XXV aniversario de la Estadística Oficial en Andalucía

Jornada técnica
El valor de la información: el reto del "Big Data"

El uso del Big data en las ciencias de la salud
José Antonio Guerrero Durán

José A. Guerrero

Data Scientist

Verified account

jaguerrerod@ono.com

Profile | Results | Scripts | Forum

Licenciado en Matemáticas (Estadística e Investigación Operativa)

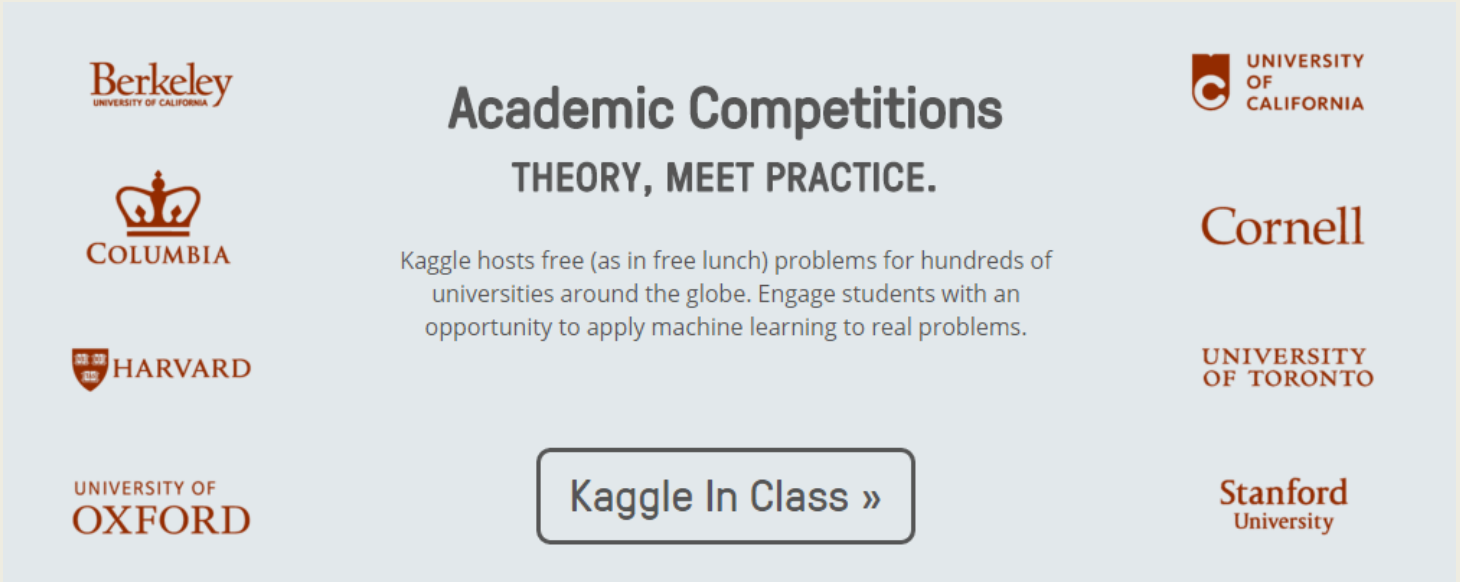
Experiencia en sector sanitario durante más de 25 años (epidemiología, investigación, tecnologías de la información y gestión).

Proyectos internacionales de análisis de datos (Plataforma Kaggle)

CEO de Datrik Intelligence

kaggle

The Home of Data Science



Singularidades del Entorno Sanitario

Complejidad productiva. Malla de procesos vs cadena lineal.

Variabilidad natural de resultados, distribuciones, rangos normales…

Alta tecnificación, tanto de medios diagnósticos como de recursos humanos

Complejidad para definir el propio producto y su medición

Datos en el Entorno Sanitario

Datos más heterogéneos

Mayor volumen de datos

Necesidad de análisis en tiempo real

Datos en el Entorno Sanitario

Datos más heterogéneos

Mayor volumen de datos

Necesidad de análisis en tiempo real

BIG DATA

Situación actual

Estudios epidemiológicos y ensayos clínicos: uso mayoritario de métodos multivariables de Estadística Clásica.

-ANOVA
-Regresión multivariable / logística
-Análisis Clúster

Problema: Los datos sanitarios suelen tener mucho ‘ruido’, outliers, observaciones perdidas, distribuciones no normales,…

Situación actual

Datos para la gestión clínica / contabilidad analítica:
Uso de herramientas de Business Intelligence (Data Warehousing y Data Marts).

Sistemas de información desagregados tipo CMBD (Conjuntos mínimos básicos de datos)

Cubos OLAP con navegación jerárquica por dimensiones: geográfica, organizativas, nivel de producto asistencial, …

Cuadros de Mandos estandarizados para el seguimiento de objetivos.

Situación actual

POCO USO DE HERRAMIENTAS BIG DATA

“Business Intelligence ayuda a encontrar respuestas a preguntas conocidas.

Big Data ayuda a encontrar las preguntas que no sabes que quieres preguntar.”

Eric D. Brown
Consultor en Tecnologías

Big Data en Entornos Sanitarios. Dificultades

Trazabilidad de individuos.

La confidencialidad obliga a anonimizar los datos.

¿Cómo trabajar con bases de datos multicéntricas si un paciente puede ser atendido en múltiples centros?

¿Sería posible una encriptación única de los datos?

¿Macro repositorios de datos a nivel nacional?
CMBD de episodios de hospitalización

Big Data en Entornos Sanitarios. Dificultades

Accesibilidad a la información.
Es una barrera de entrada importante ya que la información es parcial y se facilita a demanda concreta y justificada.

¿Es una utopía un conjunto de datos anonimizados y público para investigadores y empresas?

Filosofía ‘Not Sampling’
Big Data se basa en acceso a ‘todos’ los datos disponibles
Todos los registros y todas las variables.

Big Data en Entornos Sanitarios.
Análisis de la mortalidad estándar.

Clásico :Regresión logística con ajuste por edad, sexo, nivel de mortalidad y nivel de severidad medios (APR).

Análisis de los residuos a nivel de centros, unidades,…

Big Data: Uso de árboles de regresión aleatorios (RandomForest).

Análisis de desviaciones OOB (Out Of Bag) a nivel de centros, unidades, …

Actualmente en fase de extender el estudio a todo el SSPA.

Big Data en Entornos Sanitarios.

El promotor ofrece historia clínica electrónica gratuita a cambio de poder explotar **comercialmente** los datos.

¿Qué papel van a jugar las empresas generadoras de información?

Y los usuarios, ¿Obtendrán alguna compensación por ceder sus datos para uso comercial?

Big Data en Entornos Sanitarios.

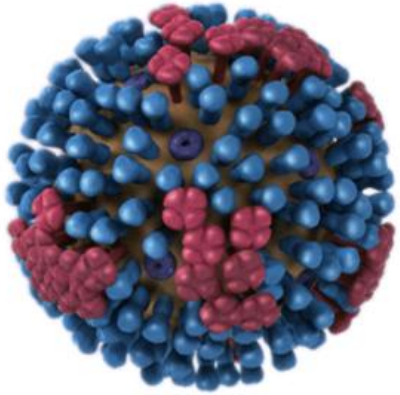
Genentech
A Member of the Roche Group

Flu Forecasting

Thu 19 Dec 2013 – Mon 3 Mar 2014 (23 months ago)

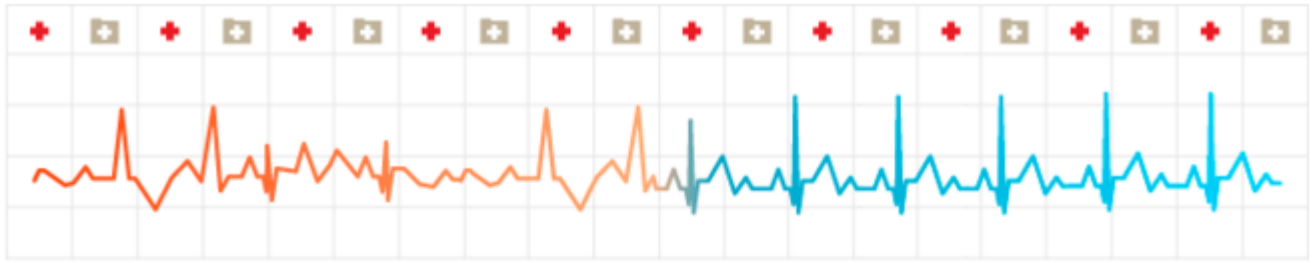
This competition is private-entry. You've been invited to participate.

Predict when, where and how strong the flu will be



Big Data en Entornos Sanitarios.

HERITAGE PROVIDER NETWORK
HEALTH PRIZE



Improve Healthcare, Win $3,000,000.

Identify patients who will be admitted to a hospital within the next year using historical claims data. (Enter by 06:59:59 UTC Oct 4 2012)

Big Data en Entornos Sanitarios.

Genentech
A Member of the Roche Group

Cervical Cancer Screening

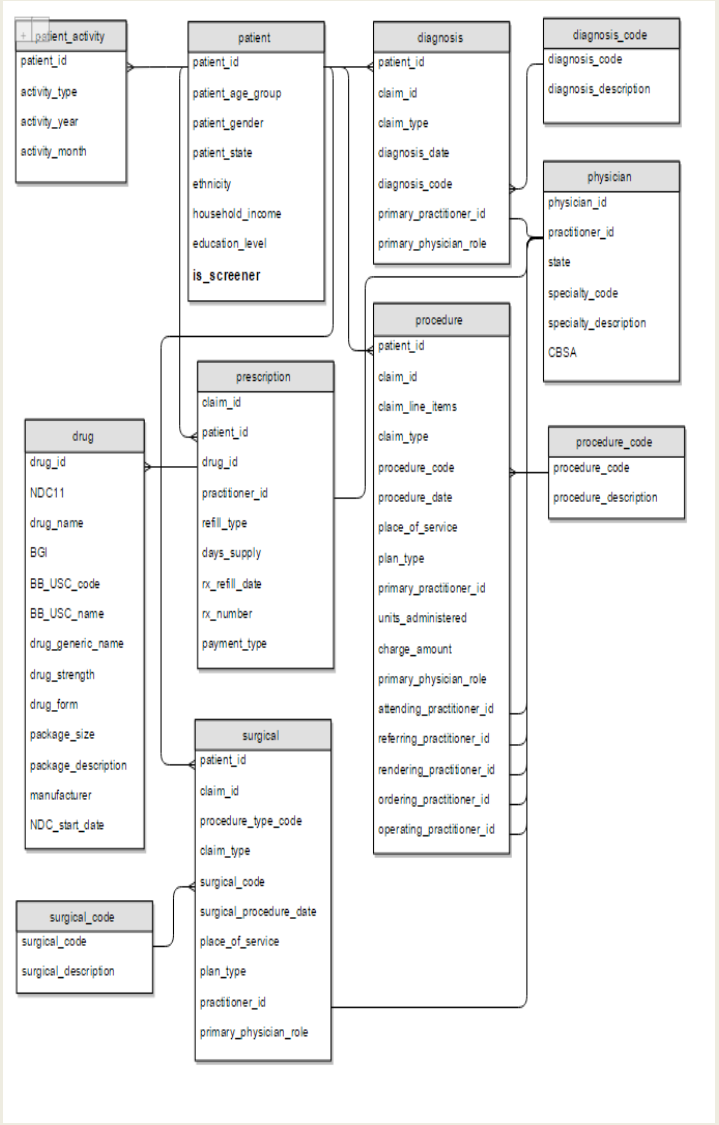
Fri 11 Dec 2015 – Mon 1 Feb 2016 (2 days ago)

Symphony™
HEALTH SOLUTIONS

This competition is private-entry. You've been invited to participate.

Help prevent cervical cancer by identifying at-risk populations

Big Data ayuda a encontrar las preguntas que no sabes que quieres preguntar



El nuevo ecosistema del Dato y la Información

Usuarios

Institutos Oficiales

Empresas generadoras de información

Empresas privadas de análisis de datos.

Fundaciones. Organizaciones privadas sin ánimo de lucro

Promotores privados con fines comerciales.

Usuarios ‘D’