

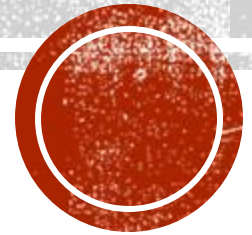
MINERÍA DE DATOS

KDD CUP 2015

Alfonso Barragán Carmona

Javier Monescillo Buitrón

Roberto Plaza Romero



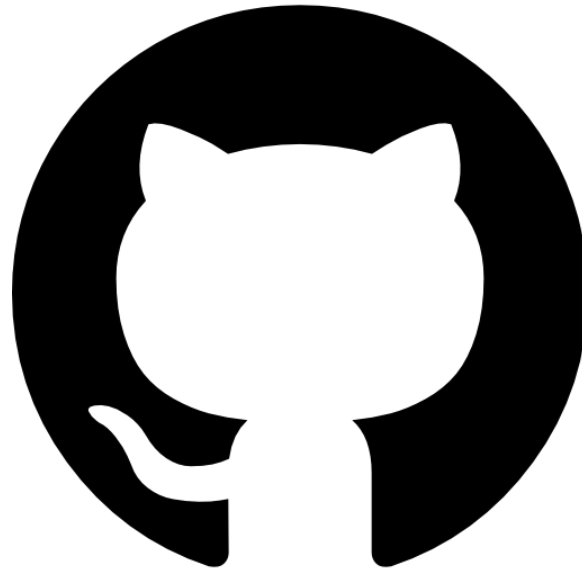
PROBLEMA: PREDICCIÓN DE ABANDONO





TECNOLOGÍAS USADAS:

- Oportunidad para hacer descubrimiento de conocimiento en bases de datos con una base de datos.





TARJETA DE DATOS

- Id curso
- número total de usuarios
- número de submódulos por tipo por curso
- número total de submódulos
- número de interacciones por tipo por curso
- número total de interacciones por curso
- número de abandonos
- número de no abandonados
- porcentaje de éxito

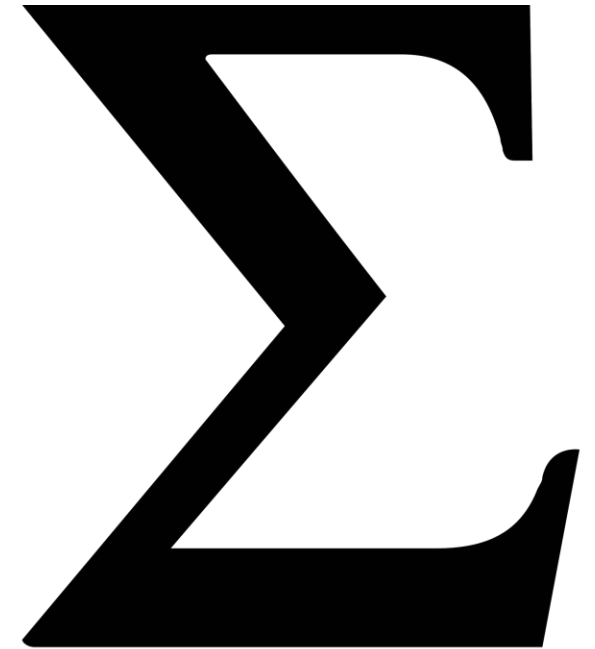
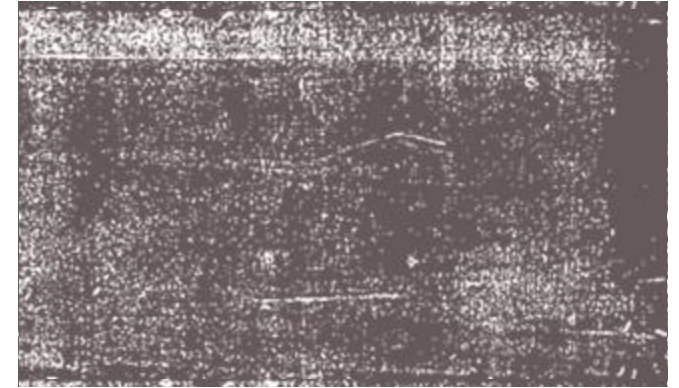
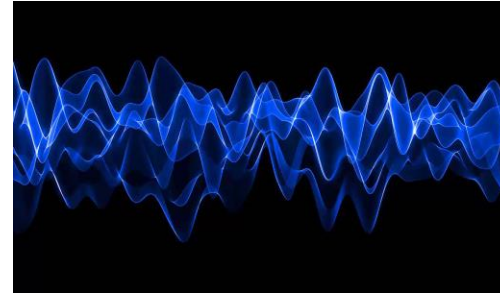


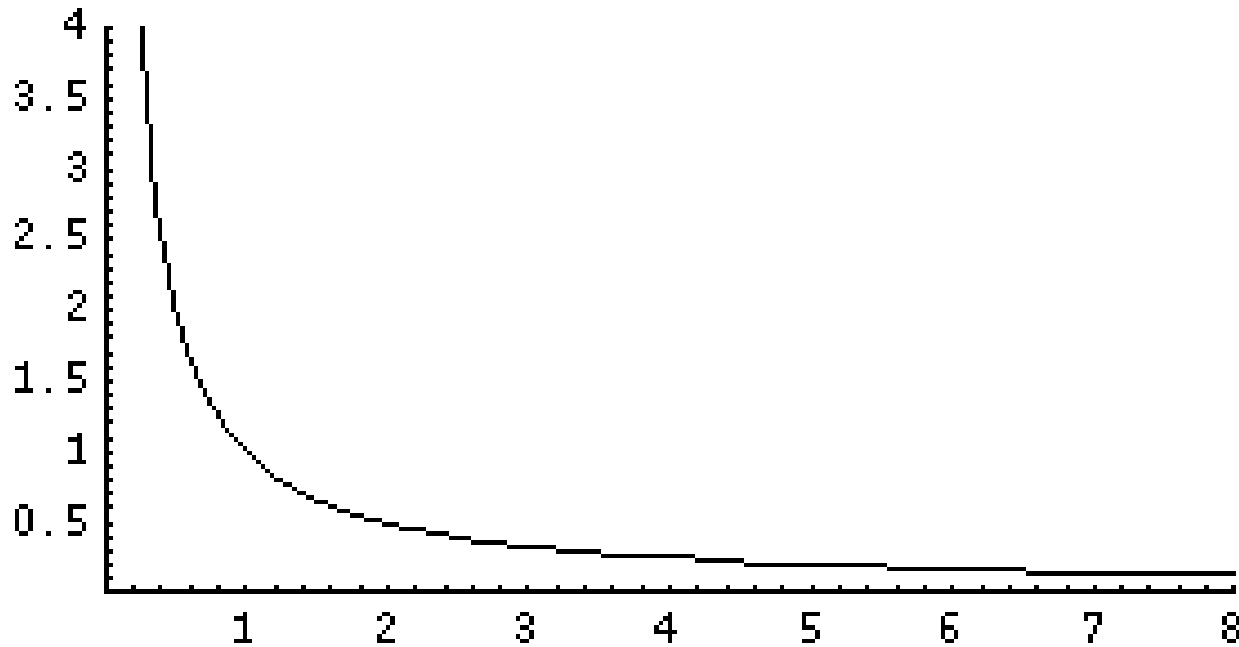
MUCHO SQL DESPUÉS



PREPROCESO ACABADO

- Hemos transformado una base de datos categórica en una tarjeta de frecuencias





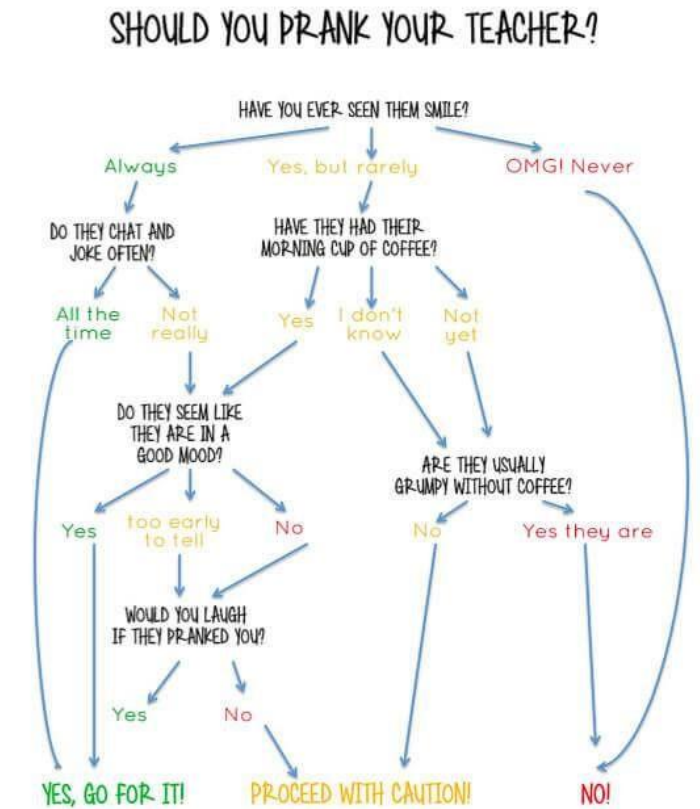
- Primero ordenar los cursos por probabilidad de abandono de mayor a menor.
- La forma que esperabamos para la distribución era una asíntota.
- También fué dividida en 3 grupos. En base a la probabilidad de abandono.

DIVISIÓN EN RANGOS



ARBOL DE DECISION

- Para este ejemplo consideramos como mejor opción un algoritmo Random Forest.
- Estamos buscando características dentro de la tarjeta de datos.
- No solo es más útil, sino que también es paralelizable y solventa el problema de representar la información.





TRANSFORMACIÓN

- Como meta final, nos gustaría transformar aquellos cursos cuya "calidad" en cuanto a prevención de abandono no queda clara, los fronterizos, en cursos menos dados al abandono.
- Por ende, el sistema ofrece sugerencias para cambiar el curso y hacer más accesible.



TRABAJO FUTURO

