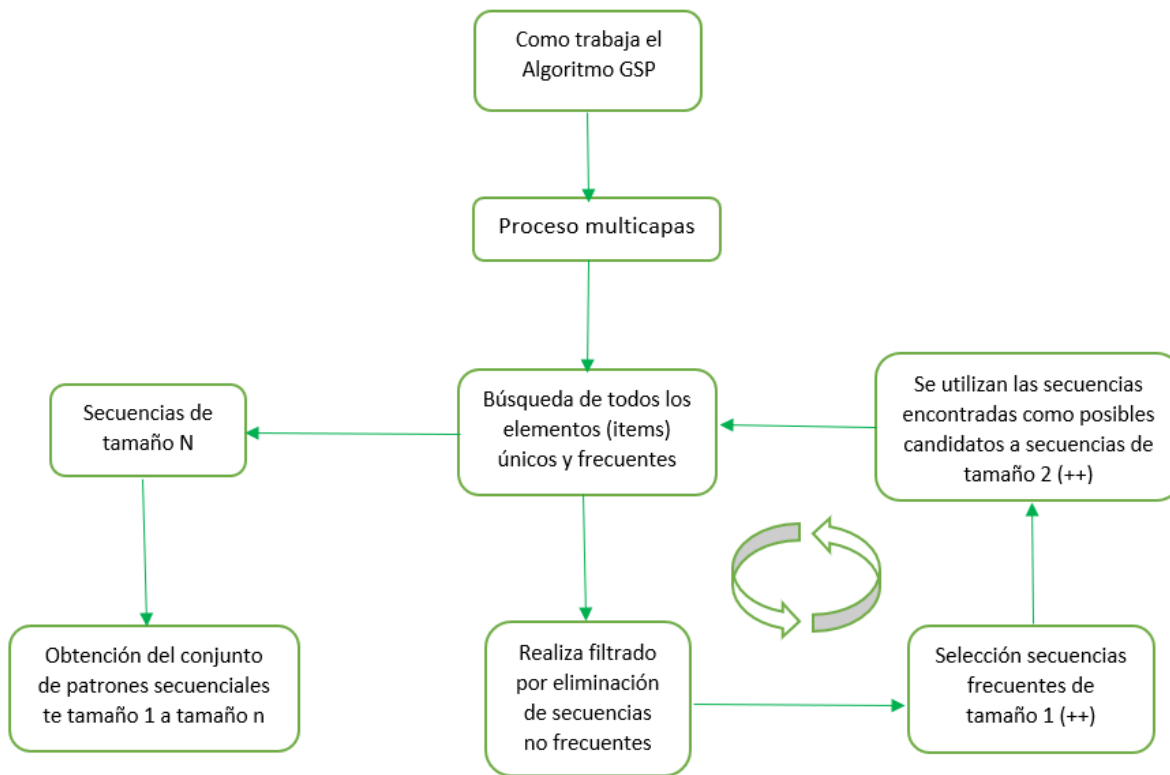
Este modelo describe el funcionamiento del algoritmo Generalización de Patrones Secuenciales (Generalized Sequential Patterns GSP) [Srikant, 2015] el cual implementa el modelo por medio de una combinación de componentes de código abierto, además de la programación del algoritmo que incorpora la lógica temporal y las funciones adjuntas del modelo de transformación de datos.

El algoritmo trabaja con algunos parámetros que se deben ir adecuando para encontrar los patrones secuenciales buscados, estos parámetros son: umbral mínimo de frecuencia, umbral de restricción GAP, umbral de soporte mínimo y el porcentaje de confianza.

Dependiendo de la configuración de los parámetros se puede llegar a la solución del algoritmo en poco tiempo, tardar mucho tiempo o incluso tener una solución enorme por la gran cantidad de patrones obtenidos, siendo que la computadora no puede procesarlos, para ello es necesario experimentar e identificar los valores adecuados para cada parámetro.

En el presente diagrama se muestra cómo es el funcionamiento interno del algoritmo GSP, en el cual podemos observar las fases por las que pasa para encontrar los patrones secuenciales, recordando que el algoritmo GSP se encarga de descubrir cómo son los patrones secuenciales

a partir de una base de datos secuencial, tomando una selección, utilizando la selección para la siguiente generación.



**Ilustración 5.** Funcionamiento del algoritmo GSP.

## 2.4 Resumen del capítulo

La metodología a emplear en esta investigación es KDD, por los resultados que ofrece, además, los algoritmos de minería de patrones pueden ser aplicados en diversos campos que tengan una base de datos formada por secuencias en eventos determinados por temporalidades, estos algoritmos trabajan acorde a parámetros establecidos que deben ser identificados para obtener los resultados deseados o los resultados con mayor eficiencia.





## 3. Estado del Arte

---

En este capítulo se describen más detalladamente los trabajos relacionados a la deserción escolar, el análisis de los trabajos sobre la deserción en la UAPT y los trabajos relacionados a la posible solución.

Existen diversos y variados estudios sobre las trayectorias académicas que los estudiantes presentan en su formación académica vista desde diversos campos de investigación, para esta investigación se recuperaron investigaciones sobre las trayectorias académicas de los estudiantes.

### *3.1 Las trayectorias académicas: dimensiones personales de una trayectoria estudiantil. Testimonio de un actor.*

En el trabajo realizado por [Guevara, 2013] se menciona que los estudios cuantitativos de la matrícula universitaria dan cuenta del rendimiento, ingreso y/o egreso, deserción, éxito o fracaso de los estudiantes; ya que la información obtenida no expresa las condiciones reales

de evolución de la matrícula ni de la compleja situación de los estudiantes; por lo que, la importancia de los valores que arrojan los estudios cualitativos hace necesario que se investigue acerca de cuáles son los factores que propician esas tendencias mencionadas.

El trabajo de Guevara ubica al estudiante en un momento de su trayectoria universitaria, observa en retrospectiva cuál ha sido su recorrido en relación al plan curricular de su carrera y el tiempo de realización del mismo, distinguiendo entre el tiempo académico y el tiempo personal del estudiante, realizando la lectura comprensiva de su trayectoria, de la dimensión simbólica e ilustración con el análisis e interpretación de un caso, presentado sus resultados obtenidos en su investigación.

Una trayectoria es la descripción de la dimensión temporal de las prácticas, que permite ver el ritmo y la duración de un proceso. Mientras que se hace mención que una trayectoria académica es un abordaje comprensivo del recorrido de los estudiantes en su contexto educativo, considerando la interacción de sus experiencias sociales y curriculares, situadas en un tiempo y espacio.

Su análisis se enfoca en el recorrido que los estudiantes realizan en el marco de las estructuras establecidas por el plan de estudios, los ritmos en que se cubren las asignaturas, las unidades de aprendizaje que se aprueban y reprueban, y las calificaciones obtenidas.

*"La trayectoria académica traduce el itinerario que los estudiantes realizan por las instituciones educativas, producto de una construcción dialéctica que se establece entre sus experiencias personales y sociales, el contexto sociocultural y la propuesta curricular de la institución (Kaplan y Fainsod, 2001)."*

Un análisis objetivo de las trayectorias académicas permite visualizar diversas características que afectan al estudiante como: la detección de aplazos y la continuidad del itinerario, la

identificación del momento, el lugar preciso y posibles razones por las que un estudiante incurre en retrasos, reprobación, repeticiones, o en un recorrido libre de esos sucesos; conociendo no sólo cuánto tiempo les toma cursar una carrera sino comprender las razones de sus logros y fracasos.

En su trabajo [Guevara, 2013] recuperó el testimonio de los estudiantes sobre sus experiencias y el significado de las mismas, asumiendo algunos principios teóricos que ofrece el enfoque biográfico. Utilizando los testimonios de estudiantes de la Universidad Nacional de San Juan, cuyos estudiantes entraron en 2007 y se encontraban en el año 2011. El relato biográfico de los estudiantes evidencia la experiencia de los individuos tal como ellos la procesan e interpretan bajo el filtro de sus creencias, representaciones y valores. Los sucesos o experiencias narradas, en forma biográfica, ubican al sujeto protagonista dentro del contexto socio-histórico.

Los resultados preliminares presentados son:

- Las trayectorias incluyen el registro de los obstáculos en las biografías personales de los estudiantes.
- Las decisiones tomadas acerca de interrupciones y "desvíos" en sus trayectorias.
- Se ubican a los estudiantes en su cuarto año de la carrera, con lo que tienen mayor conocimiento de sus recorridos.
- Realización de gráficas tomando como criterio las unidades de aprendizaje cursadas y aprobadas por curso, la relación entre unidades de aprendizaje y la cronología de estas prácticas.

Presenta varios tipos de trayectorias obtenidas de su análisis:

- **Ordenada:** El estudiante completa cada uno de los cursos académicos. La cronología de las prácticas y las especificidades de las mismas se adecuan a la matriz académico-

institucional. El 90% de este grupo ingreso entre 2007 y 2008, son quienes se encuentran en cuarto año en los tiempos académicos pautados

- **Prolija:** El estudiante completa cursos y sigue la línea disciplinar. Si bien modelizan sus prácticas siguiendo el modelo propuesto, en un momento de la carrera cambia su ritmo y patrón. Si bien el 35% de este grupo pertenece a la cohorte 2007-2008, el resto de ello se distribuye entre quienes ingresaron con anterioridad a este grupo. Los tiempos ya no son los académicos sino los que le impone cada estudiante a su práctica.
- **Estratégica:** El estudiante completa el primer curso y alterna entre el 2° y 3° priorizando la línea disciplinar. Sus actos son ordenados en un comienzo, pero a partir del segundo ciclo académico cambia y diseña su propia estrategia tomando los elementos que la institución le ofrece. Este tipo de traza en el recorrido académico lo encontramos en los estudiantes que han ingresado en distintos momentos. Su lógica es tomar todo recurso académico que le permita optimizar sus esfuerzos.
- **Táctica:** El estudiante prioriza las unidades de aprendizaje promocionales, sigue correlativas disciplinares y vencimiento de regularidad, aprovecha todas las condiciones y requisitos que ofrece el modelo, le interesa ir manejando estos acontecimientos para lograr el objetivo: aprobar las unidades de aprendizaje y avanzar. La característica de esta trayectoria es sacar provecho de cada acontecimiento es desarrollada por quienes llevan más tiempo en la institución.
- **Libre:** El estudiante sigue las correlativas y la línea disciplinar. Su recorrido se ajusta a aquellos patrones que le permiten avanzar. Este tipo de trayectoria probablemente prolongue la permanencia del estudiante en la carrera y en algunos casos al abandono de la misma.

Del mismo modo presentan los factores evaluados para cada estudiante, los cuales son [Guevara, 2013]:

- La experiencia escolar: conocimientos obtenidos antes de la evaluación.
- La figura del otro: alguna persona que intervenga, familia, conocido, amigo, etc.
- Las estrategias de estudio: estrategias usadas en el recorrido académico.

- Las exigencias del plan de estudios: el requisito de regularizar sus unidades de aprendizaje.
- El tiempo: constante en el recorrido, no manejable por estar ajustado.

### *3.2 Los estudios de trayectorias académicas y profesionales, contribuciones al estado del arte.*

Los estudios sobre las trayectorias académicas y profesionales aportan información sobre las variables que afectan o favorecen el desempeño del estudiante en su espacio académico. En su estudio [Piña, 2007] propuso como objetivo la investigación de 3 cohortes generacionales correspondientes a los años 2006 y 2007, estas 3 cohortes son los que ingresan al PUEE (Plan Único de Especialización en Enfermería) en sus diversas especialidades para identificar las características de su ingreso, trayectorias académicas, su egreso y a mediano plazo su ubicación en el mercado del trabajo.

Los estudios realizados bajo la temática académica en el nivel superior y en posgrado, centran sus investigaciones una preocupación por determinar un elemento de predicción respecto a los resultados captados a través de los filtros por los que atraviesan los estudiantes en el momento de su ingreso, la posibilidad de permanecer y concluir exitosamente el programa educativo que realizan. En este trabajo concluyen que el examen de selección no es suficiente para sostener una validez predictiva; estos estudios hacen mención a las metodologías de estudio empleadas para estudios de egresados el hacer uso de un instrumento de aplicación común que permita comparar los desempeños. En cuanto a los estudiantes que siguen estudiando se plantea llevar un seguimiento de la evolución de los estudiantes y orientarlos, apoyándolos durante su trayectoria escolar.

Piña concluye de las investigaciones realizadas sobre estudios de seguimiento de los estudiantes en su ingreso, permanencia y egreso se perfilan en 3 aspectos fundamentales:

- Recuperar los antecedentes académicos de los estudiantes a través de exámenes de ingreso, es decir, conocimientos previos para predecir el nivel académico al inicio de sus estudios y los casos de éxito.
- Las características de ingreso como las de sostenimiento durante el trayecto de formación para asegurar las condiciones idóneas para un buen rendimiento escolar.
- El destino laboral de los egresados como forma de seguimiento, los sectores de trabajo, etc.

Bajo los 3 aspectos mencionados, Piña los usa en su metodología para obtener una visión completa de todo el proceso tanto sus aciertos y desaciertos, captando al estudiante desde su ingreso y lo va siguiendo durante su trayectoria académica a lo largo de su especialidad. Por lo que para obtener su información aplica 2 cuestionarios, el primero "informativo académicamente" y el segundo "cualitativo respecto a los conocimientos adquiridos en sus especialidades", con los datos obtenidos crearon una base de datos bajo el programa estadístico SPSS de las generaciones 2006 y 2007.

Las conclusiones más significativas de la investigación realizada por Piña [Piña, 2007] son las siguientes:

- La revisión y el análisis de documentos e investigaciones para conformar un estado del arte respecto a los estudios de ingreso, seguimiento de trayectorias académicas y egresados, muestran diversas metodologías que dan un mayor conocimiento sobre quiénes son los estudiantes, cómo transitan por sus estudios y qué hacen cuando los han concluido.
- La información obtenida permite detectar aciertos que consolidan un programa educativo, como ubicar las tareas pendientes, los niveles académicos, el ambiente académico del programa educativo. La investigación de los estudiantes de las especialidades de Enfermería del PUEE, aporta información sobre los aspectos mencionados.

### *3.3 Origen social de los estudiantes y trayectorias estudiantiles en la Universidad Veracruzana.*

El artículo presentado por Casillas estudia el origen social de los estudiantes a través de la noción de capital cultural, poniéndolo en relación con los procesos de selección del primer ingreso, la trayectoria escolar y la situación escolar actual, presenta una tipología de estudiantes universitarios construida sobre la base de información estadística disponible en las instituciones.

Hace mención que la educación superior es una oportunidad social rara, que no está disponible a toda la población en general si atendemos los rasgos de desigualdad y exclusión social. Por lo que para avanzar en el conocimiento de los estudiantes universitarios se hace una clasificación cultural evidenciando las poblaciones estudiantiles que enfrentan riesgos frente a sus estudios analizando los procesos de selección de primer ingreso y la trayectoria escolar en la universidad de acuerdo con las distintas dotaciones culturales que diferencian a los estudiantes.

La metodología tiene un alcance comparativo, ya que los datos utilizados son accesibles por otros investigadores, se trata de información disponible que analizada muestra procesos asociados en la explicación de las trayectorias y experiencias estudiantiles en la universidad mexicana.

La tasa bruta de escolaridad superior se calcula poniendo en relación la población total de 20 a 24 años con el total de la matrícula en la educación superior. Están incluidas todas las modalidades del sistema de educación superior. Casillas observa que las condiciones de

oportunidad de los jóvenes mexicanos para acceder a la educación superior eran muy reducidas y a lo largo de los años se han ido aumentando gradualmente.

En la sociología de la educación hay un debate en cuanto al origen social de los estudiantes, ya que por un lado consideran que la escuela es un espacio de justicia social por otro lado critican la igualdad y la aportación de diferencias. El trabajo de Casillas se centra en responder las interrogantes: ¿quiénes son los estudiantes universitarios? ¿de qué espacios sociales provienen? ¿cuáles son sus dotaciones culturales y cómo influyen éstas en los resultados de la experiencia escolar? Para ello realiza una investigación sobre trabajos de discusión sociológica y retomando la estadística de los datos obtenidos de la universidad Veracruzana.

Casillas hace referencia a Pierre Bourdieu quien desarrolló la noción de capital cultural para tener una unidad de medida que diferencie a los estudiantes de acuerdo con sus disposiciones culturales y no sólo económicas. Casillas dice: "*Entendemos por capital familiar el conjunto de disposiciones sociales, atributos y recursos que poseen los estudiantes de acuerdo con su origen social y familiar*".

Casillas hace referencias a que el capital familiar de los estudiantes resulta de la definición de grupos de acuerdo con tres variables distintivas:

- Nivel de escolaridad de los padres, entendido como el grado máximo de escolaridad alcanzado en la familia. Se clasifica en rangos: nula (hasta primaria incompleta), básica (primaria y secundaria), media (con bachillerato o carreras técnicas) y superior (con estudios superiores hasta el posgrado).
- Nivel de prestigio de la ocupación de los padres, medido a partir de una codificación previa donde clasificamos el prestigio social. El prestigio de la ocupación de los padres puede ser bajo (corresponde a los sectores populares: trabajadores domésticos, campesinos, obreros), medio (empleados, comerciantes medios, burócratas) o alto (patrones, profesores, profesionistas, funcionarios y gerentes).



- Nivel socioeconómico de la familia, entendido como los ingresos económicos de la familia en su conjunto. Consideramos que los estudiantes provienen de ambientes culturales diversos pues viven de manera diferente de acuerdo con sus posibilidades económicas.

Casillas dice que “entendemos como capital escolar al conjunto de indicadores que nos refieren a la trayectoria de los estudiantes inmediatamente previa a la educación superior”.

Por lo que presenta la siguiente tabla que clasifica los grupos de estudiantes.

**Tabla 1.** Grupos de estudiantes por capital familiar.

<b>Grupo</b>	<b>Descripción</b>
Muy Pobre	Se caracteriza por estudiantes que combinan una situación de padres sin escolaridad, un índice socioeconómico familiar pobre y un prestigio de la ocupación de los padres bajo, aunque en ocasiones medio.
Pobre	Se caracteriza mayoritariamente por estudiantes hijos de padres con escolaridad básica y en ocasiones media, con índice socioeconómico familiar pobre, aunque muchos de ellos tienen padres cuya ocupación se ubica con un prestigio medio y alto (por la influencia de profesores y profesionistas).
Medio Alto	Se caracteriza mayoritariamente por una alta escolaridad de los padres, con índice socioeconómico medio, donde la mayoría de ellos tienen padres con ocupaciones con prestigios medios y altos.

Casillas muestra la caracterización de los grupos por su capital cultural:

- Herederos. Son originarios de los sectores medios con trayectoria escolar previa media y alta.
- Héroe. Son originarios de sectores marginales con una trayectoria escolar previa alta.
- Pobres exitosos. Son originarios de clases populares con una trayectoria escolar previa alta.

- Riesgo. Originarios de clases populares con una trayectoria escolar previa mayoritariamente media.
- Alto riesgo: Muy pobres en capital cultural.

Con base en la caracterización y definición de los grupos de estudiantes, por lo que para hacer la observación de las trayectorias de los tipos de estudiantes realiza 3 ejercicios:

- El primero consiste en hacer evidente la doble discriminación que ocurre por un lado entre los que demandan entrar a la universidad y los que efectivamente ingresan; por el otro, entre quienes ingresan y quienes han logrado sobrevivir en la universidad hasta el tercer año de los estudios de licenciatura.
- El segundo ejercicio consiste en poner en juego la tipología para caracterizar la trayectoria de quienes han sobrevivido, haciendo observable las diferentes trayectorias de cada una de las clases de estudiante.
- El tercero intenta una aproximación a la consideración sobre la situación escolar actual del presente trabajo.

Las dobles discriminaciones son: la primera discriminación dice que sólo los más dotados en capital cultural son los más capaces de lograr un espacio en la institución; la segunda discriminación ocurre entre quienes se inscriben y quienes sobreviven después de tres años de estudio.

De acuerdo a la tipología presentada las trayectorias estudiantiles pueden ser clasificadas como:

- Alta: estudiantes que tienen un alto promedio de calificaciones, aprueban sus asignaturas de manera ordinaria y promueven todos los cursos a los que se inscriben, son los más cumplidos, con mejor desempeño.

- Regular: estudiantes que se mantienen al margen aprobando asignaturas cotas aprobatorias, con un buen o regular desempeño.
- Baja: estudiantes cuya experiencia escolar está marcada por bajas notas, presentación de exámenes extraordinarios y frecuentes fracasos a la hora de promover una asignatura determinada.

En la siguiente tabla se observa la distribución de trayectorias agrupadas por tipo de estudiante:

**Tabla 2.** Trayectorias agrupadas por tipo de estudiante.

Capital cultural	Trayectorias			Total
	Alta	Regular	Baja	
Herederos	532	316	668	1516
	35.09%	20.84%	44.06%	
Héroes	225	129	228	582
	38.66%	22.16%	39.18%	
Pobres Exitosos	317	210	262	789
	40.18%	26.62%	33.21%	
Riesgo	207	195	633	1035
	20.00%	18.84%	61.16%	
Alto Riesgo	199	186	481	866
	22.98%	21.48%	55.54%	
	<b>1480</b>	<b>1036</b>	<b>2272</b>	<b>4788</b>

La situación escolar actual se refiere al número de créditos o unidades de aprendizaje que ha promovido el estudiante del total de créditos o unidades de aprendizaje requeridos del programa cursado, el índice presentado es bajo el siguiente criterio óptimo 100% para alguien que ha promovido el total de los cursos o créditos que se esperaba en un determinado tiempo establecido por el programa de estudios; Irregular es alguien que ha promovido entre el 90 y

menos del 100% de lo que debió de haber promovido; y Rezagado es alguien que debe más del 10% de las unidades de aprendizaje o créditos a los que se inscribió

**Tabla 3.** Situación escolar del estudiante de acuerdo a sus créditos aprobados.

<b>Situación Escolar</b>	<b>Frecuencia</b>	<b>Porcentaje</b>
Óptimo	2676	55.89%
Irregular	253	5.28%
Rezago	1859	38.83%
<b>Total</b>	<b>4788</b>	<b>100.00%</b>

Concluye que el acceso a la educación superior es exclusivo de una minoría que estuvo asociada a los sectores medios y altos de la población, con la expansión del sistema se han diversificado los orígenes sociales del estudiantado marcando la educación superior como reproductora de la exclusión y de la desigualdad social. Las determinaciones sociales y familiares tienen un peso específico, las poblaciones estudiantiles no son homogéneas y se muestra la diversidad de tipos de estudiantes que componen la universidad contemporánea. Con base en los tipos estudiantiles, se observan distintos modos de apropiación de las oportunidades escolares y del desarrollo de las trayectorias estudiantiles [Casillas, 2007].

### *3.4 Trayectorias académicas generacionales, constitución y diversificación del oficio académico en el caso de los bioquímicos de la Facultad de Medicina.*

La investigación realizada por Landesmann se enfocó en los procesos de constitución y diferenciación del oficio académico para una comunidad de académicos de la UNAM de una sola disciplina, explorando distintas variables: la pertenencia institucional (Facultad de Medicina) el tipo de vínculo (de tiempo completo), la afiliación disciplinaria (Bioquímica) y los periodos de inserción en la vida académica (tres generaciones entre 1957 y 1974).

Para su investigación utiliza el método biográfico que es un recurso que permite reconstruir las trayectorias socio-familiares, escolares y profesionales además de ofrecer pistas para reconocer determinismos o condicionantes sociales e institucionales y su universo simbólico, identificar las "herencias" sociales y familiares presentes en la realidad del sujeto.

El trabajo empírico desarrolla las siguientes estrategias:

- La reconstrucción de las historias de vida de diez bioquímicos que habían trabajado, por lo menos 10 años como profesores de tiempo completo, en el Departamento de Bioquímica entre 1957 y 1974.
- La recolección y análisis de la currícula de otros seis bioquímicos.
- Las consultas en fuentes primarias (expedientes de la Facultad) y secundarias relativas a la historia de la UNAM.
- La realización de algunas entrevistas a informantes clave.

La reconstrucción de las historias de vida se efectuó mediante entrevistas a profundidad de al menos 4 sesiones por persona incorporando una mezcla de elementos de orden cultural, social, científico, académico y psicológico. Para el relato biográfico se tomaron en cuenta los siguientes aspectos:

- Seleccionar de los relatos individuales los aspectos que dan cuenta de las trayectorias académicas.
- Recopilar los itinerarios y "estados mentales" compartidos por los miembros de una misma generación.
- Establecer las articulaciones que vinculan las trayectorias colectivas con su contexto.
- En la reconstrucción del relato se relacionan factores heterogéneos como motivaciones de actores y circunstancias externas.

Con los datos obtenidos de sus relatos biográficos, Landesmann identificó y clasificó las 3 generaciones estudiadas, las cuales son:

- Los “fundadores” son los primeros que ocuparon los puestos de profesores de carrera, por su función dentro de la institucionalización de la bioquímica en la Facultad de Medicina. Tienen un papel de gestión del Departamento, después de la Facultad y luego del campo universitario, en algunos casos, en el político y científico. Su trayectoria académica se reconstruye entre 1957 y 1971.
- Los “herederos” tienen una década menos que los “fundadores”, pero “heredaron”, sus aspiraciones científicas; orientados hacia la investigación, a partir de 1974, se establecen en los institutos y formando parte de la élite científica de la UNAM. Su trayectoria académica se describe entre 1957 y 1974
- Los “profesores de Facultad”, son contemporáneos de los “herederos”, tienen un ingreso más tardío en la academia y se proyectan preferentemente hacia la docencia de pregrado en la Facultad de Medicina. Su trayectoria académica se reconstruye de los años sesenta a mediados de los setenta.

Por lo que concluye que la historia de tres generaciones de profesores de carrera bioquímicos de la Facultad de Medicina muestra la compleja trama que preside la constitución del nuevo rol académico. En ella se articulan las trayectorias de acceso a esta profesión, la historia de la disciplina, el papel de los liderazgos académicos, de los grupos y comunidades como instancias socializadoras, el contexto institucional, los acontecimientos que alteran el curso de la vida institucional, las trayectorias de los grupos y los actores [Landesmann, 2001].

### *3.5 Trayectorias académicas y construcciones subjetivas de estudiantes y ex estudiantes virtuales.*

Entre 2009 y 2013 en el Programa de Educación a Distancia de la Universidad Virtual de Quilmes (UVQ), perteneciente a la Universidad Nacional de Quilmes, en Buenos Aires, Argentina.

Sepúlveda desarrolló un proyecto de investigación cuali-cuantitativo que analizó la relación entre las tutorías virtuales y las trayectorias académicas de los estudiantes, centrándose en el punto de vista de los actores que la realizaron (tutores) y la población sobre la cual se ejercía esta acción tutorial (estudiantes).

Estudió en concreto los Entornos Virtuales de Enseñanza y Aprendizaje (EVEA) y su impacto en la relación tutor – estudiante. Identificó el rol positivo del tutor respecto de las trayectorias académicas de los estudiantes a través de los relatos de éstos. Caracterizó las trayectorias de los estudiantes, la identificación de su prolongación en el tiempo y la construcción de una tipología. Identificó las cuestiones que los estudiantes y graduados destacaban como positivas, los ex estudiantes citaban motivos que los habían llevado a dejar sus estudios o abandono parcial.

Para su investigación Sepúlveda utiliza la metodología de la triangulación de métodos cualitativos y cuantitativos. En su investigación, respecto de los datos cuantitativos, la fuente de información será la base de datos de UVQ (Sistema Esmeralda). En el aspecto cualitativo realizaron entrevistas y creación de grupos focales, considerando: trayectorias y su tipología, abandono, factores de frustración, género y subjetividad, entendidas como:

- La trayectoria académica será entendida como resultado del recorrido curricular que realiza un estudiante teniendo en cuenta el tiempo de duración de la carrera, la regularidad en los estudios y el egreso.
- Tipología de trayectorias: [7]trayectoria acelerada (TA), más de seis finales por año; y trayectoria estándar (TE), entre tres y seis finales por año; trayectoria extendida en el tiempo (TET), menos de tres finales por año; y trayectoria muy extendida en el tiempo (TMET), menos de 1.5 finales por año.
- La subjetividad puede entenderse como conjunto de elementos conceptuales que resultan de utilidad para comprender la interacción entre los procesos sociales y las escenas más pequeñas de interacción cotidiana.

Sepúlveda estudia tanto a los estudiantes avanzados, los ex estudiantes y a los estudios universitarios virtuales ofrecidos por su universidad.

Sepúlveda concluye en su investigación [Sepúlveda, 2013]:

- Las entrevistas en profundidad dieron como resultado que las carreras cuyas duraciones son bastante más largas de las previstas se debían a que permitían mejorar las posiciones en sus espacios laborales actuales o consolidarse en ellos. Los factores personales destacados por los actores tuvieron que ver con el acompañamiento familiar, la perseverancia, la autodisciplina y la renuncia al descanso dada la necesidad de compatibilizar largas jornadas laborales con los estudios, la capacidad de superar las adversidades familiares y los tropiezos académicos como fallos en los exámenes y la necesidad de recurrir a asignaturas vencidas, a veces no una, sino varias veces.
- Los datos mostrados del enfoque cuantitativo son los siguientes: entre junio de 2011 y el mismo mes de 2013, un total de ocho periodos cursados, ingresaron a su universidad 5 595 estudiantes y se produjeron 1,878 bajas definitivas (33%). Respecto al estado civil como la edad: 64.9% son solteros, casados, 27.5% y unidos 14%.

### *3.6 Síntesis de los trabajos con el análisis de trayectorias académicas*

A manera de conclusión se presentan los puntos más importantes de las investigaciones presentadas en la siguiente tabla:



**Tabla 4.** Cuadro sinóptico de trabajos relacionados con el estudio de trayectorias académicas.

<b>Título</b>	<b>Enfoque</b>	<b>Herramientas</b>	<b>Resultados</b>
Las trayectorias académicas: dimensiones personales de una trayectoria estudiantil. Testimonio de un actor [Guevara, 2013].	Trayectorias académicas	Análisis objetivo, método biográfico, testimonios de estudiantes.	Una clasificación de trayectorias resultantes del análisis, factores de evaluación.
Los estudios de trayectorias académicas y profesionales, contribuciones al estado del arte [Piña, 2007].	Trayectorias académicas	Análisis cuantitativo de conocimientos, un cuestionario	Clasificación de 3 cohortes para la especialidad de PUUE, para su impacto en la vida profesional.
Origen social de los estudiantes y trayectorias estudiantiles en la Universidad Veracruzana [Casillas, 2007].	Trayectorias académicas	Análisis estadísticos del INEGI respecto a educación, análisis estadístico de variables sociales.	La educación superior como exclusiva a una minoría, origen y repercusiones sociales.
Trayectorias académicas generacionales constitución y diversificación del oficio académico. El caso de los bioquímicos de la Facultad de Medicina [Landesmann, 2001].	Trayectorias académicas	Método Biográfico, testimonios, entrevistas, análisis de currículos.	La historia académica de 3 generaciones de profesores de carrera bioquímicos. La socialización y el rol académico
Trayectorias académicas y construcciones subjetivas de estudiantes y ex estudiantes virtuales [Sepúlveda, 2013].	Trayectorias académicas	Tutorías virtuales, trayectorias académicas, entrevistas y testimonios.	Análisis estadístico de estudiantes y estudiantes egresados

Como se puede observar en los trabajos sobre trayectorias académicas presentados, no usan las trayectorias como base de investigación, es decir, no trabajan con los historiales

académicos de los estudiantes, tampoco utilizan herramientas computacionales para obtener sus resultados, específicamente herramientas de la Inteligencia Artificial. Además, los trabajos relacionados no están enfocados en el problema la deserción escolar, más bien describen como son las trayectorias académicas vistas por las experiencias de los estudiantes.

Ahora se presentan los trabajos de deserción realizados en la UAPT de la Universidad Autónoma del Estado de México (UAEM), se han realizado algunos estudios de diversos puntos de vista: [García, 2012], [García, 2013] y [Osorio, 2013].

### *3.7 Descubrimiento de factores que inciden en la eficiencia terminal de estudiantes de educación superior con árboles de decisión.*

García Lambert en su investigación, realizó un estudio completo para determinar los factores que inciden en la eficiencia terminal en estudiantes de educación superior en la UAPT [García, 2012].

Estos factores se determinaron a partir de diversos enfoques como: el nivel de conocimiento del estudiante al entrar a la carrera, datos personales, datos familiares, de entorno, etc., mediante los resultados del EXANI-II aplicado para el ingreso a la educación superior.

García utiliza una base de datos obtenidas de un cuestionario de contexto del periodo de ingreso del 2008 en la UAPT de 305 estudiantes y 206 preguntas, los estudiantes pertenecen a las 4 carreras ofertadas: Ingeniería en Plásticos, Ingeniería en Producción Industrial, Ingeniería en Software y Seguridad Ciudadana.

Al realizar diversos experimentos utilizó la base de datos presentes en la siguiente tabla, con información de estudiantes de sexto semestre ya que aún no había egresados:

**Tabla 5.** Datos de García Lambert.

Base de datos	Estudiantes	Preguntas
EXANI II	305	206
Control escolar		9
Total	305	215

García preproceso los datos obtenidos uniendo la información recabada del Exani y de control escolar para obtener una vista minable que posteriormente la aplicó con árboles de decisión.

Las conclusiones que presentó fueron:

- Las conclusiones del primer experimento dicen que las personas con las que vive el estudiante influyen en gran medida al desempeño académico; además existe dentro del entorno económico-familiar el ingreso del padre o tutor, acceso a una computadora y el número de hermanos; incluyendo la capacidad del estudiante en el entorno escolar, respuestas bajo presión.
- En el segundo experimento concluye que el idioma inglés es un factor importante, además menciona que la vida escolar del estudiante en bachillerato como: entrar a clases, promedio final, si tuvo beca, etc., son importantes para determinar si abandona el plan de estudios o no.
- Concluye en el tercer experimento que el número de créditos tomados y el número de unidades de aprendizaje tomadas al sexto semestre influyen bastante en la determinación de si el estudiante permanece o deserta, además de que es importante la trayectoria del estudiante antes de entrar a la UAPT.
- En el cuarto experimento concluye que el número de unidades de aprendizaje tomadas y el número de créditos aprobados al sexto semestre.

- En sus conclusiones finales aporta que en los experimentos algunos obtienen semejanzas por carreras y otros presentan diferencias.

### *3.8 Identificación de factores que influyen en la deserción escolar de estudiantes universitarios usando algoritmos genéticos y $K$ vecinos más cercanos*

Osorio en su estudio se enfoca en la deserción escolar tanto en la permanencia de los estudiantes [Osorio, 2013].

Osorio utiliza una base de datos proporcionada por el Centro Nacional de Evaluación para la Educación Superior (Ceneval), que fue recabada en el cuestionario de contexto del periodo de ingreso 2008 en la UAPT (Unidad académica Profesional Tianguistenco), con 305 estudiantes y 206 preguntas, además obtuvo información recabada de control escolar, en específico las trayectorias académicas de los estudiantes que trato para obtener una vista minable que se adhiriera a la información obtenida del Ceneval.

Complementando el Ceneval es una asociación civil dedicada al diseño y aplicación de instrumentos de evaluación de conocimientos, habilidades y competencias, también se dedica al análisis y a la difusión de los resultados arrojados de la evaluación [Ceneval, 2015].

Después de obtener los datos los preproceso para obtener una vista minable, Osorio utilizó un algoritmo genético para optimizar el proceso del algoritmo  $K$  vecinos más cercanos y una técnica de validación cruzada, aplicó el algoritmo a los datos obtenidos, al realizar una variación de distribución de sus variables realizó 2 experimentos significativos para su investigación.

La información obtenida fue analizada con base en los factores que influyen en las decisiones de los estudiantes de desertar de su carrera.

Osorio realizó 2 experimentos usando su método, con el cual obtuvo los siguientes resultados:

- En el primero obtuvo una lista de 29 factores sumamente importantes, tomando en cuenta solamente la información del EXANI-II.
- En el segundo obtuvo logró obtener una lista de 25 factores importantes, tomando en cuenta la información del EXANI-II más la información obtenida de control escolar.

Concluyó que aún se siguen teniendo algunos factores como en el trabajo de su estado del arte de García Lambert, como, las unidades de aprendizaje relacionadas con matemáticas, las habilidades relacionadas al uso de computadoras y habilidades del idioma inglés.

### *3.9 Reglas que describen la deserción y permanencia en los estudiantes de la UAP Tianguistenco de la UAEM.*

En la UAPT de la UAEM García, Hernández y Ledeneva [García, 2013] trataron de encontrar el conjunto de reglas de conocimiento que se pueden extraer de aquellos estudiantes que desertaron o que permanecieron en sus estudios universitarios después de 3 años de su ingreso.

Una base de datos inicial de 206 factores y 305 estudiantes de 4 licenciaturas de la UAPT de la UAEM fue utilizada para obtener árboles de decisión con la metodología de trabajo KDD (Descubrimiento de conocimiento en bases de datos).

Utilizaron el algoritmo C4.5 que genera árboles de decisión, este algoritmo lo usaron con la herramienta de software libre WEKA.

La experimentación dividida en 2 partes:

- La primera experimentación considero la información de la base de datos correspondiente al EXANI-II.
- La segunda se considera toda la base de datos (información del EXANI-II y la recabada por control escolar durante 3 años).

Los resultados obtenidos por experimento fueron:

- En el primero se encontraron 11 reglas que caracterizan a estudiantes que fueron dados de baja y 12 reglas para aquellos que no.
- En la mayoría de sus reglas tanto para los que son dados de baja como para los que no, el factor más importante fue que si el estudiante vivía con su madre.

Los resultados del segundo experimento fueron:

- Vuelve a aparecer como factor importante el si el estudiante vive con su madre.
- El porcentaje de créditos aprobados al sexto semestre es el factor que aporta más información al nodo raíz del árbol creado por algoritmo empleado.
- La trayectoria académica que el estudiante genere en el transcurso de su vida académica es fundamental para conocer si el estudiante será dado de baja o no, junto a otros factores como lo es su dominio del idioma inglés, entre otros de menor importancia.

Como conclusión usaron el algoritmo propuesto evaluando 206 factores obtenidos de información recabada por el EXANI-II y por control escolar, determinaron que con sólo 12 factores en 19 reglas se podía saber, con 82% de soporte, si un estudiante tiene riesgo de desertar o no de sus estudios en los 3 años posteriores.

Los factores más importantes obtenidos en la experimentación son 2: el primero es que si el estudiante vive con su madre y el segundo es considerar el seguimiento de la trayectoria académica [García, 2013].

### *3.10 Síntesis de los trabajos relacionados al análisis de deserción escolar en la UAP Tlanguistenco*

A manera de conclusión, los trabajos realizados en la UAPT se enfocaron en el estudio del problema de la deserción escolar y los factores que llevaban a que un estudiante desertara de su carrera, los factores encontrados fueron generales, los factores encontrados no fueron específicamente académicos.

Además, solo usaron las trayectorias para obtener cierta estadística representativa para obtener factores para determinar el porcentaje de deserción, no se estudiaron las trayectorias para fines de representación académica del estudiante.

En conclusión, las investigaciones realizadas sobre deserción en la UAPT se presentan de manera concreta en la siguiente tabla, mostrando los puntos que se tomaron como base para esta investigación:

**Tabla 6.** Cuadro sinóptico de trabajos relacionados con el estudio de trayectorias académicas en la UAPT de la UAEM.

<b>Nombre investigación</b>	<b>Enfoque de estudio</b>	<b>Herramientas usadas en estudio</b>	<b>Resultados</b>
Descubrimiento de factores que inciden en la eficiencia terminal de estudiantes de	Permanencia escolar, eficiencia terminal.	Árboles de decisión, estudiantes que terminan en tiempo (regulares), información obtenida de control escolar de la UAPT,	Encontró 54 factores que solo se enfocan en que el estudiante termine la carrera.

educación superior con árboles de decisión [García, 2012].		información del EXANI-II perteneciente al CENEVAL	
Identificación de factores que influyen en la deserción escolar de estudiantes universitarios usando algoritmos genéticos y K-vecinos más cercanos [Osorio, 2013].	Permanencia escolar, deserción escolar	Algoritmos genéticos, K vecinos más cercanos, información de estudiantes egresados obtenidos de control escolar de la UAPT, información de resultados del EXANI-II del CENEVAL.	Encontró los factores más importantes de los que García señala para deserción y permanencia, en 2 experimentos: en uno obtuvo 29 factores y en el segundo 25 factores.
Reglas que describen la deserción y permanencia en los estudiantes de la UAP Tianguistenco de la UAEM [García, 2013].	Permanencia escolar, deserción escolar	Algoritmo C4.5, árboles de decisión, información de EXANI-II, información de control escolar.	Determinó que con sólo 12 factores en 19 reglas se podía saber, si un estudiante tiene riesgo de desertar o no de sus estudios en los 3 años posteriores a su ingreso.

Ahora se presentan trabajos relacionados a la minería de patrones secuenciales.

### *3.11 Análisis de patrones secuenciales en una base de datos de estudiantes.*

Se encontró un trabajo que usó el algoritmo SPAM en una universidad de Estados Unidos: Análisis de patrones secuenciales en una base de datos de estudiantes (*Sequential patterns analysis in a student database*) [Campagni, 2015]:



En esta investigación se realizó un trabajo en que se presentó una metodología de minería de patrones secuenciales para analizar las trayectorias de los estudiantes, donde una trayectoria puede ser vista como una secuencia de estudiantes.

El modelo presentado usa un algoritmo llamado análisis de minería de patrones secuenciales (SPAM Sequential Pattern Analysis Mining).

Campagni menciona a la minería de datos educativos que es un área de investigación emergente e interesante que produce regularidades útiles, previamente desconocidas de bases de datos educativas, estas ayudan a comprender mejor y mejorar el rendimiento educativo y la evaluación del proceso de aprendizaje del estudiante. Por lo que esta investigación se realizó con dicho enfoque.

Campagni utiliza un enfoque denominado "Carrera ideal", en el cual una carrera ideal correspondiente a un estudiante que ha tomado cada examinación (examen) justo después de terminar su respectivo curso, sin tener retrasos, es decir, sin que lo tome un semestre posterior, aprobando cada curso con calificaciones regulares o buenas.

Los patrones frecuentes identificados por el algoritmo son comparados con la carrera ideal del estudiante, los patrones más interesantes son usados para refinar el análisis usando técnicas de clasificación. Finalmente, Campagni aplicó su metodología a un caso de estudio e interpretó sus resultados.

En su experimentación Campagni utilizó la herramienta de WEKA, ya que es un framework propuesto para la minería de datos educativos.

La metodología de minería de patrones secuenciales considera como eventos a los exámenes que fueron tomadas por los estudiantes; la información temporal es el semestre en el que el examen fue tomado o el retraso con el que este fue tomado.

Campagni estudió una organización de una universidad en la que los estudiantes toman exámenes en diferentes sesiones antes de terminar un curso. Para estudiar la base de datos creada, se utilizó el algoritmo SPAM, en una implementación que puede ser encontrada, usando la implementación de clasificación de K-means perteneciente a weka.

La base de datos se construyó con 2 conjuntos de datos, pertenecientes a los años 2001 -2007:

- Información anónima del estudiante: identificador de estudiante, estudiante, fecha de inscripción, fecha, fecha de la marca, calificación final del examen final.
- Información de los exámenes: para cada examen, el identificador del estudiante, el estudiante, id de examen, examen, fecha de examen, la marca correspondiente, calificación.

Tras ejecutar el algoritmo se encontró un subconjunto de 1,271 patrones, en el que el valor máximo de los eventos de un patrón fue de 5, se obtuvieron 12 secuencias de 5 eventos o tamaño 5, 5 secuencias eran de tamaño 4 complementarias a las secuencias de tamaño 5, 1 única secuencia de tamaño 5, casi 200 secuencias de tamaño 4 y muchas otras con un número menor de exámenes. Por lo que la carrera ideal para su caso de estudio fue de un tamaño 25.

Como conclusión para investigar y analizar una base de datos de estudiantes con el algoritmo SPAM, para encontrar la carrera ideal, Campagni siguió 3 pasos:

1. Generación de los patrones frecuentes usando el algoritmo SPAM.

2. Selección de los patrones más interesantes.
3. El análisis de clasificación basado en los resultados del paso anterior.

De acuerdo al análisis, en el caso de estudio los patrones correspondientes al semestre y al retraso como información temporal, divide a los estudiantes como buenos y no tan buenos, de acuerdo al grado final y a la longitud de la carrera.

Incluyendo además patrones muy frecuentes correspondientes a las unidades de aprendizaje tomadas respecto a si pasaron o no, además de ser en un espacio de tiempo determinado, son muy específicos correspondientes al caso de estudio [Campagni, 2015].

### *3.12 Minería secuencial: análisis de algoritmos y patrones*

En la investigación Sequential Mining: Patterns and algorithms analysis [Slimani, 2013], se realizó un estudio comparativo y analítico de los diversos algoritmos de minería de patrones secuenciales existentes hasta ese momento.

Una investigación donde se estudiaron diversos algoritmos de minería de patrones secuenciales perteneciente a Slimani [Slimani, 2013], se llevó a cabo una evaluación de 5 clasificaciones de algoritmos, evaluando su eficiencia y eficacia.

De acuerdo a Slimani en su trabajo de Análisis de Algoritmos y Patrones de Minería de Datos Secuenciales, realizó un trabajo analizando el funcionamiento de los algoritmos existentes hasta ese momento de minería de patrones secuenciales en cuanto a eficacia de la minería efectuada y a la eficiencia de recursos en PC, con lo cual realizó 5 clasificaciones de algoritmos y determinó que el algoritmo GSP fue el mejor en cuanto a eficacia y eficiencia tomando los dos aspectos.

Para ello, con una base de datos genérica se implementó para cada uno de los 15 algoritmos, evaluando la eficiencia en cuanto a cuánto tiempo de ejecución tardaban, en cuanto a la calidad de resultados obtenidos y en cuanto a cómo eran los tipos de resultados.

Slimani realizó una evaluación y clasificación de los algoritmos tomando como base sus resultados, con lo que obtuvo 5 clasificaciones. Los algoritmos evaluados son los siguientes: Algoritmos A priori like, GSP, SPADE, MFS, FreeSpan, PrefixSpan, SPAM, CloSpan, BIDE, SuffixTree, FASTUP, ISE, ISM, IncSP, InsSpan.

Como resultado de ello uno de los mejores algoritmos para encontrar patrones secuenciales presentes en la base de datos, sin la necesidad de crecimiento fue el algoritmo GSP, el cual mostraba muy buenos resultados de eficiencia siempre y cuando los ajustes a los parámetros fueran los óptimos.

### *3.13 Síntesis de los trabajos relacionados a la aplicación de minería de datos secuenciales a estudiantes*

A manera de conclusión de los puntos más importantes se presenta la siguiente tabla.

**Tabla 7.** Cuadro sinóptico de trabajos relacionados con la minería de datos secuenciales.

<b>Nombre investigación</b>	<b>Enfoque de estudio</b>	<b>Herramientas usadas en estudio</b>	<b>Resultados</b>
Sequential patterns analysis in a student database [Campagni, 2015].	"Carrera ideal", número de unidades de aprendizaje tomadas al semestre	Algoritmo SPAM ( <i>Sequential Pattern Mining Algorithm</i> ), base de datos de estudiantes egresados generación 2001-2007.	1271 patrones generados, identificando mejores secuencias para toma de exámenes.

<i>Sequential Mining: Patterns and algorithms analysis</i> [Slimani, 2013]	Análisis algoritmos secuenciales	Algoritmos A priori like, GSP, SPADE, MFS, FreeSpan, PrefixSpan, SPAM, CloSpan, BIDE, SuffixTree, FASTUP, ISE, ISM, IncSP, InsSpan.	Evaluación de 5 clasificaciones de algoritmos, evaluando su eficiencia y eficacia.
--	----------------------------------	---	--

Como se observa en la tabla, el primer trabajo está enfocado en “la carrera ideal”, estudiando bases de datos de historiales académicos en específico el número de unidades de aprendizaje tomadas y aprobadas por semestre, contabilizando el número de exámenes aprobados por semestre, identificando las mejores secuencias para la toma de exámenes para que el estudiante egrese de su carrera.

En cuanto al segundo trabajo presentado en la tabla anterior, se enfoca en la evaluación de eficiencia y eficacia de algoritmos de minería de patrones secuenciales, clasificándolos, evaluándolos y comparando resultados obtenidos, en cuyas conclusiones podemos observar los algoritmos adecuados a las bases de datos de estudiantes, en comparación con el primer estudio presentado.

### *3.14 Resumen del capítulo*

En los estudios realizados en la UAPT sobre deserción escolar se enfocaron en la búsqueda e identificación de factores que propician la deserción o permanencia escolar, en estos trabajos en algo que coinciden y concluyen implícitamente en que es necesario estudiar las trayectorias académicas de los estudiantes, Osorio indica que la trayectoria académica hasta el sexto semestre es muy influyente en si un estudiante deserta o no [García, 2012], [Osorio, 2013].

Es necesario mencionar que no se han encontrado trabajos cuyo objeto de estudio sean sus trayectorias académicas o historiales académicos; el trabajo más parecido fue el de

Campagni [Campagni, 2015] en que trata de definir la carrera ideal del estudiante para que logre egresar, estudiando solo las trayectorias sobre el número de exámenes tomados por semestre y aprobados.

Por otro lado, el plan de estudios de la UAPT mostrado en los mapas curriculares de las carreras ofertadas [UAEM-OE, 2016], es flexible, ya que cada carrera contiene unidades de aprendizaje que son seriadas, otras unidades de aprendizaje que no son seriadas lo que da lugar a que existan distintas trayectorias académicas reflejadas en su historial académico, cada estudiante tiene su propia trayectoria académica.

Los planes curriculares ofertados por la UAPT se encuentran en el apartado de anexos, que son de apoyo para la identificación de las unidades de aprendizaje y una visualización de las trayectorias académicas que podría tomar un estudiante a lo largo de su carrera.

Además, no se ha encontrado registro de ningún sistema (programa de software) utilizado para evaluar las trayectorias académicas de los estudiantes, o algún sistema que provea información útil y relevante sobre trayectorias que llevan a un estudiante a desertar o egresar, por lo que no se cuenta con el sistema para proveer conocimiento para la toma de decisiones.



## 4. Propuesta de solución

---

En este capítulo se presenta la propuesta de solución, así como la metodología para el desarrollo de esta tesis.

### *4.1 Propuesta*

Recordando el problema planteado el cual es el siguiente:

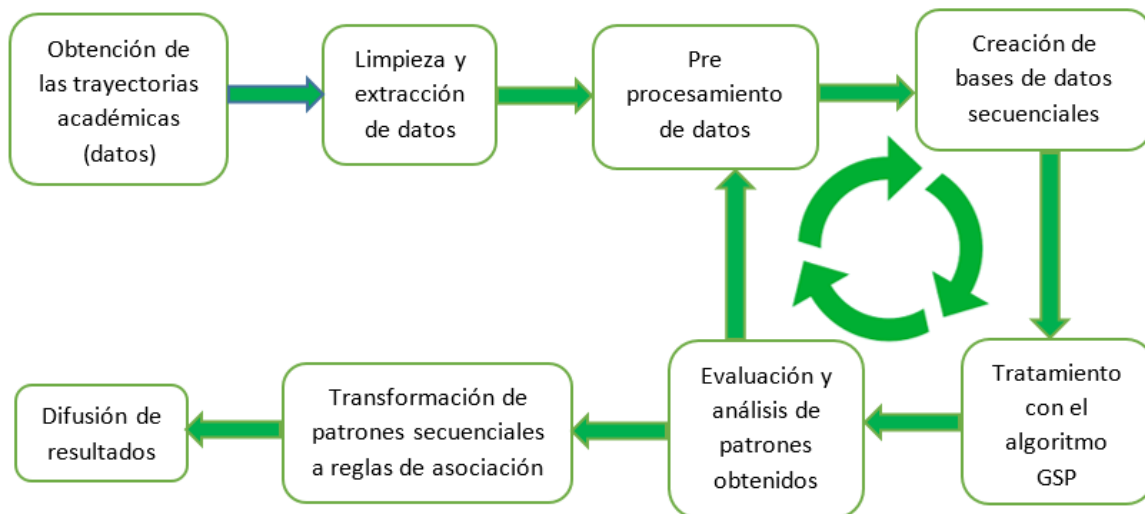
¿Cuáles son las trayectorias académicas comunes que distinguen a los estudiantes que egresan de los que desertan a partir de los historiales académicos de los estudiantes que permanecen estudiando, estudiantes que egresan y estudiantes que desertan de sus carreras en la UAPT de la UAEM, para ayudar en la detección temprana de la deserción?

La hipótesis planteada:

Si se aplica minería de datos secuenciales usando un algoritmo que encuentre trayectorias (secuencias) frecuentes en un conjunto de datos, a una base de datos de trayectorias académicas comunes, será posible identificar las trayectorias comunes que influyen en si un estudiante deserta o egresa de su carrera.

## 4.2 Método propuesto

A continuación, se presenta un diagrama con el método propuesto para esta investigación:

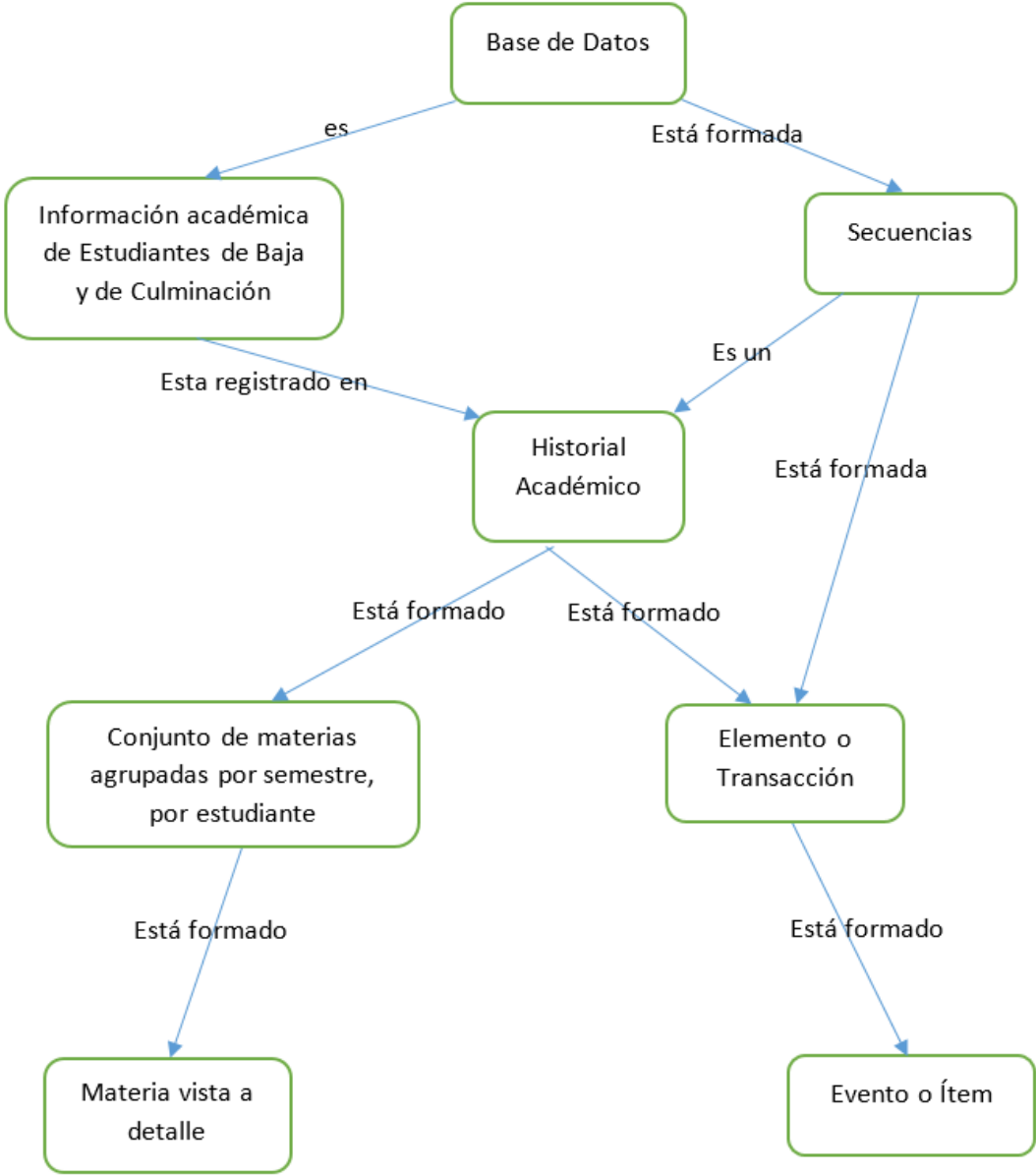


**Ilustración 6.** Diagrama de método propuesto.

En el diagrama se puede observar que hay un ciclo iterativo entre algunos puntos, esto se debe a la variante de los parámetros utilizados, a los resultados obtenidos y en la búsqueda de los resultados deseados.



Se realiza un proceso de transformación de los patrones secuenciales (previamente separados y agrupados) a reglas de asociación, en la que cada regla corresponde a una trayectoria que siguen los estudiantes. Para ver el proceso más detallado se muestra el siguiente mapa conceptual:



**Ilustración 7.** Esquema detallado de construcción de las Bases de Datos.

Retomando los conceptos de umbral mínimo de frecuencia, soporte, confianza y GAP, estos son parámetros que utiliza el algoritmo para obtener resultados semi-controlados, iterativos y acotados:

En la Minería de Patrones Secuenciales, una secuencia de datos se considera frecuente si ésta se encuentra en al menos un cierto número de registros a esto se le conoce como **umbral mínimo de frecuencia** [García, 2007].

El **soporte de la regla** es el porcentaje de transacciones en un conjunto que contienen la secuencia, además, determina cuanta frecuencia de la regla es aplicable en la transacción del conjunto [Liu, 2011].

La **confianza** determina la predictibilidad de la regla. Se le conoce como umbral de confianza a la frecuencia con que aparece una secuencia en un evento dentro de un registro de datos, dicho umbral hace mención a un mínimo número de veces que aparece [Liu, 2011].

El **GAP** es una restricción de separación máxima entre los eventos dentro de los cuales están las secuencias frecuentes [García, 2007].

Estos parámetros utilizados por el algoritmo GSP deben de estar en constante ajuste de acuerdo a los patrones secuenciales obtenidos, ya que, de otra forma, los resultados obtenidos a la hora de ejecución podrían ser nulos o en su caso entrar en un ciclo infinito de búsqueda, además de que el equipo de cómputo utilizado para tal proceso no contaría con los recursos necesarios para este, acorde a estos puntos expresados los resultados están delimitados y acotados a este conjunto de características.

### *4.3 Metodología*

El desarrollo de la investigación se realizó acorde a la metodología KDD presentada en el capítulo anterior, siguiendo paso a paso cada fase de KDD.

La metodología de KDD consta de varias etapas [Hernández, 2004]:

1. Integración y recopilación
2. Selección y limpieza
3. Minería de Datos
4. Evaluación e interpretación
5. Difusión y Uso

Para cada etapa se trabajó de la siguiente manera:

En la primera etapa se realiza una búsqueda para recopilación de historiales académicos de las primeras generaciones de la UAPT, que son el objeto de estudio. Además, de los trabajos relacionados a la deserción escolar en la UAPT y a los trabajos relacionados a la solución presentada.

En la segunda etapa se realiza una depuración de los historiales, obteniendo los datos necesarios para la manipulación de estos mismos, los programas para depuración y creación de bases de datos, así como los pre procesamientos y post procesamientos necesarios.

En la tercera etapa se realiza el desarrollo del método computacional especificado para resolver el problema planteado. Para esta fase se propone usar el algoritmo de minería de patrones secuenciales llamado Generalización de Patrones Secuenciales por sus siglas en

ingles GSP, además de desarrollar un método computacional para la transformación de patrones en reglas que permitan observar las trayectorias comunes.

En la cuarta etapa se realiza la interpretación de los datos obtenidos a partir de la tercera etapa por medio de los resultados obtenidos mediante gráficas, diagramas y tablas de resultados obtenidas de la experimentación.

En la quinta etapa se realiza la difusión de los datos más relevantes para poder darle un uso posterior por aquellos interesados.

A manera de conclusión, el alcance de la investigación, solo llega a la parte de encontrar conocimiento útil que pueda ser evaluado por los expertos y aplicado para la toma de decisiones.



## 5. Experimentación

---

En este capítulo se presenta como se trabajó en cada fase correspondiente a la metodología KDD, además de hacer mostrar el caso de estudio.

A continuación, se presentan las fases de KDD más desarrolladas y descritas, presentando todo lo que se ha hecho respecto a el caso de estudio.

### *5.1 Caso de Estudio*

En cuanto al caso de estudio, la investigación se centró en los historiales académicos de los estudiantes de la UAPT, perteneciente a la UAEM, se pretendió analizar y minar las trayectorias escolares de los estudiantes de la UAPT, comprendiendo un periodo de tiempo desde el 2008 hasta el 2015.

La agenda estadística del año 2015 presentada en la página oficial de la UAEM, muestra los siguientes resultados sobre la UAPT respecto a eficiencia terminal en la siguiente tabla [UAEM-AE, 2016]:

**Tabla 8.** Estadística de la UAP Tianguistenco.

<b>Nivel</b>	<b>1<sup>er</sup> ingreso</b>	<b>Matrícula 2014 – 2015</b>	<b>Egresados</b>	<b>Titulados / Graduados</b>
<b>Subtotal</b>	<b>319</b>	<b>1 246</b>	<b>180</b>	<b>48</b>
<b>Licenciatura</b>				
Ingeniería en Plásticos	58	232	36	11
Ingeniería en Producción Industrial	92	305	37	11
Ingeniería en Software	68	289	35	21
Licenciatura en Seguridad Ciudadana	101	420	72	3
<b>Maestría</b>				
Maestría en Ciencias de la Computación	<b>8</b>	<b>10</b>	<b>2</b>	<b>2</b>
<b>Total</b>	8	10	2	2

En la tabla anterior se muestra la cantidad de estudiantes que ingresaron, los estudiantes que egresaron y los estudiantes graduados y/o titulados, observando que, la cantidad de estudiantes que no concluyen con sus estudios de manera satisfactoria van decrecientando, en concreto las ingenierías poseen el menor número de estudiantes titulados.

Respecto a deserción escolar presenta los siguientes resultados [UAEM-AE, 2016]:

**Tabla 9.** Índices de deserción UAEM

Matrícula 2014 – 2015	Egresados 2014 – 2015	Nuevo ingreso 2014 – 2015	Matrícula 2015 – 2016	Índice de abandono escolar
1 247	180	319	1 246	11.2

El índice se calculó con la siguiente fórmula:

$$\text{Índice de abandono escolar} = 1 - \left( \frac{(\text{matricula total } n+1) - (\text{nuevo ingreso}^{\text{o}} n+1) + (\text{egresados } n)}{\text{matricula total } n} \right)$$

En la UAPT, en el periodo escolar 2014 – 2015 [UAEM-AE, 2016]:

- Matrícula 1, 247
- Egresados 180
- Abandono escolar 11.2%

En la UAPT, el año 2014 – 2015 la deserción fue de 11.2%, de la cual:

- Ing. en Plásticos: 13.6%
- Ing. en Producción Industrial: 10.1%
- Ing. en Software: 16.6%
- Lic. en Seguridad Ciudadana: 8.7%

Se puede observar que la carrera de Ingeniería de Software es la que tiene mayor índice de deserción, por lo que se presentaron más experimentos en sus bases de datos correspondientes.

Las trayectorias académicas de los estudiantes se encuentran en documentos digitales conocidos como historiales académicos, el cual contiene información detallada de cada trayectoria académica de un solo estudiante, como: número de cuenta, nombre del estudiante, plan de estudios, periodo de ingreso, último periodo, unidades de aprendizaje tomadas, calificaciones de estas unidades de aprendizaje, oportunidad en que pasaron las unidades de aprendizaje, entre otras.

## *5.2 Fase 1: Integración y recopilación*

En la presente fase se realizó la documentación que contiene información acerca de la recopilación, extracción de datos, la construcción de la base de datos, algunos ejemplos, etc.

Está redactado de tal forma que contenga todo lo que se ha visto en el análisis, diseño e implementación de los datos.

### **Conjunto de Datos**

Los datos que se ocupan para realizar la investigación, son los historiales académicos de los estudiantes que han egresado o que han desertado de sus respectivas carreras en la UAPT, perteneciente a la UAEM durante un tiempo determinado, en específico desde el año 2008 hasta el año 2015. El número total de los historiales académicos extraídos de control escolar fueron 2,240, cada generación con un aproximado de 300 estudiantes.

Los historiales académicos contienen información como:

- Datos de la **escuela**: nombre, clave de unidad, encabezados, entre otros.
- Datos del **estudiante**: nombre, número de cuenta, carrera.
- Datos de **control** escolar: cuando inicio la carrera, situación escolar, promedio general, promedio de último semestre, entre otros.
- Datos de las **unidades de aprendizaje**: nombres de las unidades de aprendizaje, calificaciones, periodo en que las tomó, entre otros.



Los datos a utilizar son los referentes a las unidades de aprendizaje, incluyendo las calificaciones, clave de la materia, semestre en que la cursaron, calificación y oportunidad en que la paso.

A continuación, se presenta un historial académico tomado aleatoriamente correspondiente a la carrera de Ing. de Plásticos, en el cual se observa a detalle los elementos que lo constituyen, así como la información que contiene.

Historial académico tomado al azar para fines demostrativos. Sobre cómo está constituido, la información que almacena digitalmente la institución y la que le es entregada al estudiante semestre a semestre, por parte de control escolar.

20 M

UNIVERSIDAD AUTONOMA DEL ESTADO DE MEXICO  
DIRECCION DE CONTROL ESCOLAR  
UNIDAD ACADEMICA PROFESIONAL TIANGUISTENCO  
TRAYECTORIA ACADEMICA

CLAVE : 0 . . . . 9  
NOMBRE :  
PLAN : LICENCIADO DE INGENIERO EN PLASTICOS - F1

Clave de acta	Per	NB	NU	Opt	Nombre de la materia	Cal.	Crđ	D
233-809-P2-1-12/06/2014	2014A	01	B	OBL	PROBABILIDAD Y ESTADISTICA	S/D	-	
233-809-P2-2-26/06/2014	2014A	01	B	OBL	PROBABILIDAD Y ESTADISTICA	S/D	-	
233-809-P2-3-08/07/2014	2014A	01	B	OBL	PROBABILIDAD Y ESTADISTICA	S/D	-	
233-900-PA-1-03/12/2013	2013B	01	B	OBL	ALGEBRA	070	8	A
233-902-P1-1-18/06/2014	2014A	01	B	OBL	CALCULO DIFERENCIAL E INTEGRAL	S/D	-	
233-902-P1-2-25/06/2014	2014A	01	B	OBL	CALCULO DIFERENCIAL E INTEGRAL	S/D	-	
233-902-P1-3-07/07/2014	2014A	01	B	OBL	CALCULO DIFERENCIAL E INTEGRAL	S/D	-	
233-906-P2-1-11/06/2014	2014A	01	B	OBL	ESTATICA	028		R
233-906-P2-2-23/06/2014	2014A	01	B	OBL	ESTATICA	S/D	-	
233-906-P2-3-30/06/2014	2014A	01	B	OBL	ESTATICA	S/D	-	
233-908-P1-1-06/12/2013	2013B	01	B	OBL	GEOMETRIA ANALITICA	060	8	A
233-909-P1-1-05/12/2013	2013B	01	B	OBL	LABORATORIO DE QUIMICA	063	2	A
233-910-P1-1-04/12/2013	2013B	01	B	OBL	QUIMICA GENERAL	020		R
233-910-P1-2-17/12/2013	2013B	01	B	OBL	QUIMICA GENERAL	S/D	-	
233-910-P1-3-13/01/2014	2013B	01	B	OBL	QUIMICA GENERAL	031		R
233-922-P2-1-18/06/2014	2014A	02	S	OBL	DIBUJO DE DETALLE	N/P	-	
233-922-P2-2-20/06/2014	2014A	02	S	OBL	DIBUJO DE DETALLE	S/D	-	
233-922-P2-3-07/07/2014	2014A	02	S	OBL	DIBUJO DE DETALLE	S/D	-	
233-924-PA-1-17/06/2014	2014A	02	S	OBL	HERRAMIENTAS COMPUTACIONALES	N/P	-	
233-924-PA-2-19/06/2014	2014A	02	S	OBL	HERRAMIENTAS COMPUTACIONALES	S/D	-	
233-924-PA-3-03/07/2014	2014A	02	S	OBL	HERRAMIENTAS COMPUTACIONALES	S/D	-	
233-939-PA-1-09/12/2013	2013B	02	S	OBL	PROGRAMACION BASICA	075	6	A
233-952-P1-1-10/12/2013	2013B	03	I	OBL	COMUNICACION Y RELACIONES HUMANAS	085	6	A
233-954-P2-1-17/06/2014	2014A	03	I	OBL	ETICA Y LEGISLACION	S/D	-	
233-954-P2-2-26/06/2014	2014A	03	I	OBL	ETICA Y LEGISLACION	S/D	-	
233-954-P2-3-01/07/2014	2014A	03	I	OBL	ETICA Y LEGISLACION	S/D	-	
233-961-P1-1-09/12/2013	2013B	03	I	OBL	METODOLOGIA DE LA INVESTIGACION	086	6	A

20 M

ASIGNATURAS ACREDITADAS	= 6
ASIGNATURAS ACREDITADAS OBLIGATORIAS	= 6
MESES DE ABANDONO	= 27
MESES DE PERMANENCIA	= 12
NUMERO DE ABANDONOS	= 1
NUMERO DE ASIGNATURAS EN SEGUNDO CURSO	= 7
NUMERO DE CREDITOS ACUMULADOS	= 36
NUMERO DE CREDITOS ACUMULADOS OBLIGATORIOS	= 36
NUMERO DE EXAMENES REPROBADOS	= 3
NUMERO DE INSCRIPCIONES	= 2
PORCENTAJE DE GRADO DE AVANCE	= 8.46
PORCENTAJE DE REZAGO POR TIEMPO DE PERMANENCIA	= 11.54
PRIMER PERIODO DE ESTUDIOS	= 2013B
PROMEDIO ANUAL	= 6.71
PROMEDIO DEL ULTIMO PERIODO	= N/A
PROMEDIO GENERAL	= 6.7
SITUACION DE REGULAR O IRREGULAR	= IRREGULAR
ULTIMO PERIODO DE ESTUDIOS	= 2014A

**Ilustración 8.** Historial académico de prueba, tomado aleatoriamente.

Existen estudiantes que cuentan con más de un historial académico con el mismo número de cuenta, por lo que se cambiaba el primer dígito del número de cuenta y se guardaba el nombre del archivo con el nuevo número. El que tengan más de un historial académico

significa que el estudiante fue dado de baja de la carrera de su primera elección, posteriormente entró a otra carrera en el mismo espacio académico (la UAPT), teniendo de entre 2 y 4 historiales académicos como máximo.

### *5.3 Fase 2: Selección y limpieza*

En esta fase se realizó la extracción y recopilación de datos, que posteriormente se utilizarían para realizar las pruebas.

Los datos a utilizar, historiales, se obtuvieron de la página de control escolar de la UAEM, de la UAPT. Los historiales se dividieron en 2 carpetas:

- **Culminación:** contienen los historiales de los estudiantes que han concluido su carrera, ya sea en el tiempo programado (5 años) o en un tiempo posterior, además incluye los historiales de los estudiantes que siguen estudiando pero que aún no egresan, también contiene los historiales de los estudiantes que no se inscriben un semestre o dos pero que siguen estudiando al siguiente.
- **Baja:** contiene los historiales de los estudiantes que no concluyen sus estudios en su carrera, aquellos estudiantes que son dados de baja por reprobado más exámenes de los que podía, los estudiantes que son dados de baja por no inscribirse en un tiempo determinado, estudiantes que piden su baja voluntaria, estudiantes que no pasan alguna(s) materia(s) y se les acaban las oportunidades de pasarla, etc.

Los historiales obtenidos comprenden los periodos de entre 2008B y 2015B, ya que el plan curricular cambió para la generación que ingreso en 2016B, es decir, el análisis se realizó para 8 generaciones de estudiantes. Con un total de fueron 2,240, más los historiales que se eliminaron ya que no tenían información de unidades de aprendizaje o correspondían a estudiantes de maestrías.

### 5.3.1 Construcción de la base de datos

Para la construcción de la Base de Datos (BD) se utilizó información encontrada dentro de los historiales, para ello se realiza una depuración del historial original, obteniendo la información que nos interesa.

En la siguiente ilustración se muestra la información obtenida (datos depurados) a partir de los historiales académicos depurados:

233-809-P2-1-12/06/2014	2014A	01	B	OBL	PROBABILIDAD Y ESTADISTICA	S/D	-
233-809-P2-2-26/06/2014	2014A	01	B	OBL	PROBABILIDAD Y ESTADISTICA	S/D	-
233-809-P2-3-08/07/2014	2014A	01	B	OBL	PROBABILIDAD Y ESTADISTICA	S/D	-
233-900-PA-1-03/12/2013	2013B	01	B	OBL	ALGEBRA	070	8 A
233-902-P1-1-18/06/2014	2014A	01	B	OBL	CALCULO DIFERENCIAL E INTEGRAL	S/D	-
233-902-P1-2-25/06/2014	2014A	01	B	OBL	CALCULO DIFERENCIAL E INTEGRAL	S/D	-
233-902-P1-3-07/07/2014	2014A	01	B	OBL	CALCULO DIFERENCIAL E INTEGRAL	S/D	-
233-906-P2-1-11/06/2014	2014A	01	B	OBL	ESTATICA	028	R
233-906-P2-2-23/06/2014	2014A	01	B	OBL	ESTATICA	S/D	-
233-906-P2-3-30/06/2014	2014A	01	B	OBL	ESTATICA	S/D	-
233-908-P1-1-06/12/2013	2013B	01	B	OBL	GEOMETRIA ANALITICA	060	8 A
233-909-P1-1-05/12/2013	2013B	01	B	OBL	LABORATORIO DE QUIMICA	063	2 A
233-910-P1-1-04/12/2013	2013B	01	B	OBL	QUIMICA GENERAL	020	R
233-910-P1-2-17/12/2013	2013B	01	B	OBL	QUIMICA GENERAL	S/D	-
233-910-P1-3-13/01/2014	2013B	01	B	OBL	QUIMICA GENERAL	031	R
233-922-P2-1-18/06/2014	2014A	02	S	OBL	DIBUJO DE DETALLE	N/P	-
233-922-P2-2-20/06/2014	2014A	02	S	OBL	DIBUJO DE DETALLE	S/D	-
233-922-P2-3-07/07/2014	2014A	02	S	OBL	DIBUJO DE DETALLE	S/D	-
233-924-PA-1-17/06/2014	2014A	02	S	OBL	HERRAMIENTAS COMPUTACIONALES	N/P	-
233-924-PA-2-19/06/2014	2014A	02	S	OBL	HERRAMIENTAS COMPUTACIONALES	S/D	-
233-924-PA-3-03/07/2014	2014A	02	S	OBL	HERRAMIENTAS COMPUTACIONALES	S/D	-
233-939-PA-1-09/12/2013	2013B	02	S	OBL	PROGRAMACION BASICA	075	6 A
233-952-P1-1-10/12/2013	2013B	03	I	OBL	COMUNICACION Y RELACIONES HUMANAS	085	6 A
233-954-P2-1-17/06/2014	2014A	03	I	OBL	ETICA Y LEGISLACION	S/D	-
233-954-P2-2-26/06/2014	2014A	03	I	OBL	ETICA Y LEGISLACION	S/D	-
233-954-P2-3-01/07/2014	2014A	03	I	OBL	ETICA Y LEGISLACION	S/D	-
233-961-P1-1-09/12/2013	2013B	03	I	OBL	METODOLOGIA DE LA INVESTIGACION	086	6 A

**Ilustración 9.** Datos de depuración del historial académico.

Los datos que se muestran en la ilustración fueron tomados como base para crear las variaciones de las BD.

El siguiente es un fragmento original del historial depurado (línea 4 de la figura anterior):

**233-900-PA-1-03/12/2013 2013B 01 B OBL ALGEBRA 070 6 A**

En la siguiente tabla se muestra detalladamente los datos contenidos del fragmento tomado:

**Tabla 10.** Composición del fragmento del historial.

<b>Fragmento</b>	<b>Indicativo</b>
<b>233</b>	Clave del centro educativo donde curso y/o aprobó la materia.
<b>900</b>	Clave única de la materia.
<b>PA</b>	Clave del grupo donde tomó la materia.
<b>1</b>	Indicador de la oportunidad en que aprobó, reprobó o tomó la materia. 1 = Ordinario, 2 = Extraordinario, 3 = Título de suficiencia.
<b>03/12/2013</b>	Fecha en la que se publicó la calificación de la materia.
<b>2013B</b>	Periodo en el que se tomó la materia.
<b>01</b>	Núcleo base
<b>B</b>	Núcleo base
<b>OBL</b>	Núcleo obligatorio
<b>ALGEBRA</b>	Nombre de la materia tomada.
<b>070</b>	Calificación obtenida.
<b>6</b>	Créditos de la materia.
<b>A</b>	Indicador de si aprobó (A), reprobó (R) o no presentó (-).

### *5.3.2 Inconvenientes para la construcción:*

Los inconvenientes que se tuvieron para la construcción de las bases de datos fueron las siguientes:

- La extracción de los historiales fue tardada debido al sistema utilizado por control escolar para la extracción por lotes, se realizó de uno a uno.

- Algunos historiales no contienen nada de información académica debido a que los estudiantes se inscribieron en la carrera, pero no tomaron ninguna materia.
- En algunos casos el nombre del archivo (número de cuenta del estudiante) no coincidía con el número de cuenta dentro del historial, por lo que se realizó una verificación automática para posteriormente renombrarlos uno por uno.
- Ya que existen algunos estudiantes con más de un historial, se modificaron los nombres de los archivos, siendo que aparecían diferentes en la verificación de números de cuenta, en estos casos especiales, los nombres se dejaron como estaban originalmente después de descargarlos.
- Los historiales descargados originales, tenían distinta codificación, algunos tenían MS-DOS y la mayoría UTF-8, por lo que algunos caracteres no se encontraban o aparecían como extraños, debido a esto al momento de tratarlos se cambiaron a formato UTF-8, además de reemplazar los caracteres extraños con los que deberían ser, los que aparecen en los historiales originales.
- Al principio se trató de cambiar el nombre de la materia por la clave de esta, sin embargo, no contábamos con la lista de claves de las unidades de aprendizaje, por lo que la obtuvimos de los informes originales, posteriormente al hacer el cambio, notamos que algunos nombres de unidades de aprendizaje tenían algunos signos de más (\*, "", .), por lo que se realizó el cambio de esa manera.
- Se realizó un nuevo análisis del historial con la información obtenida hasta ahora y se encontró que, en el apartado de la información a utilizar, que venían las claves de las unidades de aprendizaje, por lo que se utilizaron estas para la construcción, sin embargo, no se cuenta con la lista de claves de las unidades de aprendizaje y sus respectivos nombres.

Para la creación de cada base de datos se realizó un pre procesamiento de datos, limpiando los datos extra que no aportaban información útil para su implementación.

### 5.4 Fase 3: Minería de datos

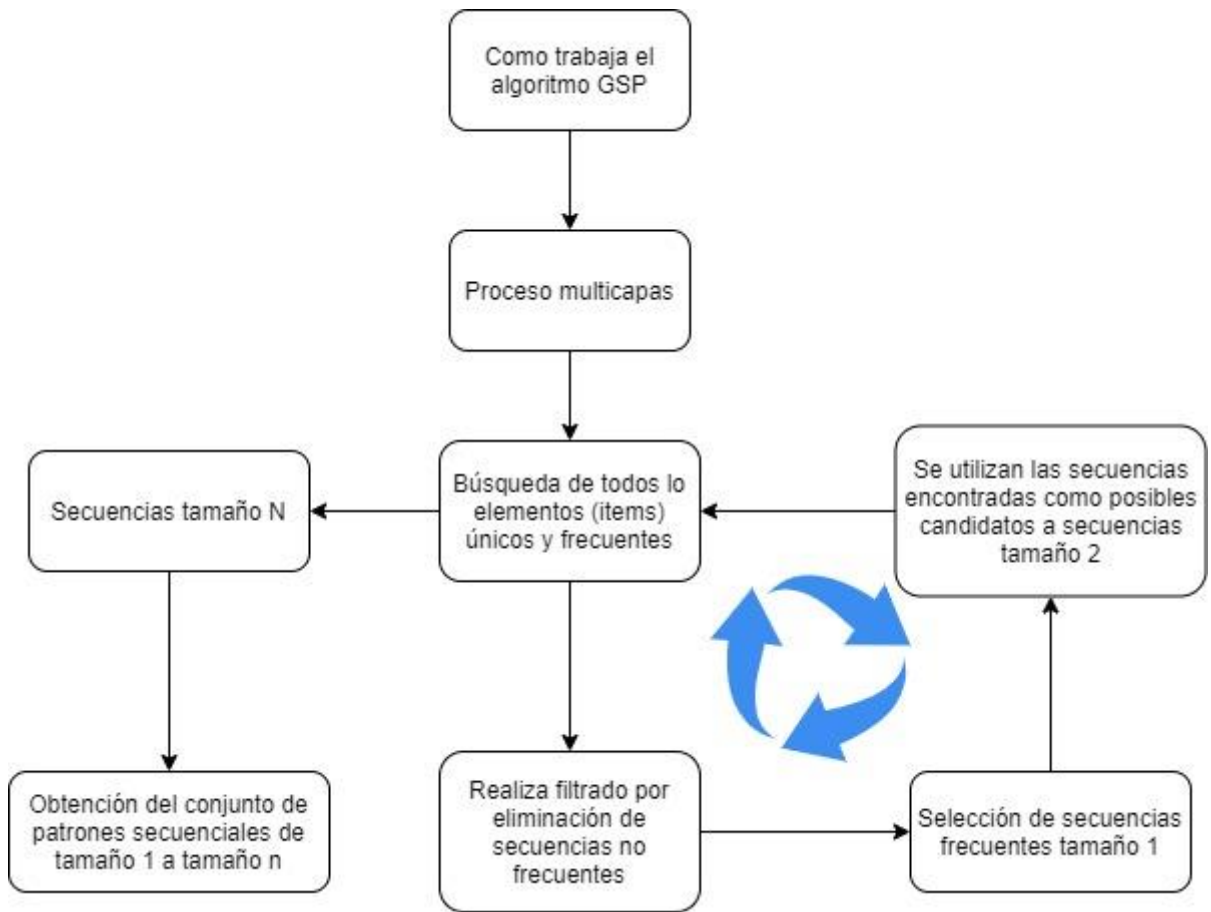
Para esta fase el funcionamiento de cómo se trabajó en esta fase, se muestra en el siguiente diagrama:



**Ilustración 10.** Trabajo fase 3.

El algoritmo GSP con el que se trabaja indica que su base de datos debe contener por línea la totalidad de un archivo dividido entre paréntesis de acuerdo al tiempo en que ocurre el suceso y este a su vez debe estar subdividido por sus elementos que contiene el suceso y separados por comas.

Retomando el diagrama del marco teórico del funcionamiento del algoritmo GSP, se observa en general el proceso de desarrollo del algoritmo, observando como se pasa de fase a fase hasta encontrar los resultados deseados:



**Ilustración 11.** Funcionamiento del algoritmo GSP.

Tomando como guía el diagrama se describen los siguientes pasos correspondientes a la construcción de las bases de datos, tanto generales como específicas, divididas por categorización.

Observando que puede haber un número creciente de iteraciones que recorren el ciclo de búsqueda e identificación de patrones secuenciales que, al ser identificados en su totalidad, pasan a una fase de transformación.



Dado a lo anterior, se construye de la siguiente forma la base de datos:

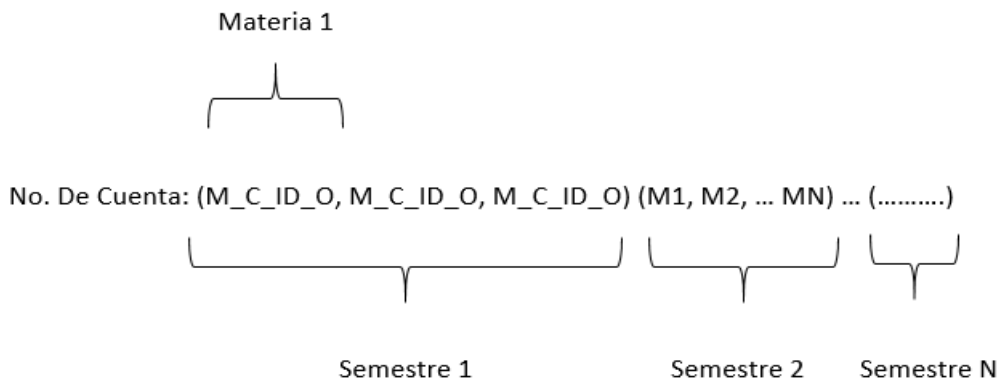
**Tabla 11.** Elementos de la Base de Datos.

<b>Elemento</b>	<b>Significado</b>
Clave de materia (M)	Número de 3 dígitos irrepitible para cada materia
Calificación de materia (C)	Número de 3 dígitos (100, 075, 060, S/D, N/P)
Identificador (ID)	Identificador de aprobado (A), reprobado (R), baja o cuando aparece guion (B)
Oportunidad (O)	Oportunidad en que la tomó: Ordinario (O), Extraordinario (E) o Título (T)

Los datos presentados en la tabla son los que se utilizaron para dar formato a las bases de datos antes de pre procesamiento, fase anterior a la minería, además, de dar el formato a los resultados obtenidos, con lo que se construyó un formato, el cual se modificó acorde a los análisis de los experimentos realizados, es decir, que algunos datos se actualizaron o modificaron acorde al requerimiento de cada experimentación.

**Formato general conceptual:**

En la ilustración se puede observar la notación sobre cómo se construye la base de datos, a la que con el orden de secuencias se puede aplicar el algoritmo.



**Ilustración 12.** Formato de Base de Datos.

Aplicando los datos obtenidos de los historiales con el anterior formato, se aplica de la siguiente forma (fragmento de un historial transformado):

0011085:

(900\_S/D\_B\_O,900\_N/P\_B\_E,900\_N/P\_B\_T,906\_N/P\_B\_O,906\_N/P\_B\_E,906\_N/P\_B\_T,908\_N/P\_B\_O,908\_N/P\_B\_E,908\_N/P\_B\_T,909\_N/P\_B\_O,909\_N/P\_B\_E,909\_N/P\_B\_T,910\_N/P\_B\_O,910\_N/P\_B\_E,910\_N/P\_B\_T,939\_N/P\_B\_O,939\_N/P\_B\_E,939\_N/P\_B\_T,952\_N/P\_B\_O,952\_N/P\_B\_E,952\_N/P\_B\_T)

(906\_083\_A\_O)

(900\_N/P\_B\_O,900\_N/P\_B\_E,900\_075\_A\_T,908\_N/P\_B\_O,908\_080\_A\_E,909\_079\_A\_O,910\_060\_A\_O,922\_081\_A\_O,939\_087\_A\_O,952\_089\_A\_O)

(902\_S/D\_B\_O,902\_005\_R\_E,902\_N/P\_B\_T,912\_S/D\_B\_O,912\_S/D\_B\_E,912\_S/D\_B\_T,913\_048\_R\_O,913\_N/P\_B\_E,913\_N/P\_B\_T,931\_S/D\_B\_O,931\_S/D\_B\_E,931\_S/D\_B\_T,954\_065\_A\_O,961\_N/P\_B\_O,961\_N/P\_B\_E,961\_N/P\_B\_T)

(809\_S/D\_B\_O,809\_N/P\_B\_E,809\_S/D\_B\_T,902\_N/P\_B\_O,902\_S/D\_B\_E,902\_024\_R\_T,926\_S/D\_B\_O,926\_S/D\_B\_E,926\_S/D\_B\_T,933\_S/D\_B\_O,933\_S/D\_B\_E,933\_S/D\_B\_T,940\_S/D\_B\_O,940\_S/D\_B\_E,940\_S/D\_B\_T,962\_S/D\_B\_O,962\_S/D\_B\_E,962\_S/D\_B\_T)

Este ejemplo es la información obtenida de un solo historial académico, por lo que hay muchos más, ya que se descargaron 8 generaciones de aproximadamente 300 historiales por cada una.

Por cada generación se realizó una o más bases de datos. Se organizaron y clasificaron de acuerdo a los siguientes criterios. Algunos historiales estaban vacíos en cuestión académica, es decir, que los estudiantes se inscribieron a la carrera pero no asistieron a ninguna clase.

La experimentación se realizó con bases de datos generales y específicas, se dividió de la siguiente forma:

- Por generación
- Por generación en 2 culminación y dado de baja
- General de culminación.
- General de datos de baja

#### *5.4.1 Método de transformación*

Para la creación del método de transformación se tomaron los patrones secuenciales obtenidos y se transformaron a la misma notación que tiene la base de datos. De esta forma se pueden transformar a reglas los datos obtenidos por el programa se almacenan en un archivo de texto, posteriormente se transforman a una notación más entendible que es similar a la que se encuentra en los historiales académicos.

El último paso es transformar los archivos de texto a la notación de reglas de asociación. Para lo cual los datos ya procesados se almacenan en un archivo de Excel para realizar un reemplazo de datos que da espaciado y formato.

Al transcurrir los pasos mencionados se obtienen reglas de asociación perfectamente entendibles por cualquier persona.

## *5.5 Fase 4: Evaluación e Interpretación*

En esta fase se realizó la evaluación de cada experimentación, para ello se dividió en varias fases: se realizó un análisis previo a la experimentación, un análisis posterior después de la experimentación, una evaluación de resultados, una interpretación de patrones obtenidos.

### *5.5.1 Experimentación*

En este trabajo se realizaron 4 **normalizaciones** (ediciones de las bases de datos que van generalizando los datos obtenidos) para la experimentación, de las cuales se obtuvieron resultados y si no eran los que se buscaban se iteró otra vez para obtener el resultado deseado.

Para cada experimentación realizada la llamaremos **normalización** de acuerdo a cada iteración que se hace siguiendo el método y la metodología, cada normalización corresponde a una vuelta siguiendo desde la fase 2 de selección y limpieza hasta la fase 4 de evaluación e interpretación.

#### *Normalización 0. Prueba general, base de datos completa*

En esta Normalización se optó por crear la base de datos general, en la que se incluye la información de todos los estudiantes que fueron dados de baja, los que culminaron y los que aún seguían estudiando, la información de todas las carreras, es decir, es una base de datos general que se incluye toda la información anterior.

La notación de la base de datos era la siguiente:

0011085:

(900\_S/D\_Baja\_O,900\_N/P\_Baja\_E,900\_N/P\_Baja\_T,906\_N/P\_Baja\_O,906\_N/P\_Baja\_E,906\_N/P\_Baja\_T,908\_N/P\_Baja\_O,908\_N/P\_Baja\_E,908\_N/P\_Baja\_T,909\_N/P\_Baja\_O,909\_N/P\_Baja\_E,909\_N/P\_Baja\_T,910\_N/P\_B\_O,910\_N/P\_Baja\_E,910\_N/P\_Baja\_T,939\_N/P\_Baja\_O,939\_N/P\_B\_E,939\_N/P\_B\_T,952\_N/P\_B\_O,952\_N/P\_Baja\_E,952\_N/P\_Baja\_T) (906\_083\_Aprobo\_O)

En el cual la notación era la del formato general.

Se realizaron las bases de datos así para poder conocer los patrones que arrojaba el algoritmo con datos reales, también se hizo para conocer algunos patrones y evaluar al menos uno para tener el dato en comparación a la siguiente experimentación.

Fue la primera vuelta del ciclo perteneciente a KDD. Posteriormente se realizó un análisis para mejorar la siguiente normalización.

Los resultados obtenidos fueron 376 patrones obtenidos, de esos patrones 340 de tamaño 1 y 36 de tamaño 2, lo que indica que solo obtenía una materia por patrón y en muy pocas dos unidades de aprendizaje ya que, en otros patrones se repetía la misma regla con diferente calificación y oportunidad.

Se observó que los patrones obtenidos mostraban una relación de las unidades de aprendizaje, las calificaciones y la oportunidad presentada, sin embargo, era muy difícil saber cuáles de estos patrones eran aplicables para cada situación escolar de un estudiante.

*Normalización 1. Prueba general con síntesis de nota, base de datos dividida por semestre y condición académica*

En esta Normalización se optó por crear las bases de datos por generación, en la que se incluyen los estudiantes que fueron dados de baja, los que culminaron y los que aún siguen estudiando, así como también contiene todas las carreras.

La notación de las bases de datos fue la siguiente:

0011085:

(900\_S/D\_B\_O,900\_N/P\_B\_E,900\_N/P\_B\_T,906\_N/P\_B\_O,906\_N/P\_B\_E,906\_N/P\_B\_T,908\_N/P\_B\_O,908\_N/P\_B\_E,908\_N/P\_B\_T,909\_N/P\_B\_O,909\_N/P\_B\_E,909\_N/P\_B\_T,910\_N/P\_B\_O,910\_N/P\_B\_E,910\_N/P\_B\_T,939\_N/P\_B\_O,939\_N/P\_B\_E,939\_N/P\_B\_T,952\_N/P\_B\_O,952\_N/P\_B\_E,952\_N/P\_B\_T)  
(906\_083\_A\_O)

Utilizando la notación en formato general, quitando el prefijo correspondiente al semestre.

En esta normalización las bases de datos se hicieron muy específicas dividiéndose de la siguiente manera:

Por generación se hicieron:

- BD de baja: incluye todos los estudiantes de todas las carreras que fueron dados de baja ya sea voluntaria o académicamente.
- BD de baja académica: incluye todos los estudiantes de todas las carreras que fueron dados de baja académicamente.
- BD de baja voluntaria: incluye todos los estudiantes de todas las carreras que fueron dados de baja voluntariamente.
- BD de siguen estudiando: incluye todos los estudiantes de todas las carreras que siguen estudiando en sus respectivas carreras, es decir, que aún no egresan ni han sido dados de baja.
- BD de culminación sus estudios: incluye todos los estudiantes de todas las carreras que egresaron de sus carreras y concluyeron sus estudios de licenciatura.

- BD de culminación: incluye todos los estudiantes de todas las carreras que egresaron y de aquellos que siguen estudiando.

Se realizó una separación para las 4 normalizaciones:

- En estas bases de datos se incluyeron todos los estudiantes solo separados por baja o culminación, con lo que la base de datos usada tenía un mayor número de registros.

Se realizaron las bases de datos así para poder conocer los patrones que arrojaba el algoritmo con datos muy específicos por generación separados por su estatus académico, además al poner todos los estudiantes de todas las carreras para tener un mayor número de registros que tiene que analizar el algoritmo, este número incremento al hacerlo para las 4 primeras generaciones para ampliar la búsqueda del algoritmo ya que estas ya contienen estudiantes egresados y por su puesto estudiantes dados de baja.

Fue la segunda vuelta del ciclo perteneciente a KDD para las bases de datos por generación.

Fue la tercera vuelta del ciclo perteneciente a KDD para las bases de datos de las 4 generaciones.

Los resultados obtenidos fueron:

- Para las bases de datos de las 4 generaciones se obtuvieron un total de 179 patrones, de los cuales 111 de tamaño 1 y 68 de tamaño 2.
- Para las bases de datos específicas:
  - Baja: 48967 patrones obtenidos, todos de tamaño 1.
  - Baja académica: 17891 patrones obtenidos, de los cuales 13 de tamaño 4, 26 de tamaño 3, 52 de tamaño 2 y 17800 de tamaño 1.
  - Baja voluntaria: 16802 patrones obtenidos, todos de tamaño 1

- Siguen estudiando: 1190 patrones obtenidos, de los cuales 24 de tamaño 5, 126 de tamaño 4, 162 de tamaño 3, 117 de tamaño 2 y 761 de tamaño 1.
- Culminación estudios: 234 patrones obtenidos, de los cuales 59 de tamaño 2 y 175 de tamaño 1
- Culminación: 23078 patrones obtenidos, de los cuales 191 de tamaño 2 y 22887 de tamaño 1.

Los patrones obtenidos mostraban que cada patrón correspondía a una trayectoria de un estudiante o a más trayectorias comunes de varios estudiantes. Posteriormente se realizó un análisis para mejorar la siguiente normalización, con lo que se observó que los patrones eran muy pequeños para determinar cómo es una trayectoria académica de un estudiante de baja o de culminación.

### *Normalización 2. Prueba de asignación de periodos, base de datos con división de semestres*

En esta Normalización se optó por crear las bases de datos incluyendo la información de todas las carreras, en la que se incluyen los estudiantes que fueron dados de baja, los que culminaron y los que aún siguen estudiando.

La notación era la siguiente:

0011085:(2008B\_L40900\_S/D\_O)(2008B\_L40900\_N/P\_E)(2008B\_L40900\_N/P\_T)(2008B\_L40902\_N/P\_O)(2008B\_L40902\_N/P\_E)(2008B\_L40902\_N/P\_T)(2008B\_L40903\_N/P\_O)(2008B\_L40903\_N/P\_E)(2008B\_L40903\_N/P\_T)

En el cual la notación era: Semestre\_ clave de materia\_ calificación\_ oportunidad.

Para la creación de bases de datos se realizaron 3 iteraciones en esta normalización.



- Para la creación de las bases de datos se dividieron por generación, posteriormente por baja y culminación para la primera iteración, incluyendo la información de todas las carreras.
- Para la segunda iteración se utilizó la información de los historiales de la carrera de software únicamente, se dividieron por generación, después por baja y culminación.
- Para la tercera iteración se utilizó la información de los historiales de la carrera de software únicamente, se dividieron por generación, después por baja y culminación, sin embargo, en esta iteración el prefijo correspondiente al semestre se recorrió al final de la notación.

Las creaciones de estas bases de datos para cada una de las iteraciones de esta normalización se diseñaron para comparar los resultados de cada una, así como para obtener un análisis para la siguiente normalización.

Fue la cuarta, quinta y sexta vuelta del ciclo perteneciente a KDD por cada iteración.

Los resultados obtenidos fueron:

- Iteración 1: 105 patrones obtenidos, de los cuales 3 de tamaño 4, 9 de tamaño 3, 20 de tamaño 2 y 74 de tamaño 1.
- Iteración 2: 219 patrones obtenidos, de los cuales 2 de tamaño 4, 14 de tamaño 3, 47 de tamaño 2 y 156 de tamaño 1.
- Iteración 3: 1124 patrones obtenidos, de los cuales 1 de tamaño 8, 12 de tamaño 7, 53 de tamaño 6, 128 de tamaño 5, 210 de tamaño 4, 274 de tamaño 3 y 292 de tamaño 2.

Los patrones obtenidos eran de difícil clasificación para saber a cuál carrera pertenecían ya que las claves de cada materia estaban recortadas en las historias académicas de los estudiantes de las cuales se obtuvieron las bases de datos, por lo que se realizaron nuevas bases de datos completando posterior y automáticamente después de obtenidos los patrones.

Se observó que los patrones obtenidos ya mostraban como es la trayectoria de un estudiante que deserta o culmina la carrera, sin embargo, se necesitaba realizar una validación de claves de unidades de aprendizaje para la siguiente normalización.

### *Normalización 3. Síntesis recopilatoria de resultados anteriores, base de datos con agrupación resultados*

En esta Normalización se optó por crear las bases de datos más específicas para demostrar la funcionalidad y veracidad de los patrones secuenciales, para ello se incluyó solo la información de los historiales de la carrera de software, en la que se incluyen los estudiantes que fueron dados de baja, los que, si culminaron y los que aún siguen estudiando.

Fue la última normalización, cuya notación es la siguiente:

0012088:(L40800\_O\_A)(L40801\_O\_R)(L40801\_E\_R)(L40801\_T\_A)(L40803\_O\_A)(L40804\_O\_A)(L40802\_O\_B)(L40802\_E\_B)(L40802\_T\_B)(L40806\_O\_A)(L40812\_O\_A)(L40807\_O\_A)(L40808\_O\_A)(L40802\_O\_B)(L40802\_E\_B)

En la cual la notación es: Clave de materia\_ Oportunidad\_ Identificador

Para la creación de la base de datos se utilizó los historiales académicos de la carrera de software separados por generación, por baja y culminación, también se hicieron 2 generales de baja y culminación.

Séptima vuelta del ciclo de KDD.

Ya que se consideró que estos patrones obtenidos responden a la hipótesis planteada, se presentan los resultados más detallados. Los patrones obtenidos se transformaron con un método computacional para realizar la identificación, los patrones se cambiaron a la notación de la base de datos entrada, posteriormente se realizó un cambio

Se muestra el total de las reglas obtenidas en la siguiente tabla:

**Tabla 12.** Registro total de reglas obtenidas.

<b>Tamaño</b>	<b>Baja</b>	<b>Culminación</b>	<b>Total encontradas</b>
2	278	497	775
3	469	1473	1942
4	552	5011	5563
5	462	17810	18272
6	251	64223	64474
7	77	227771	227848
8	12	0	12
9	1	0	12
<b>Total reglas</b>	2102	316785	318898

Ya que este registro de reglas obtenidas aún no se encuentra depurada, las reglas depuradas, eliminación de reglas que aparecían en ambos extremos (Baja y Culminación) se presentan en la siguiente tabla:

**Tabla 13.** Registro de reglas depuradas.

<b>Tamaño</b>	<b>Baja</b>	<b>Culminación</b>	<b>Total diferenciables</b>	<b>No diferenciables</b>
---------------	-------------	--------------------	-----------------------------	--------------------------

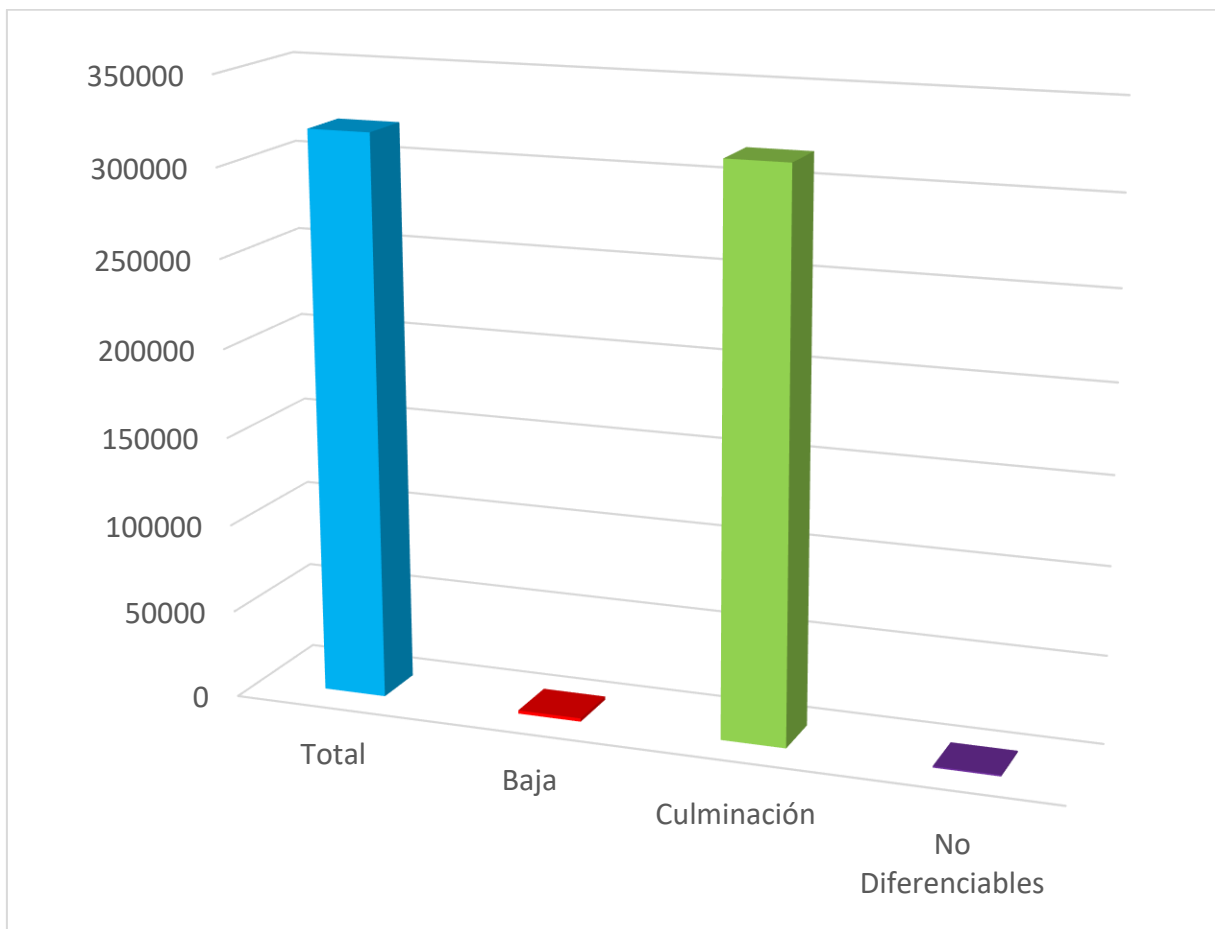
2	180	399	579	196
3	391	1395	1786	547
4	515	4974	5489	74
5	454	17803	18257	15
6	251	64223	64764	0
7	77	227771	227848	0
8	12	0	12	0
9	1	0	1	0
<b>Total reglas</b>	1881	316565	318736	832

Como se puede observar las reglas que están presentes en ambos casos y fueron depuradas, son perfectamente diferenciables, además, conforme aumenta el número de tamaño de las reglas, el número de reglas depuradas disminuye.

A manera de conclusión de los resultados obtenidos, se presenta la siguiente tabla y gráfica de datos:

**Tabla 14.** Número de reglas separadas.

<b>Total</b>	<b>Baja</b>	<b>Culminación</b>	<b>No diferenciables</b>
<b>318,898</b>	1,881	316,565	832



**Ilustración 13.** Gráfica de resultados obtenidos.

Como se puede observar en la ilustración anterior que la cantidad de reglas de baja es mucho menor en comparación a las reglas de culminación, esto se debe a que se tiene un número mayor de historiales de estudiantes que egresaron y culminaron, que de estudiantes que fueron dados de baja.

En cuanto a la cantidad de reglas no diferenciables se observa de igual manera corresponden a una mínima, siendo estas menores a la cantidad de reglas de baja.

### *5.5.2 Evaluación*

Se realizaron experimentaciones con las bases de datos generadas, se realizaron por separado, es decir, se hicieron bases de datos más particulares, como por generación, por estudiantes dados de baja y por aquellos que terminan.

Al evaluar un par de patrones secuenciales resultantes en contra parte con algunos historiales escogidos al azar se encontró que algunas unidades de aprendizaje de diferentes carreras tenían la misma clave, es decir, que la clave de una materia era la misma que en otra, siendo estas distintas y de diferentes carreras, sin embargo, esto no afecta a la experimentación, ni a los resultados evaluados.

Las primeras evaluaciones de las Normalizaciones 0 a 2 solo se realizaron mediante un análisis de secuencias contrastando con los datos de 10 historiales académicos seleccionados al azar.

Retomando la tabla de experimentación que contiene todos los patrones secuenciales obtenidos se realizó un proceso de Clasificación con el cual se dividieron en 2 clases: Baja y Culminación respectivamente, para eliminar los patrones que estaban en ambos casos y que no son representativas, se puede observar en las tablas anteriores (tabla 12 y tabla 13), correspondientes a la Experimentación de la Normalización 3.

Se realizó un proceso de transformación de los patrones secuenciales a reglas de asociación, en la que cada regla corresponde a una trayectoria que siguen los estudiantes, estas pueden variar dependiendo del tamaño de cada regla.

Con las reglas de asociación se puede describir más a detalle cómo es la trayectoria académica común que se repite más en los historiales académicos de los estudiantes que son dados de Baja o que Culminan su carrera universitaria.

La evaluación final con las reglas mejor descritas se realizó de la siguiente forma:

- Obtención de 10 historiales académicos de culminación tomados aleatoriamente
- Obtención de 10 historiales académicos de baja tomados aleatoriamente
- Comparación de las reglas obtenidas de la experimentación por tamaño (tamaño 2 a tamaño 9) respecto de los historiales académicos
- Comparación de reglas de baja respecto a historiales académicos de baja
- Comparación de reglas de baja respecto a historiales académicos de culminación
- Comparación de reglas de culminación respecto a historiales académicos de culminación
- Comparación de reglas de culminación respecto a historiales académicos de baja
- Realización de tabla comparativa de resultados en cuanto a número de reglas obtenidas, sumatoria de la confianza de las reglas obtenidas en el historial académico, sumatoria de las reglas obtenidas en el historial académico respecto al 100% de reglas obtenidas en general.
- Clasificación de los historiales en dos clases: culminación o baja.
- Realización de matriz de confusión.

La evaluación mostró que las reglas de baja y culminación son perfectamente diferenciables entre sí, ya que las reglas no aparecen en diferente clase, como se observa en la tabla 13. Las reglas obtenidas pueden describir las trayectorias que toma un estudiante para que egrese o deserte de su carrera.

Ahora se muestra el listado de reglas para la normalización 3, con la base de datos separada por situación académica, con el agrupamiento de notas por materia, por oportunidad en que aprobaron o causaron baja. Corresponden a los resultados que se buscaban, por lo que se realizó una clasificación y agrupación.

Se presentan las reglas de mayor confianza y mayor frecuencia encontradas desde tamaño 2 hasta tamaño 9, correspondientes a baja y culminación.

### 5.5.3 Listado de reglas con mayor confianza

A continuación, se muestran las primeras 10 reglas de mayor confianza ordenadas por tamaños. Tablas de las 10 reglas de mayor confianza obtenidas diferenciables para **Baja** y para **Culminación**, ordenadas por tamaño. Además, la ilustración representativa de cada tabla para una mejor apreciación visual.

Nota: las unidades de aprendizaje correspondientes al núcleo básico, al núcleo sustantivo y al núcleo integral, son obligatorias a cursar y acreditar.

Las unidades de aprendizaje de tronco común para los estudiantes son las unidades de aprendizaje de núcleo básico y las unidades de aprendizaje que están seriadas, es decir, unidades de aprendizaje que se deben aprobar y que sirven para tener el conocimiento necesario antes de tomar la materia que le sigue en cuanto a dificultad de conocimiento.

A continuación, se presentan las reglas de culminación de tamaño 2:

**Tabla 15.** Reglas de culminación tamaño 2.

Frecuencia	Confianza	Reglas de Culminación, tamaño 2, GAP 4, SUP 0.1
186	76.22%	(Requisitos y Especificación de Software_ Ordinario_ Aprobó) (Análisis y Diseño de Software_ Ordinario_ Aprobó)
164	67.21%	(Métodos y Modelos de Desarrollo de Software_ Ordinario_ Aprobó) (Arquitectura de Software_ Ordinario_ Aprobó)
163	66.80%	(Programación de Microcontroladores_ Ordinario_ Aprobó) (Requisitos y Especificación de Software_ Ordinario_ Aprobó)



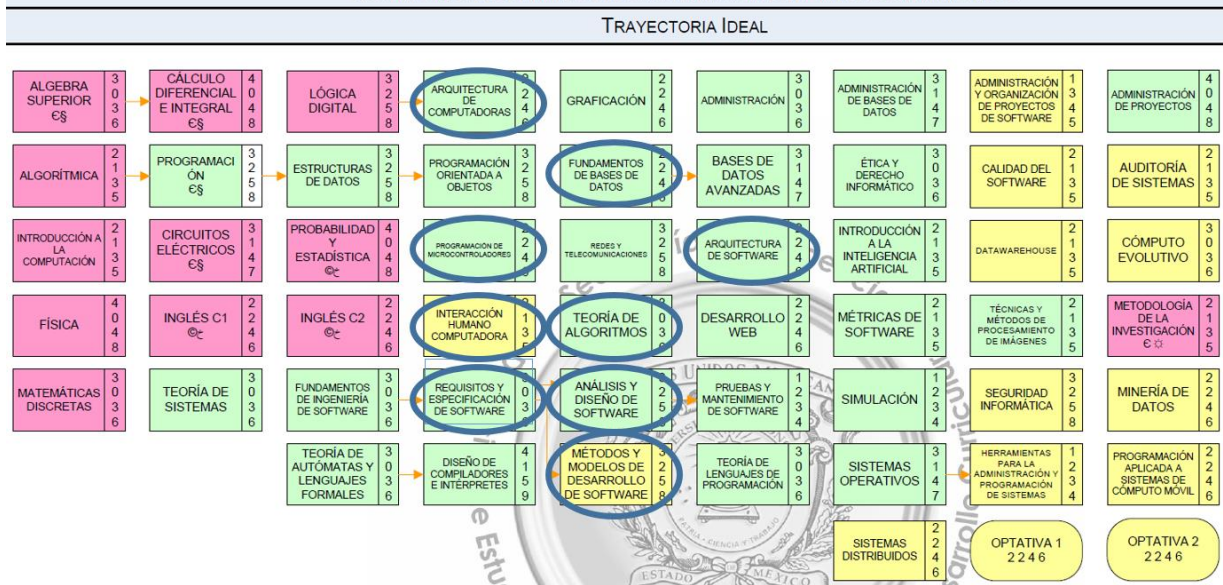
161	65.98%	(Análisis y Diseño de Software_ Ordinario_ Aprobó) (Fundamentos de Base de Datos_ Ordinario_ Aprobó)
160	65.57%	(Fundamentos de Base de Datos_ Ordinario_ Aprobó) (Métodos y Modelos de Desarrollo de Software_ Ordinario_ Aprobó)
154	63.11%	(Fundamentos de Base de Datos_ Ordinario_ Aprobó) (Teoría de Algoritmos_ Ordinario_ Aprobó)
154	63.11%	(Teoría de Algoritmos_ Ordinario_ Aprobó) (Arquitectura de Software_ Ordinario_ Aprobó)
153	62.70%	(Análisis y Diseño de Software_ Ordinario_ Aprobó) (Teoría de Algoritmos_ Ordinario_ Aprobó)
153	62.70%	(Programación de Microcontroladores_ Ordinario_ Aprobó) (Interacción Humano Computadora_ Ordinario_ Aprobó)
152	62.29%	(Análisis y Diseño de Software_ Ordinario_ Aprobó) (Métodos y Modelos de Desarrollo de Software_ Ordinario_ Aprobó)

Las reglas de Culminación para tamaño 2 cumplen con que el estudiante aprobó sus unidades de aprendizaje en la modalidad de ordinario, todas las unidades de aprendizaje corresponden del 4° al 6° semestre.

La regla más fuerte corresponde a 2 unidades de aprendizaje aprobadas en ordinario, las cuales son unidades de aprendizaje seriadas, pertenecen al 4° y 5° semestre. La regla indica que las unidades de aprendizaje fueron aprobadas en curso normal por el estudiante.

Las unidades de aprendizaje aparecen marcadas en la siguiente ilustración:

3.8 MAPA CURRICULAR DE LA LICENCIATURA EN INGENIERÍA DE SOFTWARE 2008



**Ilustración 14.** Reglas de culminación tamaño 2.

A continuación, se presentan las reglas de baja de tamaño 2:

**Tabla 16.** Reglas de baja tamaño 2.

Frecuencia	Confianza	Reglas de Baja, tamaño 2, GAP 4, SUP 0.1
94	38.52%	(Física_ Título_ Suspendido) (Matemáticas Discretas_ Extra_ Suspendido)
78	31.96%	(Física_ Título_ Suspendido) (Matemáticas Discretas_ Título_ Suspendido)
74	30.32%	(Algorítmica_ Extra_ Suspendido) (Algorítmica_ Título_ Suspendido)
74	30.32%	(Lógica Digital_ Extra_ Suspendido) (Lógica Digital_ Título_ Suspendido)
68	27.86%	(Algorítmica_ Título_ Suspendido) (Física_ Título_ Suspendido)
68	27.86%	(Teoría de Autómatas y Lenguajes Formales_ Extra_ Suspendido) (Teoría de Autómatas y Lenguajes Formales_ Título_ Suspendido)
66	27.04%	(Algorítmica_ Título_ Suspendido) (Física_ Extra_ Suspendido)

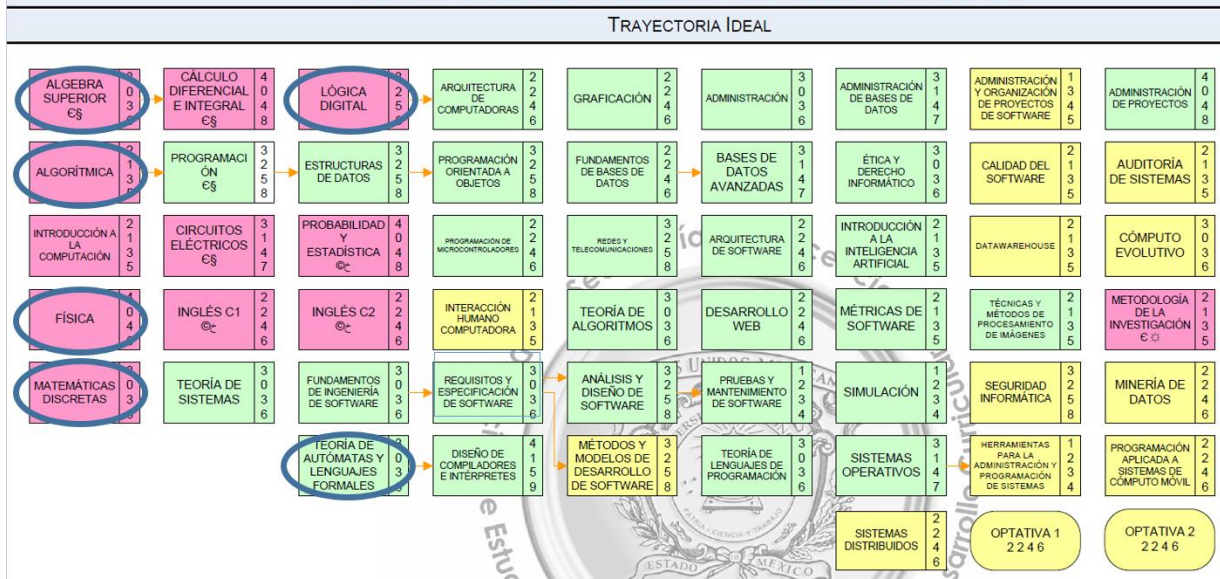
64	26.22%	(Algorítmica_Extra_Suspendido) (Física_Título_Suspendido)
63	25.81%	(Algebra Superior_Extra_Suspendido) (Algebra Superior_Título_Suspendido)
63	25.81%	(Algorítmica_Extra_Suspendido) (Física_Extra_Suspendido)

Las reglas corresponden a unidades de aprendizaje que están presentes en el 1° y en 3° semestre, todas suspendidas, lo cual se refiere a que dieron de baja la materia o que no tienen calificación asentada.

Las unidades de aprendizaje correspondientes a la regla más fuerte son del núcleo básico, corresponden al primer semestre, tienen la etiqueta de suspendido. La regla indica que no se le asentó calificación porque no tenía derecho a presentar examen por acumulación de faltas, o porque el estudiante no presentó el examen, el estudiante realizó la acción en curso normal o en segundo curso (recicle).

Las unidades de aprendizaje aparecen marcadas en la siguiente ilustración:

3.8 MAPA CURRICULAR DE LA LICENCIATURA EN INGENIERÍA DE SOFTWARE 2008



**Ilustración 15.** Reglas de baja tamaño 2.

A continuación, se presentan las reglas de culminación de tamaño 3:

**Tabla 17.** Reglas de culminación tamaño 3.

Frecuencia	Confianza	Reglas de Culminación, tamaño 3, GAP 4, SUP 0.1
157	64.34%	(Requisitos y Especificación de Software_ Ordinario_ Aprobó) (Interacción Humano Computadora_ Ordinario_ Aprobó) (Análisis y Diseño Software_ Ordinario_ Aprobó)
150	61.47%	(Requisitos y Especificación de Software_ Ordinario_ Aprobó) (Análisis y Diseño Software_ Ordinario_ Aprobó) (Fundamentos de Base de Datos_ Ordinario_ Aprobó)
146	59.83%	(Análisis y Diseño Software_ Ordinario_ Aprobó) (Fundamentos de Base de Datos_ Ordinario_ Aprobó) (Métodos y Modelos de Desarrollo de Software_ Ordinario_ Aprobó)
145	59.42%	(Análisis y Diseño Software_ Ordinario_ Aprobó) (Teoría de Algoritmos_ Ordinario_ Aprobó) (Métodos y Modelos de Desarrollo de Software_ Ordinario_ Aprobó)

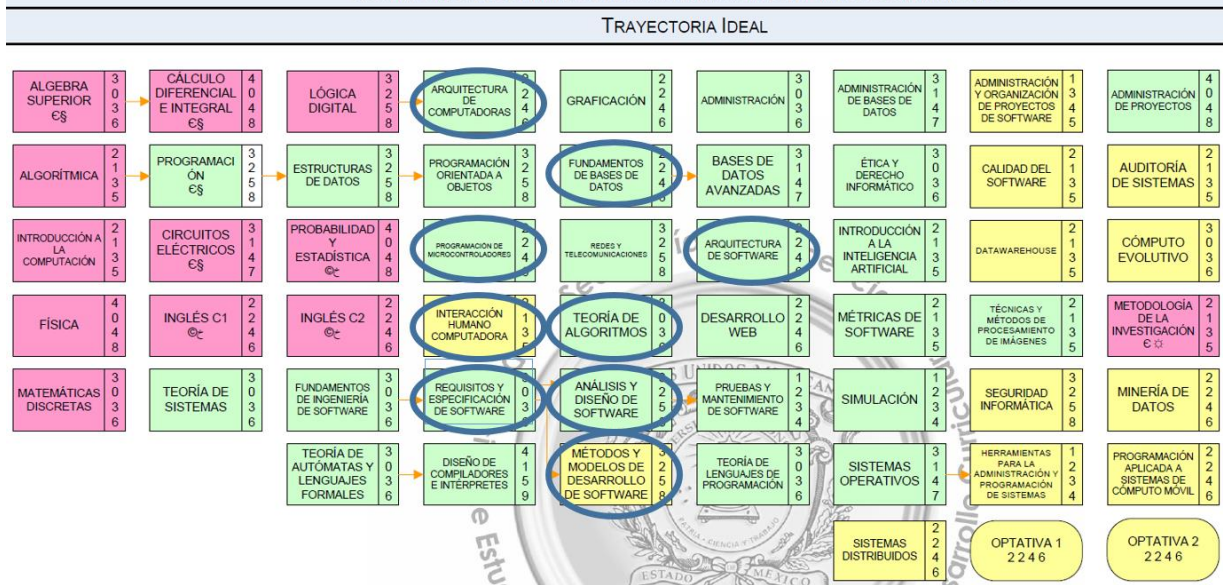
145	59.42%	(Teoría de Algoritmos_ Ordinario_ Aprobó) (Métodos y Modelos de Desarrollo de Software_ Ordinario_ Aprobó) (Arquitectura de Software_ Ordinario_ Aprobó)
144	59.01%	(Fundamentos de Base de Datos_ Ordinario_ Aprobó) (Métodos y Modelos de Desarrollo de Software_ Ordinario_ Aprobó) (Arquitectura de Software_ Ordinario_ Aprobó)
144	59.01%	(Programación de Microcontroladores_ Ordinario_ Aprobó) (Requisitos y Especificación de Software_ Ordinario_ Aprobó) (Análisis y Diseño Software_ Ordinario_ Aprobó)
144	59.01%	(Requisitos y Especificación de Software_ Ordinario_ Aprobó) (Análisis y Diseño Software_ Ordinario_ Aprobó) (Teoría de Algoritmos_ Ordinario_ Aprobó)
143	58.60%	(Fundamentos de Base de Datos_ Ordinario_ Aprobó) (Teoría de Algoritmos_ Ordinario_ Aprobó) (Métodos y Modelos de Desarrollo de Software_ Ordinario_ Aprobó)
142	58.19%	(Requisitos y Especificación de Software_ Ordinario_ Aprobó) (Análisis y Diseño Software_ Ordinario_ Aprobó) (Métodos y Modelos de Desarrollo de Software_ Ordinario_ Aprobó)

Las reglas obtenidas corresponden al 4º, 5º y 6º semestre, todas aprobadas en ordinario.

Las unidades de aprendizaje de la regla más fuerte corresponden al 4º y 5º semestre, dos de ellas serializadas, todas aprobadas en ordinario. La regla indica que las unidades de aprendizaje fueron aprobadas en curso normal por el estudiante.

Las unidades de aprendizaje aparecen marcadas en la siguiente ilustración:

3.8 MAPA CURRICULAR DE LA LICENCIATURA EN INGENIERÍA DE SOFTWARE 2008



**Ilustración 16.** Reglas de culminación tamaño 3.

A continuación, se presentan las reglas de baja de tamaño 3:

**Tabla 18.** Reglas de baja tamaño 3.

Frecuencia	Confianza	Reglas de Baja, tamaño 3, GAP 4, SUP 0.1
92	37.70%	(Física_ Título_ Suspendido) (Matemáticas Discretas_ Extra_ Suspendido) (Matemáticas Discretas_ Título_ Suspendido)
85	34.83%	(Física_ Extra_ Suspendido) (Física_ Título_ Suspendido) (Matemáticas Discretas_ Extra_ Suspendido)
73	29.91%	(Física_ Ordinario_ Reprobó) (Física_ Extra_ Suspendido) (Física_ Título_ Suspendido)
64	26.22%	(Algorítmica_ Título_ Suspendido) (Física_ Extra_ Suspendido) (Física_ Título_ Suspendido)
64	26.22%	(Física_ Extra_ Suspendido) (Física_ Título_ Suspendido) (Matemáticas Discretas_ Título_ Suspendido)
63	25.81%	(Algorítmica_ Extra_ Suspendido) (Algorítmica_ Título_ Suspendido) (Física_ Título_ Suspendido)

62	25.40%	(Algorítmica_ Extra_ Suspendido) (Algorítmica_ Título_ Suspendido) (Física_ Extra_ Suspendido)
62	25.40%	(Física_ Título_ Suspendido) (Introducción a la Computación_ Ordinario_ Aprobó) (Matemáticas Discretas_ Título_ Suspendido)
62	25.40%	(Física_ Título_ Suspendido) (Matemáticas Discretas_ Ordinario_ Reprobó) (Matemáticas Discretas_ Título_ Suspendido)
60	24.59%	(Algorítmica_ Extra_ Suspendido) (Física_ Extra_ Suspendido) (Física_ Título_ Suspendido)

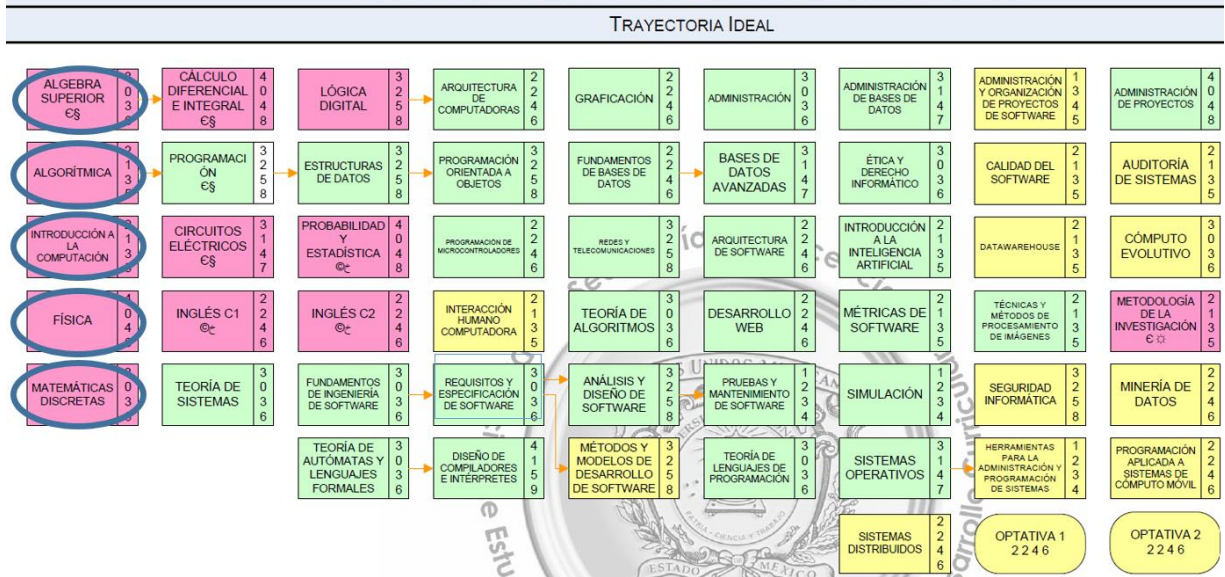
Las reglas obtenidas corresponden al 1° semestre, todas suspendidas.

La regla más fuerte contiene unidades de aprendizaje del 1° semestre, se observa que suspendió las unidades de aprendizaje en extra y título. La regla indica que no se le asentó calificación porque no tenía derecho a presentar examen por acumulación de faltas, o porque el estudiante no presentó el examen, el estudiante realizó la acción en curso normal o en segundo curso (recicle).

Las unidades de aprendizaje aparecen marcadas en la siguiente ilustración:



3.8 MAPA CURRICULAR DE LA LICENCIATURA EN INGENIERÍA DE SOFTWARE 2008



**Ilustración 17.** Reglas de baja tamaño 3.

A continuación, se presentan las reglas de culminación de tamaño 4:

**Tabla 19.** Reglas de culminación tamaño 4.

Frecuencia	Confianza	Reglas de Culminación, tamaño 4, GAP 4, SUP 0.1
137	56.14%	(Requisitos y Especificación de Software_ Ordinario_ Aprobó) (Análisis y Diseño de Software_ Ordinario_ Aprobó) (Teoría de Algoritmos_ Ordinario_ Aprobó) (Métodos y Modelos de Desarrollo de Software_ Ordinario_ Aprobó)
136	55.73%	(Requisitos y Especificación de Software_ Ordinario_ Aprobó) (Interacción Humano Computadora_ Ordinario_ Aprobó) (Fundamentos de Base de Datos_ Ordinario_ Aprobó) (Métodos y Modelos de Desarrollo de Software_ Ordinario_ Aprobó)
133	54.50%	(Requisitos y Especificación de Software_ Ordinario_ Aprobó) (Interacción Humano Computadora_ Ordinario_ Aprobó) (Análisis y Diseño de Software_ Ordinario_ Aprobó) (Fundamentos de Base de Datos_ Ordinario_ Aprobó)
132	54.09%	(Requisitos y Especificación de Software_ Ordinario_ Aprobó) (Análisis y Diseño de Software_ Ordinario_ Aprobó) (Fundamentos



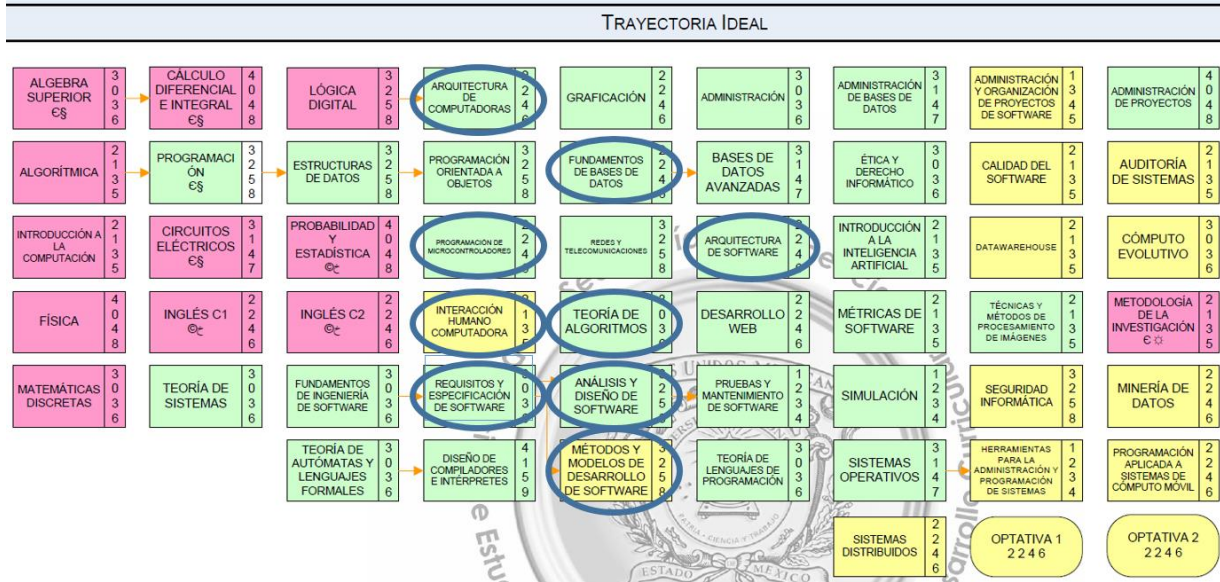
		de Base de Datos_ Ordinario_ Aprobó) (Teoría de Algoritmos_ Ordinario_ Aprobó)
131	53.68%	(Análisis y Diseño de Software_ Ordinario_ Aprobó) (Fundamentos de Base de Datos_ Ordinario_ Aprobó) (Métodos y Modelos de Desarrollo de Software_ Ordinario_ Aprobó) (Arquitectura de Software_ Ordinario_ Aprobó)
131	53.68%	(Análisis y Diseño de Software_ Ordinario_ Aprobó) (Fundamentos de Base de Datos_ Ordinario_ Aprobó) (Teoría de Algoritmos_ Ordinario_ Aprobó) (Métodos y Modelos de Desarrollo de Software_ Ordinario_ Aprobó)
131	53.68%	(Fundamentos de Base de Datos_ Ordinario_ Aprobó) (Teoría de Algoritmos_ Ordinario_ Aprobó) (Métodos y Modelos de Desarrollo de Software_ Ordinario_ Aprobó) (Arquitectura de Software_ Ordinario_ Aprobó)
131	53.68%	(Interacción Humano Computadora_ Ordinario_ Aprobó) (Análisis y Diseño de Software_ Ordinario_ Aprobó) (Fundamentos de Base de Datos_ Ordinario_ Aprobó) (Métodos y Modelos de Desarrollo de Software_ Ordinario_ Aprobó)
131	53.68%	(Interacción Humano Computadora_ Ordinario_ Aprobó) (Análisis y Diseño de Software_ Ordinario_ Aprobó) (Teoría de Algoritmos_ Ordinario_ Aprobó) (Métodos y Modelos de Desarrollo de Software_ Ordinario_ Aprobó)
131	53.68%	(Interacción Humano Computadora_ Ordinario_ Aprobó) (Fundamentos de Base de Datos_ Ordinario_ Aprobó) (Métodos y Modelos de Desarrollo de Software_ Ordinario_ Aprobó) (Arquitectura de Software_ Ordinario_ Aprobó)

Las reglas obtenidas corresponden al 4º, 5º y 6º semestre, todas aprobadas en ordinario.

La regla más fuerte contiene 1 materia del 4º semestre y 3 unidades de aprendizaje del 5º semestre, todas aprobadas en ordinario. La regla indica que las unidades de aprendizaje fueron aprobadas en curso normal por el estudiante.

Las unidades de aprendizaje aparecen marcadas en la siguiente ilustración:

3.8 MAPA CURRICULAR DE LA LICENCIATURA EN INGENIERÍA DE SOFTWARE 2008



**Ilustración 18.** Reglas de culminación tamaño 4.

A continuación, se presentan las reglas de baja de tamaño 4:

**Tabla 20.** Reglas de baja tamaño 4.

Frecuencia	Confianza	Reglas de Baja, tamaño 4, GAP 4, SUP 0.1
83	34.01%	(Física_Extra_Suspendido) (Física_Título_Suspendido) (Matemáticas Discretas_Extra_Suspendido) (Matemáticas Discretas_Título_Suspendido)
60	24.59%	(Algorítmica_Extra_Suspendido) (Algorítmica_Título_Suspendido) (Física_Extra_Suspendido) (Física_Título_Suspendido)
53	21.72%	(Algorítmica_Título_Suspendido) (Física_Título_Suspendido) (Matemáticas Discretas_Extra_Suspendido) (Matemáticas Discretas_Título_Suspendido)
52	21.31%	(Algebra Superior_Extra_Suspendido) (Algebra Superior_Título_Suspendido) (Física_Extra_Suspendido) (Física_Título_Suspendido)

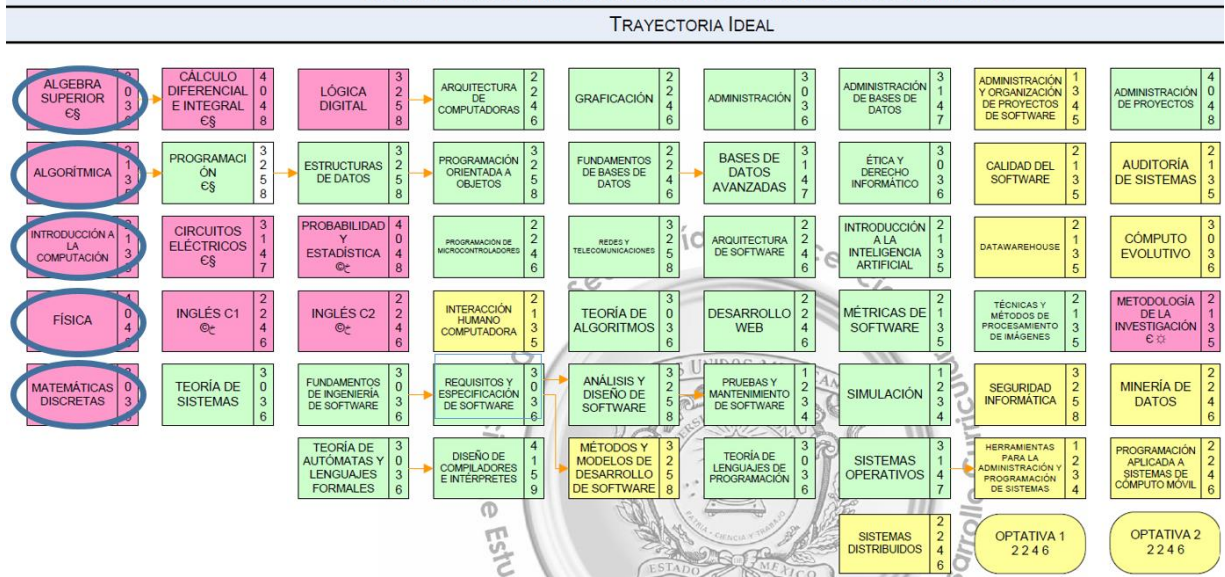
52	21.31%	(Algebra Superior_ Título_ Suspendido) (Física_ Extra_ Suspendido) (Física_ Título_ Suspendido) (Matemáticas Discretas_ Extra_ Suspendido)
52	21.31%	(Algorítmica_ Título_ Suspendido) (Física_ Extra_ Suspendido) (Física_ Título_ Suspendido) (Matemáticas Discretas_ Extra_ Suspendido)
52	21.31%	(Física_ Extra_ Suspendido) (Física_ Título_ Suspendido) (Matemáticas Discretas_ Ordinario_ Reprobó) (Matemáticas Discretas_ Título_ Suspendido)
51	20.90%	(Algorítmica_ Extra_ Suspendido) (Algorítmica_ Título_ Suspendido) (Física_ Título_ Suspendido) (Matemáticas Discretas_ Extra_ Suspendido)
51	20.90%	(Algorítmica_ Extra_ Suspendido) (Física_ Título_ Suspendido) (Matemáticas Discretas_ Extra_ Suspendido) (Matemáticas Discretas_ Título_ Suspendido)
51	20.90%	(Física_ Extra_ Suspendido) (Física_ Título_ Suspendido) (Introducción a la Computación_ Ordinario_ Aprobó) (Matemáticas Discretas_ Título_ Suspendido)

Las reglas obtenidas corresponden al 1° semestre, todas suspendidas.

La regla más fuerte contiene unidades de aprendizaje del 1° semestre, se observa que suspendió las unidades de aprendizaje en extra y título. La regla indica que no se le asentó calificación porque no tenía derecho a presentar examen por acumulación de faltas, o porque el estudiante no presentó el examen, el estudiante realizó la acción en curso normal o en segundo curso (recicle).

Las unidades de aprendizaje aparecen marcadas en la siguiente ilustración:

3.8 MAPA CURRICULAR DE LA LICENCIATURA EN INGENIERÍA DE SOFTWARE 2008



**Ilustración 19.** Reglas de baja tamaño 4.

A continuación, se presentan las reglas de culminación de tamaño 5:

**Tabla 21.** Reglas de culminación tamaño 5.

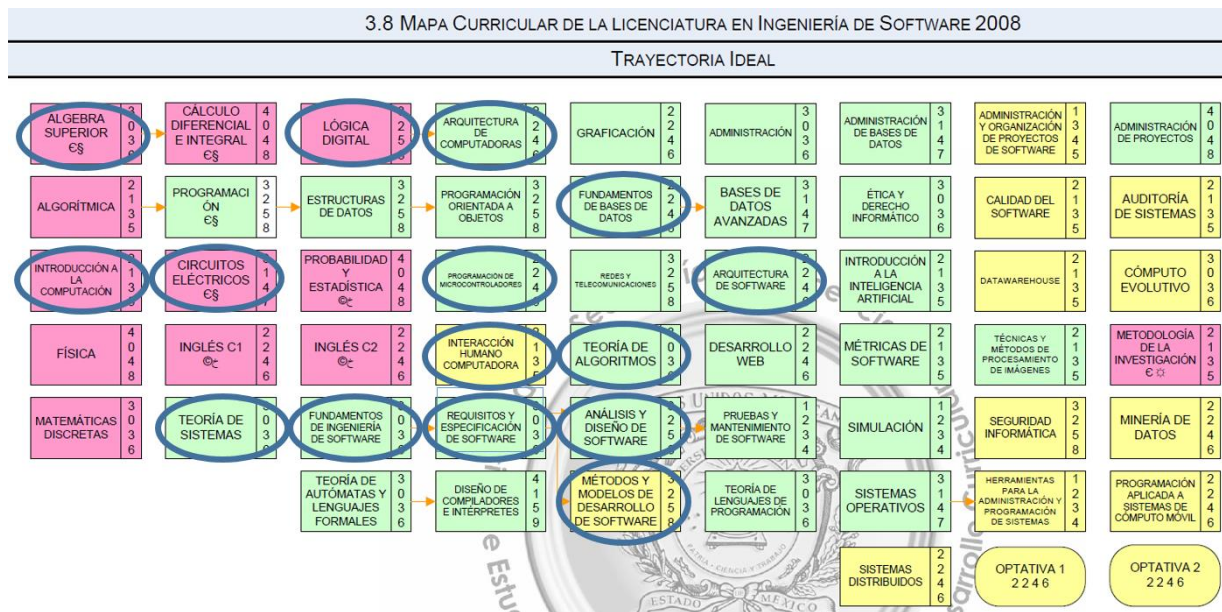
Frecuencia	Confianza	Reglas de Culminación, tamaño 5, GAP 4, SUP 0.1
127	52.04%	(Requisitos y Especificación de Software_ Ordinario_ Aprobó) (Análisis y Diseño de Software_ Ordinario_ Aprobó) (Fundamentos de Base de Datos_ Ordinario_ Aprobó) (Teoría de Algoritmos_ Ordinario_ Aprobó) (Métodos y Modelos de Desarrollo de Software_ Ordinario_ Aprobó)
127	52.04%	(Requisitos y Especificación de Software_ Ordinario_ Aprobó) (Interacción Humano Computadora_ Ordinario_ Aprobó) (Análisis y Diseño de Software_ Ordinario_ Aprobó) (Fundamentos de Base de Datos_ Ordinario_ Aprobó) (Métodos y Modelos de Desarrollo de Software_ Ordinario_ Aprobó)
127	52.04%	(Requisitos y Especificación de Software_ Ordinario_ Aprobó) (Interacción Humano Computadora_ Ordinario_ Aprobó) (Fundamentos de Base de Datos_ Ordinario_ Aprobó) (Métodos y Modelos de Desarrollo de Software_ Ordinario_ Aprobó) (Arquitectura de Software_ Ordinario_ Aprobó)

126	51.63%	(Introducción a la Computación_ Ordinario_ Aprobó) (Circuitos Eléctricos_ Ordinario_ Aprobó) (Teoría de Sistemas_ Ordinario_ Aprobó) (Lógica Digital_ Ordinario_ Aprobó) (Fundamentos de Ingeniería de Software_ Ordinario_ Aprobó)
126	51.63%	(Requisitos y Especificación de Software_ Ordinario_ Aprobó) (Análisis y Diseño de Software_ Ordinario_ Aprobó) (Fundamentos de Base de Datos_ Ordinario_ Aprobó) (Métodos y Modelos de Desarrollo de Software_ Ordinario_ Aprobó) (Arquitectura de Software_ Ordinario_ Aprobó)
125	51.22%	(Requisitos y Especificación de Software_ Ordinario_ Aprobó) (Interacción Humano Computadora_ Ordinario_ Aprobó) (Análisis y Diseño de Software_ Ordinario_ Aprobó) (Teoría de Algoritmos_ Ordinario_ Aprobó) (Métodos y Modelos de Desarrollo de Software_ Ordinario_ Aprobó)
125	51.22%	(Requisitos y Especificación de Software_ Ordinario_ Aprobó) (Interacción Humano Computadora_ Ordinario_ Aprobó) (Fundamentos de Base de Datos_ Ordinario_ Aprobó) (Teoría de Algoritmos_ Ordinario_ Aprobó) (Métodos y Modelos de Desarrollo de Software_ Ordinario_ Aprobó)
123	50.40%	(Requisitos y Especificación de Software_ Ordinario_ Aprobó) (Análisis y Diseño de Software_ Ordinario_ Aprobó) (Teoría de Algoritmos_ Ordinario_ Aprobó) (Métodos y Modelos de Desarrollo de Software_ Ordinario_ Aprobó) (Arquitectura de Software_ Ordinario_ Aprobó)
122	50.00%	(Algebra Superior_ Ordinario_ Aprobó) (Introducción a la Computación_ Ordinario_ Aprobó) (Circuitos Eléctricos_ Ordinario_ Aprobó) (Teoría de Sistemas_ Ordinario_ Aprobó) (Lógica Digital_ Ordinario_ Aprobó)
122	50.00%	(Interacción Humano Computadora_ Ordinario_ Aprobó) (Análisis y Diseño de Software_ Ordinario_ Aprobó) (Fundamentos de Base de Datos_ Ordinario_ Aprobó) (Métodos y Modelos de Desarrollo de Software_ Ordinario_ Aprobó) (Arquitectura de Software_ Ordinario_ Aprobó)

Las reglas obtenidas corresponden del 1º al 6º semestre, todas aprobadas en ordinario.

La regla más fuerte contiene 1 materia del 4º semestre y 4 unidades de aprendizaje del 5º semestre, todas aprobadas en ordinario. La regla indica que las unidades de aprendizaje fueron aprobadas en curso normal por el estudiante.

Las unidades de aprendizaje aparecen marcadas en la siguiente ilustración:



**Ilustración 20.** Reglas culminación tamaño 5

A continuación, se presentan las reglas de baja de tamaño 5:

**Tabla 22.** Reglas de baja tamaño 5.

Frecuencia	Confianza	Reglas de Baja, tamaño 5, GAP 4, SUP 0.1
51	20.90%	(Algorítmica_Título_Suspendido) (Física_Extra_Suspendido) (Física_Título_Suspendido) (Matemáticas Discretas_Extra_Suspendido) (Matemáticas Discretas_Título_Suspendido)

50	20.49%	(Algebra Superior_ Título_ Suspendido) (Física_ Extra_ Suspendido) (Física_ Título_ Suspendido) (Matemáticas Discretas_ Extra_ Suspendido) (Matemáticas Discretas_ Título_ Suspendido)
50	20.49%	(Algorítmica_ Extra_ Suspendido) (Algorítmica_ Título_ Suspendido) (Física_ Título_ Suspendido) (Matemáticas Discretas_ Extra_ Suspendido) (Matemáticas Discretas_ Título_ Suspendido)
49	20.08%	(Algorítmica_ Extra_ Suspendido) (Algorítmica_ Título_ Suspendido) (Física_ Extra_ Suspendido) (Física_ Título_ Suspendido) (Matemáticas Discretas_ Extra_ Suspendido)
49	20.08%	(Algorítmica_ Extra_ Suspendido) (Física_ Extra_ Suspendido) (Física_ Título_ Suspendido) (Matemáticas Discretas_ Extra_ Suspendido) (Matemáticas Discretas_ Título_ Suspendido)
48	19.67%	(Algebra Superior_ Extra_ Suspendido) (Algebra Superior_ Título_ Suspendido) (Física_ Extra_ Suspendido) (Física_ Título_ Suspendido) (Matemáticas Discretas_ Extra_ Suspendido)
44	18.03%	(Física_ Extra_ Suspendido) (Física_ Título_ Suspendido) (Matemáticas Discretas_ Ordinario_ Reprobó) (Matemáticas Discretas_ Extra_ Suspendido) (Matemáticas Discretas_ Título_ Suspendido)
43	17.62%	(Física_ Extra_ Suspendido) (Física_ Título_ Suspendido) (Introducción a la Computación_ Ordinario_ Aprobó) (Matemáticas Discretas_ Extra_ Suspendido) (Matemáticas Discretas_ Título_ Suspendido)
43	17.62%	(Física_ Ordinario_ Suspendido) (Física_ Extra_ Suspendido) (Física_ Título_ Suspendido) (Matemáticas Discretas_ Extra_ Suspendido) (Matemáticas Discretas_ Título_ Suspendido)
43	17.62%	(Física_ Ordinario_ Reprobó) (Física_ Extra_ Suspendido) (Física_ Título_ Suspendido) (Matemáticas Discretas_ Extra_ Suspendido) (Matemáticas Discretas_ Título_ Suspendido)

Las reglas obtenidas corresponden al 1º semestre, todas suspendidas.

La regla más fuerte contiene unidades de aprendizaje del 1º semestre, se observa que suspendió las unidades de aprendizaje en extra y título. La regla indica que no se le asentó calificación porque no tenía derecho a presentar examen por acumulación de faltas, o



porque el estudiante no presentó el examen, el estudiante realizó la acción en curso normal o en segundo curso (recicle).

Las unidades de aprendizaje aparecen marcadas en la siguiente ilustración:



**Ilustración 21.** Reglas de baja tamaño 5.

A continuación, se presentan las reglas de culminación de tamaño 6:

**Tabla 23.** Reglas de culminación tamaño 6.

Frecuencia	Confianza	Reglas de Culminación, tamaño 6, GAP 4, SUP 0.1
118	48.36%	(Requisitos y Especificación de Software_ Ordinario_ Aprobó) (Interacción Humano Computadora_ Ordinario_ Aprobó) (Análisis y Diseño de Software_ Ordinario_ Aprobó) (Fundamentos de Base de Datos_ Ordinario_ Aprobó) (Métodos y Modelos de Desarrollo de Software_ Ordinario_ Aprobó) (Arquitectura de Software_ Ordinario_ Aprobó)



117	47.95%	(Requisitos y Especificación de Software_ Ordinario_ Aprobó) (Interacción Humano Computadora_ Ordinario_ Aprobó) (Fundamentos de Base de Datos_ Ordinario_ Aprobó) (Teoría de Algoritmos_ Ordinario_ Aprobó) (Métodos y Modelos de Desarrollo de Software_ Ordinario_ Aprobó) (Arquitectura de Software_ Ordinario_ Aprobó)
116	47.54%	(Requisitos y Especificación de Software_ Ordinario_ Aprobó) (Análisis y Diseño de Software_ Ordinario_ Aprobó) (Fundamentos de Base de Datos_ Ordinario_ Aprobó) (Teoría de Algoritmos_ Ordinario_ Aprobó) (Métodos y Modelos de Desarrollo de Software_ Ordinario_ Aprobó) (Arquitectura de Software_ Ordinario_ Aprobó)
116	47.54%	(Requisitos y Especificación de Software_ Ordinario_ Aprobó) (Interacción Humano Computadora_ Ordinario_ Aprobó) (Análisis y Diseño de Software_ Ordinario_ Aprobó) (Fundamentos de Base de Datos_ Ordinario_ Aprobó) (Teoría de Algoritmos_ Ordinario_ Aprobó) (Métodos y Modelos de Desarrollo de Software_ Ordinario_ Aprobó)
116	47.54%	(Requisitos y Especificación de Software_ Ordinario_ Aprobó) (Interacción Humano Computadora_ Ordinario_ Aprobó) (Fundamentos de Base de Datos_ Ordinario_ Aprobó) (Métodos y Modelos de Desarrollo de Software_ Ordinario_ Aprobó) (Arquitectura de Software_ Ordinario_ Aprobó) (Bases de Datos Avanzadas_ Ordinario_ Aprobó)
115	47.13%	(Algebra Superior_ Ordinario_ Aprobó) (Introducción a la Computación_ Ordinario_ Aprobó) (Circuitos Eléctricos_ Ordinario_ Aprobó) (Teoría de Sistemas_ Ordinario_ Aprobó) (Lógica Digital_ Ordinario_ Aprobó) (Fundamentos de Ingeniería de Software_ Ordinario_ Aprobó)
115	47.13%	(Requisitos y Especificación de Software_ Ordinario_ Aprobó) (Interacción Humano Computadora_ Ordinario_ Aprobó) (Análisis y Diseño de Software_ Ordinario_ Aprobó) (Teoría de Algoritmos_ Ordinario_ Aprobó) (Métodos y Modelos de Desarrollo de Software_ Ordinario_ Aprobó) (Arquitectura de Software_ Ordinario_ Aprobó)
113	46.31%	(Requisitos y Especificación de Software_ Ordinario_ Aprobó) (Interacción Humano Computadora_ Ordinario_ Aprobó) (Fundamentos de Base de Datos_ Ordinario_ Aprobó) (Teoría de Algoritmos_ Ordinario_ Aprobó) (Métodos y Modelos de

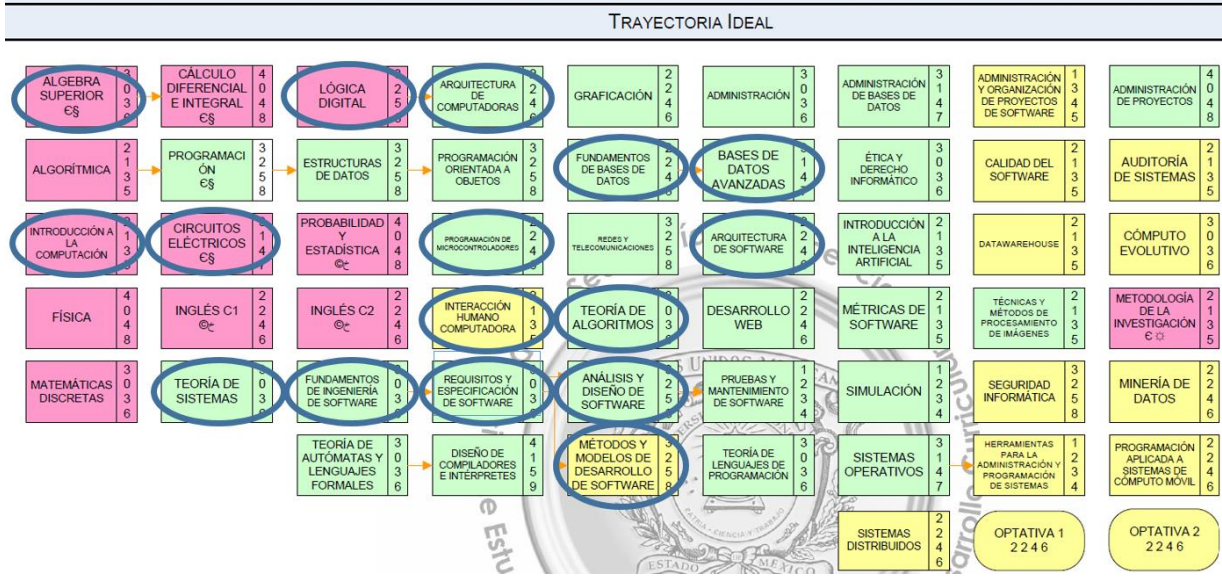
		Desarrollo de Software_ Ordinario_ Aprobó) (Bases de Datos Avanzadas_ Ordinario_ Aprobó)
112	45.90%	(Interacción Humano Computadora_ Ordinario_ Aprobó) (Fundamentos de Base de Datos_ Ordinario_ Aprobó) (Teoría de Algoritmos_ Ordinario_ Aprobó) (Métodos y Modelos de Desarrollo de Software_ Ordinario_ Aprobó) (Arquitectura de Software_ Ordinario_ Aprobó) (Bases de Datos Avanzadas_ Ordinario_ Aprobó)
112	45.90%	(Requisitos y Especificación de Software_ Ordinario_ Aprobó) (Análisis y Diseño de Software_ Ordinario_ Aprobó) (Fundamentos de Base de Datos_ Ordinario_ Aprobó) (Métodos y Modelos de Desarrollo de Software_ Ordinario_ Aprobó) (Arquitectura de Software_ Ordinario_ Aprobó) (Bases de Datos Avanzadas_ Ordinario_ Aprobó)

Las reglas obtenidas corresponden del 1° al 6° semestre, todas aprobadas en ordinario.

La regla más fuerte contiene 1 materia del 4° semestre, 4 unidades de aprendizaje del 5° semestre y 1 materia del 6° semestre, todas aprobadas en ordinario. La regla indica que las unidades de aprendizaje fueron aprobadas en curso normal por el estudiante.

Las unidades de aprendizaje aparecen marcadas en la siguiente ilustración:

3.8 MAPA CURRICULAR DE LA LICENCIATURA EN INGENIERÍA DE SOFTWARE 2008



**Ilustración 22.** Reglas de culminación tamaño 6.

A continuación, se presentan las reglas de baja de tamaño 6:

**Tabla 24.** Reglas de baja tamaño 6.

Frecuencia	Confianza	Reglas de Baja, tamaño 6, GAP 4, SUP 0.1
48	19.67%	(Algorítmica_ Extra_ Suspendido) (Algorítmica_ Título_ Suspendido) (Física_ Extra_ Suspendido) (Física_ Título_ Suspendido) (Matemáticas Discretas_ Extra_ Suspendido) (Matemáticas Discretas_ Título_ Suspendido)
47	19.26%	(Algebra Superior_ Extra_ Suspendido) (Algebra Superior_ Título_ Suspendido) (Física_ Extra_ Suspendido) (Física_ Título_ Suspendido) (Matemáticas Discretas_ Extra_ Suspendido) (Matemáticas Discretas_ Título_ Suspendido)
39	15.98%	(Algebra Superior_ Extra_ Suspendido) (Algebra Superior_ Título_ Suspendido) (Algorítmica_ Extra_ Suspendido) (Algorítmica_ Título_ Suspendido) (Física_ Extra_ Suspendido) (Física_ Título_ Suspendido)
36	14.75%	(Algebra Superior_ Título_ Suspendido) (Algorítmica_ Título_ Suspendido) (Física_ Extra_ Suspendido) (Física_ Título_ Suspendido)

		Suspendido) (Matemáticas Discretas_ Extra_ Suspendido) (Matemáticas Discretas_ Título_ Suspendido)
36	14.75%	(Algebra Superior_ Extra_ Suspendido) (Algebra Superior_ Título_ Suspendido) (Algorítmica_ Título_ Suspendido) (Física_ Extra_ Suspendido) (Física_ Título_ Suspendido) (Matemáticas Discretas_ Extra_ Suspendido)
35	14.34%	(Algebra Superior_ Título_ Suspendido) (Algorítmica_ Extra_ Suspendido) (Algorítmica_ Título_ Suspendido) (Física_ Extra_ Suspendido) (Física_ Título_ Suspendido) (Matemáticas Discretas_ Extra_ Suspendido)
35	14.34%	(Algebra Superior_ Extra_ Suspendido) (Algebra Superior_ Título_ Suspendido) (Algorítmica_ Título_ Suspendido) (Física_ Título_ Suspendido) (Matemáticas Discretas_ Extra_ Suspendido) (Matemáticas Discretas_ Título_ Suspendido)
35	14.34%	(Algebra Superior_ Extra_ Suspendido) (Algebra Superior_ Título_ Suspendido) (Algorítmica_ Extra_ Suspendido) (Algorítmica_ Título_ Suspendido) (Física_ Título_ Suspendido) (Matemáticas Discretas_ Extra_ Suspendido)
35	14.34%	(Algebra Superior_ Extra_ Suspendido) (Algebra Superior_ Título_ Suspendido) (Algorítmica_ Extra_ Suspendido) (Física_ Extra_ Suspendido) (Física_ Título_ Suspendido) (Matemáticas Discretas_ Extra_ Suspendido)
35	14.34%	(Algebra Superior_ Extra_ Suspendido) (Algorítmica_ Título_ Suspendido) (Física_ Extra_ Suspendido) (Física_ Título_ Suspendido) (Matemáticas Discretas_ Extra_ Suspendido) (Matemáticas Discretas_ Título_ Suspendido)

Las reglas obtenidas corresponden al 1º semestre, todas suspendidas.

La regla más fuerte contiene unidades de aprendizaje del 1º semestre, se observa que suspendió las unidades de aprendizaje en extra y título. La regla indica que no se le asentó calificación porque no tenía derecho a presentar examen por acumulación de faltas, o porque el estudiante no presentó el examen, el estudiante realizó la acción en curso normal o en segundo curso (recicle).

Las unidades de aprendizaje aparecen marcadas en la siguiente ilustración:



**Ilustración 23.** Reglas de baja tamaño 6.

A continuación, se presentan las reglas de culminación de tamaño 7:

**Tabla 25.** Reglas de culminación tamaño 7.

Frecuencia	Confianza	Reglas de Culminación, tamaño 7, GAP 4, SUP 0.1
110	45.08%	(Requisitos y Especificación de Software_ Ordinario_ Aprobó) (Interacción Humano Computadora_ Ordinario_ Aprobó) (Fundamentos de Base de Datos_ Ordinario_ Aprobó) (Teoría de Algoritmos_ Ordinario_ Aprobó) (Métodos y Modelos de Desarrollo de Software_ Ordinario_ Aprobó) (Arquitectura de Software_ Ordinario_ Aprobó) (Bases de Datos Avanzadas_ Ordinario_ Aprobó)
108	44.26%	(Requisitos y Especificación de Software_ Ordinario_ Aprobó) (Interacción Humano Computadora_ Ordinario_ Aprobó) (Análisis y Diseño de Software_ Ordinario_ Aprobó) (Fundamentos de Base de Datos_ Ordinario_ Aprobó) (Teoría de Algoritmos_ Ordinario_ Aprobó) (Métodos y Modelos de Desarrollo de Software_

		Ordinario_ Aprobó) (Arquitectura de Software_ Ordinario_ Aprobó)
107	43.85%	(Requisitos y Especificación de Software_ Ordinario_ Aprobó) (Análisis y Diseño de Software_ Ordinario_ Aprobó) (Fundamentos de Base de Datos_ Ordinario_ Aprobó) (Teoría de Algoritmos_ Ordinario_ Aprobó) (Métodos y Modelos de Desarrollo de Software_ Ordinario_ Aprobó) (Arquitectura de Software_ Ordinario_ Aprobó) (Bases de Datos Avanzadas_ Ordinario_ Aprobó)
107	43.85%	(Requisitos y Especificación de Software_ Ordinario_ Aprobó) (Interacción Humano Computadora_ Ordinario_ Aprobó) (Análisis y Diseño de Software_ Ordinario_ Aprobó) (Fundamentos de Base de Datos_ Ordinario_ Aprobó) (Métodos y Modelos de Desarrollo de Software_ Ordinario_ Aprobó) (Arquitectura de Software_ Ordinario_ Aprobó) (Bases de Datos Avanzadas_ Ordinario_ Aprobó)
105	43.03%	(Requisitos y Especificación de Software_ Ordinario_ Aprobó) (Interacción Humano Computadora_ Ordinario_ Aprobó) (Análisis y Diseño de Software_ Ordinario_ Aprobó) (Fundamentos de Base de Datos_ Ordinario_ Aprobó) (Teoría de Algoritmos_ Ordinario_ Aprobó) (Métodos y Modelos de Desarrollo de Software_ Ordinario_ Aprobó) (Bases de Datos Avanzadas_ Ordinario_ Aprobó)
105	43.03%	(Requisitos y Especificación de Software_ Ordinario_ Aprobó) (Interacción Humano Computadora_ Ordinario_ Aprobó) (Fundamentos de Base de Datos_ Ordinario_ Aprobó) (Métodos y Modelos de Desarrollo de Software_ Ordinario_ Aprobó) (Arquitectura de Software_ Ordinario_ Aprobó) (Bases de Datos Avanzadas_ Ordinario_ Aprobó) (Pruebas y Mantenimiento de Software_ Ordinario_ Aprobó)
104	42.62%	(Interacción Humano Computadora_ Ordinario_ Aprobó) (Análisis y Diseño de Software_ Ordinario_ Aprobó) (Fundamentos de Base de Datos_ Ordinario_ Aprobó) (Teoría de Algoritmos_ Ordinario_ Aprobó) (Métodos y Modelos de Desarrollo de Software_ Ordinario_ Aprobó) (Arquitectura de Software_ Ordinario_ Aprobó) (Bases de Datos Avanzadas_ Ordinario_ Aprobó)
103	42.21%	(Algebra Superior_ Ordinario_ Aprobó) (Algorítmica_ Ordinario_ Aprobó) (Introducción a la Computación_ Ordinario_ Aprobó) (Circuitos Eléctricos_ Ordinario_ Aprobó) (Teoría de Sistemas_

		Ordinario_ Aprobó) (Lógica Digital_ Ordinario_ Aprobó) (Fundamentos de Ingeniería de Software_ Ordinario_ Aprobó)
103	42.21%	(Programación de Microcontroladores_ Ordinario_ Aprobó) (Requisitos y Especificación de Software_ Ordinario_ Aprobó) (Interacción Humano Computadora_ Ordinario_ Aprobó) (Fundamentos de Base de Datos_ Ordinario_ Aprobó) (Métodos y Modelos de Desarrollo de Software_ Ordinario_ Aprobó) (Arquitectura de Software_ Ordinario_ Aprobó) (Bases de Datos Avanzadas_ Ordinario_ Aprobó)
103	42.21%	(Programación de Microcontroladores_ Ordinario_ Aprobó) (Requisitos y Especificación de Software_ Ordinario_ Aprobó) (Interacción Humano Computadora_ Ordinario_ Aprobó) (Fundamentos de Base de Datos_ Ordinario_ Aprobó) (Teoría de Algoritmos_ Ordinario_ Aprobó) (Métodos y Modelos de Desarrollo de Software_ Ordinario_ Aprobó) (Arquitectura de Software_ Ordinario_ Aprobó)

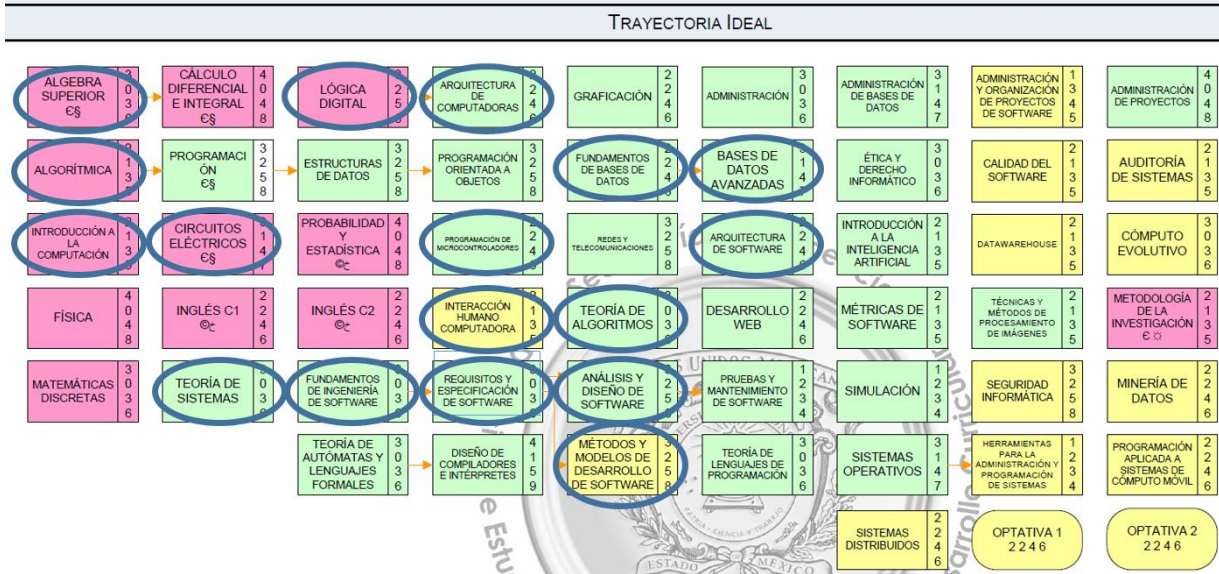
Las reglas obtenidas corresponden del 1° al 6° semestre, todas aprobadas en ordinario.

La regla más fuerte contiene 1 materia del 4° semestre, 4 unidades de aprendizaje del 5° semestre y 2 unidades de aprendizaje del 6° semestre, todas aprobadas en ordinario. La regla indica que las unidades de aprendizaje fueron aprobadas en curso normal por el estudiante.

Las unidades de aprendizaje aparecen marcadas en la siguiente ilustración:



3.8 MAPA CURRICULAR DE LA LICENCIATURA EN INGENIERÍA DE SOFTWARE 2008



**Ilustración 24.** Reglas de culminación tamaño 7.

A continuación, se presentan las reglas de baja de tamaño 7:

**Tabla 26.** Reglas de baja tamaño 7.

Frecuencia	Confianza	Reglas de Baja, tamaño 7, GAP 4, SUP 0.1
35	14.34%	(Algebra Superior_ Extra_ Suspendido) (Algebra Superior_Titulo_ Suspendido) (Algorítmica_Titulo_ Suspendido) (Física_ Extra_ Suspendido) (Física_Titulo_ Suspendido) (Matemáticas Discretas_ Extra_ Suspendido) (Matemáticas Discretas_Titulo_ Suspendido)
35	14.34%	(Algebra Superior_ Extra_ Suspendido) (Algebra Superior_Titulo_ Suspendido) (Algorítmica_ Extra_ Suspendido) (Algorítmica_Titulo_ Suspendido) (Física_ Extra_ Suspendido) (Física_Titulo_ Suspendido) (Matemáticas Discretas_ Extra_ Suspendido)
34	13.93%	(Algebra Superior_Titulo_ Suspendido) (Algorítmica_ Extra_ Suspendido) (Algorítmica_Titulo_ Suspendido) (Física_ Extra_ Suspendido) (Física_Titulo_ Suspendido) (Matemáticas Discretas_ Extra_ Suspendido) (Matemáticas Discretas_Titulo_ Suspendido)
34	13.93%	(Algebra Superior_ Extra_ Suspendido) (Algebra Superior_Titulo_ Suspendido) (Algorítmica_ Extra_ Suspendido) (Algorítmica_Titulo_ Suspendido) (Física_Titulo_ Suspendido) (Matemáticas



		Discretas_Extra_Suspendido) (Matemáticas Discretas_Titulo_Suspendido)
34	13.93%	(Algebra Superior_Extra_Suspendido) (Algebra Superior_Titulo_Suspendido) (Algorítmica_Extra_Suspendido) (Física_Extra_Suspendido) (Física_Titulo_Suspendido) (Matemáticas Discretas_Extra_Suspendido) (Matemáticas Discretas_Titulo_Suspendido)
34	13.93%	(Algebra Superior_Extra_Suspendido) (Algorítmica_Extra_Suspendido) (Algorítmica_Titulo_Suspendido) (Física_Extra_Suspendido) (Física_Titulo_Suspendido) (Matemáticas Discretas_Extra_Suspendido) (Matemáticas Discretas_Titulo_Suspendido)
32	13.11%	(Algorítmica_Ordinario_Suspendido) (Algorítmica_Extra_Suspendido) (Algorítmica_Titulo_Suspendido) (Física_Extra_Suspendido) (Física_Titulo_Suspendido) (Matemáticas Discretas_Extra_Suspendido) (Matemáticas Discretas_Titulo_Suspendido)
32	13.11%	(Algebra Superior_Ordinario_Suspendido) (Algebra Superior_Extra_Suspendido) (Algebra Superior_Titulo_Suspendido) (Física_Extra_Suspendido) (Física_Titulo_Suspendido) (Matemáticas Discretas_Extra_Suspendido) (Matemáticas Discretas_Titulo_Suspendido)
30	12.29%	(Algorítmica_Extra_Suspendido) (Algorítmica_Titulo_Suspendido) (Física_Ordinario_Suspendido) (Física_Extra_Suspendido) (Física_Titulo_Suspendido) (Matemáticas Discretas_Extra_Suspendido) (Matemáticas Discretas_Titulo_Suspendido)
30	12.29%	(Algebra Superior_Ordinario_Suspendido) (Algebra Superior_Extra_Suspendido) (Algebra Superior_Titulo_Suspendido) (Algorítmica_Extra_Suspendido) (Algorítmica_Titulo_Suspendido) (Física_Extra_Suspendido) (Física_Titulo_Suspendido)

Las reglas obtenidas corresponden al 1º semestre, todas suspendidas.

La regla más fuerte contiene unidades de aprendizaje del 1º semestre, se observa que suspendió las unidades de aprendizaje en extra y título. La regla indica que no se le asentó calificación porque no tenía derecho a presentar examen por acumulación de faltas, o porque el estudiante no presentó el examen, el estudiante realizó la acción en curso normal o en segundo curso (recicle).

Las unidades de aprendizaje aparecen marcadas en la siguiente ilustración:



**Ilustración 25.** Reglas de baja tamaño 7.

A continuación, se presentan las reglas de baja de tamaño 8:

**Tabla 27.** Reglas de baja tamaño 8.

Frecuencia	Confianza	Reglas de Baja, tamaño 8, GAP 4, SUP 0.1
34	13.93%	(Algebra Superior_ Extra_ Suspendido) (Algebra Superior_ Título_ Suspendido) (Algorítmica_ Extra_ Suspendido) (Algorítmica_ Título_ Suspendido) (Física_ Extra_ Suspendido) (Física_ Título_ Suspendido) (Matemáticas Discretas_ Extra_ Suspendido) (Matemáticas Discretas_ Título_ Suspendido)
26	10.65%	(Algebra Superior_ Extra_ Suspendido) (Algebra Superior_ Título_ Suspendido) (Física_ Extra_ Suspendido) (Física_ Título_ Suspendido) (Introducción a la Computación_ Extra_ Suspendido) (Introducción a la Computación_ Título_ Suspendido)

		Suspendido) (Matemáticas Discretas_ Extra_ Suspendido) (Matemáticas Discretas_ Título_ Suspendido)
26	10.65%	(Algorítmica_ Extra_ Suspendido) (Algorítmica_ Título_ Suspendido) (Física_ Extra_ Suspendido) (Física_ Título_ Suspendido) (Introducción a la Computación_ Extra_ Suspendido) (Introducción a la Computación_ Título_ Suspendido) (Matemáticas Discretas_ Extra_ Suspendido) (Matemáticas Discretas_ Título_ Suspendido)
26	10.65%	(Algorítmica_ Ordinario_ Suspendido) (Algorítmica_ Extra_ Suspendido) (Algorítmica_ Título_ Suspendido) (Física_ Extra_ Suspendido) (Física_ Título_ Suspendido) (Introducción a la Computación_ Ordinario_ Suspendido) (Introducción a la Computación_ Extra_ Suspendido) (Introducción a la Computación_ Título_ Suspendido)
26	10.65%	(Algebra Superior_ Ordinario_ Suspendido) (Algebra Superior_ Título_ Suspendido) (Algorítmica_ Extra_ Suspendido) (Algorítmica_ Título_ Suspendido) (Física_ Extra_ Suspendido) (Física_ Título_ Suspendido) (Matemáticas Discretas_ Extra_ Suspendido) (Matemáticas Discretas_ Título_ Suspendido)
26	10.65%	(Algebra Superior_ Ordinario_ Suspendido) (Algebra Superior_ Extra_ Suspendido) (Algebra Superior_ Título_ Suspendido) (Algorítmica_ Título_ Suspendido) (Física_ Extra_ Suspendido) (Física_ Título_ Suspendido) (Matemáticas Discretas_ Extra_ Suspendido) (Matemáticas Discretas_ Título_ Suspendido)
26	10.65%	(Algebra Superior_ Ordinario_ Suspendido) (Algebra Superior_ Extra_ Suspendido) (Algebra Superior_ Título_ Suspendido) (Algorítmica_ Extra_ Suspendido) (Algorítmica_ Título_ Suspendido) (Física_ Título_ Suspendido) (Matemáticas Discretas_ Extra_ Suspendido) (Matemáticas Discretas_ Título_ Suspendido)
26	10.65%	(Algebra Superior_ Ordinario_ Suspendido) (Algebra Superior_ Extra_ Suspendido) (Algebra Superior_ Título_ Suspendido) (Algorítmica_ Extra_ Suspendido) (Algorítmica_ Título_ Suspendido) (Física_ Extra_ Suspendido) (Física_ Título_ Suspendido) (Matemáticas Discretas_ Extra_ Suspendido)
26	10.65%	(Algebra Superior_ Ordinario_ Suspendido) (Algebra Superior_ Extra_ Suspendido) (Algebra Superior_ Título_ Suspendido) (Algorítmica_ Extra_ Suspendido) (Algorítmica_ Título_ Suspendido) (Física_ Extra_ Suspendido) (Física_ Título_ Suspendido) (Matemáticas Discretas_ Extra_ Suspendido) (Matemáticas Discretas_ Título_ Suspendido)

26	10.65%	(Algebra Superior_Ordinario_Suspendido) (Algebra Superior_Extra_Suspendido) (Algorítmica_Extra_Suspendido) (Algorítmica_Título_Suspendido) (Física_Extra_Suspendido) (Física_Título_Suspendido) (Matemáticas Discretas_Extra_Suspendido) (Matemáticas Discretas_Título_Suspendido)
----	--------	--

Las reglas obtenidas corresponden al 1º semestre, todas suspendidas.

La regla más fuerte contiene unidades de aprendizaje del 1º semestre, se observa que suspendió las unidades de aprendizaje en extra y título. La regla indica que no se le asentó calificación porque no tenía derecho a presentar examen por acumulación de faltas, o porque el estudiante no presentó el examen, el estudiante realizó la acción en curso normal o en segundo curso (recicle).

Las unidades de aprendizaje aparecen marcadas en la siguiente ilustración:



**Ilustración 26.** Reglas de baja tamaño 8.

A continuación, se presentan las reglas de baja de tamaño 9:

**Tabla 28.** Reglas de baja tamaño 9.

Frecuencia	Confianza	Reglas de Baja, tamaño 9, GAP 4, SUP 0.1
26	10.65%	(Algebra Superior_ Ordinario_ Suspendido) (Algebra Superior_ Extra_ Suspendido) (Algebra Superior_ Título_ Suspendido) (Algorítmica_ Extra_ Suspendido) (Algorítmica_ Título_ Suspendido) (Física_ Extra_ Suspendido) (Física_ Título_ Suspendido) (Matemáticas Discretas_ Extra_ Suspendido) (Matemáticas Discretas_ Título_ Suspendido)

Las reglas obtenidas corresponden al 1º semestre todas suspendidas.

La regla más fuerte contiene unidades de aprendizaje del 1º semestre se observa que suspendió las unidades de aprendizaje en extra y título. La regla indica que no se le asentó calificación porque no tenía derecho a presentar examen por acumulación de faltas, o porque el estudiante no presentó el examen, el estudiante realizó la acción en curso normal o en segundo curso (recicle).

Las unidades de aprendizaje aparecen marcadas en la siguiente ilustración:

### 3.8 MAPA CURRICULAR DE LA LICENCIATURA EN INGENIERÍA DE SOFTWARE 2008

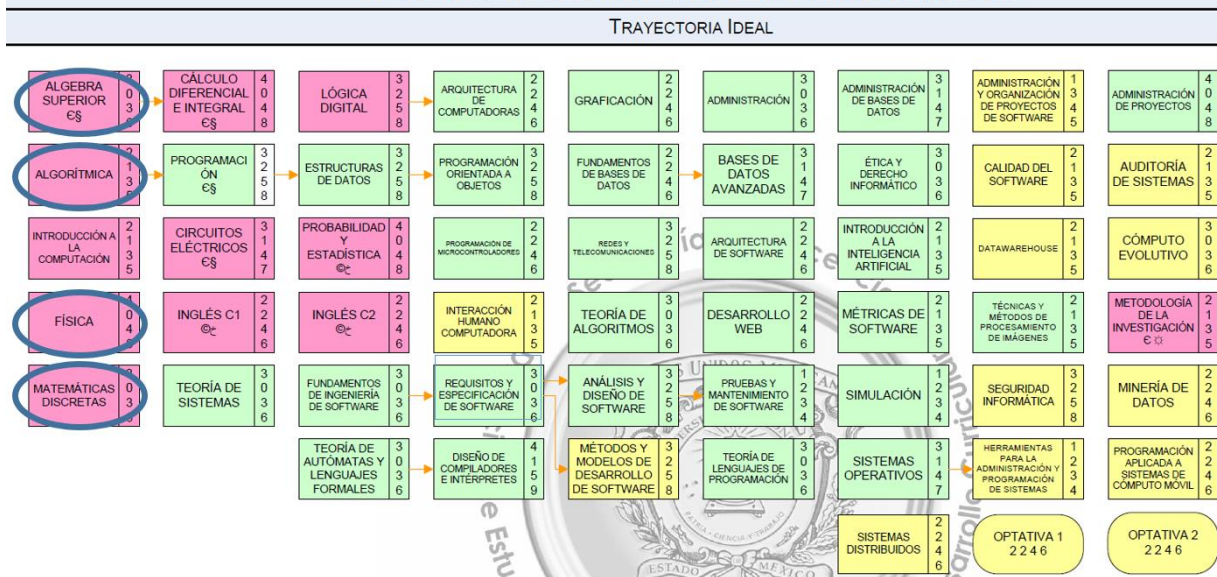


Ilustración 27. Regla de baja tamaño 9.

#### 5.5.4 Resultados Evaluación

Se evaluaron 20 historiales tomados al azar, 10 correspondientes de Baja y 10 de Culminación. Para la clasificación por la sumatoria porcentual de reglas cumplidas respecto al 100% total de reglas obtenidas.

$$x = \frac{\sum CRO}{\sum CRT}$$

Como se observa en la fórmula se divide la sumatoria de la confianza obtenida del número de reglas obtenidas en un historial académico, entre la sumatoria de confianza de las reglas totales obtenidas de las bases de datos, correspondientes por tamaño.

Retomando, la **confianza** determina la predictibilidad de la regla. Se le conoce como umbral de confianza a la frecuencia con que aparece una secuencia en un evento dentro de un

registro de datos, dicho umbral hace mención a un mínimo número de veces que aparece [Liu, 2011]. Por lo que, la sumatoria de confianza corresponde a la suma de la confianza de cada regla que aparece dentro del historial evaluado.

**Evaluación** de 10 historiales académicos de **CULMINACIÓN** tomados al azar del compendio de historiales académicos.

Total, de reglas de **Culminación** que cumplen los 10 historiales de **culminación** y total de reglas por tamaño:

**Tabla 29.** Sumatorias reglas obtenidas.

Estudiante	T2	T3	T4	T5	T6	T7	T8	T9	Sumatoria Reglas
1	226	702	2897	9443	29858	93459	0	0	136585
2	219	703	2478	7457	21889	63784	0	0	96530
3	210	674	2619	9219	32013	106701	0	0	151436
4	244	901	3843	11439	36743	113837	0	0	167007
5	231	746	2565	7794	23529	70217	0	0	105082
6	238	819	3311	11583	39458	130374	0	0	185873
7	146	413	1152	270	6533	14819	0	0	23333
8	104	236	504	970	1840	3366	0	0	7020
9	86	204	412	818	1578	2927	0	0	6025
10	203	702	2801	9700	33006	108647	0	0	155059

Total, de reglas cumplidas con la sumatoria de la confianza de las mismas reglas obtenidas:

**Tabla 30.** Sumatorias confianzas de reglas.

Estudiante	Sumatoria Reglas	Sumatoria Confianza
1	136585	2406892.76%
2	96530	1673349.35%

3	151436	2685126.71%
4	167007	2886997.43%
5	105082	1893524.39%
6	185873	3228596.71%
7	23333	459661.82%
8	7020	83027.72%
9	6025	107855.63%
10	155059	2717788.78%

Porcentaje total alcanzado de la cantidad de reglas obtenidas respecto al 100% de reglas obtenidas en general:

**Tabla 31.** Sumatorias aplicando fórmula.

<b>Estudiante</b>	<b>Sumatoria Reglas</b>	<b>Sumatoria reglas del total</b>
1	136585	43.1459574%
2	96530	30.4929477%
3	151436	47.837253%
4	167007	52.7559901%
5	105082	33.1944466%
6	185873	58.7155876%
7	23333	7.37068217%
8	7020	2.21755406%
9	6025	1.90324262%
10	155059	48.9817257%

Total, de reglas de Baja que cumplen los 10 historiales de culminación y total de reglas por tamaño:



**Tabla 32.** Sumatorias de reglas.

Estudiante	T2	T3	T4	T5	T6	T7	T8	T9	Sumatoria Reglas
1	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0
3	10	16	6	0	0	0	0	0	32
4	4	15	9	1	0	0	0	0	29
5	2	0	0	0	0	0	0	0	2
6	2	0	0	0	0	0	0	0	2
7	28	34	15	2	0	0	0	0	79
8	7	2	1	0	0	0	0	0	10
9	24	36	9	1	0	0	0	0	70
10	0	0	0	0	0	0	0	0	0

Total, de reglas cumplidas con la sumatoria de la confianza de las mismas reglas obtenidas:

**Tabla 33.** Sumatorias de confianza de reglas.

Estudiante	Sumatoria Reglas	Sumatoria Confianza
1	0	0.00%
2	0	0.00%
3	32	434.28%
4	29	443.51%
5	2	21.30%
6	2	21.30%
7	79	1057.00%
8	10	115.52%
9	70	1092.31%
10	0	0.00%

Porcentaje total alcanzado de la cantidad de reglas obtenidas respecto al 100% de reglas obtenidas en general:

**Tabla 34.** Sumatorias de reglas.

Estudiante	Sumatoria Reglas	Sumatoria reglas del total
1	0	0%
2	0	0%
3	32	0.01010851%
4	29	0.00916084%
5	2	0.00063178%
6	2	0.00063178%
7	79	0.02495538%
8	10	0.00315891%
9	70	0.02211236%
10	0	0%

**Evaluación** de 10 historiales académicos de **BAJA** tomados al azar del compendio de historiales académicos.

Total, de reglas de **Baja** que cumplen los 10 historiales de **Baja** y total de reglas por tamaño:

**Tabla 35.** Sumatoria de reglas.

Estudiante	T2	T3	T4	T5	T6	T7	T8	T9	Sumatoria Reglas
1	33	51	28	4	0	0	0	0	116
2	23	51	25	4	0	0	0	0	103
3	2	1	0	0	0	0	0	0	3
4	60	121	122	74	31	8	1	0	417

5	35	91	78	35	10	1	0	0	250
6	21	65	59	27	6	0	0	0	178
7	6	7	1	0	0	0	0	0	14
8	76	112	108	83	48	13	1	0	441
9	29	43	20	3	0	0	0	0	95
10	63	92	52	15	2	0	0	0	224

Total, de reglas cumplidas con la sumatoria de la confianza de las mismas reglas obtenidas:

**Tabla 36.** Sumatorias de confianza de reglas.

Estudiante	Sumatoria Reglas	Sumatoria Confianza
1	116	1522.38%
2	103	1386.40%
3	3	40.56%
4	417	5934.54%
5	250	3697.23%
6	178	2784.06%
7	14	193.79%
8	441	6312.29%
9	95	1361.45%
10	224	3124.73%

Porcentaje total alcanzado de la cantidad de reglas obtenidas respecto al 100% de reglas obtenidas en general:

**Tabla 37.** Sumatorias de total de reglas.

Estudiante	Sumatoria Reglas	Sumatoria reglas del total
1	116	6.16693248%

2	103	5.47581074%
3	3	0.15948963%
4	417	22.169059%
5	250	13.2908028%
6	178	9.46305157%
7	14	0.74428495%
8	441	23.4449761%
9	95	5.05050505%
10	224	11.9085593%

Total, de reglas de **Culminación** que cumplen los 10 historiales de **Baja** y total de reglas por tamaño:

**Tabla 38.** Sumatorias de reglas.

Estudiante	T2	T3	T4	T5	T6	T7	T8	T9	Sumatoria Reglas
1	19	57	59	34	12	1	0	0	182
2	13	19	10	5	1	0	0	0	48
3	195	576	1848	5025	13446	35908	0	0	56998
4	9	13	2	1	0	0	0	0	25
5	25	67	77	55	23	4	0	0	251
6	59	193	403	641	974	1452	0	0	3722
7	67	185	384	678	1259	2417	0	0	4990
8	0	0	0	0	0	0	0	0	0
9	18	39	42	24	7	1	0	0	131
10	12	13	1	0	0	0	0	0	26

Total, de reglas cumplidas con la sumatoria de la confianza de las mismas reglas obtenidas:

**Tabla 39.** Sumatorias de confianzas de reglas.

<b>Estudiante</b>	<b>Sumatoria Reglas</b>	<b>Sumatoria Confianza</b>
1	182	3264.78%
2	48	720.28%
3	56998	991393.39%
4	25	315.45%
5	251	3695.76%
6	3722	62771.09%
7	4990	91805.47%
8	0	0.00%
9	131	2521.56%
10	26	486.36%

Porcentaje total alcanzado de la cantidad de reglas obtenidas respecto al 100% de reglas obtenidas en general:

**Tabla 40.** Sumatorias de total de reglas.

<b>Estudiante</b>	<b>Sumatoria Reglas</b>	<b>Sumatoria reglas del total</b>
1	182	0.05749214%
2	48	0.01516276%
3	56998	18.005149%
4	25	0.00789727%
5	251	0.07928861%
6	3722	1.1757459%
7	4990	1.57629555%
8	0	0%

9	131	0.04138171%
10	26	0.00821316%

Dado a la evaluación se realizó una clasificación de los historiales académicos respecto a los resultados obtenidos, se muestran en la siguiente tabla:

**Tabla 41.** Evaluación y clasificación de historiales al azar.

<b>Historiales</b>	<b>Clasificación Baja</b>	<b>Clasificación Culminación</b>
Baja	8	2
Culminación	0	10

De acuerdo a la conclusión presentada de los resultados, los 2 historiales de baja que fueron clasificados como de culminación se debe a que los estudiantes estaban llevando su carrera como cualquier otro estudiante regular, sin embargo, la clasificación se debe a que los estudiantes dejaron de asistir a la universidad en 7° y 9° semestre sin retorno, por lo que sus trayectorias se asemejan más a un historial de culminación, dándose la confusión.

## *5.6 Fase 5: Difusión y uso*

Los resultados de las experimentaciones realizadas se evaluaron por parte de los expertos, asesor y tesista, evaluando los patrones secuenciales más frecuentes, posteriormente cambiándolos a una notación más entendible para un grupo de expertos (maestros, coordinadores, etc.) para darle un mayor uso al conocimiento encontrado, es decir, transformación de los patrones secuenciales a reglas de asociación.

Las trayectorias académicas comunes obtenidas en forma de reglas de asociación, están contenidas en un archivo de Excel, el cual contiene todas las reglas clasificadas por tamaño, separadas por la clase a la que pertenecen (Baja y Culminación), para un análisis posterior por parte de las personas interesadas y por para un análisis por parte de los expertos de la comunidad universitaria de la UAEM.

Una parte de los resultados obtenidos se encuentra en la sección de anexos, en los cuales se agregaron las reglas que contienen mayor confianza y mayor frecuencia, están separadas por tamaños y por su situación de Baja o Culminación.



## 6. Conclusiones y Trabajo Futuro

---

### 6.1 Conclusiones

La conclusión de esta investigación es que, sí se cumple la hipótesis planteada, es decir, que al aplicar el algoritmo GSP de minería de patrones secuenciales a bases de datos de secuencias, siguiendo la metodología de KDD, fue posible obtener patrones secuenciales que transformados y posteriormente evaluados dieron como resultado reglas que por sí solas corresponden a una trayectoria académica que fueron tomadas por distintos estudiantes.

Cada trayectoria académica está debidamente clasificada en 2 clases, la primera llamada Baja para estudiantes que fueron dados de baja académicamente o voluntariamente, para estudiantes que dejaron de asistir a la escuela; la segunda clase llamada Culminación para estudiantes que egresaron y estudiantes que siguen estudiando con buen historial académico o con mayores probabilidades de egresar.

De acuerdo a las conclusiones los trabajos del estado del arte referentes a las investigaciones en la UAEM, queda demostrado que para conocer si un estudiante deserta o egresa se debe



poner especial énfasis en los semestres 2° a 6°, académicamente es donde se presentan las unidades de aprendizaje que sirven como filtro para los estudiantes que desertan o egresan.

## *6.2 Trabajo Futuro*

El trabajo futuro propuesto está dividido en varias partes, derivadas de las experimentaciones y a causa de la falta de tiempo para llevarlas a cabo:

- Obtener los historiales académicos de la UAPT actualizados a la última fecha de las carreras ofertadas, realizar nuevas bases de datos, teniendo en cuenta que los estudiantes han tenido cambios en sus historiales académicos, algunos estudiantes otros han terminado sus respectivas carreras o en otros casos fueron dados de baja.
- Realizar la separación de los historiales por carrera y realizar sus respectivas bases de datos. Posteriormente realizar la experimentación por carrera y documentar los resultados.
- Aplicar el estudio en otro plantel de la UAEM, para rectificar la validez del estudio enfocado a otras bases de datos. Evaluada por los expertos de ese plantel académico, documentando los resultados, testimonios del experto, así como la calificación que se le daría al programa realizado.
- Realizar experimentos con una mayor cantidad de datos, de otras instituciones educativas y/o niveles educativos, recomendablemente nivel medio superior y superior.
- Realizar un sistema experto para una mayor comprensión del usuario. Mejorar el sistema desarrollando una interfaz más gráfica orientada al usuario que lo manejará.

# Referencias Bibliográficas

- [Abril, 2008] Abril Valdez Elba, Román Pérez Rosario, Cubillas Rodríguez María José, Moreno Celaya Icela; ¿Deserción o autoexclusión? Un análisis de las causas de abandono escolar en estudiantes de educación media superior en Sonora, México; Revista Electrónica de Investigación Educativa, 3 de marzo 2008; Consultado: 24 de agosto de 2015, URL: <http://redie.uabc.mx/index.php/redie/article/viewFile/183/319>
- [Agrawal, 2015] Agrawal Rakesh, Ramakrishnan Skrikant; *Mining Sequential Patterns*; Centro de Investigación Almaden IBM; Consultado 14 de agosto de 2015; URL: <http://rakesh.agrawal-family.com/papers/icde95seq.pdf>
- [Campagni, 2015] Campagni R, Merlini D., Sprugnoli R.; *Sequential patterns analysis in a student database*; Departamento de Sistemas e Informática, Florencia, Italia; Consultado el 09 de diciembre de 2015; URL: [http://dicode-project.eu/sites/default/files/paper\\_2%20-%20Sequential%20patterns%20analysis%20in%20a%20student%20database%20.pdf](http://dicode-project.eu/sites/default/files/paper_2%20-%20Sequential%20patterns%20analysis%20in%20a%20student%20database%20.pdf)
- [Casillas, 2007] Casillas Miguel, Chain Ragueb, Jácome Nancy; Origen social de los estudiantes y trayectorias estudiantiles en la Universidad Veracruzana; Abril – Junio 2007, Universidad Veracruzana, México; URL: [http://www.scielo.org.mx/scielo.php?script=sci\\_arttext&pid=S0185-27602007000200001](http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0185-27602007000200001)
- [Ceneval, 2015] Centro Nacional de Evaluación para la Educación Superior, A. C. (Ceneval); Consultado: 28 de agosto de 2015, URL: <http://www.ceneval.edu.mx>
- [Díaz, 2008] Díaz Peralta Christian; Modelo Conceptual para la deserción Estudiantil Universitaria Chilena; Universidad Católica de la Santísima Concepción 2008; Consultado el 25 de agosto de 2015, URL: <http://www.scielo.cl/pdf/estped/v34n2/art04.pdf>

- [Dzay, 2012] Dzay Chulim Floricely; La deserción escolar desde la perspectiva estudiantil; diciembre 2012, Editorial Manda; Consultado: 27 de agosto de 2015, URL:  
<http://www.uv.mx/personal/onarvaez/files/2013/02/La-desercion-escolar.pdf>
- [Edel, 2003] Edel Navarro Rubén; El rendimiento académico: Concepto, Investigación y Desarrollo; REICE: Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación, 2003; URL:  
<http://www.ice.deusto.es/rinace/reice/vol1n2/Edel.pdf>
- [ENCCCH, 2018] Escuela Nacional Colegio de Ciencias y Humanidades; Trayectoria escolar para padres de familia; UNAM, CCH, 2018, Url:  
[https://www.cch.unam.mx/padres/trayectoria\\_escolar](https://www.cch.unam.mx/padres/trayectoria_escolar)
- [Espíndola, 2002] Espíndola Ernesto, León Arturo; La deserción escolar en América Latina: un tema prioritario para la agenda regional; Revista Iberoamericana de Educación, diciembre 2002; URL: <http://www.rieoei.org/rie30f.htm>
- [Espinosa, 2014] Espinosa Enrique, Antonio Alejandrina, Medina Juana, Meza Isis, Cruz Zahira; Determinación de la correlación entre factores socioeconómicos, rendimiento académico y valores existentes en los alumnos del programa educativo de ingeniería industrial en la Universidad Politécnica de Altamira; Congreso Interdisciplinario de Cuerpos Académicos, Educational Handbook, 2014.
- [Fierros, 2014] Fierros Rosita, Gómez Lorena, Sánchez María; Factores que determinan la Deserción Escolar en la Licenciatura de Trabajo Social de la Universidad de Guadalajara en los calendarios escolares A y B del 2012; Congreso Interdisciplinario de Cuerpos académicos, *Educational Handbook*, 11 de septiembre de 2014; URL:  
<http://dialnet.unirioja.es/servlet/articulo?codigo=4888344>
- [García, 2012] García Lambert Guillermo; Descubrimiento de factores que inciden en la eficiencia terminal de estudiantes de educación superior con árboles de decisión; Texcoco, Estado de México, diciembre 2012, Centro Universitario UAEM Texcoco.

- [García, 2013] García Lambert Guillermo, García Hernández Rene Arnulfo, Ledeneva Yulia; Reglas que describen la deserción y permanencia en los estudiantes de la UAP Tlanguistenco de la UAEM; Revista Iberoamericana sobre calidad, eficacia y cambio en educación, 13 de diciembre de 2013; Consultado: 7 de diciembre de 2015; Uri: <http://www.redalyc.org/articulo.oa?id=10431177003>
- [García, 2007] García Hernández Rene Arnulfo; Desarrollo de Algoritmos para el descubrimiento de patrones secuenciales maximales; Repositorio institucional INAOE – 2007; URL: <http://www.inaoe.repositorioinstitucional.mx/bitstream/GarciaHRA>
- [GobBC, 2015] Gobierno de Baja California, Sistema Educativo Estatal, Subsecretaría de Educación Media Superior, Superior, Formación Docente y Evaluación; Deserción escolar en el nivel medio superior; Consultado el 22 de septiembre de 2015; URL: [http://www.educacionbc.edu.mx/departamentos/investigacion/publicaciones/desercion/1/interior\\_1.pdf](http://www.educacionbc.edu.mx/departamentos/investigacion/publicaciones/desercion/1/interior_1.pdf)
- [Goicovic, 2002] Goicovic Donso Igor; Educación, deserción escolar e integración laboral juvenil; Centro de estudios sociales, Valparaíso Chile, 16 de marzo de 2002; URL: <http://www.redalyc.org/articulo.oa?id=19501602>
- [Gorgas, 2009] Gorgas Javier, Cardiel Nicola´s, Zamorano Jaime; Estadística Básica para Estudiantes; Universidad Complutense de Madrid, Febrero 2009; Consultado el 4 de Noviembre de 2016.
- [Guevara, 2013] Guevara Hilda M., Belelli Sandra E.; Las trayectorias académicas: dimensiones personales de una trayectoria estudiantil. Testimonio de un actor.; Instituto de Investigaciones Socio Económicas, Facultad de Ciencias Sociales. Universidad Nacional de San Juan, Argentina; 2013; URL: <http://www.ojs.unsj.edu.ar/index.php/reviise/article/view/40>
- [Gutiérrez, 2014] Gutiérrez Ana, Amador María; Estudio para mejorar la capacidad de gestión en los estudiantes emprendedores de la Facultad de Ciencias Contables de la Universidad Nacional de San Marcos y Centro Universitario de Lagos de la Universidad de Guadalajara; Congreso Interdisciplinario de Cuerpos Académicos, Educational Handbook, 2014.

- [Hassanbeigi, 2010] Hassanbeigi A., Askari J.; A study of the most important risk factors of motivational deficiencies in university students, 4 de marzo de 2010; Consultado: 14 de Agosto de 2015, URL:  
<http://www.sciencedirect.com/science/article/pii/S1877042810017726>
- [Hernández, 2004] Hernández Orallo José, Ramírez Quintana María José, Ferri Ramírez César; Introducción a la minería de datos; Pearson: Prentice Hall, Madrid 2004.
- [Landesmann, 2001] Landesmann Monique; Trayectorias académicas generacionales constitución y diversificación del oficio académico. El caso de los bioquímicos de la Facultad de Medicina; Revista Mexicana de Investigación Educativa, Distrito Federal México, Enero - Abril 2001; Url:  
<http://www.redalyc.org/articulo.oa?id=14001104>
- [Liu, 2011] Liu Bing; Association Rules and Sequential Patterns; Springer, Web Data Mining, 15 de Abril de 2011; URL:  
[https://link.springer.com/chapter/10.1007%2F978-3-642-19460-3\\_2](https://link.springer.com/chapter/10.1007%2F978-3-642-19460-3_2)
- [López, 2017] López Gómez Ernesto, 2017; el concepto y las finalidades de la tutoría universitaria: una consulta a expertos; Revista española de orientación y psicopedagogía, España, Agosto 2017; URL:  
<http://www.redalyc.org/articulo.oa?id=338253221004>
- [Méndez, 2006] Méndez Héctor, Tesoro Rafael, Tiranti Federico; El rol del tutor como puente entre la familia y escuela; Buenos Aires, Argentina 2006; URL:  
[www.terras.edu.ar/biblioteca/16/16El-tutor-como-puente.pdf](http://www.terras.edu.ar/biblioteca/16/16El-tutor-como-puente.pdf)
- [Motegaonkar, 2014] Motegaonkar S. Vishal, Vaidya V. Madhav; A Survey on Sequential Patterns Mining Algorithms; *International Journal of Computer Science and Information Thechnology*, Vol 5 2014; Consultado: 05 de abril de 2016; URL:  
<http://www.ijcsit.com/docs/Volume%205/vol5issue02/ijcsit20140502344.pdf>
- [Nicuesta, 2015] Nicuesta Maite; Formación académica; Definición ABC, 2015, Url:

<https://www.definicionabc.com/general/formacion-academica.php>

- [Osorio, 2013] Osorio de Jesús Nayely; Identificación de factores que influyen en la deserción escolar de estudiantes universitarios usando algoritmos genéticos y k vecinos más cercanos; UAEM, Unidad Académica Profesional Tianguistenco, Julio 2013, Tianguistenco Estado de México.
- [Pajares, 2010] Pajares Martínsanz Gonzalo, Santos Peñas Matilde; Inteligencia Artificial e Ingeniería del conocimiento; Alfaomega, Madrid España, febrero 2010.
- [Palacios, 2007] Palacios Delgado Jorge Raúl, Andrade Palos Patricia; Desempeño académico y conductas de riesgo en adolescentes; UNAM, Facultad de Psicología, Revista de Educación y Desarrollo, diciembre 2007; URL: [http://www.cucs.udg.mx/revistas/edu\\_desarrollo/anteriores/7/007\\_Palacios.pdf](http://www.cucs.udg.mx/revistas/edu_desarrollo/anteriores/7/007_Palacios.pdf)
- [Pande, 1997] Pande Joshi Karuna; Analysis of Data Mining Algorithms; Marzo 1997; URL: <http://ebiquity.umbc.edu/paper/html/id/458/Analysis-of-Data-Mining-Algorithms>
- [Piña, 2007] Piña Jiménez Irma, Balán Gleavez Cristina; Los estudios de trayectorias académicas y profesionales, contribuciones al estado del arte; Revista enfermería Universitaria, UNAM, Septiembre – Diciembre 2007 Url: [www.revistas.unam.mx/index.php/reu/article/viewFile/30282/28129](http://www.revistas.unam.mx/index.php/reu/article/viewFile/30282/28129)
- [REA, 2015] Real Academia Española; Diccionario en línea; Consultado: 29 de septiembre de 2015; URL: <http://www.rae.es/>
- [Rodríguez, 2014] Rodríguez Fernández Noemí; Fundamentos del proceso educativo enseñanza, aprendizaje y evaluación a distancia; Universidad Particular Técnica de Loja, Revista Iberoamericana de Educación a Distancia, 2014; URL: <http://www.redalyc.org/articulo.oa?id=331431248005>
- [SEP-DGPPEE, 2016] Secretaría de Educación Pública, Dirección General de Planeación, Programación y Estadística Educativa página oficial, Información estadística e indicadores educativos, Publicación "Principales Cifras. Sistema Educativo de los Estados Unidos Mexicanos"; Consultado: 10 de mayo de 2016; Url:

- <http://planeacion.sep.gob.mx/estadistica-e-indicadores/estadisticas-e-indicadores>
- [SEP-SEMS, 2016] Secretaría de Educación Pública, Subsecretaría de Educación Media Superior (SEMS), Reporte de la Encuesta Nacional de Deserción en la Educación Media Superior; Consultado: 4 de mayo de 2016; Url: [http://www.sems.gob.mx/en\\_mx/sems/encuesta\\_nacional\\_desercion\\_e\\_ms](http://www.sems.gob.mx/en_mx/sems/encuesta_nacional_desercion_e_ms)
- [Sepúlveda, 2013] Sepúlveda Patricia Graciela; Trayectorias académicas y construcciones subjetivas de estudiantes y ex estudiantes virtuales; Universidad Nacional de Quilmes, Buenos Aires, Argentina, Octubre 2013; Url: <http://www.udgvirtual.udg.mx/apertura/index.php/apertura/article/view/438/344>
- [Slimani, 2013] Slimani Thabet, Lazzez Amor; *Sequential Mining: Patterns and algorithms analysis*; Universidad Taif, Arabia Saudita, 2 de noviembre 2013; URL: <http://arxiv.org/abs/1311.0350>
- [Srikant, 2015] Srikant Ramakrishnan, Agrawal Rakesh; Mining Sequential Patterns: Generalizations and Performance Improvements; Centro de Investigación Almaden IBM; Consultado: 14 de Agosto de 2015; Url: <http://rakesh.agrawal-family.com/papers/edbt96seq.pdf>
- [SSP, 2011] Secretaría de Seguridad Pública; Deserción escolar y conductas de riesgo en adolescentes; Dirección General de Prevención y Participación Ciudadana, junio 2011; URL: <http://www.ssp.gob.mx/portalWebApp/ShowBinary?nodeId=/BEA%20Repositorio/1214166//archivo>
- [Tinto, 1989] Tinto Vincent Definir la deserción: una cuestión de perspectiva; Publicaciones ANUIES, Revista de la educación superior No. 71, 18 de septiembre de 1989; URL: <http://publicaciones.anuies.mx/revista/71/1/3/es/definir-la-desercion-una-cuestion-de-perspectiva>
- [UAEM-AE, 2016] Universidad Autónoma del Estado de México, Secretaría de Planeación y Desarrollo Institucional, Agenda Estadística; Consultada: 2 de abril de 2016; Url: <http://www.uaemex.mx/planeacion/Numeros.html>

- [UAEM-OE, 2016] Universidad Autónoma del Estado de México, Secretaria de Docencia Dirección de Estudios Profesionales; Oferta educativa; Página oficial de oferta académica de la UAEM; Consultado: 15 marzo de 2016; URL: [http://dep.uaemex.mx/curricular/sitio\\_index.php](http://dep.uaemex.mx/curricular/sitio_index.php)
- [UAPT, 2015] Universidad Autónoma del Estado de México, Unidad Académica Profesional Tianguistenco, página oficial; consultada 20 de octubre de 2015; URL: <http://www.uaemex.mx/UAPTianguistenco/>
- [Valero, 2009] Valero Orea Sergio; Aplicación de técnicas de minería de datos para predecir la deserción; Universidad Tecnológica de Izúcar de Matamoros; Consultado: 20 de agosto de 2015; URL: <http://www.utim.edu.mx/~svalero/docs/MineriaDesercion.pdf>
- [Vargas, 2016] Vargas Garbanzo, M. Guiselle; Desarrollo organizacional y los procesos de cambio en las instituciones educativas, un reto de la gestión de la educación; Revista Educación, volumen 40, 2016, Universidad de Costa Rica; URL: <http://www.redalyc.org/articulo.oa?id=44043204005>
- [Zermeño, 2009] Zermeño Rodríguez Irene Victoria; Deserción escolar en adolescentes a nivel básico de la comunidad de Zaragoza; Universidad Veracruzana, Minatitlán, Veracruz, 2009; Consultado 4 de septiembre de 2015; URL: <http://cdigital.uv.mx/bitstream/12345678/217/2/IreneVictoriaZermeño.pdf>



# ANEXOS

Planes curriculares de las carreras de la UAPT: Plan Curricular Ingeniería en Plásticos



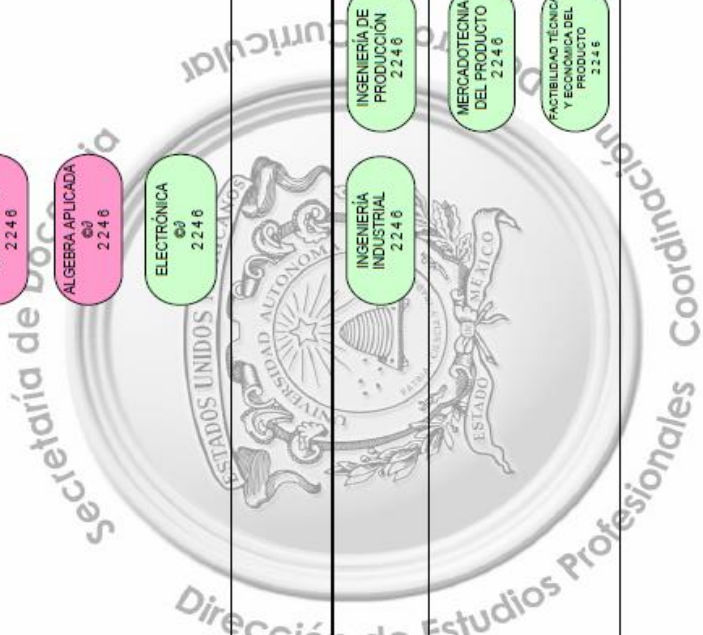
DESCRIPCIÓN DE LAS UNIDADES DE APRENDIZAJE OPTATIVAS DE LA LICENCIATURA EN INGENIERÍA DE PLÁSTICOS 2008

DESARROLLO PROFESIONAL	<p>ARTE Y PLÁSTICOS 2.246</p> <p>PSICOLOGÍA INDUSTRIAL 2.246</p> <p>DESARROLLO SUSTENTABLE 2.246</p> <p>PSICOLOGÍA DEL PRODUCTO 2.246</p> <p>DESARROLLO EMPRESARIAL 2.246</p>
DISCIPLINAS BÁSICAS	<p>GEOMETRÍA DESCRIPTIVA 2.246</p> <p>ÁLGEBRA APLICADA <sup>CS</sup> 2.246</p> <p>ELECTRÓNICA <sup>CS</sup> 2.246</p>
DISEÑO POR COMPUTADORA	<p>MANUFACTURAS ASISTIDA POR COMPUTADORA 2.246</p> <p>INGENIERÍA ASISTIDA POR COMPUTADORA 2.246</p>
INGENIERÍA DE MANUFACTURA	<p>INGENIERÍA INDUSTRIAL 2.246</p> <p>INGENIERÍA DE PRODUCCIÓN 2.246</p> <p>DISEÑO DE MOLDES 2.246</p>
INGENIERÍA DEL PRODUCTO	<p>MERCADOTECNIA DEL PRODUCTO 2.246</p> <p>DESARROLLO DE PROTOTIPOS 2.246</p> <p>FACTIBILIDAD TÉCNICA Y ECONÓMICA DEL PRODUCTO 2.246</p>
INVESTIGACIÓN	<p>AVANCES TECNOLÓGICOS Y SU APLICACIÓN 2.246</p> <p>SEMINARIO DE INVESTIGACIÓN 2.246</p>

**SIMBOLOGÍA**

- HT= HORAS TEÓRICAS
- HP= HORAS PRACTICAS
- TH= TOTAL DE HORAS
- CR= CRÉDITOS
- <sup>CS</sup> = COMÚN ENTRE: INGENIERÍA DE PLÁSTICOS E INGENIERÍA DE PRODUCCIÓN INDUSTRIAL
- <sup>E</sup> = EQUIVALENTE ENTRE: INGENIERÍA DE PLÁSTICOS E INGENIERÍA DE PRODUCCIÓN INDUSTRIAL

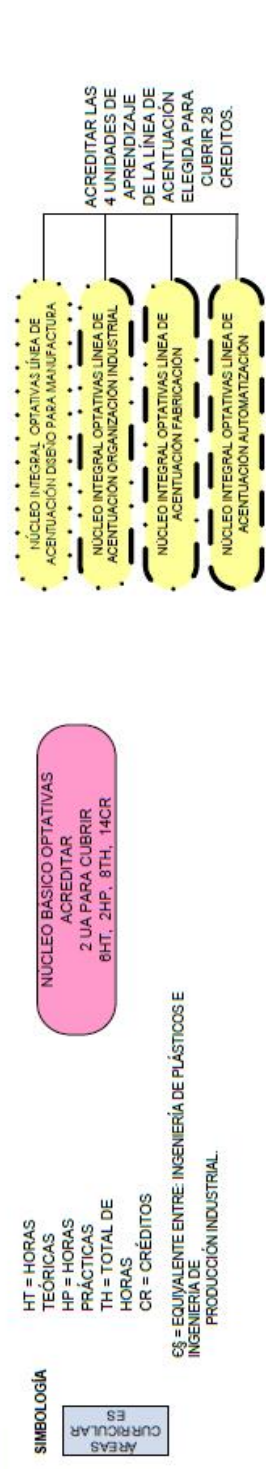
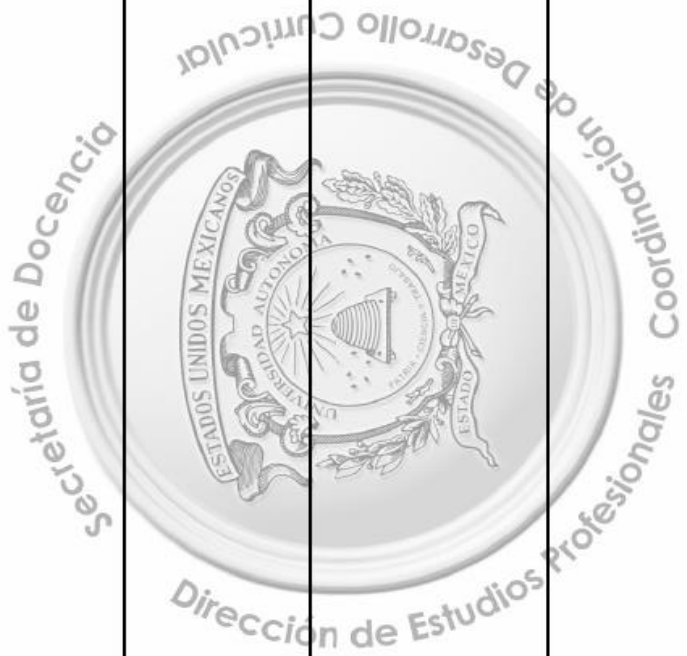
AREA CURRICULAR  
 NUCLEO BÁSICO OPTATIVAS ACREDITAR 2 UA PARA CUBRIR 4HT., 4HP., 8TH., 12CR  
 NUCLEO SUSTANTIVO OPTATIVAS ACREDITAR 4 UA PARA CUBRIR 8HT., 8HP., 16TH., 24CR  
 NUCLEO INTEGRAL OPTATIVAS ACREDITAR 3 UA PARA CUBRIR 6HT., 6HP., 12TH., 18CR







DESCRIPCIÓN DE LAS UNIDADES DE APRENDIZAJE OPTATIVAS DE LA LICENCIATURA EN INGENIERÍA DE PRODUCCIÓN INDUSTRIAL 2008	
CIENCIAS BÁSICAS	<p>TALLER DE MATEMÁTICAS 3.147</p> <p>PSICOLOGÍA INDUSTRIAL 3.147</p> <p>TALLER DE MECÁNICA 3.147</p>
AUTOMATIZACIÓN AVANZADA	<p>MANUFACTURA SIGMA FOR COMPUTADORA E3 3.147</p> <p>PROGRAMACIÓN DE MÁQUINAS DE CONTROL NUMÉRICO 3.147</p> <p>ROBÓTICA 3.147</p> <p>SISTEMAS DE MANUFACTURA FLEXIBLE 3.147</p>
FABRICACIÓN	<p>DISEÑO DEL PROCESO DE FABRICACIÓN 3.147</p> <p>MECANIZADO DE METALES 3.147</p> <p>TRATAMIENTOS TÉRMICOS 3.147</p> <p>VIBRACIONES EN MÁQUINAS 3.147</p>
ORGANIZACIÓN INDUSTRIAL	<p>CONTROL AMBIENTAL 3.147</p> <p>CONTROL ESTADÍSTICO 3.147</p> <p>BALANCEO DE LINEAS DE PRODUCCIÓN 3.147</p> <p>PLANEO DE RECURSOS PARA MANUFACTURA 3.147</p> <p>TOPICOS DE CALIDAD 3.147</p>
DISEÑO PARA MANUFACTURA	<p>"DISEÑO DEL PRODUCTO" 3.147</p> <p>FUNDAMENTOS DE ELEMENTO FINITO 3.147</p> <p>INGENIERÍA DE MATERIALES 3.147</p> <p>PROYECTO DE DISEÑO DE MÁQUINAS 3.147</p> <p>RESISTENCIA DE MATERIALES AVANZADA 3.147</p>





DESCRIPCIÓN DE LAS UNIDADES DE APRENDIZAJES OPTATIVAS DE LA LICENCIATURA EN INGENIERÍA DE SOFTWARE 2008

PROGRAMACIÓN E INGENIERÍA DE SOFTWARE	<div style="display: flex; flex-wrap: wrap; justify-content: space-around;"> <div style="border: 1px solid black; border-radius: 15px; padding: 5px; width: 45%; text-align: center;">                 Programación paralela 2 2 4 0             </div> <div style="border: 1px solid black; border-radius: 15px; padding: 5px; width: 45%; text-align: center;">                 Redes neuronales y lógica difusa 2 2 4 0             </div> <div style="border: 1px solid black; border-radius: 15px; padding: 5px; width: 45%; text-align: center;">                 Sistemas de información geográfica 2 2 4 0             </div> <div style="border: 1px solid black; border-radius: 15px; padding: 5px; width: 45%; text-align: center;">                 Software embebido 2 2 4 0             </div> </div>
TRATAMIENTO DE INFORMACIÓN	<div style="display: flex; flex-wrap: wrap; justify-content: space-around;"> <div style="border: 1px solid black; border-radius: 15px; padding: 5px; width: 45%; text-align: center;">                 Administración de empresas de desarrollo de software 2 2 4 6             </div> <div style="border: 1px solid black; border-radius: 15px; padding: 5px; width: 45%; text-align: center;">                 Bases de datos orientadas a objetos 2 2 4 0             </div> <div style="border: 1px solid black; border-radius: 15px; padding: 5px; width: 45%; text-align: center;">                 Desarrollo multimedia 2 2 4 0             </div> <div style="border: 1px solid black; border-radius: 15px; padding: 5px; width: 45%; text-align: center;">                 Reconocimiento de patrones 2 2 4 0             </div> </div>

SIMBOLOGÍA

AREA CURRICULAR

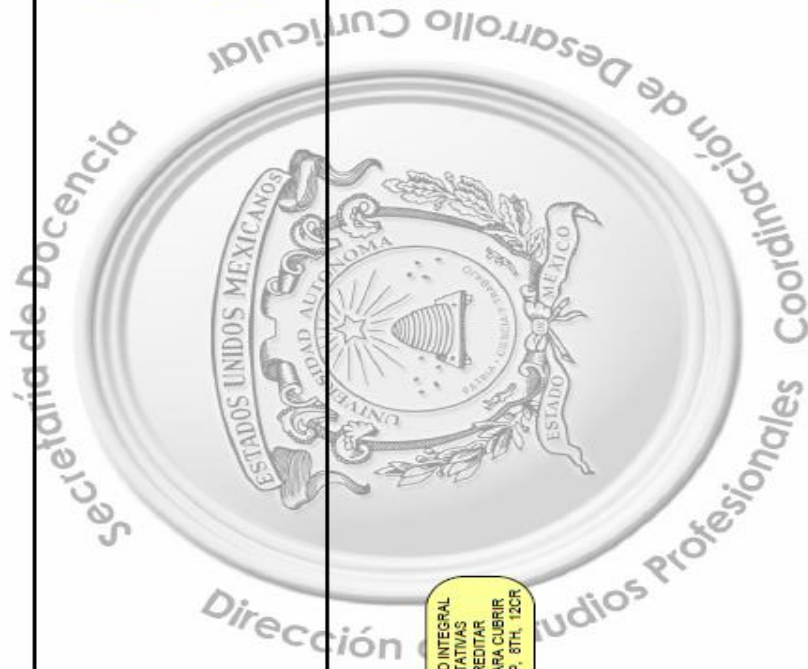
HT = Horas teóricas

HP = Horas prácticas

TH = Total de horas

CR = Créditos

NUCLEO INTEGRAL OPTATIVAS ACREDITAR  
2 UA PARA CUBRIR  
4HT, 4HP, 8TH, 12CR

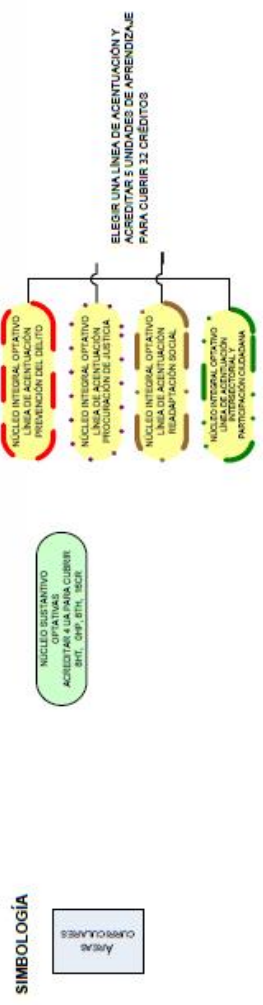
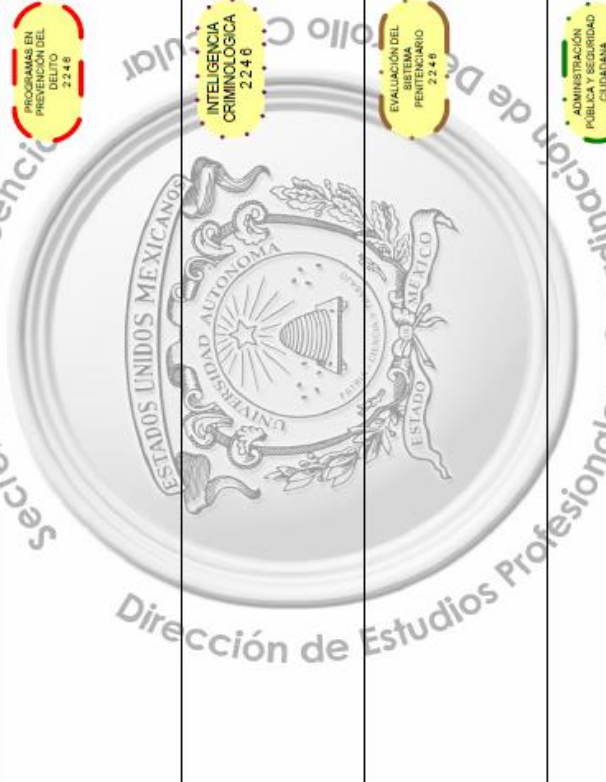






DESCRIPCIÓN DE LAS UNIDADES DE APRENDIZAJE OPTATIVAS DE LA LICENCIATURA EN SEGURIDAD CIUDADANA 2008

CIENCIAS SOCIALES	ANTROPOLOGÍA SOCIAL 4 048	GRUPOS VULNERABLES 4 048
CIENCIAS ECONÓMICAS	DISEÑO Y EVALUACIÓN DE PROYECTOS 4 048	
ADMINISTRACIÓN DE JUSTICIA	MARCO NORMATIVO EN ADMINISTRACIÓN DE JUSTICIA 4 048	
PREVENCIÓN DE DELITO	PROGRAMAS EN PREVENCIÓN DEL DELITO 2248	CRIMINALIDAD EN MEXICO 2248
PROCURACIÓN DE JUSTICIA	INTELIGENCIA CRIMINOLOGICA 2248	DIAGNÓSTICO Y PREVENCIÓN DE LA DELINCUENCIA 2248
READAPTACIÓN SOCIAL	EVALUACIÓN DEL SISTEMA PENITENCIARIO 2248	MARCO NORMATIVO EN PREVENCIÓN DEL DELITO 4048
INTERSECTORIAL Y PARTICIPACIÓN CIUDADANA	INVESTIGACIÓN PÚBLICA SEGURIDAD CIUDADANA 2248	MÉTODOS DE INVESTIGACIÓN MINISTERIAL 2 2248
		MARCO NORMATIVO EN PROCURACIÓN DE JUSTICIA 4048
		MÉTODOS DE INVESTIGACIÓN MINISTERIAL 1 2248
		PROGRAMAS EN READAPTACIÓN SOCIAL 2248
		MONITOREO DE LA REINTEGRACIÓN SOCIAL Y LABORAL 2248
		PROBLEMAS CONTEMPORÁNEOS EN SEGURIDAD 2248
		PARTICIPACIÓN SOCIAL EN LA SEGURIDAD CIUDADANA 4048



En el presente anexo se agrega una tabla con las reglas con mayor confianza, tanto para reglas de Baja como para reglas de Culminación, se anexan 30 reglas de cada tamaño, desde tamaño 2 a tamaño 7, dado que, para las reglas de tamaño 8 y 9 las reglas obtenidas son solo de Baja, y son reglas con menor número.

Confianza	Regla tamaño 2 Baja
38.52%	Física Título Suspendido, Matemáticas Discretas Extra Suspendido
31.96%	Física Título Suspendido, Matemáticas Discretas Título Suspendido
30.32%	Algorítmica Extra Suspendido, Algorítmica Título Suspendido
30.32%	Lógica Digital Extra Suspendido, Lógica Digital Título Suspendido
27.86%	Algorítmica Título Suspendido, Física Título Suspendido
27.86%	Teoría de Automatas y Lenguajes Formales Extra Suspendido, Teoría de Automatas y Lenguajes Formales Título Suspendido
27.04%	Algorítmica Título Suspendido, Física Extra Suspendido
26.22%	Algorítmica Extra Suspendido, Física Título Suspendido
25.81%	Algebra Superior Extra Suspendido, Algebra Superior Título Suspendido
25.81%	Algorítmica Extra Suspendido, Física Extra Suspendido
25.00%	Algebra Superior Título Suspendido, Física Extra Suspendido
25.00%	Lógica Digital Ordinario Suspendido, Lógica Digital Extra Suspendido
24.59%	Circuitos Eléctricos Extra Suspendido, Circuitos Eléctricos Título Suspendido
24.59%	Matemáticas Discretas Ordinario Suspendido, Matemáticas Discretas Extra Suspendido
24.59%	Matemáticas Discretas Ordinario Suspendido, Matemáticas Discretas Título Suspendido
24.18%	Lógica Digital Ordinario Suspendido, Lógica Digital Título Suspendido
23.36%	Física Extra Suspendido, Matemáticas Discretas Extra Suspendido
23.36%	Física Extra Suspendido, Matemáticas Discretas Título Suspendido
22.95%	Teoría de Automatas y Lenguajes Formales Ordinario Suspendido, Teoría de Automatas y Lenguajes Formales Extra Suspendido
22.54%	Programación de Microcontroladores Extra Suspendido, Programación de Microcontroladores Título Suspendido
21.72%	Algorítmica Ordinario Suspendido, Algorítmica Extra Suspendido
21.72%	Algorítmica Ordinario Suspendido, Algorítmica Título Suspendido
21.72%	Probabilidad y Estadística Extra Suspendido, Probabilidad y Estadística Título Suspendido
21.72%	Probabilidad y Estadística Ordinario Suspendido, Probabilidad y Estadística Título Suspendido
21.31%	Circuitos Eléctricos Ordinario Suspendido, Circuitos Eléctricos Extra Suspendido
21.31%	Física Título Suspendido, Matemáticas Discretas Ordinario Suspendido
20.90%	Física Extra Suspendido, Matemáticas Discretas Ordinario Suspendido
20.90%	Teoría de Automatas y Lenguajes Formales Ordinario Suspendido, Teoría de Automatas y Lenguajes Formales Título Suspendido
20.49%	Física Título Suspendido, Introducción a la Computación Ordinario Suspendido
20.49%	Introducción a la Computación Ordinario Suspendido, Matemáticas Discretas Título Suspendido
Confianza	Regla tamaño 2 Culminación
76.22%	Requisitos y Especificación de Software Ordinario Aprobó, Análisis y Diseño de Software Ordinario Aprobó
67.21%	Métodos y Modelos de Desarrollo de Software Ordinario Aprobó, Arquitectura de Software Ordinario Aprobó
66.80%	Programación de Microcontroladores Ordinario Aprobó, Requisitos y Especificación de Software Ordinario Aprobó
65.98%	Análisis y Diseño de Software Ordinario Aprobó, Fundamentos de Base de Datos Ordinario Aprobó
65.57%	Fundamentos de Base de Datos Ordinario Aprobó, Métodos y Modelos de Desarrollo de Software Ordinario Aprobó
63.11%	Fundamentos de Base de Datos Ordinario Aprobó, Teoría de Algoritmos Ordinario Aprobó
63.11%	Teoría de Algoritmos Ordinario Aprobó, Arquitectura de Software Ordinario Aprobó
62.70%	Análisis y Diseño de Software Ordinario Aprobó, Teoría de Algoritmos Ordinario Aprobó
62.70%	Programación de Microcontroladores Ordinario Aprobó, Interacción Humano Computadora Ordinario Aprobó

62.29%	Análisis y Diseño de Software Ordinario Aprobó, Métodos y Modelos de Desarrollo de Software Ordinario Aprobó
61.47%	Interacción Humano Computadora Ordinario Aprobó, Fundamentos de Base de Datos Ordinario Aprobó
60.65%	Pruebas y Mantenimiento de Software Ordinario Aprobó, Teoría de Lenguajes y Programación Ordinario Aprobó
60.24%	Arquitectura de Software Ordinario Aprobó, Bases de Datos Avanzadas Ordinario Aprobó
59.42%	Arquitectura de Computadoras Ordinario Aprobó, Programación de Microcontroladores Ordinario Aprobó
57.78%	Redes y Telecomunicaciones Ordinario Aprobó, Métodos y Modelos de Desarrollo de Software Ordinario Aprobó
57.37%	Administración y Organización de Proyectos de Software Ordinario Aprobó, Herramientas para la Administración y Programación de Sistemas Ordinario Aprobó
57.37%	Arquitectura de Software Ordinario Aprobó, Desarrollo Web Ordinario Aprobó
57.37%	Arquitectura de Software Ordinario Aprobó, Pruebas y Mantenimiento de Software Ordinario Aprobó
57.37%	Fundamentos de Base de Datos Ordinario Aprobó, Redes y Telecomunicaciones Ordinario Aprobó
57.37%	Métodos y Modelos de Desarrollo de Software Ordinario Aprobó, Bases de Datos Avanzadas Ordinario Aprobó
56.96%	Requisitos y Especificación de Software Ordinario Aprobó, Fundamentos de Base de Datos Ordinario Aprobó
56.14%	Administración y Organización de Proyectos de Software Ordinario Aprobó, Calidad de Software Ordinario Aprobó
56.14%	Análisis y Diseño de Software Ordinario Aprobó, Redes y Telecomunicaciones Ordinario Aprobó
55.32%	Redes y Telecomunicaciones Ordinario Aprobó, Teoría de Algoritmos Ordinario Aprobó
54.91%	Teoría de Algoritmos Ordinario Aprobó, Bases de Datos Avanzadas Ordinario Aprobó
54.50%	Bases de Datos Avanzadas Ordinario Aprobó, Pruebas y Mantenimiento de Software Ordinario Aprobó
54.09%	Herramientas para la Administración y Programación de Sistemas Ordinario Aprobó, Seguridad Informática Ordinario Aprobó
53.68%	Desarrollo Web Ordinario Aprobó, Pruebas y Mantenimiento de Software Ordinario Aprobó
53.27%	Desarrollo Web Ordinario Aprobó, Teoría de Lenguajes y Programación Ordinario Aprobó
53.27%	Métricas de Software Ordinario Aprobó, Simulación Ordinario Aprobó
Confianza	Regla tamaño 3 Baja
37.70%	Física Título Suspendido, Matemáticas Discretas Extra Suspendido, Matemáticas Discretas Título Suspendido
34.83%	Física Extra Suspendido, Física Título Suspendido, Matemáticas Discretas Extra Suspendido
29.91%	Física Ordinario Reprobó, Física Extra Suspendido, Física Título Suspendido
26.22%	Algorítmica Título Suspendido, Física Extra Suspendido, Física Título Suspendido
26.22%	Física Extra Suspendido, Física Título Suspendido, Matemáticas Discretas Título Suspendido
25.81%	Algorítmica Extra Suspendido, Algorítmica Título Suspendido, Física Título Suspendido
25.40%	Algorítmica Extra Suspendido, Algorítmica Título Suspendido, Física Extra Suspendido
25.40%	Física Título Suspendido, Introducción a la Computación Ordinario Aprobó, Matemáticas Discretas Título Suspendido
25.40%	Física Título Suspendido, Matemáticas Discretas Ordinario Reprobó, Matemáticas Discretas Título Suspendido
24.59%	Algorítmica Extra Suspendido, Física Extra Suspendido, Física Título Suspendido
24.59%	Física Extra Suspendido, Física Título Suspendido, Matemáticas Discretas Ordinario Reprobó
24.18%	Algebra Superior Título Suspendido, Física Extra Suspendido, Física Título Suspendido
24.18%	Lógica Digital Ordinario Suspendido, Lógica Digital Extra Suspendido, Lógica Digital Título Suspendido
22.95%	Matemáticas Discretas Ordinario Suspendido, Matemáticas Discretas Extra Suspendido, Matemáticas Discretas Título Suspendido
22.54%	Física Extra Suspendido, Matemáticas Discretas Extra Suspendido, Matemáticas Discretas Título Suspendido
22.13%	Algebra Superior Extra Suspendido, Algebra Superior Título Suspendido, Física Extra Suspendido
22.13%	Algorítmica Título Suspendido, Física Título Suspendido, Matemáticas Discretas Extra Suspendido
21.72%	Algorítmica Ordinario Aprobó, Física Extra Suspendido, Física Título Suspendido
21.72%	Física Extra Suspendido, Matemáticas Discretas Ordinario Reprobó, Matemáticas Discretas Título Suspendido









