

Solución de inteligencia de Negocios con minería de datos supervisada en datos de ambientes de aprendizaje virtual.

Hugo Santiago Hernández Limas <hs.hernandez@uniandes.edu.co>

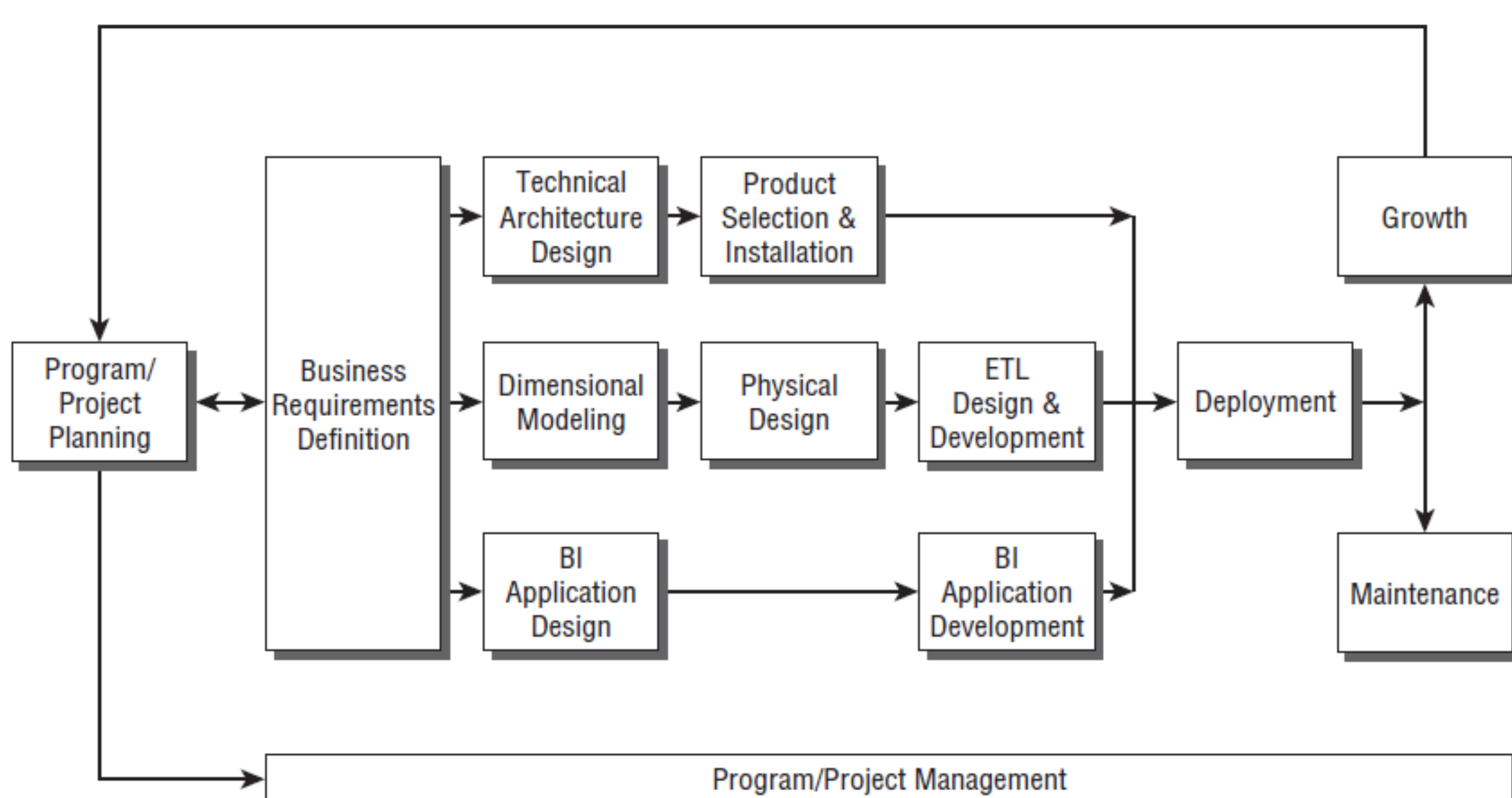
Este proyecto contiene el desarrollo de una solución de inteligencia de negocios para el Centro de Innovación en Tecnología y Educación de la Universidad de los Andes (Conecta-TE). Esto, con el objetivo de poder determinar la efectividad de los diferentes recursos desarrollados para los diferentes cursos apoyados por Conecta-TE. Con la evaluación del proyecto se busca predecir el resultado final de los estudiantes. Para esto, se tiene en cuenta variables como la cantidad de interacciones que tiene por tipo de recurso, la cantidad total interacciones después de empezado el curso y la cantidad de interacciones antes de iniciado el curso.

La información almacenada por Conecta-TE es de gran valor para el negocio en cuanto esta data es explotada. Tomando los datos de las interacciones de los estudiantes con los diferentes recursos ofrecidos por una materia, es posible implementar un algoritmo de árboles de decisión que busque predecir el resultado final de los estudiantes con base en cómo fue su compartamiento con los recursos ofrecidos. Aunque el proyecto es para Conecta-TE, el modelo fue desarrollado de manera genérica hacia cualquier fuente de información de interacción con recursos. Se especifica la fuente de datos del proyecto Open University Learning Analytics.

Para este proyecto se definieron 2 objetivos generales:

- Propuesta de un data mart que aloje los datos presentes en la base de datos de Conecta-TE referentes al uso de los ambientes de aprendizaje virtual particularmente interacciones de los estudiantes.
- Implementar un algoritmo de predicción que sea capaz de tomar(que aparte de las interacciones de los estudiantes) inputs de interacciones y retornar correctamente una predicción con el hecho si el estudiante aprueba o no el curso.

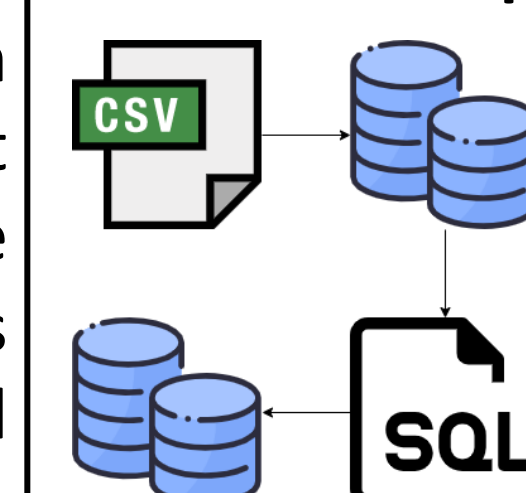
Para lograr el primer objetivo, se siguió la metodología de kimball basada en el ciclo de vida del negocio para la creación de una solución de inteligencia de negocio. Se enfatiza en el proceso central donde se inició por un proceso de entendimiento del negocio y modelado dimensional.



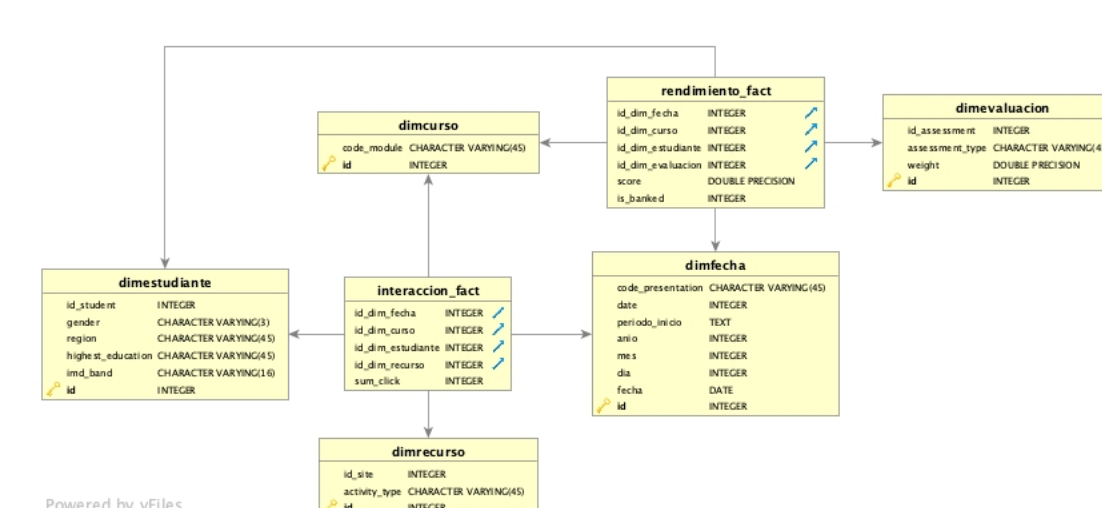
Fuente de datos: Los datos fueron obtenidos del proyecto Open University Learning Analytics Dataset (OULAD), contiene información de estudiantes, cursos y sus interacciones en ambientes de aprendizaje virtual (VLE).

Proceso de limpieza y cague de datos:

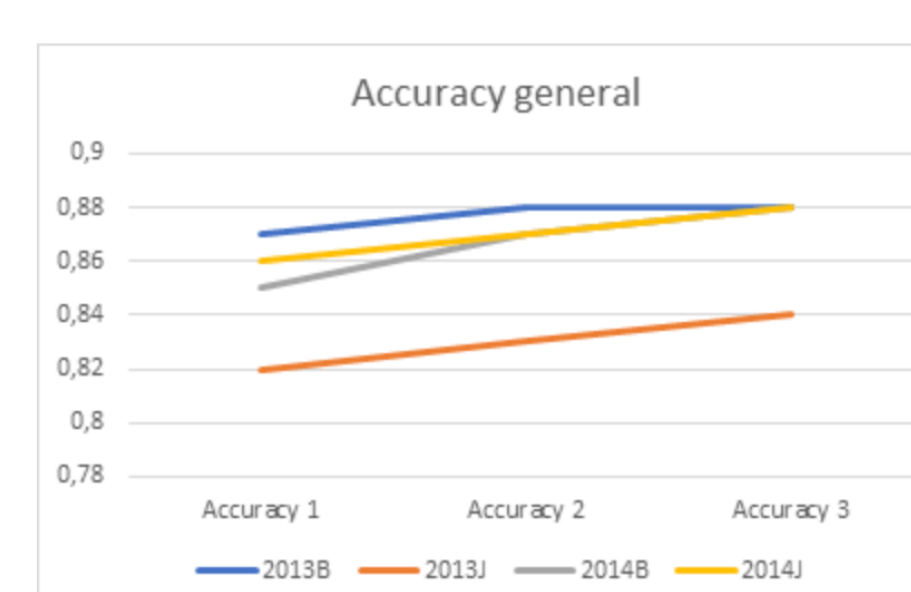
El data mart fue contruido utilizando un servicio de base de datos de postgres y alojada en AWS



Modelo dimensional:

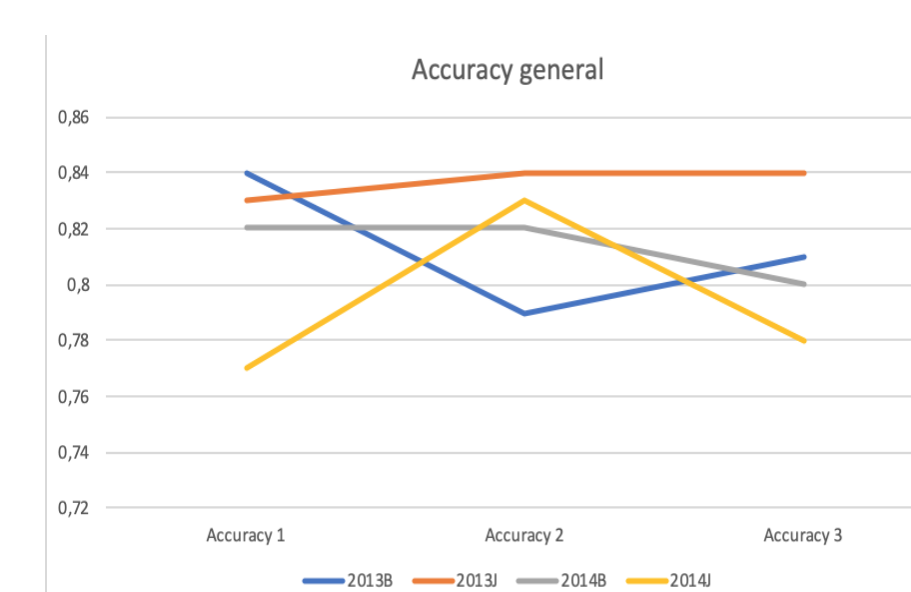


El modelo dimensional consta de 2 tablas de hechos: La primera con la que se puede analizar la interacción de los estudiantes con los distintos recursos y la segunda para analizar el rendimiento que han tenido los estudiantes a lo largo del curso



Con los datos ya preparados se empezaron a realizar las pruebas. Con un total de 8 pruebas y 32 árboles de decisión entrenados se obtienen los siguientes resultados:

- Con clases balanceadas se obtiene una precisión mayor al 82% llegando en ciertos modelos al 88%. Por otro lado, reduciendo el porcentaje sobre el cual una columna es binarizada se aumenta la precisión del modelo (fig.1).
- Con las clases no balanceadas se obtienen resultados variados donde no se encuentra una correlación entre la disminución del porcentaje y la precisión del modelo (fig. 2).



Con el objetivo de ampliar las pruebas de concepto con los datos de interacciones con los que cuenta Conecta-Se te obtienen resultados positivos. Si bien los datos con los que las pruebas se realizaron no son los mismo con los que cuenta Conecta-TE, los datos utilizados y el modelo dimensional diseñado fueron pensados de una manera general para cualquier ambiente virtual de aprendizaje. Con esta nueva aproximación Conecta-TE puede darle un valor agregado a los datos con los que cuenta. Una vez el implementado el modelo, le permitirá a Conecta-TE entender como los diferentes recursos disponibles para los estudiantes afectan su proceso educativo.

Referencias

- Kuzilek J., Hlostá M., Zdrahal Z. Open University Learning Analytics dataset Sci. Data 4:170171 doi: 10.1038/sdata.2017.171 (2017).