

Alineando Data Science con objetivos de negocios

A machine learning process from the Watson project

Ayudando a su equipo de Data Science

- Data Science puede intimidar
 - Terminología especializada
 - Complejidad técnica
- Sin embargo, en Data Science se aprende
 - Sobre machine learning y estadística
 - No sobre su negocio
- Conocimiento específico sobre su negocio es clave para recuperar la inversión en Data Science

Sacandole el jugo a Data Science



No haciéndolo bien

- Tiramos la fruta dentro
- Analizamos la composición química del jugo
- Probamos cambios en la juguera

Donde hacer cambios

- Elegir mejor la fruta: improved data acquisition
- Elegir las mejores partes de la fruta: improved “features”
- Probar el jugo: custom error metrics
- Preguntar a expertos en jugos (no en ML):
 - Qué frutas son buenas.
 - Cuáles son las mejores partes de las frutas.
 - Qué propiedades del jugo son importantes.

Haga preguntas

- ¿Qué datos se están usando?
 - ¿Son válidos? ¿Cómo se obtuvieron?
 - ¿Son representativos de los datos actuales?
- ¿Qué se asume en los resultados actuales?
- ¿Cómo se midieron los resultados?
 - ¿Considera que todos los errores son igualmente problemáticos?
 - Compare: color vs. sabor

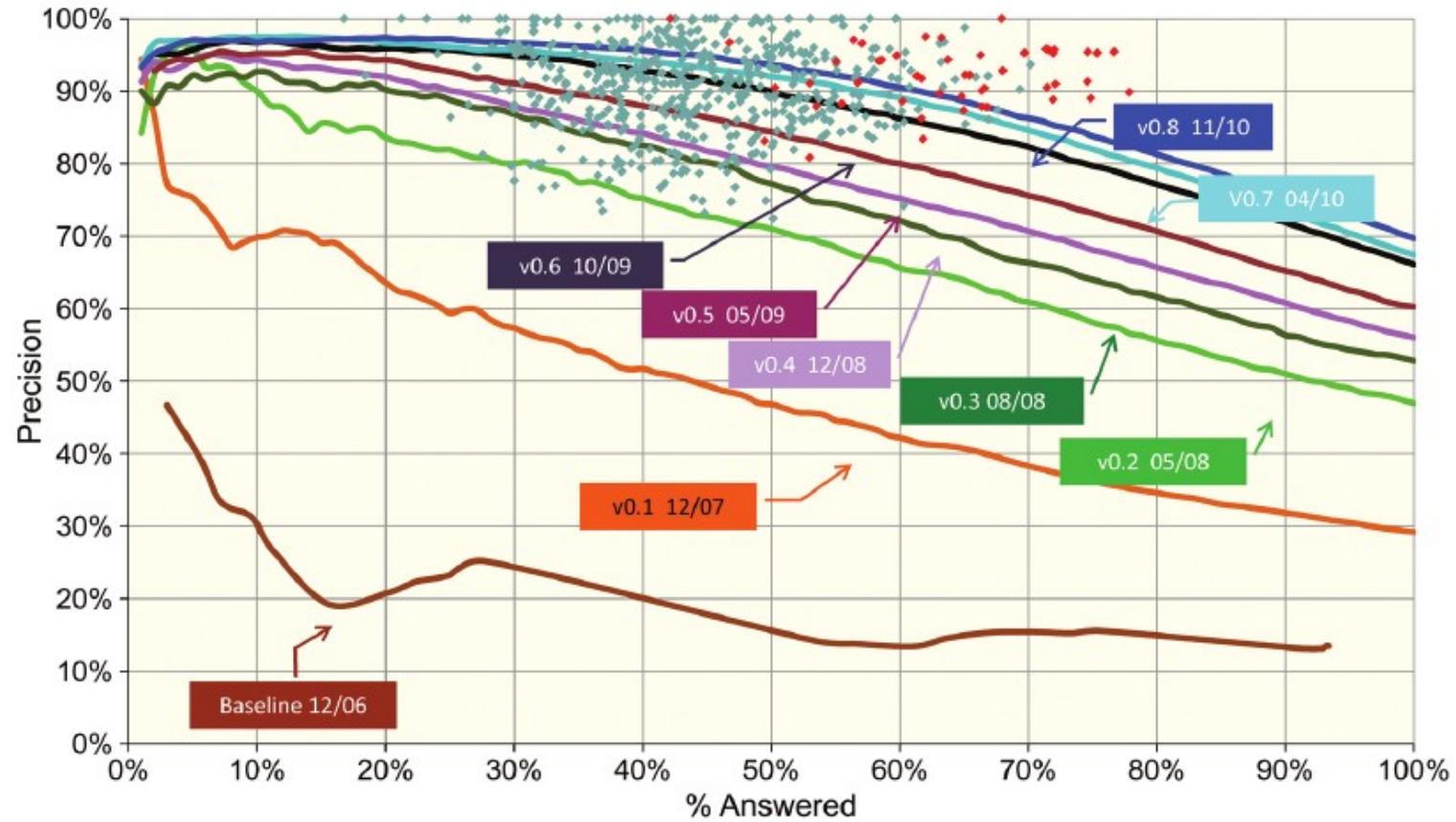
El porqué de esta presentación

- IBM Jeopardy! Show
- Consultoría
- FaMAF 2014 course (<http://aprendizajengrande.net>)
- Libro de texto de ML en la Cambridge University Press

Riesgos de invertir en IA

- Invertir en IA es como cualquier inversión
 - Tiene que ser evaluada en términos de ROI **en conjunto** con los riesgos
 - IA puede generar un ROI increíble
 - Fácil olvidarse de los riesgos
- Los riesgos están ligados con el nivel de inversión
 - Pequeña inversión es más riesgosa
 - Las compañías grandes pueden seguir invirtiendo hasta que funcione

Inversión en el proyecto Jeopardy!



The Art of Feature Engineering

- Técnicas para mejorar los resultados de los modelos de aprendizaje
- Disponible en pre-orden en Amazon (ISBN: **1108709389**)
 - En edición, saldrá Feb/Abr 2020
- 7 casos de estudio en feature engineering
 - 10,000 líneas de Python notebooks
- Un proceso inspirado en el Jeopardy! Show
 - Exploratory data analysis
 - Error analysis

CLEI 2019

- Impacto del dialecto del español
 - Spanish NLP using Deep Learning
- 10 años de debates del congreso
 - Argentina
 - Costa Rica
- Datos y código disponible en <https://github.com/IE4OpenData/clei2019dialect>

Textualization Software Ltd.

- Consultoría en Inteligencia Artificial
 - Machine Learning
 - Deep Learning
 - Natural Language Processing
- Clientes internacionales
 - Seed / Series A startups
- Bootstrapping productos propios: Textualization Highlighter

Conclusiones

- RECAP
 - El éxito del proyecto en Data Science depende también de los expertos dentro de la empresa que conocen el negocio de la misma
- FaMAF Network group
 - <https://www.famaf.unc.edu.ar/investigacion/areas-de-investigacion/computacion-ofi/analisis-y-procesamiento-de-grandes-redes-sociales-y-semanticas/>