**2020/03/06**

**SISTEMAS DE RECOMENDACIÓN**

**Juan Arévalo**

Esta clase va a ser una especie de practica enfocada al TFM

fuentes de datos:

PUBLIC DATASETS

Statistical organizations

Banco Mundial, ONS

APIs

Scraping: recurso cuando no hay una API

Kaggle: sólo como última opción.

PRIVATE DATASETS

Contacting a company. Cuanto más pequeña la compañía más fácil es.

Generar datos: pedirlos a gente con encuestas, hacer fotos, …

Handling private data: cuiado si se maneja información identificadora, nunca publicar esos datos. Sí puedes publicar el proyecto y sus conclusiones pero no los datos.

Diapo 7:

Minimum Viable Prototype (MVP): hacer uno en 2 semanas para ver si funciona

Debemos hacer un proyecto rapidito en 2 semanas para ver si funciona pero simplicando al máximo todas las fases.

Y luego ir mejorarndo cada una de las partes del proyecto. Esto ayuda a darte cuenta de cómo afectan al producto final la mejora de cada una de las partes del proyecto.

Esta estrategia implica que en todo momento es entregable, en caso de que haya retrasos o imprevistos.

Tener un bucle de iteración ligero. Que no tenga que hacer / ejecutar mil cosas cada vez que haga cambios. Lo ideal es que el proyecto se pueda correr de principio a fin con un clic

Ponerte parcial deadlines. Fail fast = identificar las cosas más inciertas o partes más débiles y tirarse a solucionarlas lo primero.

Pensar un par de alternativas de proyecto por si el modelo principal resulta no ser viable.

Evaluating Models

Generar una baseline, un punto de compación de sentido comun. Algo muy simple, la media, o el valor del mes anterior, etc. Hay que definir qué es bueno y qué es malo, primero comparando con un baseline y en base a los requerimientos propios del proyecto.

Vigilar de manera continua su rendimiento a lo largo del tiempo.

Tipos de proyectos:

Exploratorios: como investigación científica. Tengo un fenomeno o hipótesis. Contrastarla o confirmar en base a unos datos. No tiene mucho riesgo, pero no suelen salir cosas interesantes. El resultado sería como un paper científicico, una publicación

Modelling/Predictive: el riesgo es que los datos no sean suficiente para determinar la variable de salida. Testear rápido para ver si son viables.

Clustering/market segmentation: al profe le dejan fríos. No tienen muchas profundidad porque no hay gran cosa que puedas hacer con eso.

All or nothing. Los proyectos de Deep learning suelen caer en esto. El clasificador funcionan o no, si no funciona estás jodido.

Recomendadores: el problema es que es muy difícil evaluarlos. Es obligado montar un front end, una web básica donde metas un dato y te devuelva in item.

ES OBLIGATORIO HACER UN FRONT END

Permite enfatizar los puntos buenos del proyecto, vendiendo lo que mola. No dando simplemente un tocho de código. Con flask se puede hacer una web fácil. Tableau es otro front end.

En el caso de un white paper style , el front end es el propio end.

QLIKVIEW es otra herramienta para hacer un front end.

**herokuapp**

**bookdown**: manera fácil de escribir un lbro con R a partir de un .Rmd. Puedes meter código y todo

Memoria vs Front End

La diferencia en el publico objetivo. La memoria es para otro data scientist. El front end es para el público final.