

BENEMÉRITA UNIVERSIDAD AUTÓNOMA DE PUEBLA

FACULTAD DE CIENCIAS DE LA COMPUTACIÓN



**SISTEMA PARA EL SOPORTE DE DECISIONES  
DE LA SECRETARÍA ACADÉMICA DE LA FCC**

TESIS PRESENTADA PARA OBTENER EL TÍTULO DE:

LICENCIATURA EN INGENIERÍA EN CIENCIAS DE LA COMPUTACIÓN

PRESENTA: MARIO ALBERTO POSADA ZAMORA

ASESOR: DR. ABRAHAM SÁNCHEZ LÓPEZ

COASESOR: M.C. YALÚ GALICIA HERNÁNDEZ

Diciembre 2015



Dedicado para mis padres, hermana, sobrino y amigos



## RESUMEN

---

En las tareas cotidianas de la planeación académica de cursos dentro de una universidad poblada, es necesario tener una buena idea sobre un cierto número preliminar de cursos a trabajar y con esto obtener una buena distribución; todo esto para saber la cantidad de recursos humanos y mobiliarios que son necesarios con la única finalidad de que la carga académica sea la adecuada, cubriendo varios criterios específicos.

En el presente trabajo de tesis se analiza lo complejo de la situación de elegir el número adecuado de cursos, los problemas que surgen al realizarlo y así también una serie de propuestas para poder enfrentar este problema haciendo uso de herramientas de software y de modelos matemáticos estadísticos para llegar a una buena estimación en una planeación escolar.

Se presenta la situación actual y algunos modelos estadísticos simples con ejemplos prácticos para demostrar la eficiencia de cada método, así también como la manera en la que se debe aplicar cada método en un conjunto de datos específico: las series de tiempo.

Posteriormente se abordan temas de aplicación sobre la teoría y en qué situaciones funciona mejor cada modelo y las características que los datos de entrada deben tener. De igual manera se aborda todo esto con ejemplos gráficos producidos con la herramienta propuesta en el presente trabajo.

Finalmente se realizan pruebas sobre un conjunto de datos aplicándole simultáneamente varias técnicas para contrastar la que se adecúa más a cada situación particular, se presentan casos específicos que denotan toda una clase entera de datos similares que pueden ser tratados de la misma manera, así también como un caso especial encontrado en el desarrollo del trabajo.

Por último se presentan las conclusiones al lector junto con una propuesta de la segunda etapa en el desarrollo de la propuesta: la generación y distribución de horarios. Dicha propuesta utilizaría los datos calculados a través del presente trabajo para realizar la planeación académica completa.

Dentro del desarrollo del sistema se encuentran algunas propuestas interesantes como el sistema de confianza que permite saber en qué medida un usuario es confiable para un sistema de acuerdo a una serie de reglas personalizadas.



# ÍNDICE GENERAL

---

II	
Resumen .....	I
Índice general.....	III
Índice de figuras .....	V
Índice de tablas .....	VI
1 Introducción .....	1
1.1 Problemática general .....	2
1.2 definición del problema.....	3
1.3 estado del arte.....	4
1.3.1 Software de planificación de horarios .....	4
1.3.2 Análisis de la demanda .....	5
1.4 Justificación .....	6
1.5 Contribución del trabajo.....	7
1.6 Organización del escrito de tesis.....	8
2 Sistemas de soporte de decisiones .....	9
2.1 Juicio humano en la toma de decisiones .....	9
2.1.1 Heurística del juicio en la toma de decisiones.....	10
2.1.1.1 Heurísticas representativas.....	10
2.1.1.2 Heurísticas de disponibilidad.....	10
2.1.1.3 Heurísticas de anclaje y ajuste .....	11
2.1.1.4 Heurística afectiva.....	11
2.2 ¿Qué es el soporte de decisiones? .....	12
2.2.1 Modelo de una decisión .....	12
2.2.2 Componentes de un modelo de decisión.....	12
2.3 sistemas de soporte de decisiones.....	13
2.3.1 Componentes de un sistema de soporte de decisiones.....	14
2.3.2 Tipos de sistemas de soporte de decisiones.....	15
2.4 Sistemas de soporte de decisiones basados en decisiones analíticas .....	16
2.4.1 Sistemas de soporte de decisiones basados en ecuaciones.....	16
3 Modelo matemático para la propuesta.....	19
3.1 Introducción al análisis de datos históricos.....	19

3.2	Modelos de predicción .....	20
3.3	Series de tiempo.....	21
3.4	Suavizado de una serie de tiempo .....	22
3.4.1	Promedio Móvil.....	25
3.4.2	Suavizados exponenciales.....	26
3.4.2.1	Suavizado simple .....	27
3.4.2.2	Predicción con suavizados.....	29
3.4.2.3	Suavizado doble .....	29
3.4.2.4	Suavizado triple.....	32
4	Diseño de la propuesta del sistema.....	35
4.1	Esquema general del sistema.....	35
4.2	Módulo de análisis histórico.....	36
4.2.1	Open Forecast .....	41
4.3	Módulo de encuesta.....	42
4.3.1	Struts 2 .....	45
4.4	Conjunción de resultados.....	46
5	Pruebas y resultados .....	47
5.1	Pruebas sobre cursos invariantes .....	47
5.2	Pruebas sobre cursos con rango de variación moderada .....	48
5.3	Pruebas sobre cursos con rango de variación alta .....	49
5.3.1	Caso particular: Formación Humana y Social.....	50
6	Conclusiones y trabajo futuro.....	53
7	Bibliografía .....	55

## ÍNDICE DE FIGURAS

---

FIGURA 1-1 VENTANA DE TRABAJO DE FET (FREE TIMETABLING SOFTWARE).....	6
FIGURA 2-1 ENTRADAS Y SALIDAS DE UN SISTEMA DE SOPORTE DE DECISIONES.....	14
FIGURA 2-2 COMPONENTES Y ACTORES DE UN SISTEMA DE SOPORTE DE DECISIONES.....	15
FIGURA 3-1 EJEMPLO DE UNA SERIE DE TIEMPO.....	22
FIGURA 3-2 EJEMPLO DE UNA SERIE DE TIEMPO CON POCA FLUCTUACIÓN EN VALORES.....	24
FIGURA 3-3 APLICACIÓN DE UN PROMEDIO MÓVIL SOBRE TRES DATOS CONTIGUOS.....	26
FIGURA 3-4 APLICACIÓN DE SUAVIZADO EXPONENCIAL SIMPLE EN UNA SERIE DE TIEMPO CON POCA VARIACIÓN.....	28
FIGURA 3-5 APLICACIÓN DE SUAVIZADO EXPONENCIAL DOBLE EN UNA SERIE DE TIEMPO CON VARIACIÓN MEDIA Y TENDENCIA.....	30
FIGURA 3-6 APLICACIÓN DE SUAVIZADO EXPONENCIAL TRIPLE EN UNA SERIE DE TIEMPO CON TEMPORALIDAD.....	34
FIGURA 4-1 COMPONENTES PRINCIPALES DEL SISTEMA PROPUESTO.....	36
FIGURA 4-2 SERIE DE TIEMPO SIN NINGÚN TIPO DE VARIACIÓN.....	37
FIGURA 4-3 SERIE DE TIEMPO CON RANGO DE VARIACIÓN PEQUEÑO.....	38
FIGURA 4-4 CAPTURA DE PANTALLA DEL MÓDULO DE ANÁLISIS HISTÓRICO.....	38
FIGURA 4-5 SERIE DE TIEMPO CON TENDENCIA NOTABLE.....	39
FIGURA 4-6 SERIE DE TIEMPO CON TENDENCIA LINEAL.....	40
FIGURA 4-7 SERIE DE TIEMPO CON TEMPORALIDAD MARCADA.....	41
FIGURA 4-8 APLICACIÓN DE PRUEBA UTILIZANDO OPENFORECAST.....	42
FIGURA 4-9 CAPTURA DE PANTALLA DEL MAPA DE ING. EN CIENCIAS DE LA COMPUTACIÓN PARA EL MÓDULO DE ENCUESTAS.....	44
FIGURA 4-10 FLUJO DE ACTIVIDADES DEL FRAMEWORK STRUTS 2.....	45
FIGURA 5-1 COMPARACIÓN DE TÉCNICAS EN UNA SERIE DE TIEMPO POCO VARIANTE.....	48
FIGURA 5-2 COMPARACIÓN DE TÉCNICAS EN UNA SERIE DE TIEMPO CON VARIACIÓN MEDIA.....	49
FIGURA 5-3 COMPARACIÓN DE TÉCNICAS EN UNA SERIE DE TIEMPO CON VARIACIÓN ALTA Y TEMPORALIDAD.....	50
FIGURA 5-4 COMPARACIÓN DE TÉCNICAS EN UNA SERIE DE TIEMPO CON VARIACIÓN MEDIA.....	51

## ÍNDICE DE TABLAS

---

TABLA 2-1 CAUSAS DE MUERTES Y SUS CIFRAS DE ACUERDO A DATOS REALES Y ESTIMADOS POR PERSONAS. ....	11
TABLA 2-2 COMPARACIÓN ENTRE SOPORTE DE DECISIONES Y ADMINISTRACIÓN DE DECISIONES. ....	13
TABLA 3-1 TABLA DE VALORES PARA EL CURSO COMPUTABILIDAD. ....	23
TABLA 3-2 TABLA DE VALORES PARA EL CURSO TRANSMISIÓN Y COMUNICACIÓN DE DATOS. ....	25
TABLA 3-3 TABLA DE VALORES PARA EL CURSO INNOVACIÓN Y TALENTO EMPRENDEDOR.....	28
TABLA 3-4 TABLA DE VALORES PARA EL CURSO ADMINISTRACIÓN DE PROYECTOS.....	30
TABLA 3-5 TABLA DE VALORES PARA EL CURSO MATEMÁTICAS ELEMENTALES.....	33

# 1 INTRODUCCIÓN

---

En la actualidad, la importancia de los sistemas de soporte de decisiones ha aumentado a lo largo de las últimas décadas, desde la creación de los conceptos esenciales.

Tomar decisiones en entornos muy complejos (tales como procesos industriales, operaciones organizacionales, unidades de control militar o control nuclear) generalmente pone a prueba nuestras capacidades cognitivas. Incluso aunque las interacciones individuales de todas las variables de un sistema pueden ser bien comprendidas, predecir como el funcionamiento organizacional va a reaccionar a una manipulación externa; como un cambio de políticas, es extremadamente difícil.

Hay una buena cantidad de evidencia empírica que indica que el juicio intuitivo humano y la toma de decisiones puede ser lo más distante a ser lo óptimo, y se deteriora aún más con la complejidad de las consecuencias y el estrés del ambiente. Debido a que en muchos de los entornos, la calidad de las decisiones es muy importante, ayudar a las deficiencias en el juicio humano ha tenido el foco de la ciencia durante varios años. Disciplinas como la estadística, la economía y la investigación de operaciones han ideado varios métodos para realizar elecciones racionales. Más recientemente estos métodos; ayudados por la ciencia de la información, la psicología cognitiva y la inteligencia artificial, han sido implementados en forma de programas de computadora. Estos entornos digitales reciben el nombre de Sistemas de Soporte de Decisión (DDS por su abreviación en inglés).

El concepto de estos Sistemas de Soporte de Decisión; en adelante SSD, es extremadamente extenso, muchos autores manejan diferentes puntos de vista de acuerdo a la especialidad de los mismos, sin embargo la mayoría coincide en que básicamente es un software que asiste al usuario en el juicio y la elección de actividades. Algunas personas utilizan los SSD como un sinónimo de los *sistemas basados en conocimiento*, que refieren así a sus intentos de formalizar el conocimiento de dominio para el razonamiento mecanizado.

En este trabajo de tesis se expone el caso de estudio de una parte de la planeación de cursos de la Facultad de Ciencias de la Computación de la Benemérita Universidad Autónoma de Puebla, la cual tiene lugar específicamente en el departamento de Secretaría Académica. Dicho caso de estudio se enfoca específicamente en el análisis del progreso de los estudiantes para poder ofertar un buen número de secciones por materia, de tal manera que la falta de cupos se minimice y se puedan asignar de una manera eficiente los recursos académicos.

El proyecto es un sistema que recibe información de diversas fuentes, el cual permite reflejar una predicción del número de secciones recomendado por cada curso para cada periodo, a través de algunas técnicas de análisis de datos y estadística, que al final permite asistir a la toma de decisiones de la Secretaría Académica.

Durante el presente escrito se tocarán temas en profundidad como los antes mencionados Sistemas de Soporte de Decisiones, el análisis de datos, y en particular, de datos históricos,

series de tiempo y algunas técnicas para trabajar con ellas, además de modelos de predicción para este tipo de estructuras.

### 1.1 PROBLEMÁTICA GENERAL

Debido al modelo académico que se ejerce en la Benemérita Universidad Autónoma de Puebla, el actual sistema de créditos, no es raro encontrar que dicho sistema presente un verdadero reto para la planeación académica de cada unidad de estudios. Para profundizar en ello haré una breve descripción del sistema.

El sistema de créditos básicamente consiste en que cada curso dentro de la currícula del estudiante tiene un determinado valor en función de alguna característica del curso (particularmente se toma el número de horas totales a la semana como créditos). Para que el estudiante culmine sus estudios es necesario que cubra la totalidad de los créditos mínimos establecidos para cada programa.

Anteriormente se manejaba un cierto límite de número de créditos por cada periodo escolar, por lo que el estudiante no tenía permitido colocar más de 30 horas a la semana, sin importar el número de materias para conformar dicha restricción. El sistema actual está restringido por cierto número de materias que establece cada unidad académica.

Aunque el sistema haya sufrido modificaciones, la base es la que provoca el problema del que queremos tratar. El estudiante puede no tener un orden en los cursos que elige para cada periodo, debido a que puede seleccionar cualquier curso siempre y cuando cumpla con los prerrequisitos que le exige (cursos precedentes o cierto porcentaje de progreso). Esto al principio puede no significar mucho, sin embargo debido al gran número de combinaciones que se pueden formar una vez se avanza hasta casi la mitad de los créditos totales, es bastante pesado realizar una planeación de cursos de tal manera que satisfaga las necesidades de todos los estudiantes.

Dicho problema está presente en las unidades académicas, generando además de disconformidad, atraso y rezago en estudiantes que quizá no tengan una disponibilidad de horario completa para dedicar a sus estudios.

El problema se agrava cuando los encargados de la planeación académica no toman en cuenta el número de candidatos potenciales que podrían aspirar a cierto curso, lo cual podría resultar en una falta de cupos para las secciones que se hayan aperturado en el periodo, lo cual a su vez también puede provocar una sobrepoblación en los cursos.

El tamaño de los grupos es un factor determinante en la calidad de los cursos impartidos de manera presencial. Para un profesor, esto complica la aplicación de las técnicas y procedimientos de enseñanza activa tales como el seguimiento individual, la participación uniforme, la resolución de dudas, además de no poder uniformizar la motivación estudiantil y de resultado obtener un grupo totalmente heterogéneo [1].

Si bien existen diversas discusiones acerca de técnicas que podrían funcionar con grupos numerosos (más de 45 estudiantes), el problema que genera esta sobrepoblación no está

completamente resuelto, por lo que esta es una solución temporal. Sin embargo, contando el número de profesores disponibles es una opción que se podría considerar.

De igual manera algo que podría complementar el problema es el número de estudiantes admitidos cada año, lo cual si no se cuenta con la suficiente infraestructura, podría llevarnos a desencadenar avances dispares y muchos más atrasos y rezagos de los esperados.

## 1.2 DEFINICIÓN DEL PROBLEMA

Cada generación de alumnos a lo largo de su estancia en la Facultad de Ciencias de la Computación sufre de un importante rezago en cuanto al avance de sus créditos cursados de por lo menos el 60%, estando la gran parte estancados cerca de la mitad del total de cursos destinados a cada carrera (colocar porcentajes).

Una de las principales causas de este problema es la mala planeación, tal como se ha planteado en la sección anterior. Específicamente el problema se suscitó en el momento donde entra un nuevo plan de estudios con necesidad de aperturar cursos diferentes a los que se han estado ofertando.

La planeación de los cursos no es una tarea nada sencilla, puesto que es una tarea de satisfactibilidad de múltiples criterios, ya sea desde los alumnos, espacios disponibles y profesores. Dicha tarea es una de las tareas más complejas en muchas áreas, incluida la computación, en la que este tipo de problemas pertenece a la familia de problemas NP-Complejos. Probablemente para poder realizar la mejor planeación es necesario explorar todas las posibilidades, sin embargo, no podemos comenzar la exploración si desde un inicio no se toman en cuenta el número correcto de cursos a ofertar, lo cual es el problema al que intenta proponer una solución el presente trabajo de tesis.

En ciertas ocasiones, existen algunos factores que pueden alterar una planeación una vez implementada, como el abandono de los cursos por parte de los alumnos, suceso que es muy común en dicha facultad. Otro de los fenómenos, sin embargo más raros, es la ausencia del profesor en el curso por diversos motivos propios de él.

Cuando sucede lo primero, causa que el número de secciones sea mayor a la demanda real, debido a que dicha demanda disminuyó sobran espacios y tanto se pudo haber previsto un menor número de secciones para aperturar otras, o poder acomodar esta sección en otro espacio más acorde al número de inscritos que permanecerán hasta el final del periodo.

Si ocurre lo segundo, es mucho más problemático, debido a que se debe buscar profesores que impartan el curso vacío, sin embargo aquí ya no se busca en criterio de calidad, sino en ver quién se encuentra disponible en la franja horaria en la que fue programado dicho curso.

No es posible que ninguno de estos problemas arriba descritos sea subsanado de manera eficaz en la fase de planeación, debido a que son sucesos no planeados, o incertidumbres, de los cuales no existen modelos matemáticos que determinen con exactitud lo que pasará con ellos, sin embargo sí existen modelos que sirven, en base a lo ocurrido y a un factor de probabilidad, intentar aproximar de la mejor manera lo que podría ocurrir con cierto grado de incertidumbre [2].

Es por esto el propósito del presente trabajo, en base a la observación de planeaciones anteriores, además de otras fuentes externas, se determina una posible incertidumbre sobre la planeación anterior y tratar de determinar el número más adecuado de secciones que se deben ofertar en cada periodo.

### 1.3 ESTADO DEL ARTE

Antes de hablar del estado del arte, es preciso aclarar algunos puntos acerca del objetivo de esta tesis, para explicar mejor la clasificación de los trabajos relacionados al área.

Dentro de la planeación participan dos fases muy importantes, que bien actualmente se han estado haciendo de manera conjunta, esta vez se pretende que dichas fases se procesen de manera separada.

La primera fase es la selección del número adecuado de secciones por ofertar, la cual se decidía mientras se iban planeando los cursos.

La segunda fase es la asignación de un horario y un espacio físico de acuerdo a la disponibilidad global de la facultad.

Comenzaremos por los hablar sobre el estado del arte en la segunda fase de planeación para posteriormente pasar al tema que nos compete, la primera fase.

#### 1.3.1 Software de planificación de horarios

Existe una gran variedad de productos que permiten realizar de manera sencilla una planeación de horario escolar directamente implementable.

Algunos ejemplos son:

- Timetable Web
- Automatic Timetable
- Make Your Time Table
- Prime Table
- Asc Timetables
- Free Timetabling Software (FET)<sup>1</sup>

Si bien el número de ejemplos es reducido, existe una enorme variedad de sistemas de este tipo que son auxiliares en el trabajo de la planeación de horarios. La mayoría herramientas de pago, aunque se encuentran algunas gratuitas de manera inicial o permanente, e incluso libres, como el caso de la última herramienta.

En primera instancia, si alguien está dispuesto a utilizar este tipo de herramientas, ya debe conocer la demanda actual de los cursos. Es decir el número de cursos a ofertar ya debe ser conocido de antemano, además de los recursos humanos y espacios físicos para poderlos impartir.

---

<sup>1</sup> FET Free Timetable Software. [10]/

Lo que hacen todos estos programas es asociar estos recursos entre sí y acomodarlos de tal manera que no existan empalmes entre ellos, además de cumplir ciertas restricciones particulares al dominio de la institución.

Algunas desventajas que se pueden enlistar sobre este tipo de software, excluyendo un poco al último ejemplo son:

- Número de profesores pequeño
- Organización por grupos de estudiantes
- Número pequeño de grupos
- Horarios bien definidos
- Es necesario en algunos que los estudiantes permanezcan en las aulas

Con estas restricciones y dada la descripción del problema de la facultad, es posible que este tipo de softwares sirvan más a una institución de nivel básico, medio superior, o superior pero con grupos organizados de manera similar a los niveles anteriores.

Por observación, se denota que nuestro problema no se acota por este tipo de software, que además solo realiza una parte del trabajo completo del proceso de planeación.

En el caso de la última herramienta (FET) es posible realizar algo más acorde a nuestras necesidades, sin embargo, para poder adaptarla al problema es necesario modificar el software a nivel de programación, lo cual básicamente haría que realizáramos más trabajo adaptándolo que lo que realmente nos tomaría desarrollar una plataforma hecha a la medida.

Cabe destacar que para que estos productos de software funcionen es necesario que ya se conozca el número de cursos que se deben aperturar, lo cual suponemos en un principio que es un valor desconocido.

### 1.3.2 Análisis de la demanda

Hemos visto algunos tipos de software que asisten el proceso pero hasta ahora no hemos podido ver que alguno ataque realmente el problema raíz acerca del número de secciones correcto para cada periodo.

Dentro de la facultad, ya se han llevado a cabo algunas estrategias para atacar este problema de manera más eficaz.

En el año 2011 se realizó una herramienta web la cual pretendía encuestar a los estudiantes acerca de los cursos que pretendían tomar para cada periodo con el objetivo de poder obtener estadísticas que fueran auxiliares en el proceso de decisión del número de secciones.

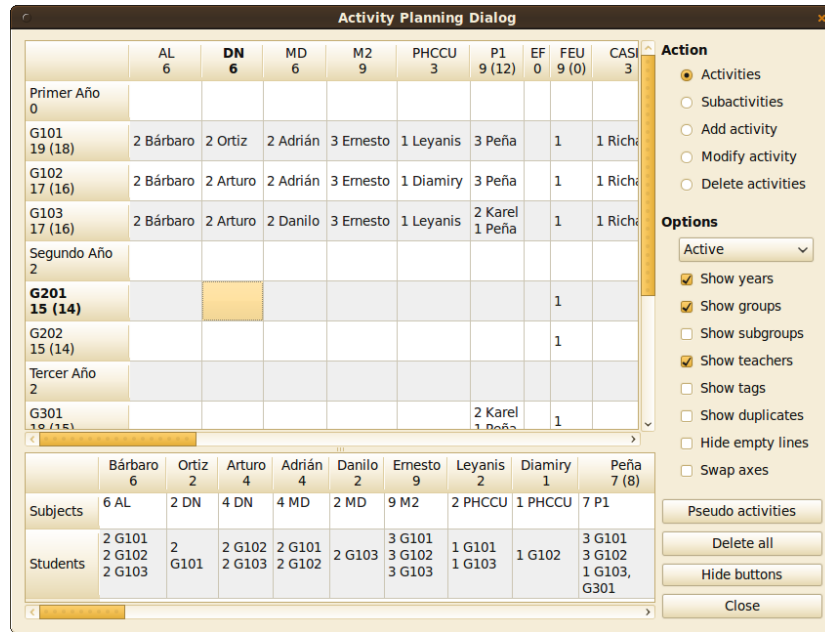


Figura 1-1 Ventana de trabajo de FET (Free Timetabling Software).

Este proyecto funcionó durante un periodo, sin embargo, por diversos motivos no se le dio continuidad, por lo que pronto el trabajo quedó descartado.

Esta aproximación fue una buena idea, sin embargo, para el autor del presente trabajo, esta herramienta carecía de algunas medidas robustas que permiten hacer la información resultante más fiable.

El primer escenario que analizaremos es en el que la participación de la herramienta fuera baja (lo cual ocurrió), al final esto determina una estadística bruta que realmente no sirve de mucho, debido a que representa menos de la mitad de la comunidad estudiantil.

El segundo escenario es acerca de la exactitud de los datos registrados, donde se puede dar el caso en el que los estudiantes puedan mentir sobre sus planeaciones, lo cual concluiría en información completamente errónea y poco útil para el proceso, además de poner en riesgo al mismo.

A pesar de todo lo que pudiera no haber salido bien, se decidió tomar en cuenta este trabajo inicial de hace varios años para implementarlo de nuevo, sin embargo con un enfoque diferente el cual asegura aminorar estas fallas, lo cual establece una base más segura para el análisis de las estadísticas de la facultad.

## 1.4 JUSTIFICACIÓN

La finalidad de este trabajo es auxiliar el proceso de selección de cursos por periodo, como ya se había descrito en secciones anteriores, aunque la finalidad real es la de continuar con el trabajo realizado en el año 2011, implementando mejoras en él, o más bien rediseñarlo de manera que sea más útil y eficiente para la Secretaría Académica de la facultad.

El principal aporte es el de poder dar una herramienta no solo al personal administrativo, sino también a los estudiantes para poder expresar realmente sus necesidades, como lo era el objetivo del trabajo anterior, pero esta vez con los medios para poder filtrar la información útil de la que no lo es.

La idea de retomar un trabajo anterior surge por la observación del autor a la disconformidad actual de los estudiantes, del cual fue parte. Dicha disconformidad ha disminuido a lo largo de los años con el cambio del personal a cargo del proceso, sin embargo, todavía se pueden seguir realizando esfuerzos para mejorar este proceso con el fin de ofrecer a los estudiantes una flexibilidad de estudio que se adapte a sus necesidades cotidianas.

Otro de los problemas que se pretende erradicar es el de la realización de esta tarea de forma manual, debido a que propiciaba errores, además de ser un proceso demasiado exhaustivo y cansado para quienes lo realizan, pudiendo perdurar varios días hasta poder sacar una versión aceptable de horarios. Finalmente ese es un enfoque de la computación en general, el poder brindar una herramienta que sustituya el trabajo manual, que aparte de tedioso, pueda ser difícil o imposible de realizarse por un humano y pueda propiciar error

A la vez del interés del autor por aplicar conocimientos obtenidos a lo largo de su formación para subsanar un problema del cual formaba parte día con día, lo cual guiado por la empatía permite realizar esta propuesta.

## 1.5 CONTRIBUCIÓN DEL TRABAJO

En este trabajo se realizaron algunas investigaciones sobre la recolección, el manejo de datos y el análisis de los mismos, todo ello para dar como resultado un sistema de soporte de decisiones para la asistencia de la planeación escolar, por lo que trataremos temas en cada capítulo de esta tesis relacionados a cada fase del desarrollo.

- **Soporte de decisiones** enfocado a la educación y particularmente a la planeación de recursos escolares y horarios.
- **Sistemas de soporte de decisiones** y sus diversos tipos de acuerdo a su función o sus estrategias.
- **Análisis de datos históricos** que describen el comportamiento de un suceso a lo largo de un periodo de tiempo establecido.
- **Modelos de predicción** sobre sucesos anteriormente ocurridos, esto analizando datos históricos.
- **Series de tiempo** describiendo una sucesión de eventos de manera matemática con una serie de números de longitud definida por el periodo de tiempo a analizar.
- **Estrategias de predicción** sobre series de tiempo y los tipos de técnicas utilizadas.

Posteriormente se plantea el desarrollo y la aplicación de las técnicas descritas anteriormente para adaptarlas a un sistema de soporte de decisiones basado en estadística y en series de tiempo para predecir el número de secciones por periodo.

Se revisarán las propuestas del autor para el complemento del trabajo anterior que data del 2011, para intentar corregir los fallos que llevaron al abandono del proyecto.

Así mismo el trabajo se complementará con propuestas de parte del autor para la corrección de los datos basados en el modelo específico de la recolección de información para cierto porcentaje de participación.

## 1.6 ORGANIZACIÓN DEL ESCRITO DE TESIS

Para dar un breve resumen acerca de lo que se abordará a lo largo del presente trabajo se presentan a continuación un listado de cada capítulo con una breve reseña que lo describe.

- **Capítulo 1: Introducción**  
En este capítulo ya visto se tratan aspectos generales del problema, algunos alcances y la organización del trabajo de tesis.
- **Capítulo 2: Sistemas de Soporte de Decisiones**  
Se tratan temas acerca de la teoría de decisiones desde el punto de vista psicológico humano, además de tópicos generales de teorías de este tipo de sistemas y sus arquitecturas.
- **Capítulo 3: Modelo matemático para la propuesta**  
Se presentan diferentes modelos para realizar análisis en series de tiempo, además de presentar específicamente los utilizados en el presente trabajo, además de presentar ejemplos utilizando datos reales pertenecientes al problema que se desea tratar.
- **Capítulo 4: Diseño de la propuesta del sistema**  
Se expone el diseño más a fondo, enfocado a cómo toma las decisiones el sistema para determinar el número adecuado de secciones, además de presentar la arquitectura general del sistema y la descripción de los módulos principales que lo conforman.
- **Capítulo 5: Pruebas y resultados**  
Se presentan resultados más detallados obtenidos a partir de la aplicación de predicciones sobre periodos para estimar una planeación en el periodo Otoño 2015
- **Capítulo 6: Conclusiones y trabajo futuro**  
Se propone la segunda parte del presente trabajo para realizar la planeación completa de la carga académica y su alcance, además de realizar las pertinentes conclusiones a este trabajo.

## 2 SISTEMAS DE SOPORTE DE DECISIONES

---

Hablando de manera más general a partir de este capítulo, los humanos continuamente necesitan tomar decisiones a lo largo de su vida. Este proceso es natural y generalmente es basado en el juicio propio, sin embargo a veces nuestra visión limitada no nos permite ver más allá de ciertas fronteras, lo que resulta en decisiones limitadas con resultados inesperados o mediocres, lo cual hace necesario que existan modelos y sistemas que nos ayuden a contemplar mejor otros parámetros para una mejor decisión.

### 2.1 JUICIO HUMANO EN LA TOMA DE DECISIONES

Estudios teóricos en la toma de decisiones racional, notablemente en el contexto de la teoría de la probabilidad y la teoría de la decisión, han sido acompañados por la investigación empírica sobre lo que el comportamiento humano logra con esta teoría. Ha sido más convincente la demostración en numerosos estudios empíricos que el juicio humano y la toma de decisiones, ambos están basados en estrategias intuitivas, lo que es opuesto a las reglas de razonamiento teóricas que se han establecido.

Estas estrategias intuitivas, referidas como heurísticas del juicio (del inglés *judgemental heuristics*) en el contexto de la toma de decisiones, ayudan a reducir la carga de tener que razonar de manera exhaustiva todas las posibilidades, sin embargo esto puede llevarnos muy lejos de tomar la decisión óptima en cada situación y contexto.

Claramente, nuestro juicio inasistido y nuestras decisiones rompen las reglas de los axiomas de la probabilidad (conocidos como *biases*).

Uno podría esperar que las personas que han alcanzado cierto nivel de conocimientos en un dominio no sean sujetos de este juicio simple y van a acercarse a decisiones más próximas a la óptima [3]. Mientras la evidencia empírica muestra que los expertos de hecho son más acertados que los novatos dentro de su área de dominio, también demuestra que también los expertos son igual de propensos al mismo juicio simple que los novatos y demuestran errores aparentes e inconsistencias en sus decisiones.

Los profesionales como los que pertenecen al área de la salud utilizan esencialmente las mismas heurísticas de juicio y son propensos a tener preferencias erróneas, mientras que el grado que normalmente se prevé en dichos estudios parece decrecer con la experiencia.

Sin embargo, en adición a la evidencia del fenómeno, varios estudios sobre el desempeño del dominio de los expertos en situaciones tomadas de la vida real, muestran que el juicio de los expertos es menos eficaz que los modelos lineales más simples<sup>4</sup>. Por ejemplo, las predicciones del comportamiento violento futuro en pacientes psiquiátricos hecho por un panel de psiquiatras que tuvo acceso al historial del paciente, además de entrevistas que fueron hechas por este panel.

Las predicciones hechas por el grupo de expertos fueron menos certeras que las producidas por un modelo simple que sólo incluía las incidencias pasadas de los pacientes.

Las predicciones hechas por consejeros maritales sobre la felicidad en el matrimonio resultaron ser inferiores a un modelo simple que sólo tomó en cuenta la diferencia entre el rango de peleas y el rango de encuentros sexuales (De nuevo los expertos tuvieron acceso a todos los datos que solicitaron, incluso a entrevistas con las parejas).

Toda esta ejemplificación denota que en muchas ocasiones es necesario contar con un modelo que nos permita realizar predicciones para complementar nuestro juicio que generalmente se ve sesgado y no es completamente objetivo.

### 2.1.1 Heurística del juicio en la toma de decisiones

Las heurísticas, dentro de la psicología, son reglas simples que usualmente la gente utiliza para realizar y apoyar juicios y decisiones. Generalmente se denotan como atajos mentales a procesamientos duros, los cuales conllevan a la gran concentración hacia un aspecto de un problema complejo, ignorando los otros aspectos y los contextos.

En un trabajo de investigación [4] se demostró la falla de las heurísticas bajo ciertas situaciones controladas. En la sección anterior comentamos acerca de este hecho, sin embargo, está comprobado que las heurísticas no son del todo malas, debido a que ahorran recursos cerebrales y ayudan a decisiones más rápidas.

#### 2.1.1.1 Heurísticas representativas

Cuando realizamos un juicio basado en la apariencia de una persona, sin explorar más allá de ciertos rasgos deducibles de manera instantánea, solemos realizar un juicio por heurísticas representativas.

Una heurística de este tipo se da cuando utilizamos patrones y grupos que son familiares para el cerebro, catalogando y etiquetando a todo lo que se involucre en nuestras decisiones.

Mientras este tipo de heurísticas son buenas para algunas aplicaciones, tales como para juzgar la seguridad de un establecimiento en base a su apariencia, para otros tipos de juicios es posible que sólo nos concentremos en ciertas características sin tomar en cuenta cada aspecto unido como un todo y la manera en la que funcionan estos aspectos en conjunto.

#### 2.1.1.2 Heurísticas de disponibilidad

Cuando solemos dar un juicio a partir de un escenario desarrollado en nuestra mente de manera detallada estamos utilizando la heurística de disponibilidad.

Esta heurística es particular a los casos en los que la gente está más abrumada sobre las muertes ocasionadas por tornados o el terrorismo. Ambos eventos son casos raros por los cuales una persona podría morir, sin embargo, como son sucesos bastante dramáticos la gente tiende a recordarlos más, por lo que cuando estiman muerte, observan esta clase de escenarios. Por otro lado, escenarios simples y más comunes como la muerte por suicidio, infartos y diabetes.

Los primeros generalmente están más presentes en la mente debido a que la información que recibimos de fuentes externas generalmente se concentra en este tipo de eventos, lo que asegura su permanencia en nuestro subconsciente.

Todo esto explica por qué la gente se deja llevar más fácil por una historia vívida dramática a un conjunto de datos estadísticos. Este concepto se aplica al comprar un boleto de lotería, la gente sigue comprando debido a que la muy reducida minoría que obtiene miles o millones de dólares es más convincente que los miles o millones de personas que no obtienen algún premio.

De la tabla siguiente ¿Cuáles se consideran las causas de muerte más comunes de acuerdo a la información que tenemos en mente?

**2.1.1.3 Heurísticas de anclaje y ajuste**

Se consideran los pares homicidios y enfisema pulmonar, y el otro par, accidentes de tráfico y cáncer de pulmón. De entre cada pareja se le pidió a varias personas en los Estados Unidos que diera su opinión sobre la causa de muerte más común entre las dos opciones. Estos fueron los resultados.

Causas de muerte	Elección de la gente	Totales anuales en USA	Reportajes en periódicos anuales
Homicidios	55%	19000	264
Enfisema Pulmonar	45%	22000	1
Accidentes de tráfico	57%	46000	127
Cáncer de pulmón	43%	140000	3

Tabla 2-1 Causas de muertes y sus cifras de acuerdo a datos reales y estimados por personas.

Este tipo de heurísticas toman lugar al momento de estimar un número [5]. De acuerdo a la descripción original de Tversky y Kahneman, generalmente se comienza por un número que se encuentra disponible en nuestra mente, lo que se denomina como “el ancla” o “el gancho”, para después intentar ajustar este número hacia arriba o hacia abajo.

En los experimentos sobre este fenómeno, la gente no modificaba la cifra del ancla de manera significativa, por lo que se intuye que el ancla contamina el estimado, incluso si es claramente irrelevante. Se piensa de primera instancia que el ancla se elige de manera aleatoria, sin embargo algunas observaciones podrían denotar que estos números se determinan en base a experiencias, las cuales están limitadas a lo que la mente pueda recordar en ese momento.

**2.1.1.4 Heurística afectiva**

Las decisiones pueden verse afectadas por sensaciones experimentadas en las personas, tales como el miedo, el placer y la sorpresa. Básicamente se refiere a todo tipo de emociones que pueden alterar la calidad de las decisiones de una persona.

Aunque los episodios de algunas emociones son cortos, esto no detiene el hecho de poder modificar de manera instantánea y permanente una decisión que podría no estar del todo correcta.

Cuando nosotros percibimos un lugar que nos causa algún tipo de nostalgia o nos trae malos recuerdos, preferimos evitarlo, así podamos tomar una vía corta en caso de requerir movernos de un espacio a otro de manera eficiente.

## 2.2 ¿QUÉ ES EL SOPORTE DE DECISIONES?

En la sección 2.1 estudiamos aspectos generales sobre el factor humano en la toma de decisiones, todo esto para percatarnos que existen muchos factores externos a los eventos principales que nos llevan a tomar decisiones erróneas.

Con el propósito de justificar el presente trabajo, se expone la idea de un sistema que ayude a contemplar los datos que quizá podríamos estar ignorando o viendo de mala manera.

Para comenzar a indagar sobre la teoría de soporte de decisiones debemos conocer algunos conceptos fundamentales como el modelado de las decisiones y sus componentes.

### 2.2.1 Modelo de una decisión

La superioridad de modelos lineales tan simples sobre la intuición humana sugiere que una manera de mejorar la calidad de las decisiones es descomponer un problema de juicio en componentes más simples que están bien definidos y bien entendidos. Este proceso básicamente se conoce como modelado.

Construir un modelo de un problema de decisión, opuesto al razonamiento sobre un problema de una manera superficial, permite aplicar conocimiento científico que puede ser transferido entre problemas y sus dominios cruzados.

### 2.2.2 Componentes de un modelo de decisión

Mientras que matemáticamente un modelo consiste de variables y una especificación de la interacción entre ellas, desde el punto de vista de la toma de decisiones un modelo y sus variables representan los siguientes tres componentes: una medida de preferencias sobre los objetivos de la decisión, opciones de decisión disponibles, y una medida de incertidumbre sobre las variables influenciando la decisión y sus consecuencias.

La preferencia es ampliamente vista como el concepto más importante de la toma de decisiones. Los resultados de un proceso de decisión no son todos igualmente atractivos y es crucial para el experto examinar estos resultados en términos de qué tan deseables son.

Las preferencias pueden ser cuantificadas de manera simple (por ejemplo, más ganancias son preferidas a menos ganancias), pero se considera conveniente y muchas veces necesario representarlas como cantidades numéricas, especialmente si los resultados se conforman de muchos atributos que necesitan ser comparados en una escala. Expresar los resultados en una escala numérica permite expresar de manera certera diferenciar entre cifras deseables contra cifras consideradas de riesgo.

El segundo componente de los problemas de decisión son las opciones disponibles de decisión. A menudo, estas opciones pueden ser enlistadas (como una lista posible de proveedores para un establecimiento por ejemplo), pero a veces son valores continuos de variables sujetas a políticas de un proceso (como la cantidad de un material que debe ser conservado en almacén sin agotarlo). Listar todas las opciones disponibles es un elemento importante para el modelado de la estructura.

Soporte de decisiones	Administración de decisiones
Confía en que el usuario tiene experiencia y juicio	Confía en que la experiencia está correctamente embebida en sí
Debe presentar su análisis para que una persona lo use	Debe ser capaz de aplicar su análisis programáticamente
Responsivo para lecturas y aplicar regulaciones y normas	Las decisiones que toman son inherentemente fijas
Cada usuario tiene cierto grado de libertad	El proceso de decisión es aplicado todo el tiempo, de manera programada y repetible

Tabla 2-2 Comparación entre soporte de decisiones y administración de decisiones.

El tercer elemento de la toma de decisiones es la incertidumbre. La incertidumbre es un concepto que se acarrea al lado de la teoría del conocimiento, originándose de la incompletitud de información, imprecisión y modelos de aproximaciones hechos simplemente por el afán de la simplicidad. No sería una exageración decir que las decisiones reales no involucrando cierto grado de incertidumbre no existen o pertenecen a una clase muy limitada.

La toma de decisiones bajo incertidumbre puede verse como una deliberación: determinando qué acción debería tomarse que maximice la ganancia esperada. Debido a la incertidumbre no hay garantía de que el resultado de la acción va a ser el que se espera, y lo mejor que se puede esperar es maximizar la probabilidad de un resultado deseable.

### 2.3 SISTEMAS DE SOPORTE DE DECISIONES

Los sistemas de soporte de decisiones son sistemas interactivos basados en la computación, que ayudan a los usuarios en las actividades de juicio y elección. Proveen almacenamiento y consulta de datos pero mejoran el acceso a esta información con el soporte para la construcción de un modelo de razonamiento que nos indique una acción más concisa a sólo ver datos en bruto. Soportan modelado, estructurado y solución de problemas.

Las áreas de aplicación típicas de los SSD son administración y planeación en negocios, asistencia médica, la milicia y generalmente cualquier área en la cual la administración se tope con situaciones complejas de decisión. Estos sistemas son usados para decisiones tácticas y estratégicas enfocadas para administración de alto nivel, para decisiones con una baja frecuencia y con consecuencias de alto potencial, en el cual el tiempo tomado para pensar y modelar un problema recompensa de manera generosa a la larga.

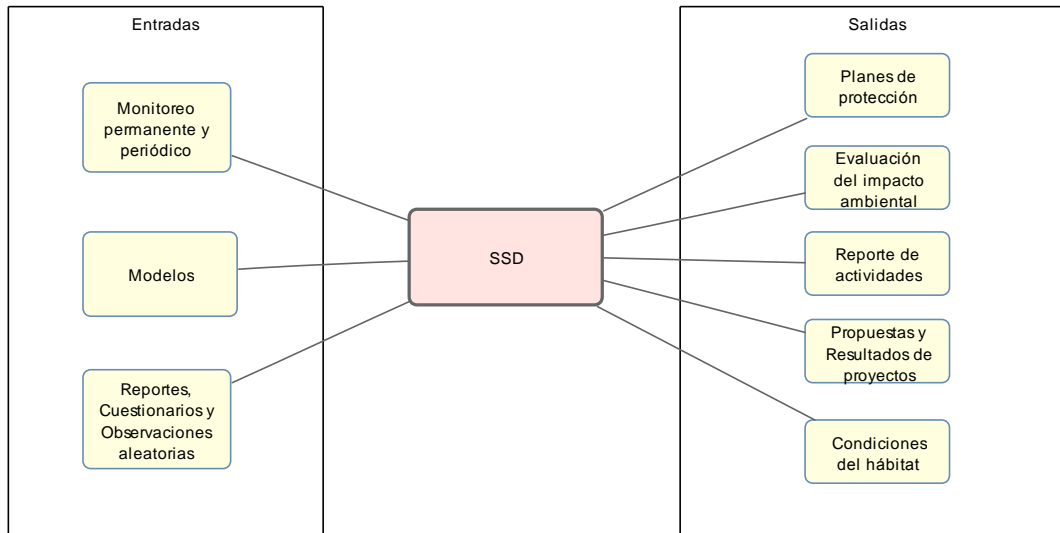


Figura 2-1 Entradas y salidas de un Sistema de Soporte de Decisiones.

### 2.3.1 Componentes de un sistema de soporte de decisiones

Existen tres componentes fundamentales que conforman todo SSD:

- Sistema administrador de base de datos (DBMS). Un DBMS sirve como un banco de datos para el sistema. Almacena grandes cantidades de datos que son relevantes a la clase de problemas por los cuales el SSD ha sido diseñado y provee estructuras lógicas de datos (opuesto a las estructuras físicas como los archivos) con la que los usuarios interactúan.
- Sistema de administración del modelo base (MBMS). El rol de los MBMS es análogo al del DBMS. Su función primaria es proveer independencia entre modelos específicos que son usados en un SSD de parte de las aplicaciones que lo usan. Transforma los datos de un DBMS en información que es útil para la toma de decisiones. Debido a que muchos problemas no poseen una estructura desde el punto de vista del usuario, el MBMS debe ser capaz de asistir al usuario con la creación de modelos.
- Sistema de administración y generación de diálogos (DGMS). Básicamente se trata del conjunto de interfaces con el usuario. Debido a que el usuario generalmente no tiene conocimientos profundos de computación, es necesario que los SSD estén equipados de interfaces intuitivas y fáciles de utilizar.

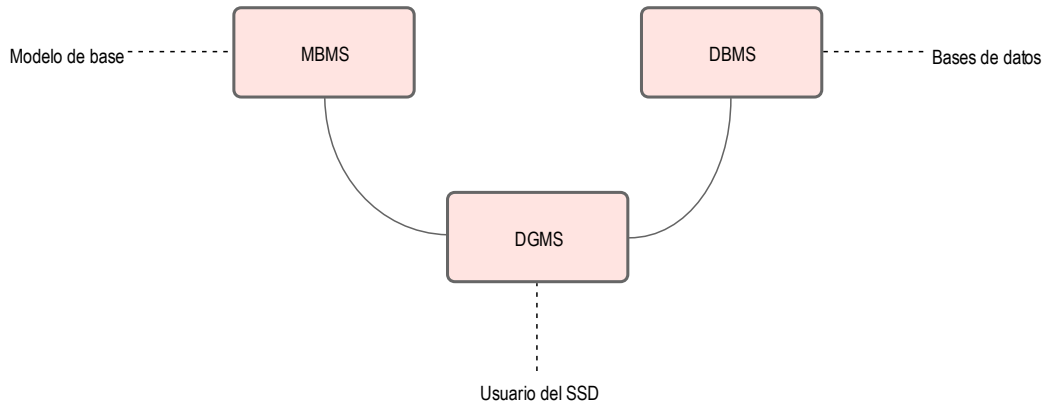


Figura 2-2 Componentes y actores de un Sistema de Soporte de Decisiones.

### 2.3.2 Tipos de sistemas de soporte de decisiones

Existen principalmente dos aproximaciones de cómo debe ser modelado un SSD, la primera es la aproximación normativa y la segunda es la aproximación descriptiva.

La aproximación normativa se enfoca en construir sistemas o procedimientos de soporte que imitan a los expertos humanos. Los miembros más prominentes de este tipo de sistemas son los *sistemas expertos*, programas computacionales basados en reglas dictadas de un por el *expertise* humano que imita el razonamiento de dicho experto en un dominio dado. Los sistemas expertos son generalmente capaces de soportar tomas de decisiones en el dominio a un nivel comparable a como lo haría un humano experto. Mientras son flexibles y continuamente aptos para agendar problemas de decisión complejos, están basados en el razonamiento de la intuición humana y la falta de sentido, además de las garantías formales con respecto a la fiabilidad teórica de sus resultados. El peligro de la aproximación de los sistemas expertos, incrementalmente apreciados por los diseñadores de SSD, es que mientras imitan el pensamiento humano y sus principios heurísticos eficientes, también podrían imitar los fallos indeseables tocados en secciones anteriores.

La segunda aproximación está basada en la asunción de que los métodos más confiables para lidiar con decisiones complejas es a través de un pequeño conjunto de principios de tipo normativo de cómo las decisiones deben ser hechas. Mientras los métodos heurísticos y los esquemas de razonamiento *ad hoc* que imitan la cognición humana pueden en muchos dominios desempeñar de manera eficiente, muchos administradores que hacen uso de estos sistemas no se fiarán mucho de ellos cuando el costo de errar es bastante alto. Para dar un ejemplo, algunas personas prefieren volar aviones contruidos usando principios heurísticos sobre aviones contruidos usando las leyes de la aerodinámica mejoradas con análisis de fiabilidad probabilística.

Una vez aclaradas estas dos aproximaciones podemos establecer cinco tipos de SSD de acuerdo a su funcionamiento:

- Basado en comunicación
- Basado en datos

- Basado en documentos
- Basado en conocimiento
- Basado en modelos

De acuerdo a la base establecida de cada SSD es el tipo en el que cae, por ejemplo los sistemas expertos están clasificados según la descripción vista arriba como basado en conocimiento, debido a que en el fondo existe una base de conocimiento que permite realizar las decisiones. En los modelos basados en comunicación y datos, la diferencia es el tipo de entrada que el sistema podría tener, además de que el primero es dinámico a las respuestas de los usuarios, mientras que el segundo es estático mientras los datos no varíen.

Los SSD basados en modelos son aquellos que operan bajo algún tipo de modelo matemático que permite obtener resultados a partir de cálculos con variables almacenadas en el sistema.

## 2.4 SISTEMAS DE SOPORTE DE DECISIONES BASADOS EN DECISIONES ANALÍTICAS

Una clase emergente de SSD conocidos como SSD basados en decisiones analíticas aplica los principios de la teoría de la decisión, teoría de la probabilidad y el análisis de las decisiones a sus modelos. La teoría de la decisión es una teoría axiomática de toma de decisiones que es construida en un conjunto pequeño de axiomas para la toma de decisiones racional.

Se expresa inciertamente en términos de las probabilidades y preferencias en términos de utilidades. Estos son combinados usando la operación de la especulación matemática. El atractivo de la teoría de la probabilidad, como un formalismo para manejar los SSD inciertos, recae en su rendimiento a largo plazo. La teoría de la probabilidad es a menudo vista como el estándar de oro para la racionalidad en el razonamiento sobre la incertidumbre. Siguiendo sus axiomas ofrece una protección para algunas inconsistencias primarias, tales como en los casos en los que existen huecos en un conjunto de información que podrían ser rellenados utilizando probabilidad.

Los sistemas normativos son usualmente basados en modelos probabilísticos gráficos, que son representados de la unión de una distribución de probabilidades sobre las variables de un modelo en términos de grafos dirigidos. Estos grafos dirigidos son conocidos como Redes Bayesianas o redes causales [6].

En general, los sistemas que utilicen modelos matemáticos que arrojan resultados numéricos como la aproximación de parámetros óptimos se engloban en un tipo especial de SSD: los sistemas basados en ecuaciones y los sistemas mixtos.

### 2.4.1 Sistemas de soporte de decisiones basados en ecuaciones

En muchos problemas de negocios e ingeniería, por hacer mención a algunas áreas, la interacción entre variables de un modelo puede ser descrita por ecuaciones tales que, cuando son resueltas simultáneamente, pueden ser usadas para predecir el efecto de las decisiones en un sistema, y como tal, realizar soporte de decisiones.

Un tipo especial de modelos de ecuaciones simultáneas es conocido como el modelo de ecuaciones estructurales, el cual ha sido un método popular de representar sistemas en economía. Una ecuación es estructural si describe un único e independiente mecanismo actor en el sistema. Las ecuaciones estructurales están basadas en el conocimiento experto del sistema combinado con consideraciones teóricas. Las ecuaciones estructurales permiten una descripción modular y natural de un sistema, cada ecuación representa un componente individual.

La ventaja principal de tener un sistema basado en ecuaciones estructurales es, como se explica en un artículo [7], es que incluye información causal y ayuda a la predicción de los efectos de intervenciones externas.

En adición, la estructura causal de un sistema de ecuaciones estructurales puede ser representada gráficamente que permite la combinación con modelos gráficos de SSDs analíticos en sistemas prácticos [8] [9].



### 3 MODELO MATEMÁTICO PARA LA PROPUESTA

---

Como ya pudimos observar en el capítulo anterior, se trataron tópicos generales sobre teoría de SSDs y sus estructuras, pero particularmente un tema de interés para el presente trabajo son los sistemas basados en ecuaciones. Para la propuesta específicamente se decidió utilizar un sistema basado en ecuaciones estructurales, las cuales describen los diferentes mecanismos que permiten decidir el número adecuado de secciones por periodo.

Particularmente se utilizó un modelo basado en series de tiempo para analizar el comportamiento de cada materia a través del tiempo. Dicho modelo utiliza ciertas técnicas que se irán describiendo a lo largo de todo el capítulo para realizar predicciones con base en sucesos precedentes. Adicionalmente, se utilizan ciertos mecanismos de probabilidad simples para afinar las propuestas que el sistema deduce.

Por último, un segundo modelo matemático plantea la posibilidad de manejar interferencias externas por parte de los alumnos a los cursos, a través de un sistema de encuestas para obtener estadísticas e intentar dar un parámetro preliminar al sistema principal para ayudar a que tome cada una de las decisiones.

En este capítulo se tratarán los tópicos que inspiraron al modelo matemático, además de presentar cómo se aplica en el sistema actualmente.

Dado que los datos que tenemos acerca del comportamiento de cada curso son datos históricos comenzaremos a hablar un poco acerca de este tipo de información, como tratarla, y además, técnicas para poder realizar inferencias sobre esos datos, específicamente predicciones que podría tomar una variable en un instante de tiempo futuro teórico.

#### 3.1 INTRODUCCIÓN AL ANÁLISIS DE DATOS HISTÓRICOS

Cuando nosotros contamos con cientos o miles o quizá millones de datos, lo que nosotros observamos a primera vista es una colección de simples cifras o palabras que quizá no tengan ningún significado relevante, sin embargo esta visión está muy lejos de ser la realidad. Si nosotros contamos con un conjunto de datos brutos sobre un suceso específico podemos emplear diversos mecanismos para convertir ese enorme conjunto de datos en información valiosa que nos permite conocer detalladamente dicho suceso e incluso en muchas ocasiones poder inferir cosas que no se tenían descritas anteriormente o hasta poder predecir un comportamiento.

Los datos históricos, en particular, son conjuntos de datos que se acumulan y se pueden etiquetar de acuerdo a un cierto periodo de tiempo, el cual consiste en cualquier cantidad que sea necesaria, siempre y cuando la cantidad sea constante, es decir, si se elige llevar un control mensual de los datos, todos los datos deben representar una entidad mensual y no cambiar el plazo del tiempo.

Como los datos históricos llevan una distribución equidistante entre cifras, es posible determinar modelos que permitan analizarlos con más facilidad que sólo analizar datos brutos de otra índole. A través de este capítulo tocaremos algunas de estas técnicas.

Una de las aplicaciones directas sobre este tipo de datos es la predicción de sucesos en un tiempo futuro, como ya había mencionado anteriormente, con el objetivo de poder tomar medidas preventivas que, aplicadas en muchas áreas, permiten mejorar la eficiencia de procesos o evitar riesgos en tomas de decisiones. Algunas de las aplicaciones a estos datos tienen lugar en analizar fenómenos climatológicos que permiten predecir en qué momento aproximado un desastre natural podría tener lugar.

En el caso particular de la facultad, el número de secciones ofertado se basa en un periodo que se repite tres veces en un año, por lo que podemos establecer un periodo regular a pesar de que los tres periodos no tienen la misma longitud, sin embargo suceden en fechas similares. Dentro de cada periodo existe para cada curso contiene una cifra que indica el número de grupos que se impartieron.

### 3.2 MODELOS DE PREDICCIÓN

La predicción, dentro de la estadística, se conoce como el proceso de realizar inferencias del futuro basadas en datos pasados y presentes y el análisis de tendencias. Un ejemplo típico de una predicción puede ser la estimación de alguna variable de nuestro interés, en nuestro caso los números de cursos a ofertar que pudieran ajustarse mejor a la demanda. Generalmente se aplican a datos que están contenidos en series de tiempo o datos longitudinales.

Aquí el riesgo y la incertidumbre son un aspecto central en las predicciones, y es generalmente considerada una buena práctica indicar el grado de incertidumbre ligado a una predicción.

Existen principalmente dos tipos de métodos para realizar predicciones estadísticamente hablando:

- Métodos cualitativos: Estas técnicas son subjetivas y se basan en la opinión y el juicio de expertos. Son apropiadas cuando no se cuenta con datos pasados.
- Métodos cuantitativos: Estos modelos se utilizan para predecir datos futuros en función de los datos pasados. Son apropiadas cuando se cuenta con datos numéricos pasados de un determinado suceso.

Dentro de los modelos matemáticos en métodos cuantitativos contamos con algunas aproximaciones que, dependiendo el conjunto de datos con el que se trabaje, obtienen resultados buenos en diferentes circunstancias. Algunas de estas aproximaciones son:

- Promedio
- Último valor obtenido
- Naïve
- Naïve con temporalidad

- Métodos con series de tiempo entre los cuales se encuentran:
  - Promedio móvil con y sin peso
  - Suavizados exponenciales
  - Extrapolación
  - Predicción lineal
  - Predicción de tendencia
- Análisis de regresión

Existen muchos más modelos que utilizan otro tipo de aproximaciones como pueden ser algunas técnicas de inteligencia artificial como las redes neuronales o las máquinas de soporte vectorial, e incluso técnicas de minería de datos o de reconocimiento de patrones.

Sin embargo nosotros nos vamos a concentrar en series de tiempo, debido a que nuestros datos se ajustan a la descripción de una.

### 3.3 SERIES DE TIEMPO

La historia se repite es el lema central de las predicciones de tipo cuantitativas en una serie de tiempo.

Una serie de tiempo es una secuencia ordenada de valores de una variable en intervalos de tiempo equidistantes. Las series de tiempo son ampliamente utilizadas en diferentes áreas como la estadística en general, la economía, procesamientos de señales, reconocimiento de patrones, entre otras más.

La manera en la que se ordena una serie de tiempo hace diferente el análisis que se puede llevar a cabo, a diferencia de estudios transversales en los cuales no se observan órdenes naturales de los datos. El análisis de series de tiempo es también distinto del análisis de datos espaciales.

Las series de tiempo pueden contener cualquier tipo de valor, que abarcan números reales, datos de valores continuos, datos de tipo discreto o datos discretos simbólicos (como las secuencias de caracteres y palabras).

Los métodos para analizar las series de tiempo pueden ser divididas en dos clases: los métodos de dominio de frecuencia y los de dominio de tiempo.

En el dominio de frecuencia se analizan funciones o señales con respecto de su frecuencia, es decir, el número de ocurrencias de un evento repetido por unidad de tiempo, y la forma que ésta toma. En esta clase se incluye el análisis espectral y los análisis de wavelets.

En el dominio de tiempo se hace énfasis a porciones de tiempo y observar cómo una función cambia en cada instante. En esta clase se incluye la auto-correlación y la correlación cruzada.

Adicionalmente, las técnicas de análisis de series de tiempo pueden ser divididas en métodos paramétricos y no paramétricos. Los métodos paramétricos asumen que el proceso estocástico estacionario tiene una cierta estructura que puede ser descrita usando un pequeño número de parámetros (como los modelos autorregresivos o los promedios

móviles). En esta aproximación la tarea principal es estimar los parámetros del modelo que describe el proceso estocástico. Por otra parte, las aproximaciones sin parámetros explícitamente estiman la covarianza o el espectro del proceso sin asumir que el proceso tiene alguna estructura particular.

**FIGUM 001 Formación Humana y Social (LCC ICC ITI)**

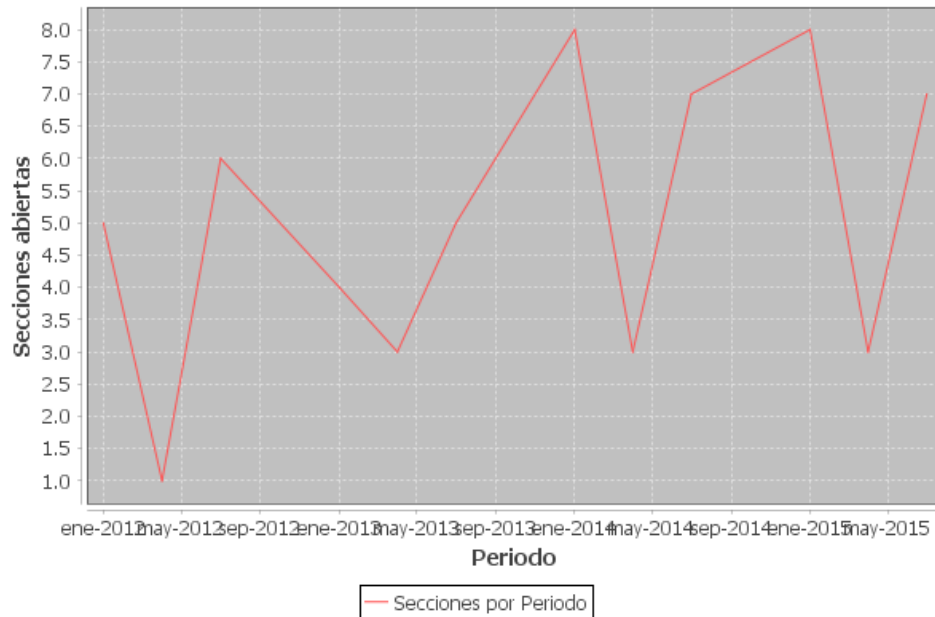


Figura 3-1 Ejemplo de una serie de tiempo.

En el objetivo de esta tesis se revisa la aproximación parametrizada debido a que supone una estructura describible a través de parámetros. Específicamente enfocamos la atención a una familia de métodos llamados Suavizados Exponenciales de Holt-Winters debido a la cantidad pequeña de datos sobre los cuales debemos trabajar. Se consideraron otros métodos como las regresiones lineales, sin embargo este método en particular trabaja mejor en grandes conjuntos de datos y con datos pequeños no puede realizar estimaciones adecuadas.

### 3.4 SUAVIZADO DE UNA SERIE DE TIEMPO

Dentro de la colecta de datos tomados en el tiempo existe alguna forma de variación aleatoria. Dicha variación puede perjudicar la manera en la que se estiman los valores en una serie de tiempo. Afortunadamente existen métodos para reducir los efectos producidos por una variación aleatoria. Una de las técnicas utilizadas a menudo en la industria es el suavizado o “smoothing” en inglés. Esta técnica bien aplicada nos puede revelar de manera más clara la tendencia oculta, la temporalidad y los componentes cíclicos.

El término filtro suele aplicarse más a un suavizado que un modelo como tal, debido a que lo que realmente hace un suavizado es ajustar una serie para poder observar mejor la estructura de la función temporal. Por instancia, si el valor suavizado para un tiempo particular es calculado como una combinación lineal de observaciones para los tiempos

aledaños, podría decirse que aplicamos un filtro lineal a los datos (no es lo mismo que decir que generan una línea recta).

Todos conocemos un método de suavizado que se puede decir que es el más simple del conjunto de suavizados: el promedio. Un promedio o media es la suma de una lista de números dividida entre el número de números en dicha lista. Generalmente se conoce como la media aritmética.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n a_i = \frac{1}{n} (a_1 + a_2 + \dots + a_n) \quad (1)$$

Existen otros tipos de media como la media geométrica o la armónica, sin embargo no vamos a indagar en cada uno de ellos ya que sólo esta aproximación es la única que nos sirve para explicar la teoría subyacente en la propuesta.

Una serie de tiempo se puede suavizar a través de la media simple de todos los datos pasados. El problema subyacente a una media simple es claramente cuando los valores oscilan mucho entre un rango distante, es decir que valores pequeños se mezclan con valores grandes y al final el suavizado nos daría como resultado un valor en medio del verdadero rango de valores, lo cual alteraría la serie y las futuras inferencias que se realizaran en ella.

Para ilustrar lo anterior tenemos el ejemplo con los datos estadísticos de la sección de Computabilidad perteneciente a la Licenciatura en Ciencias de la Computación.

Periodo	Valor
Primavera 2012	1
Verano 2012	0
Otoño 2012	0
Primavera 2013	1
Verano 2013	1
Otoño 2013	2
Primavera 2014	1
Verano 2014	1
Otoño 2014	1
Primavera 2015	1
Verano 2015	1
Otoño 2015	1

Tabla 3-1 Tabla de valores para el curso Computabilidad.

Si notamos los datos son muy similares entre sí, con fluctuaciones entre 0 y 2, donde la media aritmética de los valores redondeada nos da 1. El valor real de la media aritmética nos da 0.91, lo cual es un valor muy cercano a 1 y como necesitamos números enteros lo subimos a 1.

Como se pudo observar el método puede predecir los valores futuros como 1, lo cual para este caso es funcional, sin embargo en casos donde los valores sean demasiado alejados entre sí no va a proponer una buena estimación.

En la figura 3-2 tenemos la representación gráfica de los valores en el tiempo junto con la media de la serie de tiempo para los datos del curso de Computabilidad.

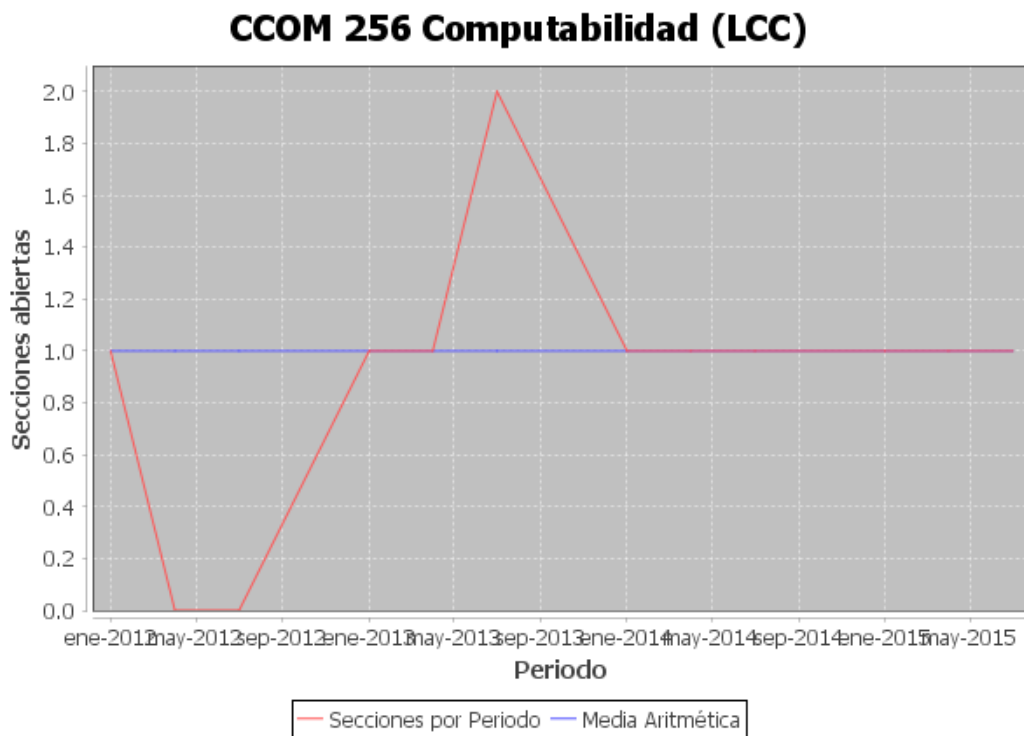


Figura 3-2 Ejemplo de una serie de tiempo con poca fluctuación en valores.

Para conocer la verdadera precisión de un método podemos utilizar el error cuadrático medio (MSE por su abreviatura en inglés). Para obtener el MSE se debe calcular la media de los errores cuadráticos de todos los valores a través de la diferencia del valor esperado entre el valor obtenido al cuadrado.

$$MSE = \frac{1}{n} \sum_{i=1}^n (v_e - v_o)^2 \quad (2)$$

Donde:

- $v_e$  es el valor esperado
- $v_o$  es el valor obtenido

Se utiliza este método de medición de errores porque es muy noble y el estimador que minimiza el MSE para un conjunto de datos aleatorios es la media.

La media o el promedio simple sirve generalmente en valores aproximados entre sí, o las fluctuaciones son escasas, sin embargo, este tipo de suavizado no reaccionaría bien si existe

una tendencia en los datos, es decir, que las cifras suban o bajen en determinada cantidad en cada instante de tiempo ya que la media le da el mismo “peso” a todos los valores.

El factor que multiplica a la media se le llama el “peso”, el cual no es exclusivo de este método, sin embargo por observación en la fórmula se puede ver que el peso de todos los valores es el mismo, además que todos los pesos suman 1.

Una alternativa para resumir los datos pasados es calcular la media de pequeños conjuntos sucesivos de números pasados. Este proceso es un suavizado en zonas. Dado un intervalo de números pequeño se calcula la media. A este proceso se le conoce como promedio móvil simple, el cual vamos a detallar posteriormente.

### 3.4.1 Promedio Móvil

Como se observó en la apertura de esta sección, el promedio o media simple pierde mucha eficiencia en valores con diferencias notables entre ellos. La aproximación de un promedio móvil es dividir la serie en partes donde quizá no existan estas diferencias y aproximar cada parte a través de la media aritmética. En su versión más simple se toma la media con esquema de pesado igual.

Si nosotros decidiéramos un rango de tres valores se obtendría la media de los primeros tres valores y lo colocaríamos como el nuevo tercer valor. Mientras se esté avanzando consecutivamente para cada valor se tomaría ese y los dos anteriores para calcular su valor ajustado.

En vez de haber calculado la media con los 2 valores anteriores pudimos escoger el valor inmediato anterior y el próximo y esto funcionaría con rangos impares, sin embargo en rangos pares el valor quedaría en una posición confusa, si fuera 4 el rango del promedio móvil el valor nuevo sería  $n+1/2$ , es decir entre dos valores. Para solucionar este problema debemos suavizar los promedios móviles con un rango de 2. Entonces estamos suavizando los valores que ya estaban suavizados. A esta técnica se le conoce como el promedio móvil centrado. Para ilustrar un poco el promedio móvil se toman en cuenta los siguientes datos donde se calcula el promedio móvil utilizando dos valores:

Periodo	Valor	P.M.
Primavera 2012	2	2
Verano 2012	1	2
Otoño 2012	3	1
Primavera 2013	2	3
Verano 2013	3	2
Otoño 2013	2	3
Primavera 2014	4	2
Verano 2014	2	4
Otoño 2014	4	2
Primavera 2015	4	4
Verano 2015	2	4
Otoño 2015	4	2

Tabla 3-2 Tabla de valores para el curso Transmisión y Comunicación de Datos.

Lo curioso que se observa es que la serie en promedio móvil es básicamente la misma pero recorrida en un instante de tiempo, es decir que para el tiempo  $t+1$  nos va a pronosticar el último valor observado en el tiempo  $t$  redondeado.

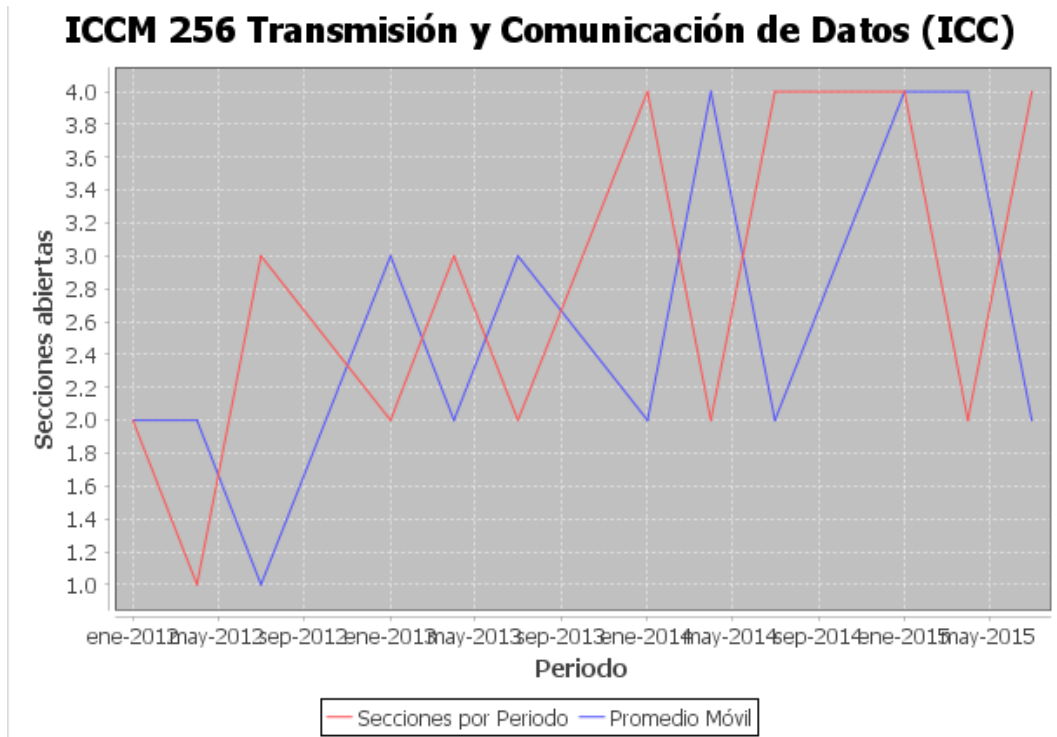


Figura 3-3 Aplicación de un promedio móvil sobre tres datos contiguos.

La ventaja aparente de esta técnica consiste en que pudiera respetar más que un promedio simple los valores de una serie, sin embargo aún falla en manejar una tendencia lineal, sobre todo para poder inferir nuevos valores de periodos de tiempo futuros.

Este tipo de técnicas pertenece a la clase de métodos sin parámetros, debido a que no analiza ningún tipo de estructura que la serie tendría. Existen otro tipo de técnicas de suavizado que se basan en el tipo de series que están procesando y reciben parámetros para ajustar las estructuras, conocidas como suavizados exponenciales.

### 3.4.2 Suavizados exponenciales

Estas técnicas son esquemas muy populares para producir series de tiempo suavizadas. Son muy similares a un promedio móvil. La diferencia radica en el esquema de pesado: el promedio móvil utiliza pesos iguales heredados de la media aritmética, mientras que un suavizado exponencial asigna pesos exponencialmente disminuidos conforme las observaciones comienzan a hacerse viejas.

En otras palabras, para un suavizado exponencial, las observaciones recientes tienen más peso en la inferencia de nuevos valores que las más viejas.

En el caso de los promedios móviles, los pesos asignados a las observaciones eran iguales y equivalen a uno entre el número de elementos. En el suavizado exponencial, existe uno o

más *parámetros de suavizado* que necesitan ser determinados (o estimados) y estas elecciones determinan los pesos asignados a las observaciones.

De acuerdo al número de parámetros y las características de cada serie se pueden utilizar uno, dos o tres parámetros para definir la estructura de una serie de tiempo, y es con base en esto que se define si el suavizado exponencial es simple, doble o triple. Debido a la naturaleza de las series con las que trabajamos estas son las técnicas que se emplearon en el modelo matemático, por ello vamos a describir cada una de ellas.

### 3.4.2.1 Suavizado simple

Este esquema de suavizado comienza por colocar  $S_2$  a  $y_1$ , donde  $S_i$  es la observación o el promedio móvil pesado exponencialmente (EWMA por sus siglas en inglés) y  $y$  es la observación original. Los subíndices refieren a los periodos de tiempo  $1, 2, \dots, n$ . Para el tercer periodo,  $S_3 = \alpha y_2 + (1 - \alpha)S_2$ , y así sucesivamente. No existe el término  $S_1$  en la serie suavizada, la serie comienza con la versión suavizada del segundo término.

Para cualquier periodo de tiempo  $t$ , el valor suavizado  $S_t$  se encuentra calculando lo siguiente:

$$S_t = \alpha y_{t-1} + (1 - \alpha)S_{t-1}, \quad 0 < \alpha \leq 1, \quad t \geq 3 \quad (3)$$

Esta es la ecuación básica del suavizado exponencial y la constante o el parámetro  $\alpha$  es llamado la constante de suavizado.

Sobre el primer valor de la serie  $S_2$ , es posible inicializarlo de diferentes maneras y cada una de ellas representa un rol importante en la generación de las demás observaciones suavizadas. El primer método que vimos en el que la primera observación se vuelve el primer valor de la serie es uno simple pero eficaz, aunque existen otros más como colocarle el valor al valor objetivo o esperado o incluso utilizar un promedio de los primeros  $k$  valores.

Para observar el porqué del nombre exponencial en esta familia de métodos vamos a observar la ecuación.

Expandamos la ecuación básica primeramente sustituyendo para  $S_{t-1}$  en la fórmula para obtener:

$$S_t = \alpha y_{t-1} + (1 - \alpha)[\alpha y_{t-2} + (1 - \alpha)S_{t-2}] \quad (4)$$

$$S_t = \alpha y_{t-1} + \alpha(1 - \alpha)y_{t-2} + (1 - \alpha)^2 S_{t-2} \quad (5)$$

Una vez que llegamos a esto seguimos realizando las sustituciones retroactivas hasta llegar a  $S_2$  con lo que obtenemos la siguiente sumatoria:

$$S_t = \left( \alpha \sum_{i=1}^{t-2} (1 - \alpha)^{i-1} y_{t-i} \right) + (1 - \alpha)^{t-2} S_2, \quad t \geq 2 \quad (6)$$

Esto a su vez ilustra el comportamiento exponencial. Los pesos  $\alpha(1 - \alpha)^t$  se disminuyen geométricamente, y la suma es tal cual como se muestra a continuación utilizando propiedades de las series geométricas.

$$\sum_{i=1}^{t-2} (1 - \alpha)^i = \alpha \left[ \frac{1 - (1 - \alpha)^t}{1 - (1 - \alpha)} \right] = 1 - (1 - \alpha)^t \quad (7)$$

De la ecuación 6 se observa que el término de adición muestra que la contribución al valor suavizado  $S_t$  se vuelve menos en cada periodo de tiempo consecutivo.

Aplicando a un ejemplo particular con los siguientes datos obtenemos un suavizado exponencial utilizando una constante de suavizado de 0.6.

Periodo	Valor	S.E.S.
Primavera 2012	4	4
Verano 2012	4	4
Otoño 2012	4	4
Primavera 2013	4	4
Verano 2013	4	4
Otoño 2013	5	4
Primavera 2014	5	5
Verano 2014	4	5
Otoño 2014	5	4
Primavera 2015	5	5
Verano 2015	4	5
Otoño 2015	5	4

Tabla 3-3 Tabla de valores para el curso Innovación y Talento Emprendedor.

**FGUM 008 Innovación y Talento Emprendedor (LCC ICC ITI)**

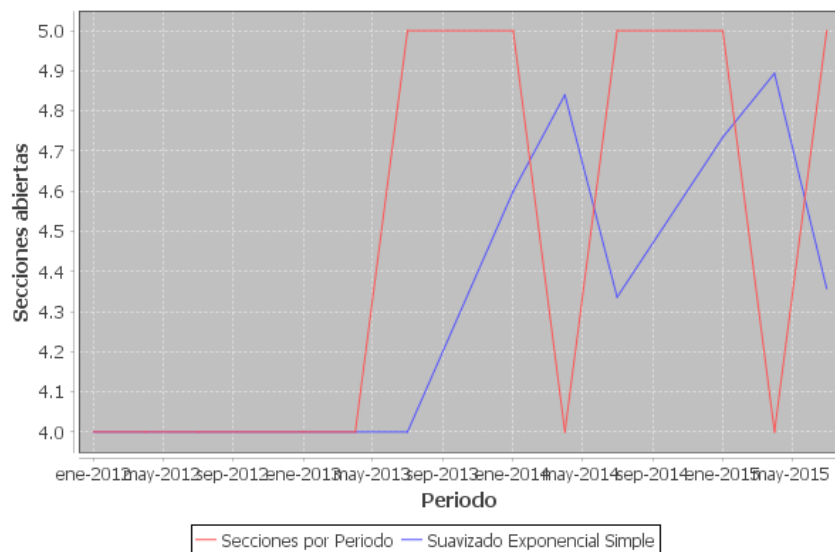


Figura 3-4 Aplicación de suavizado exponencial simple en una serie de tiempo con poca variación.

La ligera diferencia entre los promedios y el suavizado es que el suavizado se puede ajustar mejor al tipo de serie que se maneja, como vemos en la figura 3-4 al estar en un valor estancado conforme comenzaron a subir los valores el pronóstico comenzó a elevarse, aunque un poco tardío, sin embargo mejor que los promedios.

### 3.4.2.2 Predicción con suavizados

Una vez que suavizamos los valores sólo tenemos una nueva serie corregida, sin embargo aún no posee valores nuevos ni realizamos ninguna inferencia sobre ella.

Para realizar esto nos basamos en simplemente sustituir la fórmula original con el valor en el tiempo  $t+1$ .

$$S_{t+1} = \alpha y_t + (1 - \alpha)S_t, \quad 0 < \alpha \leq 1, \quad t > 0 \quad (8)$$

Que puede ser escrita como:

$$S_{t+1} = S_t + \alpha \epsilon_t \quad (9)$$

Donde  $\epsilon_t$  es el error en la predicción (la diferencia entre el valor actual y el pronóstico) para el periodo  $t$ .

En otras palabras, el nuevo valor o la predicción es el valor viejo más un ajuste por el error ocurrido en la última predicción o suavizado.

¿Qué sucede en el caso de querer predecir a partir de algún valor de origen, usualmente el último valor y no existen observaciones actuales disponibles? En esta situación necesitamos modificar la fórmula para obtener lo siguiente:

$$S_{t+1} = \alpha y_{origen} + (1 - \alpha)S_t \quad (10)$$

Donde  $y_{origen}$  permanece constante y equivale al punto de origen conocido.

Esta técnica se conoce como *bootstrapping* en inglés.

A nosotros nos permite inferir valores donde sólo tenemos los últimos periodos calculados debido a que generalmente en nuestras series de tiempo las materias optativas no tienen tantos datos como las materias de base y sólo tenemos pocos valores conocidos, que incluso no varían de valor en el tiempo.

### 3.4.2.3 Suavizado doble

El suavizado simple (nombre corto para el suavizado exponencial simple) no es muy bueno cuando existe tendencia. El coeficiente simple  $\alpha$  no es suficiente.

Podemos demostrarlo con el siguiente ejemplo:

Observamos los siguientes datos obtenidos del curso Administración de Proyectos, calculamos el suavizado doble que nos da los valores registrados en la tabla 3-4.

Esta vez los valores tienen una estructura peculiar en donde se inician con pocas secciones (dos), sin embargo, en cada rango de periodos que avanza es necesario abrir una o dos secciones más de las anteriores, el cambio se denota anualmente para estabilizarse en un rango que va de 4 a 5 actualmente.

Se utilizó para este ejemplo un valor para la constante de suavizado de 0.85 y para la constante de la tendencia 0.4.

Periodo	Valor	S.E.D.
Primavera 2012	2	2
Verano 2012	2	2
Otoño 2012	2	2
Primavera 2013	3	3
Verano 2013	3	3
Otoño 2013	5	5
Primavera 2014	4	4
Verano 2014	5	5
Otoño 2014	4	4
Primavera 2015	5	5
Verano 2015	4	4
Otoño 2015	5	5

Tabla 3-4 Tabla de valores para el curso Administración de Proyectos.

Con los valores redondeados obtuvimos los valores representados en la tabla 3-4 en la que podemos apreciar que este modelo fue más exacto con lo que estaba prediciendo, realmente acertando cada uno de los valores de manera adecuada. Utilizando los respectivos valores reales los podemos observar en la figura 3-5..

Como se observó en el ejemplo anterior, el suavizado simple no es muy bueno en datos con tendencia, sin embargo esta situación puede ser mejorada con la introducción de una segunda ecuación con una segunda constante,  $\gamma$ , que debe ir en conjunción con  $\alpha$ .

Aquí se presentan las ecuaciones del este modelo doble:

$$S_t = \alpha y_t + (1 - \alpha)(S_{t-1} + b_{t-1}), \quad 0 \leq \alpha \leq 1 \quad (11)$$

$$b_t = \gamma(S_t - S_{t-1}) + (1 - \gamma)b_{t-1}, \quad 0 \leq \gamma \leq 1 \quad (12)$$

Nótese que esta vez se utiliza el valor de la observación actual para calcular el valor suavizado de reemplazo en el suavizado doble.

Lo que se realiza es calcular un posible incremento exponencial adicional que representa la tendencia en cada periodo, representado por  $b_t$ . Posteriormente la fórmula conserva el mismo sentido que el suavizado simple, con la excepción de la suma y el valor original que se utiliza.

**IDDM 001 Administración de Proyectos (LCC ICC ITI)**

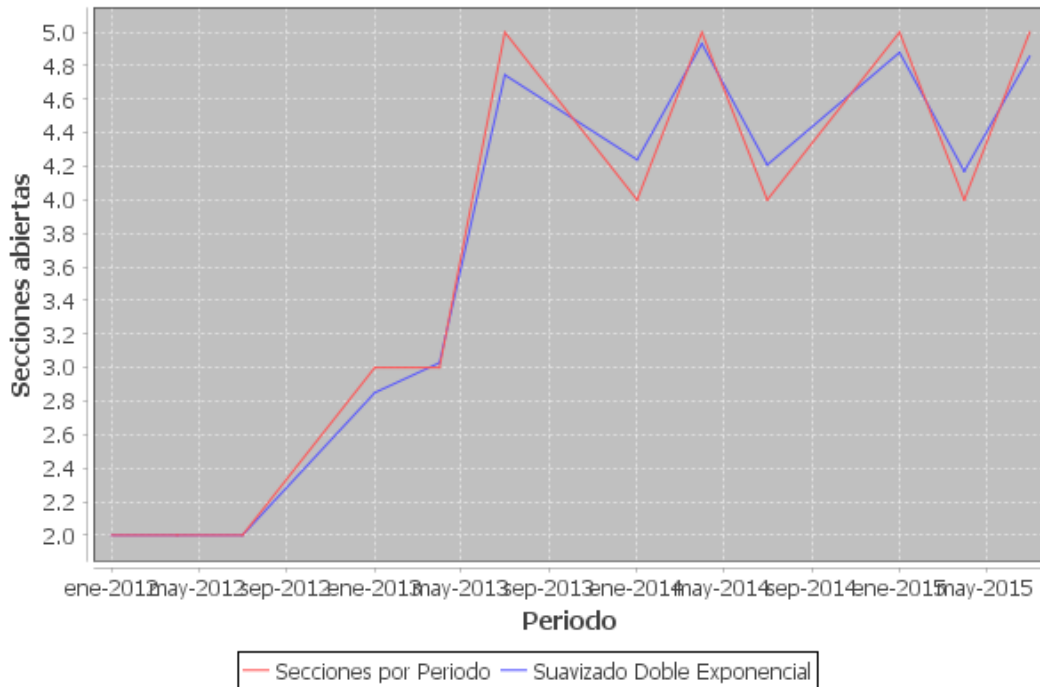


Figura 3-5 Aplicación de suavizado exponencial doble en una serie de tiempo con variación media y tendencia

Para calcular los valores iniciales de la serie, en el caso del suavizado simple se utilizaba una serie de esquemas donde el más simple era poner el primer valor de la serie para el segundo valor en el suavizado. Para el caso del suavizado doble existen varios esquemas tanto para  $S_t$  como para  $b_t$ .

$S_1$  generalmente es el valor contenido en  $y_1$ . Para  $b_1$  existen varias sugerencias..

$$b_1 = y_2 - y_1 \tag{13}$$

$$b_1 = \frac{1}{3}[(y_2 - y_1) + (y_3 - y_2) + (y_4 - y_3)] \tag{14}$$

$$b_1 = \frac{y_n - y_1}{n - 1} \tag{15}$$

Pueden existir otras sugerencias más pero se limitan a resolver problemas particulares. En nuestro caso obtenemos los valores por medio de la primera propuesta para  $b_1$ .

Las fórmulas para realizar la inferencia futura sobre la serie suavizada varían de acuerdo a qué periodo se intente predecir.

El pronóstico para un periodo siguiente está dado por:

$$F_{t+1} = S_t + b_t \tag{16}$$

El pronóstico para  $m$  periodos adelante del último periodo conocido está dado por:

$$F_{t+m} = S_t + mb_t \quad (17)$$

#### 3.4.2.4 Suavizado triple

Cuando los datos además de tendencia presentan modificadores de temporadas el suavizado doble no va a funcionar. Para resolver esto se introduce una nueva ecuación al modelo para poder manejar series de tiempo con temporadas (a veces llamada también periodicidad). El conjunto de ecuaciones resultante es llamado el método de Holt-Winters (HW) por sus inventores.

Las ecuaciones básicas para su método están dadas por el siguiente conjunto:

Para el suavizado general la fórmula es:

$$S_t = \alpha \frac{y_t}{I_{t-L}} + (1 - \alpha)(S_{t-1} + b_{t-1}) \quad (18)$$

Para el suavizado de las tendencias es la siguiente fórmula:

$$b_t = \gamma(S_t - S_{t-1}) + (1 - \gamma)b_{t-1} \quad (19)$$

Por último para la temporalidad se aplica:

$$I_t = \beta \frac{y_t}{S_t} + (1 - \beta)I_{t-L} \quad (20)$$

Para realizar inferencias futuras tenemos la siguiente fórmula:

$$F_{t+m} = (S_t + mb_t)I_{t-L+m} \quad (21)$$

Donde:

- $y$  es la observación del tiempo  $t$
- $S$  es la observación suavizada de  $y$
- $b$  es el factor de la tendencia
- $I$  es el índice de la temporalidad
- $F$  es la predicción en  $m$  periodos adelante
- $t$  es un índice denotando un periodo de tiempo

Y para este caso  $\alpha$ ,  $\gamma$  y  $\beta$  son constantes que deben ser estimadas de tal manera que el MSE del error sea minimizado. Es mejor esta parte dejársela a un software de estimación de parámetros.

Para poder utilizar este método necesitamos por lo menos una temporada completa para determinar los estimados iniciales de los índices de temporalidad  $I_{t-L}$ .

La fórmula general para estimar la tendencia inicial está dada por:

$$b = \frac{1}{L} \left( \frac{y_{L+1} - y_1}{L} + \frac{y_{L+2} - y_2}{L} + \dots + \frac{y_{L+L} - y_L}{L} \right) \quad (22)$$

Para ilustrar esta sección se tiene un ejemplo utilizando los datos del curso Matemáticas Elementales perteneciente a la Licenciatura e Ingeniería en Ciencias de la Computación.

Los valores reales y calculados a través del suavizado triple exponencial se presentan en la tabla 3-5.

Esta vez se eligió este curso debido a que tiene muy marcadas las temporalidades entre periodos, sobre todo que los valores se ven disparados en los periodos de otoño, mientras que en los periodos de verano se disminuyen drásticamente para ser recuperados en primavera en aproximadamente 40% de lo que representan en otoño.

Para calcular los valores del suavizado triple exponencial se utilizaron los valores de para la constante de suavizado, para la constante de tendencia y finalmente para la constante de la temporalidad.

Periodo	Valor	S.E.T.
Primavera 2012	6	-
Verano 2012	2	-
Otoño 2012	17	15
Primavera 2013	6	6
Verano 2013	3	3
Otoño 2013	15	15
Primavera 2014	7	6
Verano 2014	2	3
Otoño 2014	12	12
Primavera 2015	8	5
Verano 2015	3	3
Otoño 2015	12	12

Tabla 3-5 Tabla de valores para el curso Matemáticas Elementales.

Se observa en los valores y en la figura 3-6 que se tienen valores muy cercanos a lo predicho, exceptuando el periodo de primavera 2015 en donde el suavizado predijo un poco menos de lo que resultó, sin embargo la aproximación fue bastante buena. En los periodos primavera y verano del 2014 se predijo un valor inferior con diferencia de 1 con respecto del valor original.

Cabe destacar que este suavizado triple es mucho más intensivo que los otros dos debido al juego de constantes que permite realizar diferentes combinaciones tratando de minimizar el MSE.

El suavizado exponencial ha probado a través de los años ser muy útil en muchas situaciones de predicciones. Fue sugerido primeramente por C.C. Holt en el año de 1957 y fue concebido para usarse en series de tiempo sin tendencia ni temporalidad. Él mismo propuso un método en el año de 1958 que manejaba las tendencias. En el año de 1965, Winters generaliza el método para incluir la temporalidad, de ahí el nombre del “método de Holt-Winters”.

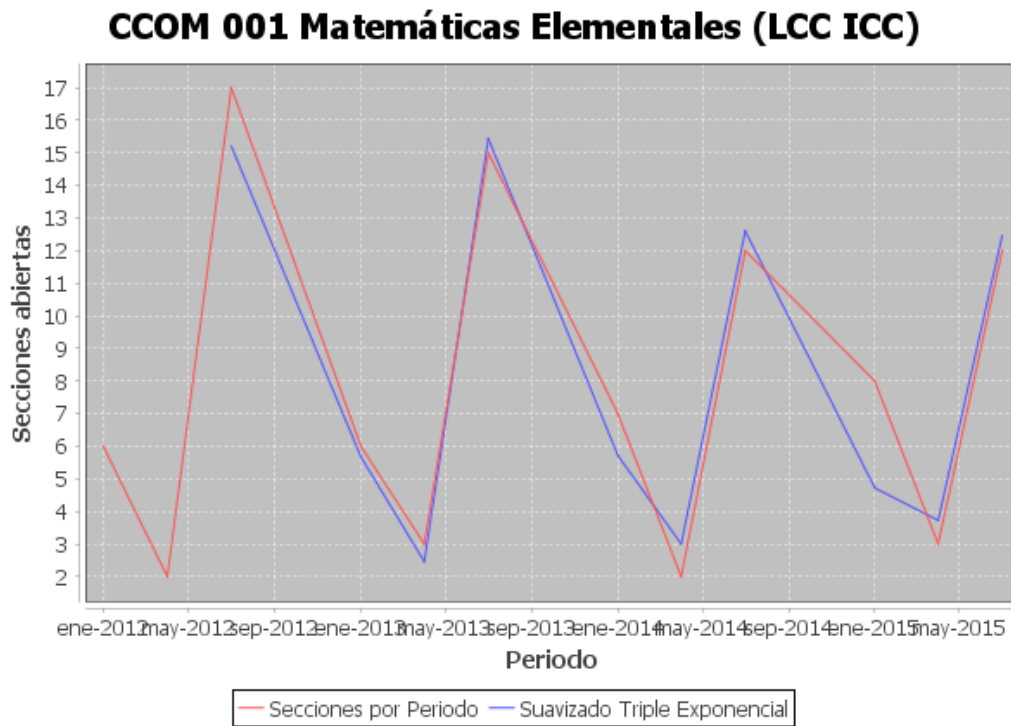


Figura 3-6 Aplicación de suavizado exponencial triple en una serie de tiempo con temporalidad.

## 4 DISEÑO DE LA PROPUESTA DEL SISTEMA

---

En el capítulo anterior vimos toda la teoría subyacente sobre inferencias futuras en una serie de tiempo utilizando técnicas de suavizado. En este capítulo veremos el diseño más enfocado a lo que se utilizó en la propuesta para resolver el problema de planeación del número de secciones.

De manera general tenemos un esquema que utilizan varias entradas que alimentan al sistema. Anteriormente sólo se describió una de ellas que son las series de tiempo del número de secciones ofertadas en periodos anteriores, sin embargo, vamos a introducir otros parámetros para nuestro sistema como son un sistema auxiliar de encuestas que permite expresar la demanda aproximada de los estudiantes para dicho periodo y por último ajustes que el usuario quiera realizar.

El capítulo primero se centrará en el modelo principal del SSD, mientras que después dedicaremos una sección para describir el sistema de encuestas que aún se encuentra en construcción.

### 4.1 ESQUEMA GENERAL DEL SISTEMA

Como ya se había comentado, el sistema recibe de parámetros varias entradas que le permiten realizar la inferencia. Por cada uno de los cursos existentes en la Facultad de Ciencias de la Computación existe una serie de tiempo que describe el comportamiento que ha tenido a lo largo de cada periodo.

Los periodos dentro de la universidad están compuestos por primavera, verano y otoño, formando tres periodos anuales. El primer periodo en el que parten todas las series de tiempo es en primavera 2012 para realizar un total de 12 periodos al momento de la redacción de este trabajo, siendo el último otoño 2015. Algunos periodos pertenecientes a la carrera de más reciente creación no contienen datos en sus series de tiempo debido a que ese curso a pesar de existir en la currícula, nunca se había ofertado debido a que ningún estudiante había avanzado lo suficiente para poder tomarla.

Para describir de manera global el funcionamiento interno del sistema completo se hace la distinción siguiente:

En la figura podemos apreciar dos componentes principales: el módulo de análisis histórico y el módulo de encuesta. Cada uno de estos modelos emplea sus propias reglas para poder realizar la inferencia que al final se ve conjunta en un tercer módulo que realiza la comparación y con base en ello realizar correcciones en el número final de secciones.

En las secciones posteriores se describen los funcionamientos detallados que utiliza cada módulo para tomar decisiones.

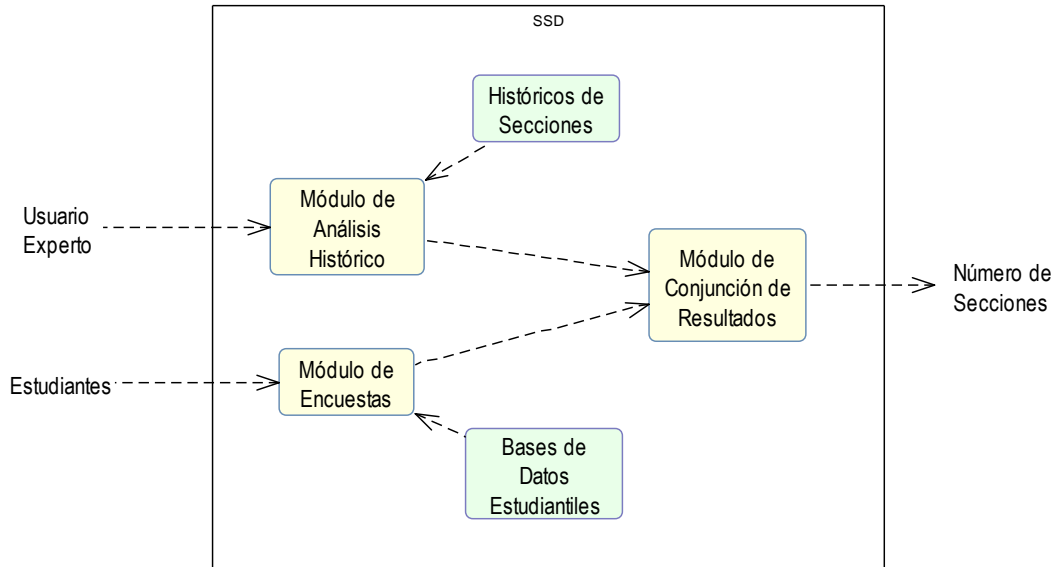


Figura 4-1 Componentes principales del sistema propuesto.

## 4.2 MÓDULO DE ANÁLISIS HISTÓRICO

Este módulo utiliza todos los modelos descritos anteriormente en el capítulo 3 sobre cómo realizar la inferencia de los datos. Específicamente vamos a presentar algunos casos donde se necesite utilizar un tipo de técnica particular.

Retomando el capítulo anterior, y en concreto las sub secciones dentro de la sección 3.4.2, se presentaron las técnicas de suavizado exponencial. Dentro de nuestro sistema se utilizaron estos tres modelos en conjunción con técnicas para verificar el tipo de serie con el que estamos trabajando.

Como se había hecho mención anteriormente, es posible que nuestros registros en una representación gráfica se observen como diferentes figuras según las características de la serie de tiempo que representan. Algunas figuras destacables que nos encontramos son líneas rectas, pendientes ascendentes o descendentes y montañas o curvas repetitivas.

Se enlistan a continuación todos los criterios que se tomaron en cuenta para decidir qué modelo encaja mejor con cada una de las series.

Las series de tiempo cuya representación gráfica sea una línea recta cuyo valor en el eje de las abscisas sea una constante se ven como un valor que nunca varía con el tiempo demuestran un curso que nunca ha cambiado su valor en el rango de tiempo que lo estamos considerando, por lo que lo más probable es que el valor se siga conservando en instantes de tiempo futuros.

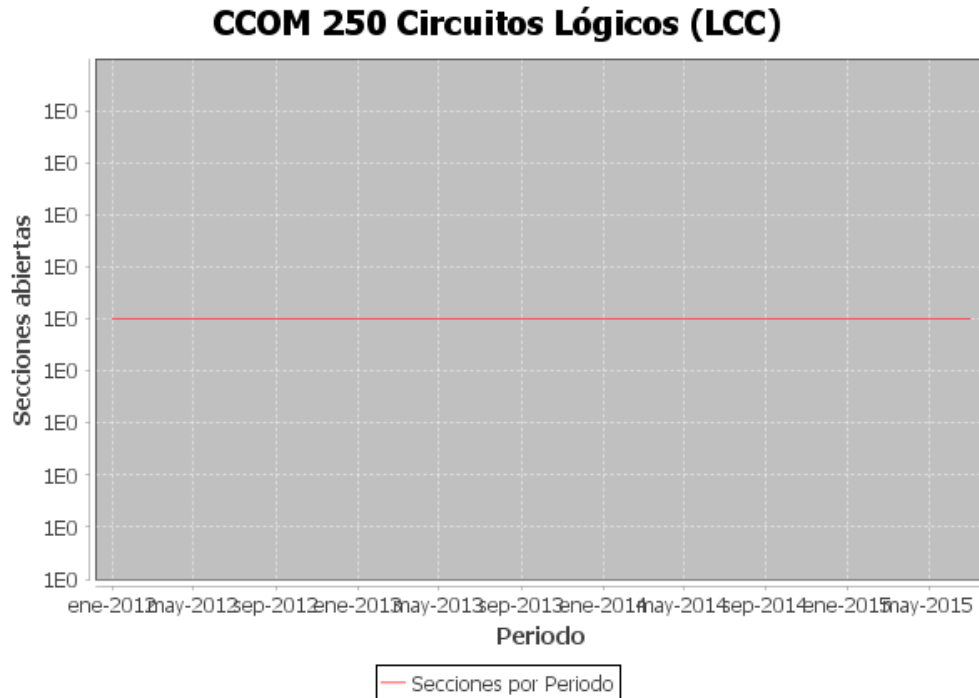


Figura 4-2 Serie de tiempo sin ningún tipo de variación.

Para deducir esto se toma en cuenta que la media aritmética de toda la serie sea la misma que cada uno de los valores en la serie, esto es:

$$\forall X_i \in S_c: X_i = \bar{X} \quad (23)$$

Lo que equivale decir que todos los valores son iguales.

Existe un caso especial en el que por excepción un número de periodos inferior a un valor de tolerancia proporcionado se tenga fluctuaciones en la línea recta no mayores a cierto rango. En este caso se puede inferir que si los últimos valores siguen perteneciendo a la media se puede utilizar tanto la media aritmética redondeada o truncada según el caso.

Si consideramos que los valores de nuestra serie están cercanos uno de otro pero no son iguales podemos emplear la técnica de suavizado exponencial simple. En esta técnica podemos manipular que los valores más viejos o más nuevos sean los predominantes en la decisión.

Cabe destacar que en caso de no querer ajustar manualmente la serie, se puede determinar que el sistema ajuste automáticamente los parámetros de acuerdo a la media del error cuadrático como se aprecia en la figura 4-4. Si se elige esta opción el algoritmo también tratará de determinar el mejor tipo de suavizado en base a algunas características de la serie de tiempo dada.



**IDDM 001 Administración de Proyectos (LCC ICC ITI)**

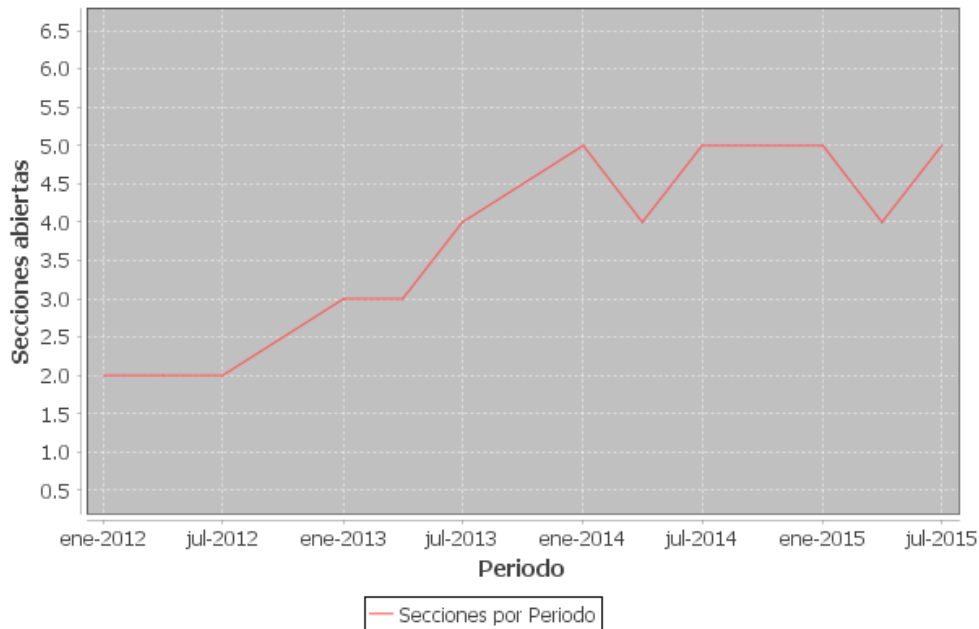


Figura 4-5 Serie de tiempo con tendencia notable.

hora bien, si los valores de nuestra serie en vez de fluctuar en un rango pequeño crecen en cierta cantidad cada cierto periodo de tiempo se puede decir que existe una tendencia en la serie. Para manejar tendencia según lo visto en el capítulo 3 se debe utilizar un suavizado doble, esta vez manejando dos constantes, una que considera los valores más viejos o más nuevos y la otra que considera en qué tanto porcentaje debe tomarse en cuenta la tendencia.

Para identificar que existe tendencia en una serie existen diferentes maneras de acuerdo al tipo de tendencia.

La tendencia más simple es la tendencia en línea recta, que aunque es la más fácil de detectar, no es la más común de encontrar en datos reales.

La diferencia entre dos valores consecutivos dentro de una serie es igual a cierto valor  $Z$ , es decir:

$$\forall X_i \in S_c: X_i - X_{i+1} = Z \quad 1 < i < N \quad (24)$$

El valor de  $Z$  se obtiene de la diferencia de los dos primeros valores.

Eso fue una tendencia lineal, en el caso de una tendencia exponencial,  $Z$  no es como tal un número simple si no un coeficiente elevado a una potencia.

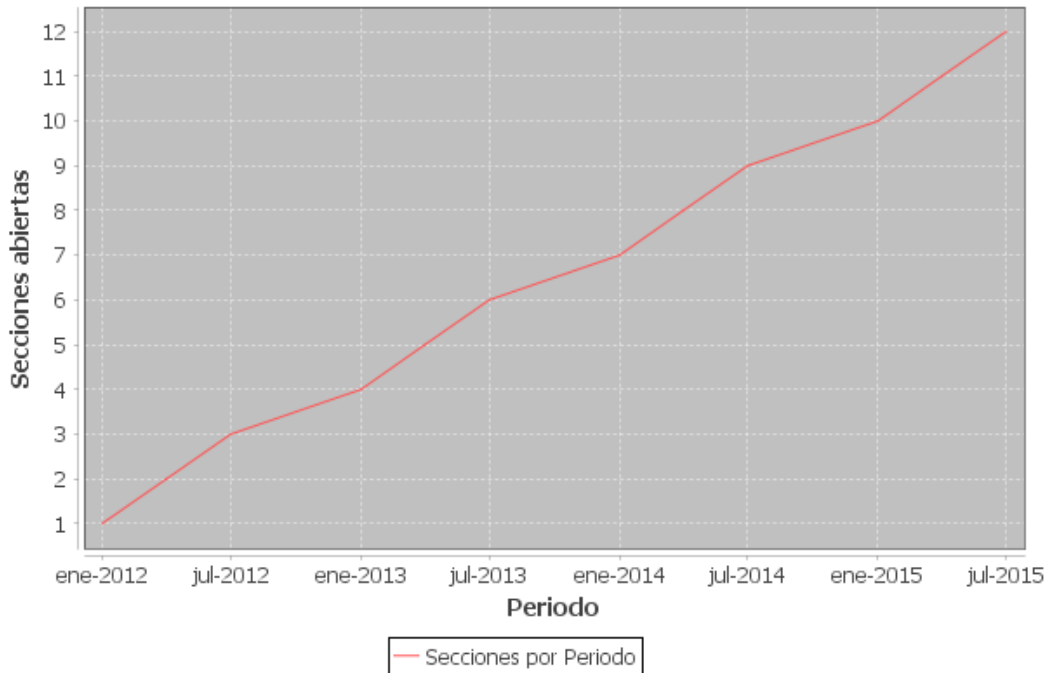
**IDDM 001 Administración de Proyectos (LCC ICC ITI)**

Figura 4-6 Serie de tiempo con tendencia lineal.

En realidad, revisar la tendencia en una serie no es una tarea fácil a través de un algoritmo, debido a que la tendencia puede o no estar bien definida, sin embargo es posible que el método de suavizado doble pueda determinar la serie de manera correcta o aproximada, por esto es que se deja abierta la selección del tipo de método que se desea emplear y los resultados que arroja, y de igual manera configurar los parámetros de manera manual o automática. En el caso de hacer automática la selección del método se eligen los parámetros de manera automática, para esto se ejecuta cada método con sus mejores parámetros y se compara la fiabilidad de cada uno para determinar al final cuál fue el mejor método.

Para los tipos de serie en los que además de poder o no existir una tendencia, existe una estructura en la que se observa que se repite un patrón cada cierto número de periodos de tiempo, en nuestro caso cada 3 periodos, se utiliza el suavizado exponencial triple.

Este tipo de casos se da más seguido en materias donde la demanda en cierto ciclo de periodos suele ser más alta o más baja que en los demás.

En el caso del ajuste de parámetros, se toman en cuenta los dos parámetros anteriores utilizados en el suavizado doble más una tercer constante que permite decidir sobre qué tanto se debe considerar el ciclo que se forma entre cada ciertos periodos.

Existen otros tipos de serie que no encajan directamente en un modelo descrito anteriormente o que el número de datos no es suficiente.

En el caso de tener una serie con ceros lo que se hace es fiarse completamente de la demanda, debido a que no se cuentan con datos históricos que apoyen el soporte de decisión.

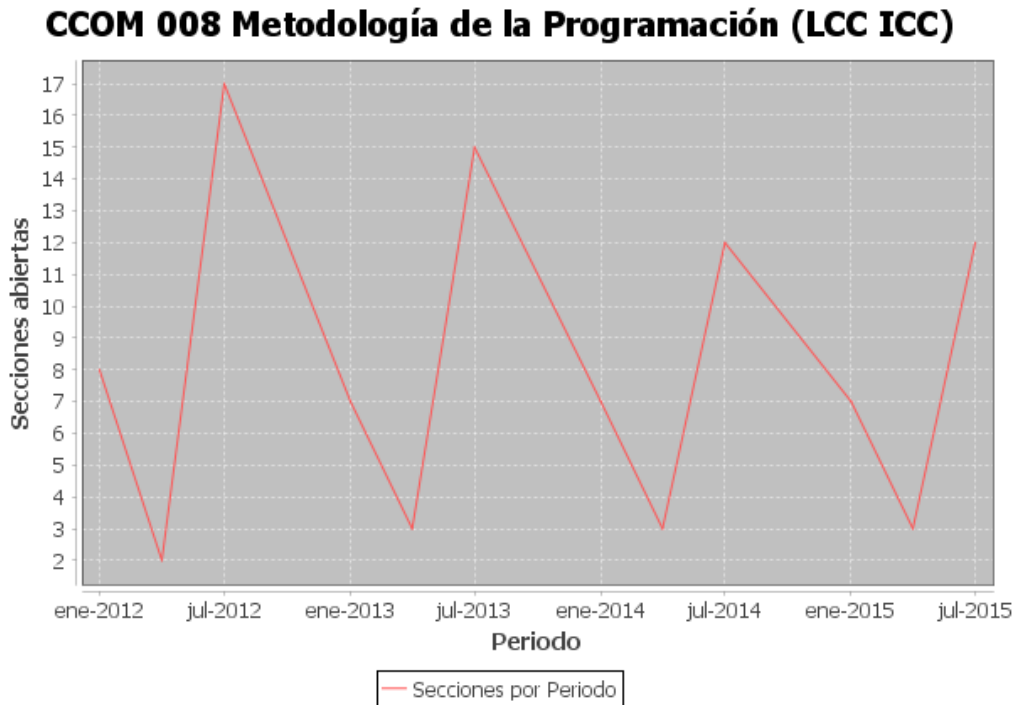


Figura 4-7 Serie de tiempo con temporalidad marcada.

Para las series donde existen ceros al inicio, se toma sólo la parte de la serie donde existen valores diferentes de cero para ser procesada.

Cabe aclarar que la demanda no es un dato bruto obtenido de manera simple, ya que se emplean ciertas fórmulas de corrección que permiten obtener el número estimado. Todo esto se detalla en la sección contigua.

Para realizar el modelo de las predicciones se utilizó un API de desarrollo para Java llamada OpenForecast que contiene la programación de módulos para realizar predicciones sobre datos estadísticos en general, aunque tiene una parte especializada en series de tiempo.

#### 4.2.1 Open Forecast

Open forecast es un paquete de software de propósito general, con modelos de predicción escritos en Java que pueden ser aplicados a cualquier conjunto de datos. Actualmente se encuentra en la versión 0.5 y al parecer su desarrollo se ha detenido desde hace algunos años, sin embargo no deja de ser una buena herramienta para la predicción e inferencia sobre datos para el futuro.

En nuestro caso se utilizaron los modelos de predicción basados en los suavizados exponenciales que vienen incluidos, donde el mismo API puede estimar los valores más adecuados de acuerdo a la media del error cuadrático.

El software fue desarrollado por Steven Gould y se puede obtener de su sitio web, además de contener guías de como incorporar este paquete en un proyecto de software en Java.

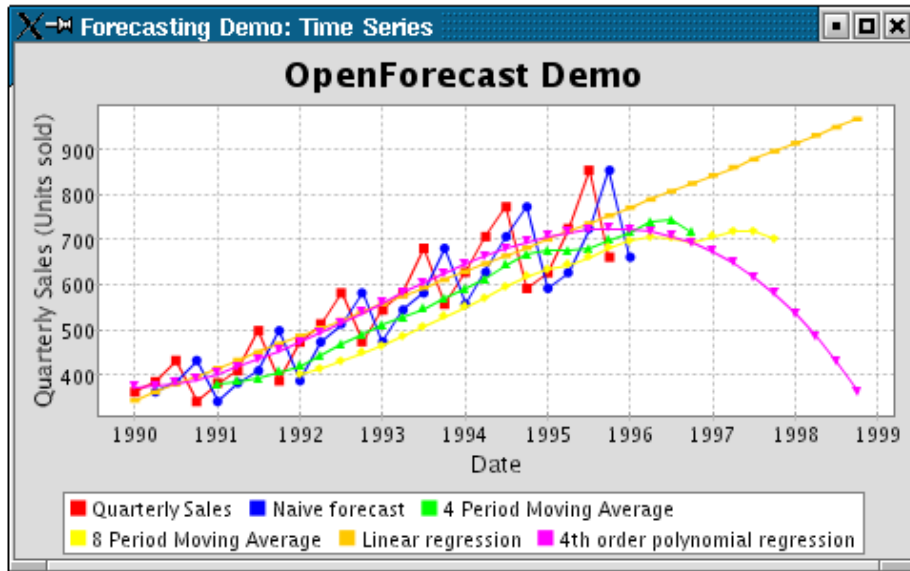


Figura 4-8 Aplicación de prueba utilizando OpenForecast.

### 4.3 MÓDULO DE ENCUESTA

Este módulo es el encargado de inferir datos sobre encuestas que se realizan a los estudiantes.

Las encuestas se infieren acerca de las materias que cada estudiante desea y debe tomar.

Para la obtención de estos datos se desarrolló un subsistema web que recopila información acerca del progreso de los estudiantes, así también como las materias que tiene previsto el estudiante tomar.

En el caso de la Secretaría Académica, a pesar de que todos estos datos existen dentro de la universidad, esta información está clasificada como confidencial, por lo que nosotros debemos construir nuestra propia base de datos. Lo único con lo que se cuenta es con el porcentaje de avance que tiene cada estudiante.

Con este porcentaje lo que se hace es preguntarle al estudiante en una primera instancia que elija las materias que ha tomado a lo largo de su carrera. Los créditos que tiene cada materia son conocidos, por lo que el sistema una vez que el estudiante ingresa su lista de materias, calcula a partir de esta selección el porcentaje resultante y lo compara con el real almacenado.

Si el porcentaje real no coincide con un pequeño rango de tolerancia, al estudiante se le pide que vuelva a introducir los datos de manera correcta.

Una vez hecho esto se le pide al estudiante que ingrese la demanda académica para el siguiente periodo a cursar.

Hasta este punto el sistema de encuestas suena simple, sin embargo existen diversas situaciones que podrían dar como resultado un número erróneo de secciones si no se toman

en cuenta. Una primera circunstancia es la de que la planeación se realiza antes de saber si los estudiantes aprobaron o no un curso. Esto es insignificante para los cursos que no requieren algún tipo de prerrequisito. Por otro lado, si un curso requiere haber aprobado otro curso anterior, es posible que se estime más de lo debido si el curso tiene un alto grado de reprobación. En este caso los problemas que nos acarrea son más secciones de las debidas para el curso con los prerrequisitos y menos secciones para el curso cuyo grado de reprobación es alto.

Para solucionar este problema se toma en cuenta una propuesta utilizando un factor de corrección. Dicho factor representa la probabilidad, o más bien el índice, de que del total de los aspirantes a un curso, cuántos de estos realmente van a poder cubrir sus prerrequisitos para poder ofertarlo.

Estos factores se calcularon mediante observación de índices de reprobación de todas las materias y con base en las series de tiempo proporcionadas.

Antes de exponer la fórmula propuesta, el sistema realiza la conversión al número de secciones viable por una constante que dicta el número de estudiantes promedio en una sección. Esta constante también varía de acuerdo al curso sobre el que se realice la inferencia.

La fórmula es la siguiente:

$$NC = \frac{N}{C} * F_a \quad (25)$$

Donde:

- $NC$  es el número de cursos a ofertar
- $N$  es el número de estudiantes que propusieron la demanda al curso
- $C$  es el número de cupos promedio cubiertos para cada sección anteriormente
- $F_a$  es el factor de corrección descrito anteriormente

Hasta ahora tenemos cubierto todo con respecto a la opinión estudiantil, sin embargo, ¿Qué sucede si la participación estudiantil es baja? Para afrontar esta situación se propone otra fórmula que ayude a “rellenar” el porcentaje de estudiantes faltante.

Para esto el sistema calcula que de todos los usuarios registrados, el número de estudiantes candidato a tomar un curso menos el número de estudiantes candidatos que hayan participado, ya sea eligiendo o no dicho curso.

Si la participación de los candidatos a un curso es nula, se estima el valor sólo con el análisis de la serie de tiempo.

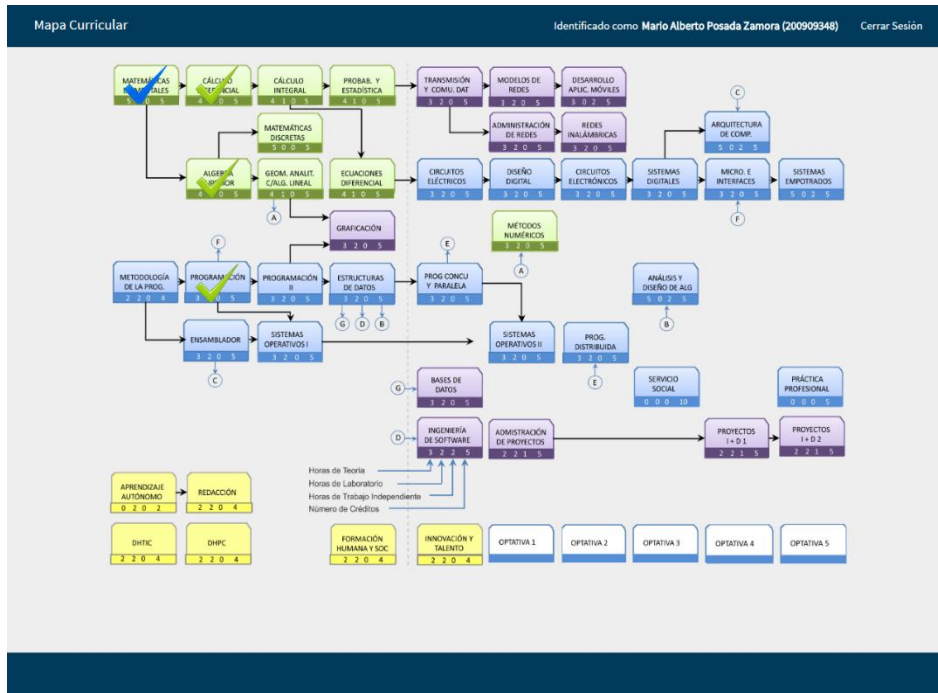


Figura 4-9 Captura de pantalla del mapa de Ing. en Ciencias de la Computación para el módulo de encuestas.

Para el caso en el que la participación no sea nula se toma en cuenta la siguiente fórmula:

$$NC = \alpha \left( \frac{N}{C} * F_a \right) + (1 - \alpha) V_{t'} \quad (26)$$

Donde:

- $NC$  es el número de cursos a ofertar
- $N$  es el número de estudiantes que propusieron la demanda al curso
- $C$  es el número de cupos promedio cubiertos para cada sección anteriormente
- $F_a$  es el factor de corrección descrito anteriormente
- $\alpha$  es un factor calculado que permite adecuar el peso de las observaciones
- $V_{t'}$  es el valor del mismo periodo anterior.

Para calcular el valor de  $\alpha$  se utiliza el nivel de participación estimado para cada materia de los posibles candidatos a la misma. El sistema revisa a los estudiantes que tengan la posibilidad de tomar esa materia y corrobora que han participado en la encuesta para el periodo que estamos analizando.

Ahora bien no se ha hecho mención en el caso de estudiantes que puedan no ser del todo honestos en su opinión. Para tratar de minimizar los ruidos causados por este tipo de interferencias en las encuestas se propone un modelo de confianza en el que cada acción en la que el sistema pueda detectar un engaño en la información y reflejar estas acciones en un coeficiente que representa la fiabilidad del estudiante.

Dicho coeficiente es una variable que comienza en valor de 1 para todos y se va decrementando con cada acción deshonesto detectable que surja. Para el cálculo de la demanda se tira una ruleta con la probabilidad del coeficiente de confianza. Si la ruleta favorece al estudiante, su opinión para cada materia será tomada en cuenta, en caso contrario se le tomará como ausente su participación.

El sistema está hecho en Java, específicamente la herramienta web en un framework de desarrollo web en MVC llamado Struts 2 perteneciente a Apache.

#### 4.3.1 Struts 2

Es un framework de código abierto para el desarrollo de aplicaciones web en Java EE. Usa y extiende el API de Servlets de Java para que se adopte una arquitectura de MVC (Modelo Vista Controlador).

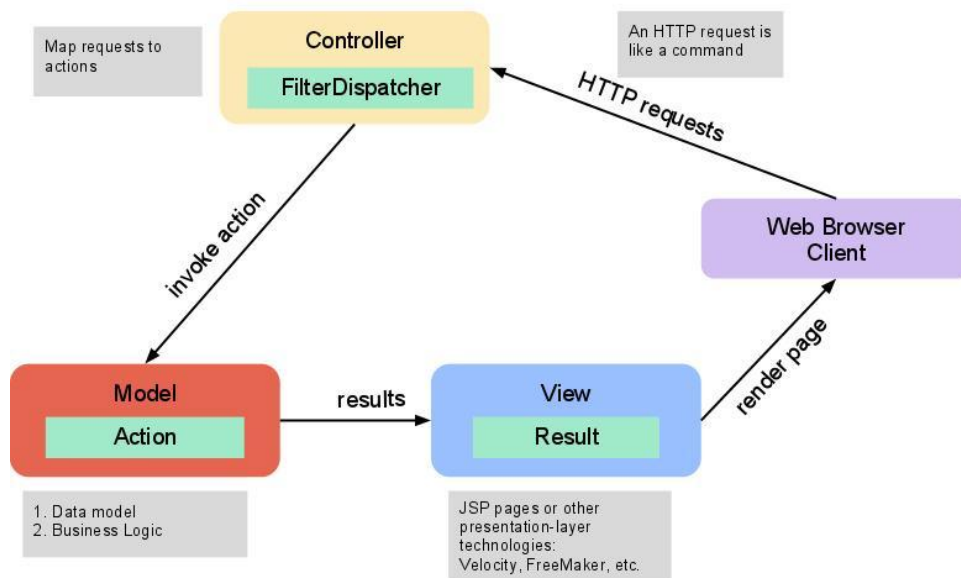


Figura 4-10 Flujo de actividades del framework Struts 2.

Su modelo de trabajo consiste en separar las capas del desarrollo web en el API de Servlets, dejando en tres capas principales toda la aplicación:

- El FilterDispatcher que equivale al controlador que gestiona las comunicaciones finales con el cliente y realiza el primer procesamiento a las entradas y el último a las salidas.
- Las clases del modelo llamadas JavaBeans las cuales son el modelo de la aplicación
- Las interfaces gráficas y los modelos de renderizado que equivalen a la vista de la aplicación, aquí se contiene todo con respecto a cómo se debe generar un resultado

Las primeras dos capas deben administrarse utilizando el lenguaje de programación Java, entre algunas otras herramientas, mientras que para la tercera capa se deben ocupar lenguajes de marcas (o etiquetas) para describir cómo debe entregarse la información

resultante de una operación. Se utiliza generalmente JSP como lenguaje de estructura de las páginas y etiquetados con frameworks como Velocity o FreeMarker para hacer la conexión entre el modelo y la vista.

En el sitio de Struts 2 se pueden consultar la documentación completa de cómo utilizar el framework y los aditamentos del mismo. Este tema queda fuera del alcance de este trabajo debido a la gran extensión para explicar el desarrollo de una aplicación.

#### 4.4 CONJUNCIÓN DE RESULTADOS

Por último, cuando el sistema determina los dos valores para cada módulo realiza una comparación entre ellos para ver si coinciden o si es necesario ajustar la demanda. Para esto verificamos diferentes casos.

En primera instancia, si los valores son iguales se determina que la demanda se ajusta a lo inferido por la parte histórica y se exhibe ese como el valor final.

En caso contrario, se elige un esquema de pesado entre los dos valores sobre qué tanto porcentaje entre los dos debe tomarse en cuenta.

$$CF_t = \alpha S_t + (1 - \alpha)E_t \quad (26)$$

Donde:

- $CF_t$  es el cálculo final del número de secciones en un periodo de tiempo  $t$ .
- $\alpha$  es la constante de pesado.
- $S_t$  es el valor de la predicción utilizando las series de tiempo hasta el periodo  $t$ .
- $E_t$  es el valor del número de secciones resultante por la encuesta.

De igual manera al usuario siempre se le presenta la opción de consultar los resultados parciales por si es necesario determinar en qué medida se debe tomar cada una de las propuestas.

La propuesta de utilizar dos valores provenientes de diferentes fuentes nos ayuda a realizar correcciones en los casos en el que la demanda rebese por mucho a lo pronosticado debido al avance natural de los estudiantes. También nos ayuda para corroborar si los datos que se obtienen entre ambos sistemas son consistentes, ya que si ambos arrojan valores muy cercanos o iguales quiere decir que vamos por una buena aproximación.

Cabe aclarar que la parte del sistema de encuestas aún entrega datos teóricos sobre las encuestas debido a que no se han aplicado a los estudiantes y es necesario que se realicen algunos ajustes al funcionamiento del sistema.

## 5 PRUEBAS Y RESULTADOS

---

En general los modelos propuestos han arrojado resultados bastante buenos, como ya se ilustró en algunos ejemplos vistos a lo largo del capítulo 3, muchos de los modelos de predicción se ajustan de manera adecuada a la estructura de las secciones que intentan calcular. Sin embargo los ejemplos fueron limitados a sólo la técnica que se estaba abordando.

En este capítulo vamos a ver ahora la comparación de diferentes técnicas sobre el mismo conjunto de datos, además del caso de una materia donde no es fácil estimarla y la razón por la cual no se ajusta a ningún tipo de estructura conocida para precisar.

Para cada prueba vamos a tomar un cierto número de cursos de cada grupo de cursos. Estos grupos están formados por los cursos que se toman de acuerdo al porcentaje de avance estimado que los estudiantes llevan, se dividen en cursos básicos (los cursos tomados en los primeros tres o cuatro cuatrimestres), cursos intermedios (los cursos programados para tomarse del cuarto al séptimo cuatrimestre), y los cursos avanzados (los cursos que se están programados para ser tomados en los últimos tres cuatrimestres).

Los cursos más fáciles son los cursos de valor invariante que consisten en el conjunto de cursos avanzados y algunos cursos intermedios de las carreras de Licenciatura en Ciencias de la Computación y de Ingeniería en Tecnologías de la Información, debido a que sus valores generalmente oscilan en un rango pequeño de valores, usualmente con un único valor de diferencia.

Los cursos más difíciles son los cursos intermedios que se ofertan sobre demanda, como el caso de algunas materias llamadas de tronco común universitario.

Comenzaremos abordando los cursos de acuerdo a su facilidad para ser analizados basados en la estructura que sus series de tiempo denotan.

### 5.1 PRUEBAS SOBRE CURSOS INVARIANTES

Los cursos avanzados tienen la peculiaridad de ser más fáciles de estimar debido a que generalmente sólo se ofertan una o dos secciones a lo mucho para cada periodo. Esto significa que generalmente con una media aritmética es mucho más que suficiente para poder estimar los resultados correctos.

Se reportan los resultados de aplicar diferentes modelos matemáticos de suavizado para el conjunto de datos del curso Lógica Matemática. En la figura 5-1 se muestran las formas gráficas de la serie original y los valores estimados con los suavizados.

Se utilizaron las constantes de suavizado en 0.95, de tendencia en 0.4 y de temporalidad en 0.5 para los modelos que lo requieran. También se contrastó con el promedio móvil de tres periodos.

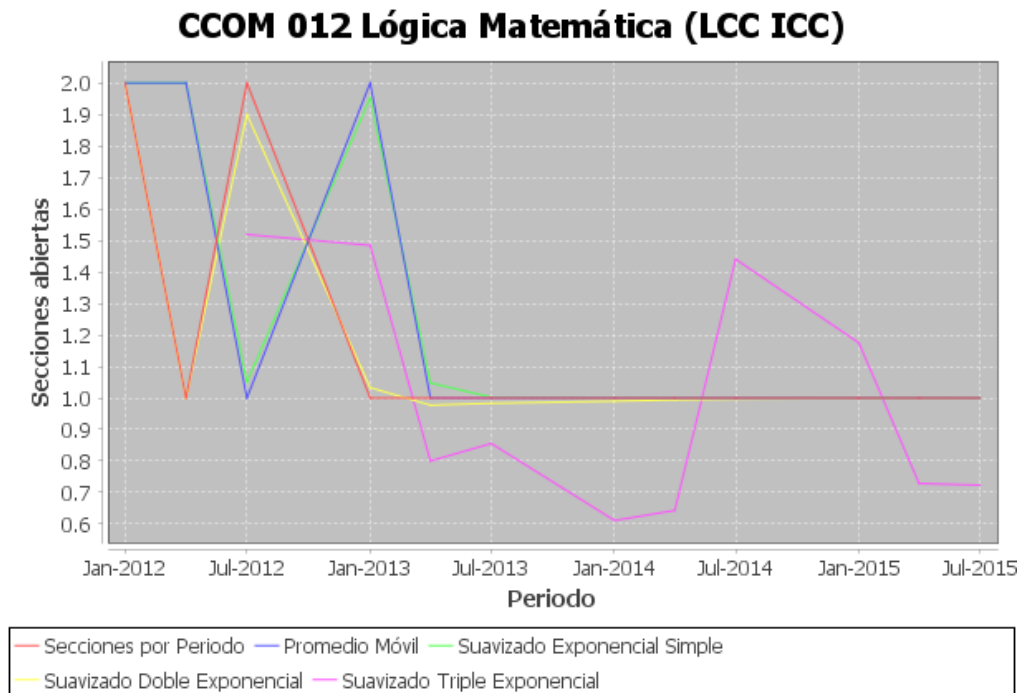


Figura 5-1 Comparación de técnicas en una serie de tiempo poco variante.

La conclusión de esta prueba es que los suavizados simple y doble fueron la mejor opción, mientras que los triple y el promedio móvil no fueron tan buenos ajustándose a los valores que fueron necesarios.

Sobre los cursos donde la variación es menor conviene ocupar modelos simples o dobles de suavizado exponencial en base a las observaciones obtenidas en diferentes pruebas.

## 5.2 PRUEBAS SOBRE CURSOS CON RANGO DE VARIACIÓN MODERADA

Para esta categoría, un rango de variación moderada consiste en cursos donde los valores oscilen entre un rango donde el valor mínimo y máximo tengan de diferencia mínima de 3 y una diferencia máxima de 5. Si ponemos un ejemplo un rango mínimo sería de 1 a 4 la diferencia, mientras que para el máximo sería de 2 a 7.

Para estos cursos la tendencia o la temporalidad se puede observar pero no se encuentra lo suficientemente bien definida como para que las técnicas dedicadas funcionen, sin embargo vamos a ilustrar esta situación con un ejemplo.

Cabe destacar que la mayoría de los cursos intermedios se encuentran en este rango de valores.

En los resultados que se pueden ver en la figura 5-2 se observa que el modelo más próximo a la línea roja (datos reales) es el perteneciente a la línea rosa (suavizado triple), además de estar otro modelo cerca que pertenece a la línea amarilla (suavizado doble).

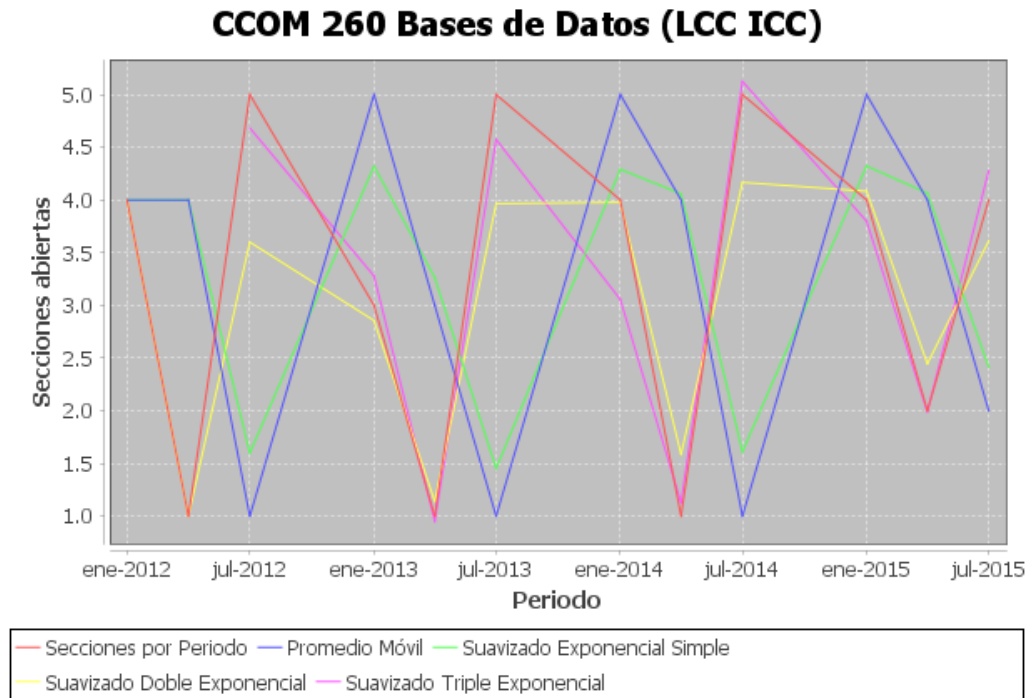


Figura 5-2 Comparación de técnicas en una serie de tiempo con variación media.

Es importante aclarar, antes de hacer un análisis detallado de lo que pasó aquí, sobre la estructura de la serie de tiempo, la cual contaba con nula tendencia pero con temporalidad muy marcada, sobre todo en los periodos de verano, lo que teóricamente debió haber salido mejor con un suavizado triple y así fue.

Las pruebas se realizaron con un valor de 0.8 para la constante de suavizado, 0.3 para la constante de tendencia y 0.5 en la constante de temporalidad.

El promedio móvil y el suavizado simple no logran una buena aproximación debido a que toman los promedios de cierta manera sin considerar que los periodos no son exactamente iguales y existen fluctuaciones si bien no tan grandes, están bien definidas. El suavizado doble puede remediar un poco la situación debido al rango moderado pero el mejor ajuste lo tuvo el suavizado triple debido a las características de la serie.

En conclusión ya se observó que en este tipo de series resalta mucho la estructura de la serie para sacar una conclusión. En las series invariantes generalmente los modelos más simples funcionan mejor que los complejos, sin embargo en esta categoría pueden haber series casi invariantes o series fluctuantes como la que vimos en la figura 5-2.

### 5.3 PRUEBAS SOBRE CURSOS CON RANGO DE VARIACIÓN ALTA

Este tipo de cursos se destaca por tener un rango de variación de sus valores mínimo y máximo en más de 5. Generalmente los cursos básicos entran en esta categoría donde la demanda en ciertos periodos es más alta que en otros, y marcan una temporalidad mucho más pronunciada. Los cursos con tendencia continua entran aquí también, sin embargo en

la práctica no se han encontrado cursos que cumplan con esta característica con un rango de variación alta.

Para estas pruebas vamos a observar el caso de un curso básico, mientras que después analizaremos un caso de estudio particular donde no parece existir una estructura conocida y los resultados que obtuvimos.

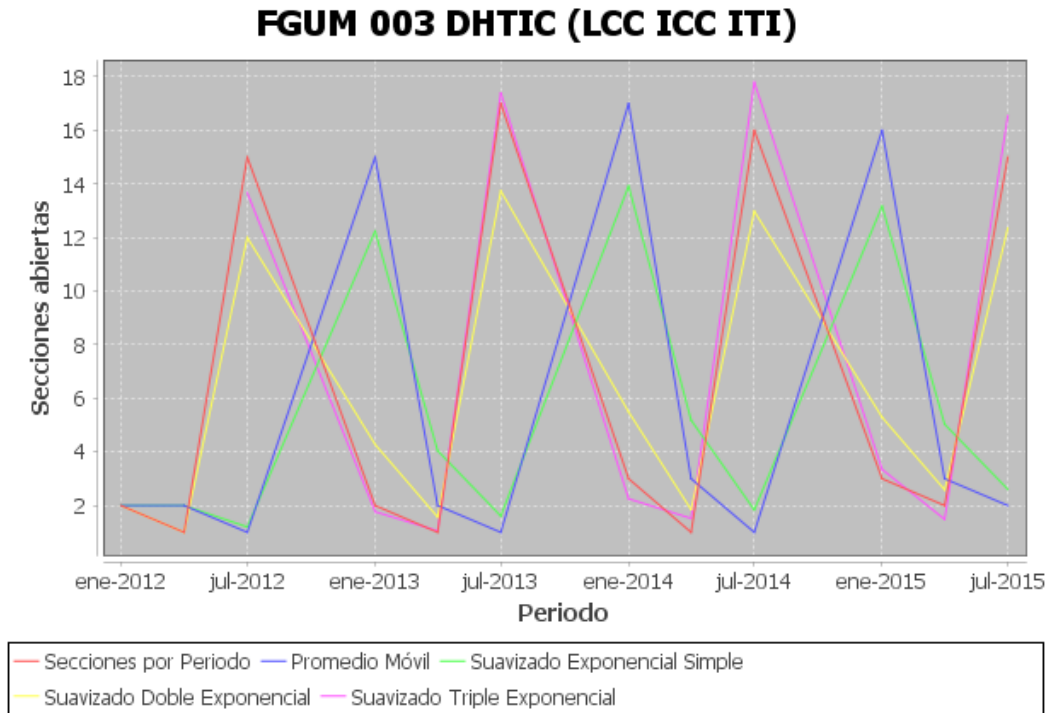


Figura 5-3 Comparación de técnicas en una serie de tiempo con variación alta y temporalidad.

En la figura anterior se aprecia que un solo modelo es el mejor para ajustarse a la serie. Es de nuevo el suavizado triple.

El ejemplo muestra una serie con una extremadamente bien definida temporalidad cuyos picos más altos ocurren en los periodos de otoño. Según las definiciones que vimos sobre los suavizados exponenciales la que se ajustaría mejor sería el suavizado triple, lo cual no ha sido sorpresa que lo haya hecho.

El ajuste de las constantes fue de 0.7 para la constante de suavizado y 0.2 para el resto de las constantes.

Hasta ahora hemos presentado ejemplos donde las series tienen una estructura conocida y de cierta manera y con el juego correcto de constantes se puede llegar a un resultado muy aproximado, sin embargo existe un caso particular que vale la pena analizar el cual parece ser que no tiene una estructura bien definida.

### 5.3.1 Caso particular: Formación Humana y Social

Como ya se ha venido mencionando desde el inicio de este capítulo, existe un caso particular, el del curso Formación Humana y Social, cuya serie de tiempo no presenta una

estructura conocida, y esto es debido a que sus valores fluctúan dentro de un rango alto y son valores que varían de manera aleatoria ya que desde la planeación este curso simplemente se oferta a manera de rellenar y no como tal analizando lo que se ha abierto anteriormente.

**FGUM 001 Formación Humana y Social (LCC ICC ITI)**

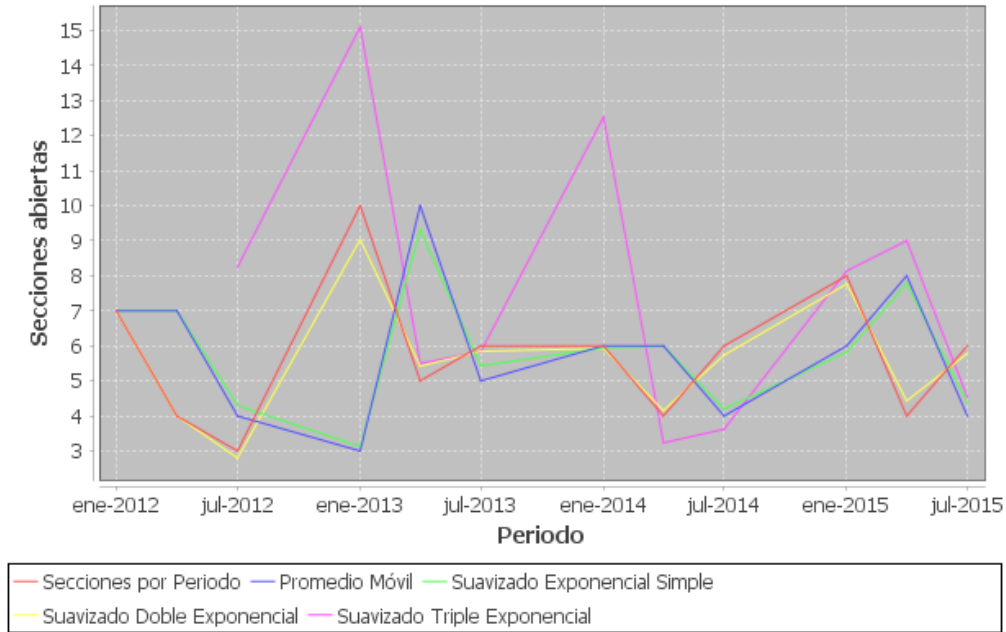


Figura 5-4 Comparación de técnicas en una serie de tiempo con variación media.

La serie que consiguió mejor comportamiento fue la que se aproximó con el suavizado doble pero con parámetros demasiado forzados. Se utilizó una constante de suavizado de 0.95 y el resto de las constantes en 0.4 aproximadamente.

Aunque en la gráfica se aprecia que el suavizado doble casi adoptó la forma de la gráfica, básicamente estamos utilizando los nuevos valores en vez de los viejos con ese valor de constante tan próximo a uno. Esto representa un problema al momento de querer pronosticar valores que no conozcamos debido a que no estaremos seguros si son del todo correctos por ese valor tan alto. De cualquier forma podría servir para dar una aproximación, pero es por este caso específico que se decidió tomar en cuenta el sistema de encuestas como un correctivo de este número.



## 6 CONCLUSIONES Y TRABAJO FUTURO

---

Dentro del marco de aplicaciones que se pueden construir para resolver problemas de la vida real utilizando la teoría vista a lo largo de la carrera se puede ver que podemos aplicar tópicos que quizá no sean tan complejos, como el caso particular de un suavizado en una serie temporal, pero que resulten ser herramientas poderosas en el auxilio de la planeación de cursos para la Facultad de Ciencias de la Computación.

Aunque pronosticar el número de secciones es un tema que parece sencillo a primera instancia, es información muy útil que se puede emplear en una corrección de la planificación dando a conocer, sobre todo con la asistencia del sistema de encuestas, los casos donde realmente no se esté prestando la suficiente atención y sea necesaria la revisión y ajuste.

Es posible que incluso utilizando la ayuda de sistemas expertos o sistemas de soporte de decisiones se pueda cumplir al pie de la letra la sugerencia, sobre todo en los casos donde las restricciones de un problema hacen que no exista tan fácilmente una solución óptima tomando en cuenta que dicha solución se estima buena de acuerdo al grado de satisfacción humana y no a parámetros rígidos que hagan más fácil su estimación.

A pesar de lo anterior, se plantea la idea como segunda parte de este sistema que se realice la distribución de los cursos de acuerdo a la cantidad de profesores candidatos a impartir una sección de acuerdo a su dominio y disponibilidad utilizando los criterios que emplea la Secretaría Académica para realizar esta tarea. Esto haría una planeación de acuerdo a los números que arroje esta propuesta hacia el otro sistema, lo que permite que la etapa desarrollada en el presente trabajo sirva para complementar la siguiente tarea.

Este tipo de colaboración entre sistemas ilustra que varias instancias pueden comunicarse siendo cada una especializada en una tarea para al final dar un resultado conjunto. En el caso particular de nuestra propuesta con la idea futura es que realmente aunque se proponga un número de secciones, éste no necesariamente debe ser fijo para ser procesado cuando se haga la distribución de la carga académica, es decir que se podrá modificar a voluntad el número de secciones ya que realmente el sistema propuesto aquí es auxiliar más no mandatorio en una decisión.



## 7 BIBLIOGRAFÍA

---

- [1] J. De los Santos, «Los procesos de permanencia y abandono escolar en educación superior.,» [En línea]. Available: <http://www.rieoei.org/deloslectores/628Santos.PDF>.
- [2] A. H., 25 02 1996. [En línea]. Available: <http://home.ubalt.edu/ntsbarsh/opre640s/spanishp.htm>.
- [3] M. Druzdzel y R. Flynn, « Decision Support System,» *Encyclopedia of Library and Information Science*, 2002.
- [4] D. Kahneman y G. Klein, «Conditions for intuitive expertise: A failure to disagree,» *American Psychologist* 64 (6), p. 515-526, 2009.
- [5] J. Baron, *Thinking and deciding* (3rd ed.), New York: Cambridge University Press, 2000.
- [6] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.*, San Mateo, CA: Morgan Kaufmann Publishers, Inc., 1988.
- [7] H. A. Simon, *Causal ordering and identifiability.*, New York, NY: John Wiley & Sons, Inc., 1953.
- [8] J. Pearl, *Causality: Models, Reasoning, and Inference.*, UK: Cambridge University Press, 2000.
- [9] M. J. Druzdzel y H. A. Simon, «Causality in Bayesian belief networks,» *In Proceedings of the Ninth Annual Conference on Uncertainty in Artificial Intelligence*, 1993.
- [10] «FET Free Timetabling Software,» 16 11 2015. [En línea]. Available: <http://lalescu.ro/liviu/fet>.
- [11] J. M. Torralba, O. Coltell y J. M. Torralba López, «Innovación en la enseñanza en grupos numerosos».
- [12] D. Brown, C. M. Davis y S. Stanlick, *Struts 2 in Action*, Greenwich, CT: Manning, 2008.
- [13] J. Duckett, *Beginning HTML, XHTML, CSS and Javascript*, Indianapolis, IN: Wiley Publishing, Inc., 2010.
- [14] NIST/SEMATECH, «e-Handbook of Statistical Methods,» Abril 2012. [En línea]. Available: <http://www.itl.nist.gov/div898/handbook/>. [Último acceso: Septiembre 2015].

- [15] E. Turban, J. E. Aronson y T.-P. Liang, *Decision Support Systems and Intelligent Systems*, New Delhi: Pearson Education, 2005.
- [16] V. T. N. Chau y N. H. Phung, *A Knowledge-Driven Educational Decision Support*.
- [17] M. Telem, «DSS in Educational Organizations,» *Computers Education*, vol. 14, pp. 61-69, 1990.
- [18] M. Telem, «Educational DSS: Potential Services, Benefits, Difficulties,» *Computers Education*, vol. 14, pp. 71-80, 1990.
- [19] D. Z. Deniz y I. Ersan, «An Academic Decision-Support System Based on Academic Performance Evaluation for Student and Program Assessment,» *Int. J. Engng*, vol. 18, pp. 236-244, 2002.
- [20] S. America, *Data Analysis and Statistical Model Selection*, SAP AG, 2003.
- [21] R. Garns, «Judgment Heuristics and Biases,» [En línea]. Available: [http://www.nku.edu/~garns/165/pptj\\_h.html](http://www.nku.edu/~garns/165/pptj_h.html). [Último acceso: 15 09 2015].
- [22] «Técnicas de Predicción,» [En línea]. Available: [http://www.ub.edu/aplica\\_infor/spss/cap8-5.htm](http://www.ub.edu/aplica_infor/spss/cap8-5.htm). [Último acceso: 10 10 2015].
- [23] «Heuristics in judgment and decision making,» [En línea]. Available: [https://en.wikipedia.org/wiki/Heuristics\\_in\\_judgment\\_and\\_decision-making](https://en.wikipedia.org/wiki/Heuristics_in_judgment_and_decision-making). [Último acceso: 25 10 2015].
- [24] «Forecasting,» [En línea]. Available: <https://en.wikipedia.org/wiki/Forecasting>. [Último acceso: 25 10 2015].
- [25] «Time Series Analysis,» [En línea]. Available: [https://en.wikipedia.org/wiki/Time\\_series](https://en.wikipedia.org/wiki/Time_series).
- [26] «Time domain,» [En línea]. Available: [https://en.wikipedia.org/wiki/Time\\_domain](https://en.wikipedia.org/wiki/Time_domain). [Último acceso: 28 10 2015].
- [27] «Struts 2,» [En línea]. Available: [struts.apache.org](http://struts.apache.org). [Último acceso: 06 11 2015].