

Sistema para la gestión y analítica de datos estadísticos para medir el rendimiento de los jugadores en un equipo de béisbol.

Trabajo Terminal No. 2023-A014

*Alumnos: Callejas Díaz Eduardo, Casas Carbajal José Mauricio**

Directores: Ocampo Botello Fabiola

**e-mail: jcacasasc1300@alumno.ipn.mx*

Resumen

Proponemos realizar un sistema que se enfoque en servir como herramienta para medir el rendimiento de los jugadores dentro de un equipo de béisbol, particularmente de la Escuela Superior de Cómputo (ESCOM) del Instituto Politécnico Nacional (IPN), haciendo uso de la sabermetría para generar métricas y aplicando técnicas de minería de datos, generar reportes que muestren información útil brindando una ayuda al mánager para la toma de decisión, tomando como referencia los resultados de entrenamiento y juegos.

Palabras clave – Analítica de Datos, Bases de Datos, Béisbol, Minería de Datos

1. Introducción

En el béisbol, los números siempre han jugado un rol importante en el juego, las estadísticas han contribuido a predecir los resultados no solo de campeonatos, sino también de series, juegos, entradas e incluso turnos al bat. Por lo que el conocimiento y análisis de las diferentes estadísticas que existen en un juego de béisbol puede ser importante para fans, jugadores, managers, dueños, etc.

El análisis de las estadísticas se ha vuelto tan importante que en el mundo del béisbol se han creado tendencias especializadas únicamente en realizar estos análisis como la sabermetría, la cual, como establecen los autores James & Wirth:

“La sabermetría ha sido definida como el análisis empírico del juego de béisbol a través del estudio de la evidencia objetiva obtenida, específicamente usando técnicas estadísticas, cuyo fin es medir de manera eficaz las actividades que se suscitan en el terreno de juego” [1]

El término derivado de SABR hace referencia a la Sociedad para la Investigación del Béisbol Americano, esta sociedad se ha encargado de crear estadísticas “sabermetrías” para medir el rendimiento de los jugadores en el terreno de juego. [2]

El detalle es que, por lo general, el análisis de estas estadísticas tiende a ser de manera manual en hojas de papel, siendo que de esta forma no se pueden apreciar de una manera más práctica los resultados de estos. En el Instituto Politécnico Nacional no existe una manera estadística de medir el rendimiento de los jugadores de béisbol, nuestra propuesta desea crear una herramienta para el mánager de un equipo que permita ingresar los resultados de los juegos de béisbol para que posteriormente estos datos sean procesados en la base de datos y genere las métricas que permitan medir el rendimiento de los jugadores, de acuerdo con la “sabermetría”.

Haciendo uso de técnicas de analítica de datos para capturar, procesar y analizar los datos pretendemos obtener conclusiones en la información sobre el rendimiento de los jugadores [3]. La analítica de datos cuenta con herramientas como la minería de datos también conocida como el proceso del KDD (Knowledge Discovery in Databases) que permite hallar patrones, anomalías y correlaciones en grandes conjuntos de datos [4]. La minería de datos nos ayudará a mostrar más información que no se ve a simple vista y que pueda servir para tomar mejores decisiones. Usando como fuente la información obtenida mediante el registro del mánager, y el resultado del procesamiento de estos datos, se aplicarán diferentes técnicas de minería de datos que nos permitan generar reportes que ayuden a la toma de decisiones.

Estado del arte

Investigando acerca de nuestra propuesta, encontramos ciertas herramientas similares a la que proponemos, pero no cuentan con las características que se requieren y que son tomadas en cuenta para la elaboración de nuestra propuesta.

Dentro de las aplicaciones encontradas existe una llamada “Basketball Scorebooks & Charts” la cual permite generar estadísticos por jugador y por periodo, siendo esta una manera más sencilla de previsualizar de los resultados, aspectos que se tomarán en cuenta en nuestra aplicación.

Además, no todas las aplicaciones están enfocadas en el béisbol, como lo está nuestro proyecto, o tienen una mala traducción que termina afectando a la experiencia y a la interpretación de los resultados. La única herramienta que integra más características consideradas en este proyecto es el uso de plantillas de Scorecards (tarjetas de puntuación) en la aplicación de Excel.

Scorecards (tarjetas de puntuación) es una aplicación de Excel que toma en cuenta diversas características consideradas en este proyecto. La desventaja del uso de Excel es que, aunque es una herramienta eficiente para el manejo y generación de datos, tiene una interacción limitada en cuanto a las consultas que los usuarios desean realizar.

Las herramientas que pretendemos utilizar para la visualización de estadísticas en el entorno del béisbol son Iscore, Ballstat Ballscore, Baseball Softball Lineups Keeper y Mis Estadísticas Baseball.

Aunque no todas ofrecen las mismas características que las demás o se muestran limitadas en lo que ofrecen. Tal es el caso de Mis estadísticas baseball que no ofrece la misma diversidad que Iscore, o ballstat ballscore, donde a pesar de que, si genera estadísticas, lo hace a través de una segunda aplicación, por lo que requiere un cierto conocimiento para el uso de ésta.

En la siguiente página, se puede observar una tabla comparativa denotadas con color verde cuando cuentan con las características deseadas y en rojo cuando no de las siguientes herramientas:

1. Iscore [5]
2. Basketball Scorebook & Charts [6]
3. Plantillas de Scorecards en Excel
4. Ballstat Ballscore [7]
5. Baseball Softball Lineups Keeper [8]
6. Mis Estadísticas Baseball [9]

Software	Características						Precio del mercado	
	Registro de datos del equipo	Registro de datos del jugador	Registro de resultados de juegos	Registro de alineaciones	Generación de estadísticas	Lenguaje en español	Técnicas adicionales de minería de datos	Gratis
Iscore				Se deben registrar en cada juego		Mala traducción	Compras integradas	
Ballstat Ballscore					Hace uso de aplicaciones adicionales			
Baseball Softball Lineups Keeper			Únicamente de resultados finales (Sin detalle)					
Mis estadísticas baseball	Aplicación para el uso personal de un jugador				Cantidad limitada de registros	Mala traducción	Contiene Anuncios	
Basketball Scorebooks & Charts								Compra única de \$23.00 MXN
Plantillas de scorecards en excel	Dependiendo el conocimiento en Excel del usuario							
Sistema para la gestión y analítica de datos estadísticos para medir el rendimiento de los jugadores en un equipo de béisbol.		Especializando en los datos manejados en la escuela						

Tabla 1 Estado del arte (elaboración propia)

2. Objetivo

Desarrollar un sistema para los mánager de los equipos de béisbol en la liga interna del IPN, que les permita registrar los datos de sus jugadores y los resultados de los juegos de béisbol, para posteriormente medir el rendimiento de estos haciendo uso de métricas y técnicas de minería de datos.

Como objetivos particulares se consideran:

1. Desarrollar un módulo para el registro de los datos de los alumnos que integran el equipo de béisbol en una temporada.
2. Desarrollar un módulo para el registro de los resultados de los juegos de béisbol.
3. Generar métricas para medir el rendimiento de los jugadores con base en la sabermetría.
4. Desarrollar un módulo para la generación de reportes con datos estadísticos descriptivos y consultas considerando ciertos parámetros y algunas técnicas de minería de datos.
5. Diseñar una interfaz intuitiva al usuario que sea fácil de manejar que facilite la gestión de su equipo de béisbol.

3. Justificación

Un problema que se tiene en los equipos de béisbol de la liga interna del IPN es la ausencia de un sistema que les permita administrar los datos de sus jugadores, así como también conocer con datos estadísticos el rendimiento de estos.

Por ejemplo, en la ESCOM, cuando un alumno está interesado en ingresar al equipo de béisbol, deberá notificar al coordinador del club o mánager del equipo para pedir informes. Posteriormente, el mánager realiza el llenado de un formulario a mano en papel para tener los datos de los miembros en el equipo de ESCOM [10]. Como se puede observar, la falta de un sistema para la gestión de estos datos puede complicar mantener un orden de los miembros actuales dentro del equipo.

Cada equipo de béisbol en la liga del IPN se maneja de manera diferente, la mayoría registra los resultados de los juegos y sus jugadores en los llamados scorecards o tarjetas de puntuación. Sin embargo, estas tarjetas no se registran en ningún sistema o base de datos por lo que las estadísticas de los jugadores son desconocidas. Si un mánager quisiera conocer las estadísticas de un jugador, deberá manualmente realizar los cálculos necesarios lo cual sería una tarea muy larga y tediosa para realizar, además de poco eficiente.

El tener acceso a las estadísticas de los jugadores dentro de un equipo, permite a los mánager tomar mejores decisiones en cuanto a movimientos en la defensiva o la ofensiva durante un juego de béisbol e incluso ha permitido a equipos con bajo presupuesto obtener mejores resultados, como se ve en el libro de Lewis, M. (2004) Moneyball [11], donde haciendo uso de estos datos estadísticos, narra cómo los Atléticos de Oakland llegaron a la Serie Mundial, siendo uno de los equipos con menor presupuesto dentro de la liga.

Con nuestra propuesta, se creará una plataforma que permita gestionar los datos de los miembros dentro de un equipo, registrar los resultados de sus jugadores después o durante cada juego, realizar los cálculos necesarios para la generación de estadísticas de cada uno estos y adicionalmente, aplicar técnicas de minería de datos para encontrar información valiosa dentro los datos obtenidos que permitan al mánager tomar mejores decisiones o anticiparse a situaciones que pudieran ocurrir en un juego.

El proyecto que se presenta no pretende sustituir ningún sistema actual, sino más bien, ofrecer una ayuda adicional a los mánager de los equipos para una mejor gestión de los datos de sus jugadores.

4. Productos o Resultados esperados

El producto final esperado es un sistema para gestión y analítica de datos, la cual sirva de ayuda al mánager de un equipo de béisbol para registrar los datos de los miembros, conocer las estadísticas de estos y mostrar información útil que no se ve a simple vista para conocer su rendimiento y poder tomar las decisiones adecuadas.

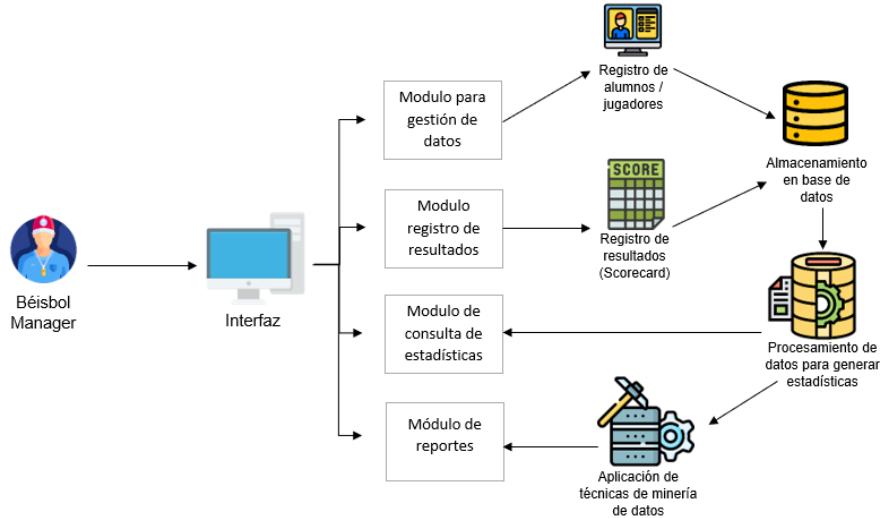


Figura 1 Diagrama del sistema (elaboración propia)

En cada módulo:

- Módulo para gestión de datos: permite gestionar los datos de los alumnos jugadores del equipo, añadir un nuevo jugador, modificar sus datos, consultar o eliminar. Permite la exportación en PDF de los datos de un jugador para su registro en torneos.
- Módulo registro de resultados: permite registrar los resultados de cada jugador en un juego de béisbol (Hits, Sencillos, Dobles, etc.). Estos datos se almacenan en la base de datos y se procesan para calcular las estadísticas necesarias.
- Módulo de consulta de estadísticas: despliega los resultados y estadísticas de los jugadores registrados en el equipo, permitiendo al mánager tener conocimiento del rendimiento de sus jugadores y ayudándolo en la toma de decisiones
- Módulo de reportes: despliega los resultados de las consultas y estadísticos descriptivos para mostrar al mánager información útil y otra generada al aplicar técnicas de minería de datos. De esta manera, el mánager tendrá un mayor entendimiento de sus jugadores, sus ventajas y desventajas, ayudándolo así en la toma de decisiones y a tomar las acciones necesarias para reforzar al equipo.

Los entregables al final del producto son:

1. Manual técnico
2. Manual de usuario
3. Código fuente del sistema en Java, Python para las técnicas de minería de datos y SQL para la base de datos.
4. Sistema funcional.

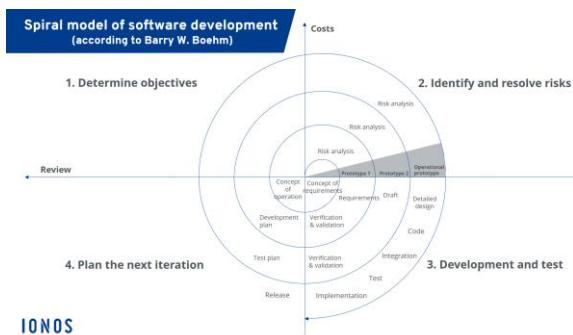
5. Metodología

El modelo en espiral es un modelo que en un principio buscaba presentarse como una alternativa al modelo en cascada, ya que, a diferencia de este último, este no parte de la base de que las tareas se desarrollen de manera lineal, sino que se pueden entender como tareas iterativas. [12]

Las fases se realizan varias veces dando forma de espiral, siendo así una repetición cíclica.

- (1) Definición de objetivos y alternativas
- (2) Valoración de las alternativas
- (3) Desarrollo
- (4) Planificación del siguiente ciclo

Figura 2 Representación Gráfica del Modelo en Espiral

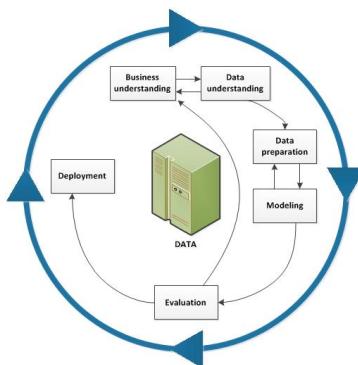


Ionos, 2020. Spiral model of software development. [Imagen] Disponible en: <<https://www.ionos.mx/startupguide/fileadmin/StartupGuide/Schaubilder/ES-spiral-modell.png>> [Último acceso 27 abril 2022].

La razón por la cual elegimos el modelo en espiral es por lo flexible que resulta ser y por las comprobaciones periódicas que se realizan para minimizar el riesgo, siendo que, de esta manera, se podrá verificar cada parte del proyecto en cada iteración.

Por otro lado, haremos uso de la metodología CRISP-DM para el proceso de minería de datos: CRISP-DM, que son las siglas de Cross-Industry Standard Process for Data Mining, es un método probado para orientar sus trabajos de minería de datos, el cual como metodología incluye descripciones de las fases normales de un proyecto, las tareas necesarias en cada fase y una explicación de las relaciones entre las tareas. [13]

Figura 3 Ciclo de vida de minería de datos



IBM, 2022. Ciclo de vida de minería de datos. [Imagen] Disponible en: <https://www.ibm.com/docs/es/SS3RA7_sub/modeler_crispdm_ddita/clementine/images/crisp_process.jpg> [Último acceso 27 abril 2022].

6. Cronograma

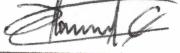
Anexo al final del documento.

7. Referencias

- [1] B. James, The Bill James historical baseball abstract., Estados Unidos: Random House Incorporated., 1988.
- [2] G. Costa, M. Huber y J. Saccoman, Understanding sabermetrics: An introduction to the science of baseball statistics., Estados Unidos: McFarland, 2019.
- [3] L. Joyanes, Inteligencia de negocios y analítica de datos, Bogota: Alfaomega, 2019.
- [4] U. Fayyad, G. Piatetsky-Shapiro y P. Smyth, «The KDD process for extracting useful knowledge from volumes of data,» Communications of the ACM, vol. 39, nº 11, pp. 27-34, 1996.
- [5] iScore Sports, «iScore Sports,» 2022. [En línea]. Disponible: <https://iscoresports.com/>. [Último acceso: 20 Abril 2022].
- [6] tricosoft, «Basketball Scorebook & Charts,» 2022. [En línea]. Disponible: https://play.google.com/store/apps/details?id=com.tricosoft.basketballscorebook.paid&hl=es_MX&gl=US. [Último acceso: 20 Abril 2022].
- [7] James Habel Software, «BallStatBallScore,» 2013. [En línea]. Disponible: <http://ballstat.com/>. [Último acceso: 20 04 2022].
- [8] Telen, «Baseball Softball Lineups Keeper,» Telen, 2022. [En línea]. Disponible: <https://play.google.com/store/apps/details?id=com.telen.easylineup>. [Último acceso: 20 Abril 2002].
- [9] Advance Software, «Mis Estadísticas Baseball,» Advance Software, 2022. [En línea]. Disponible: <https://play.google.com/store/apps/details?id=per.vmonsanto.mypersonalstats>. [Último acceso: 20 Abril 2022].
- [10] Escuela Superior de Cómputo, «Actividades Culturales y Deportivas ESCOM,» Instituto Politécnico Nacional, 2022. [En línea]. Disponible: <https://www.facebook.com/ESCOMculturaydeportes>. [Último acceso: 20 Abril 2022].
- [11] M. Lewis, MoneyBall, Estados Unidos: W. W. Norton & Company, 2003.
- [12] Á. Prieto, « Adaptación de las metodologías tradicionales cascada y espiral para la inclusión de evaluación inicial de usabilidad en el desarrollo de productos de software en México. REPOSITORIO NACIONAL CONACYT.,» 2015.
- [13] IBM Corporation, «Conceptos básicos de ayuda de CRISP-DM,» IBM, 2021. [En línea]. Disponible: <https://www.ibm.com/docs/es/spss-modeler/SaaS?topic=dm-crisp-help-overview>. [Último acceso: 27 Abril 2022].

8. Alumnos y directores

Callejas Diaz Eduardo. - Alumno de la carrera de Ing. en Sistemas Computacionales en ESCOM, Especialidad Sistemas, Boleta 2019630003: Tel. 5554120679, email ecallejas1800@alumno.ipn.mx

Firma: 

CARÁCTER: Confidencial
FUNDAMENTO LEGAL: Artículo 11 Fracc. V y Artículos 108, 113 y 117 de la Ley Federal de Transparencia y Acceso a la Información Pública.
PARTES CONFIDENCIALES: Número de boleta y teléfono.

Casas Carbajal José Mauricio. - Alumno de la carrera de Ing. en Sistemas Computacionales en ESCOM, Especialidad Sistemas, Boleta: 2014120202, Tel. 5568533735, email jcasasc1300@alumno.ipn.com.mx

Firma: 

Ocampo Botello Fabiola. - Ingeniero en Sistemas Computacionales por el Instituto Tecnológico de Ciudad Madero, Maestría en Ciencias de la Computación por el CINVESTAV-IPN y Doctorado en Educación Internacional por la Universidad Autónoma de Tamaulipas (UAT). Profesora de la Escuela Superior de Cómputo (ESCOM) desde 1995, Tel.: 57296000 Ext.: 52083, email: focampob@ipn.mx

Firma: 

CRONOGRAMA

Nombre del alumno: Callejas Diaz Eduardo

Título de TT: Sistema para la gestión y analítica de datos estadísticos para medir el rendimiento de los jugadores en un equipo de béisbol.

	Actividades	AGOSTO	SEPTIEMBRE	OCTUBRE	NOVIEMBRE	DICIEMBRE	ENERO	FEBRERO	MARZO	ABRIL	MAYO	JUNIO
Iteración 1												
1	Estudio del estado del arte											
2	Análisis de requerimientos funcionales y no funcionales											
3	Estudio y comprensión de los datos para el modulo de gestión de datos											
4	Análisis de los datos y selección de características para el modulo de gestión de datos											
Iteración 1												
5	Análisis de riesgos											
6	Análisis de costos											
7	Análisis de los datos y selección de características para el modulo de gestión de datos											
8	Modelación UML: casos de uso, descripción de casos de uso, secuencia y clases.											
9	Diseño de interfaces de los modulos											
10	Diseño de la base de datos relacional											
11	Análisis de requerimientos funcionales y no funcionales											
Iteración 1												
12	Codificación del modulo para gestión de datos: 1) Agregar datos de alumno, 2) Consultar datos, 3) Modificar datos, 4) Eliminar datos de alumno 5) Generar reportes en PDF con los datos del alumno											
13	Pruebas del modulo codificado											
Iteración 1												
14	Evaluacion TT1											
15	Retroalimentación e integración de correcciones y comentarios al proyecto											
Iteración 2												
16	Estudio y comprensión de los datos para el modulo de registro de resultados											
17	Análisis de los datos y selección de características para el modulo de registro de resultados											
18	Investigación y planeación: 1) del modulo de registro de resultados 2) de la base de datos relacional											
Iteración 2												
19	Análisis de documentos 1) del modulo de registro de resultados 2) implementación de la base de datos relacional											
Iteración 2												
20	Codificación del módulo de registro de resultados											
21	Pruebas de los modulos codificados											
Iteración 2												
22	Retroalimentación por parte de los sinodales											
23	Retroalimentación e integración de correcciones y comentarios al proyecto											
Iteración 3												
24	Estudio y comprensión de los datos para el modulo de consulta de estadísticas											
25	Análisis de los datos y selección de características para el modulo de consulta de estadísticas											
26	Investigación y planeación del módulo de consulta de estadísticas											
Iteración 3												
27	Análisis de documentos del módulo de consulta de estadísticas											
Iteración 3												
28	Codificación e implementación del módulo de consulta de estadísticas											
29	Pruebas del módulo consulta de estadísticas											
Iteración 3												
30	Retroalimentación por parte de los sinodales											
31	Retroalimentación e integración de correcciones y comentarios al proyecto											
Iteración 4												
32	Estudio y comprensión de los datos para el modulo de reportes											
33	Análisis de los datos y selección de características para el modulo de reportes											
34	Investigación y planeación del módulo de reportes											
35	Selección de modelos para desplegar											
Iteración 4												
36	Análisis de documentos del módulo de reportes											
37	Evaluación de los modelos											
Iteración 4												
38	Codificación e implementación del módulo de reportes											
39	Despliegue de los modelos											
40	Pruebas del módulo de reportes											
Iteración 4												
41	Retroalimentación por parte de los sinodales											
42	Retroalimentación e integración de correcciones y comentarios al proyecto											
43	Evaluación del TT2											



CRONOGRAMA

Nombre del alumno: Casas Carbalaj José Mauricio

Título de TT: Sistema para la gestión y analítica de datos estadísticos para medir el rendimiento de los jugadores en un equipo de béisbol.

Iteración 1	Actividades	AGOSTO	SEPTIEMBRE	OCTUBRE	NOVIEMBRE	DICIEMBRE	ENERO	FEBRERO	MARZO	ABRIL	MAYO	JUNIO
Planeación												
1	Estudio del estado del arte											
2	Análisis de requerimientos funcionales y no funcionales											
3	Estudio y comprensión de los datos para el modulo de gestión de datos											
4	Análisis de los datos y selección de características para el modulo de gestión de datos											
Análisis de riesgos												
5	Análisis de riesgos											
6	Análisis de costos											
7	Análisis de los datos y selección de características para el modulo de gestión de datos											
8	Modelación UML: casos de uso, descripción de casos de uso, secuencia y clases.											
9	Diseño de interfaces de los modulos											
10	Diseño de la base de datos relacional											
11	Análisis de requerimientos funcionales y no funcionales											
Ingeniería												
12	Codificación del modulo para gestión de datos: 1) Agregar datos de alumno, 2) Consultar datos, 3) Modificar datos, 4) Eliminar datos de alumno 5) Generar reportes en PDF con los datos del alumno											
13	Pruebas del modulo codificado											
Evaluación del software												
14	Evaluacion TT1											
15	Retroalimentación e integración de correcciones y comentarios al proyecto											
Planeación												
16	Estudio y comprensión de los datos para el modulo de registro de resultados											
17	Análisis de los datos y selección de características para el modulo de registro de resultados											
18	Investigación y planeación: 1) del modulo de registro de resultados 2) de la base de datos relacional											
Análisis de riesgos												
19	Análisis de documentos 1) del modulo de registro de resultados 2) implementación de la base de datos relacional											
Ingeniería												
20	Codificación y conexión de la base de datos relacional											
21	Pruebas de los modulos codificados											
Evaluación del software												
22	Retroalimentación por parte de los sinodales											
23	Retroalimentación e integración de correcciones y comentarios al proyecto											
Planeación												
24	Estudio y comprensión de los datos para el modulo de consulta de estadísticas											
25	Análisis de los datos y selección de características para el modulo de consulta de estadísticas											
26	Investigación y planeación del módulo de consulta de estadísticas											
Análisis de riesgos												
27	Análisis de documentos del módulo de consulta de estadísticas											
Ingeniería												
28	Codificación e implementación del procesamiento de datos para generar estadísticas											
29	Pruebas del módulo consulta de estadísticas											
Evaluación del software												
30	Retroalimentación por parte de los sinodales											
31	Retroalimentación e integración de correcciones y comentarios al proyecto											
Planeación												
32	Estudio y comprensión de los datos para el modulo de reportes											
33	Análisis de los datos y selección de características para el modulo de reportes											
34	Investigación y planeación del módulo de reportes											
35	Selección de modelos para desplegar											
Análisis de riesgos												
36	Ánalisis de documentos del módulo de reportes											
37	Evaluación de los modelos											
Ingeniería												
38	Codificación e implementación de técnicas de minería de datos											
39	Despliegue de los modelos											
40	Pruebas del módulo de reportes											
Evaluación del software												
41	Retroalimentación por parte de los sinodales											
42	Retroalimentación e integración de correcciones y comentarios al proyecto											
43	Evaluación del TT2											

