teradata.

AutoML y el Futuro de la Automatización de los Proyectos Analíticos

Luis Cajachahua

Principal Data Scientist, Teradata Americas

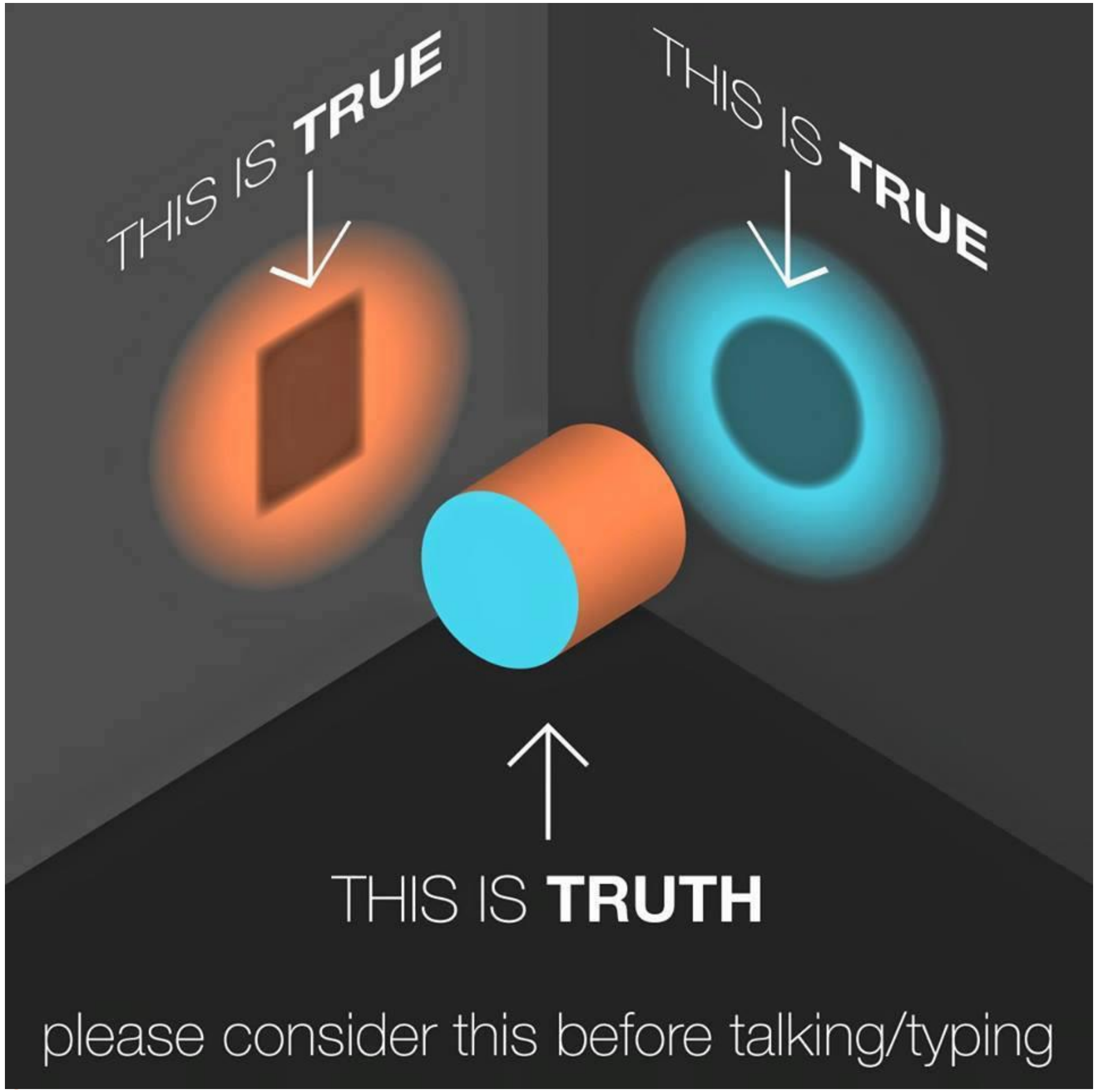
Agosto 2021

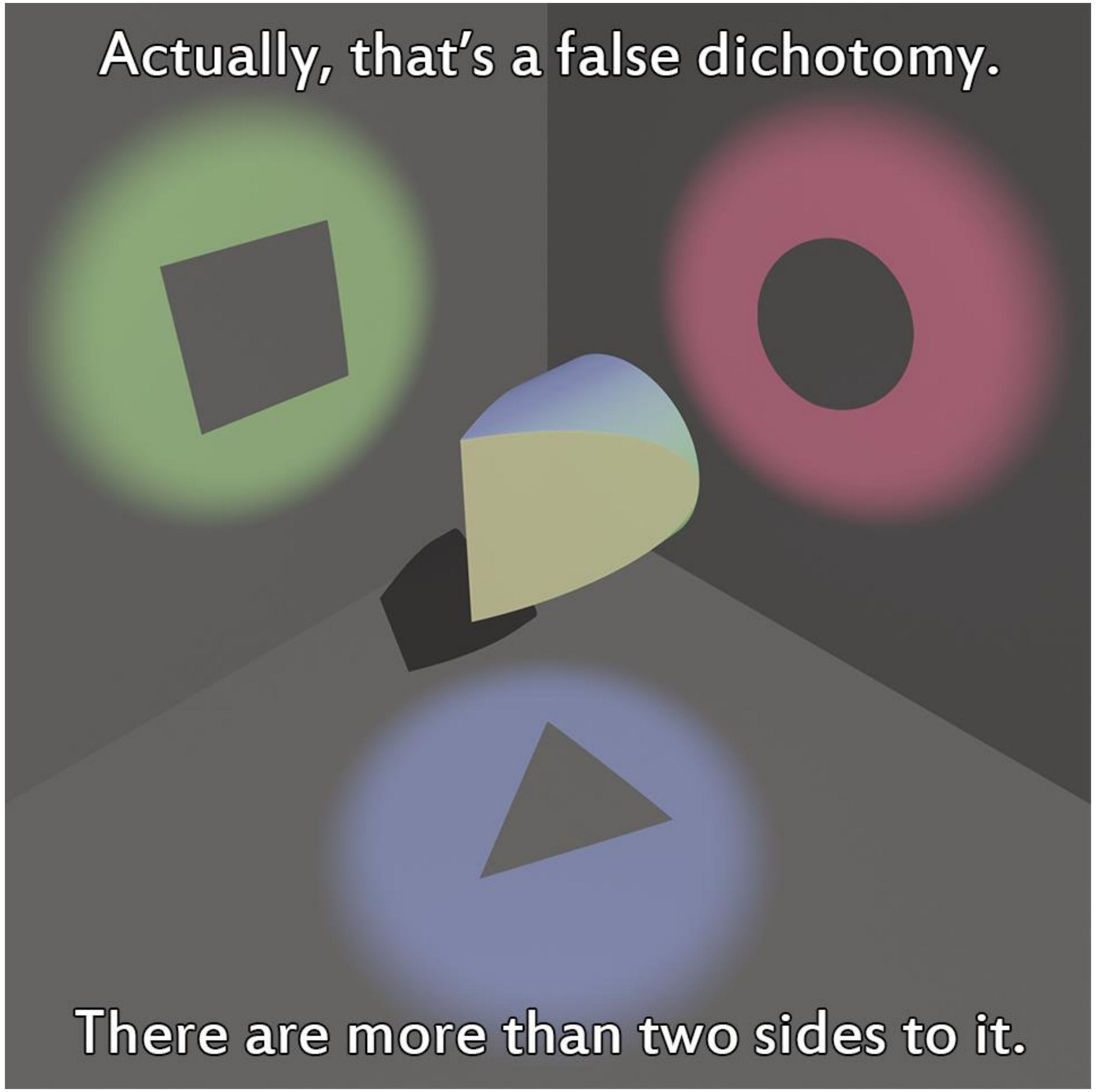
https://www.linkedin.com/in/lcajachahua/

Agenda

- ¿Qué es AutoML?
- Antecedentes
- Soluciones AutoML
- ¿Hacia dónde va AutoML?
- Recomendaciones

Antes de empezar…





Pintemos la cancha…

En esta charla trataremos de aportar algo al extenso material que ya está disponible sobre AutoML en Internet. Trataremos de cubrir puntos de los que no se ha hablado mucho.

Para saber más sobre la historia de AutoML y ver demos de las principales herramientas disponibles, revisar la sección de referencias y el repositorio de Github.

teradata.

¿Qué es AutoML?

"Automated machine learning (AutoML) is the process of automating the process of applying machine learning to real-world problems. AutoML covers the complete pipeline from the raw dataset to the deployable machine learning model. AutoML was proposed as an artificial intelligence-based solution to the ever-growing challenge of applying machine learning".
Wikipedia

El proceso de automatizar la aplicación de modelos de Machine Learning para resolver problemas del mundo real.

¿Qué es AutoML?

¿Y por qué tendría que interesarme?

Google is funding "an artificial intelligence for data science"

Derrick Harris
Dec 2, 2014 – It won't steal your jobs, yet

Google is funding a project called Automatic Statistician that bills itself as "an artificial intelligence for data science," it announced Tuesday. The project, which comes out of the University of Cambridge and is still in its early stages, aims to automate the selection, building and explanation of machine learning models.

In a nutshell, Automatic Statistician works by looking at a dataset and then determining which type of model would be best for analyzing it as well as which features, or variables, are the strongest. After the model runs, Automatic Statistician will return a text report explaining its findings in plain English — or as close as you can get when dealing with statistics.

A snippet of an Automatic Statistician report on unemployment data.

The project's homepage quotes Google research scientist Kevin Murphy, who also wrote the blog post announcing Google's funding for it, explaining the promise of Automatic Statistician like this:

[blockquote person="" attribution=""]The first problem is that current Machine Learning (ML) methods still require considerable human expertise in devising appropriate features and models. The second problem is that the output of current methods, while accurate, is often hard to understand, which makes it hard to trust. The "automatic statistician" project from Cambridge aims to address both problems, by using Bayesian model selection strategies to automatically choose good models / features, and to interpret the resulting fit in easy-to-understand ways, in terms of human readable, automatically generated reports.[/blockquote]

However, Automatic Statistician isn't the first attempt to deliver this type of service; there have, in fact, been multiple commercial attempts at doing similar things. The most accurate comparison might be to a now-defunct tool by machine learning startup Skytree called Skytree Adviser, which also automatically selected models and generated text reports of its findings. Startups including BeyondCore, Nutonian and even Ayasdi are all promising varying degrees of this functionality, as well.

As sexy as it is to talk about automating the data scientist job, though, it's a bit early to suggest any software will eliminate the need for such employees any time soon. Even if projects like Automatic Statistician or commercial tools can make it possible for relative laypersons to run machine learning models and uncover patterns, that's just a step or two down what's often a much-longer path of turning insights into real value or, possibly, products.

Data Scientists Automated and Unemployed by 2025?

Will Data Scientists be unemployed by 2025? Majority of voters in latest KDnuggets Poll expect expert-level Data Science to be automated in 10 years or less.

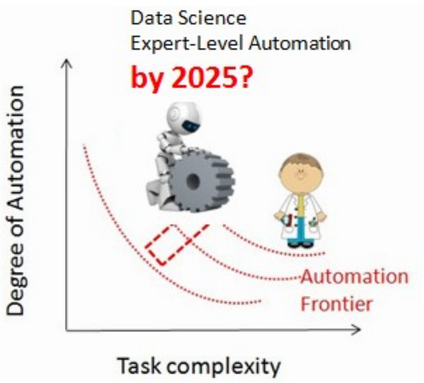
By Gregory Piatetsky, KDnuggets.

Data Scientist has been called the sexiest job of the 21st century. But perhaps the century will last only 25 years.

With even knowledge-based jobs like lawyers and accountants being automated, will Data Scientists prove to be an exception?

What predictive analytics professionals predict about the future of their profession?

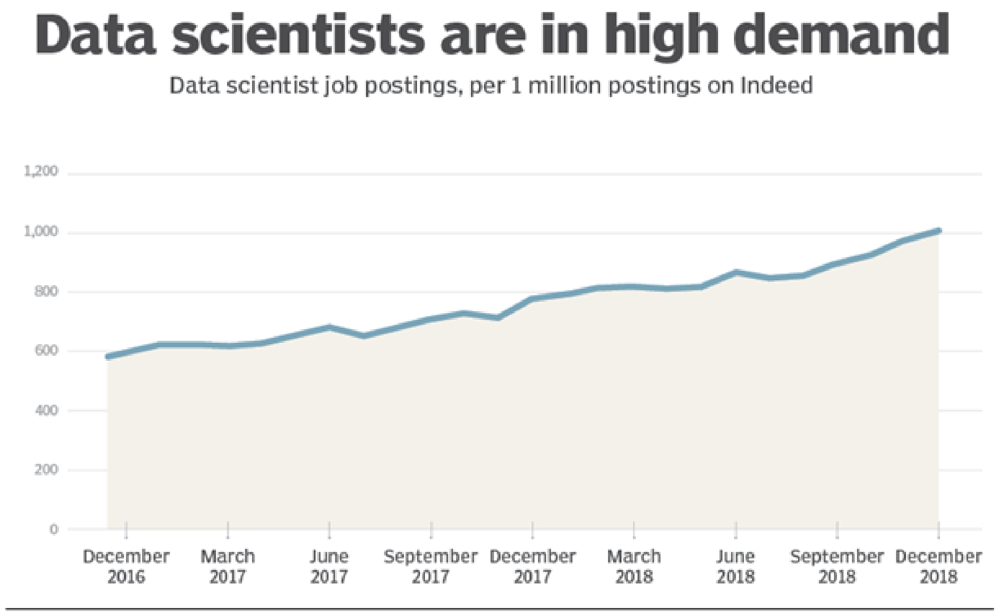
Latest KDnuggets Poll asked:

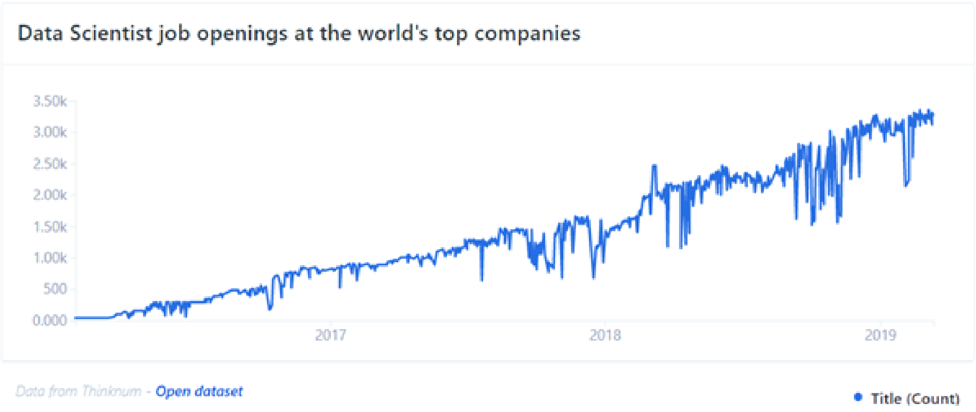


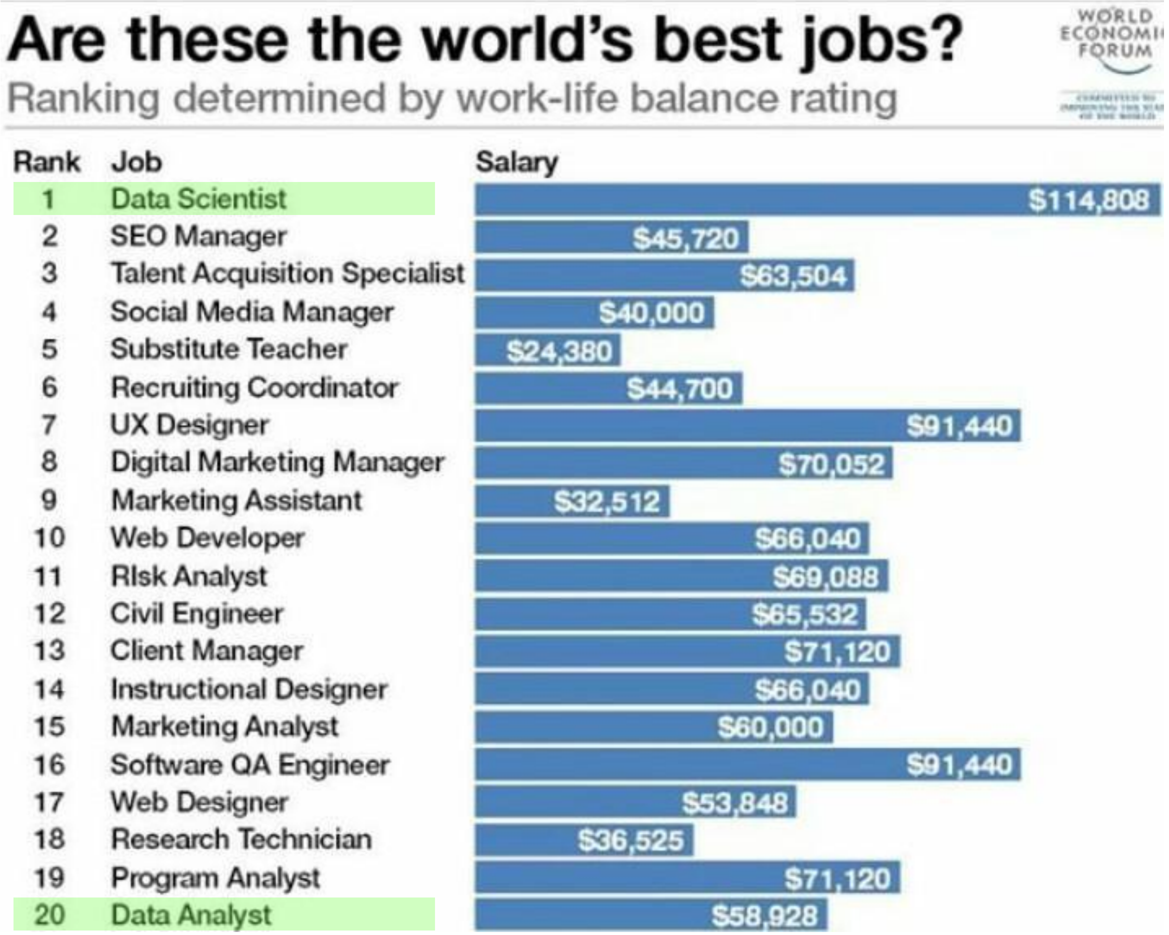
When will most expert-level Predictive Analytics/Data Science tasks - currently done by human Data Scientists - be automated?

teradata.

En realidad, todo iba muy bien…







…Hasta que cierta Pandemia empezó…

Why the year 2020 will prove to be a headache for Data Scientists

The effects of coronavirus will ripple through data science projects

Usman Gohar Mar 13, 2020 · 4 min read

Photo by Aaron: Unsplash

"Your model is as good as your data" is the most basic postulation in data science. Good data equals a good model! The coronavirus has impacted millions of lives around the globe, wreaked havoc on the airline industry and shattered equity markets globally.

The Recession's Impact on Analytics and Data Science

There has been a huge demand for data scientists in the past decade. Is that about to change?

Jeffrey D. Camm, Melissa R. Bowers, and Thomas H. Davenport • June 16, 2020

READING TIME: 7 MIN

SUBSCRIBE SHARE

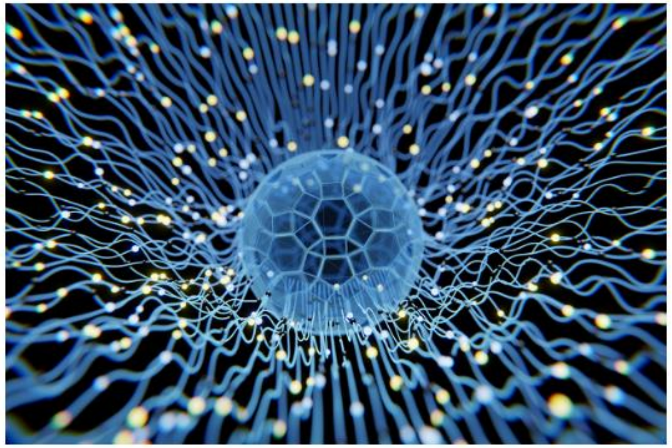
The outbreak of the COVID-19 pandemic is having a dramatic negative impact on economies in the U.S. and worldwide, and unemployment rates are soaring. Given the economic disruptions, it seems likely that many countries in the global economy will experience a recession.

Data Science, Quarantined

Companies are beginning to reboot their machine learning and analytics, which have been disrupted by the global pandemic.

Jeffrey D. Camm and Thomas H. Davenport • July 15, 2020 READING TIME: 7 MIN

SUBSCRIBE SHARE



The economic impact of COVID-19 is unprecedented, dramatically changing markets and prospects for economic growth. Supply chains, transportation, food processing, retail, e-commerce, and many other industries have transformed overnight. Unemployment in the U.S. has reached levels unknown in recent memory, and GDP is expected to fall around the world. As one economic journalist summed up the situation: "Nearly everything in the world is super-weird and disrupted right now."

Pero ya veníamos de una realidad complicada…

80%

Del tiempo empleado en las fases de recolección y preparación de los datos

15%

Sólo el 15% de los proyectos Big Data llegan a producción (Gartner).

+6 Meses

Para implementar un modelo o data product, desde la idea hasta el accionamiento.

teradata.

blogs.gartner.com/andrew_white/2019/01/03/our-top-data-and-analytics-predicts-for-...

Our Top Data and Analytics Predicts for 2019

by **Andrew White** | January 3, 2019 | Comments Off on Our Top Data and Analytics Predicts for 2019

Predicts 2019: Data and Analytics Strategy

- By 2022, 90% of corporate strategies will explicitly mention information as a critical enterprise asset and analytics as an essential competency.
- By 2023, data literacy will become an explicit and necessary driver of business value, demonstrated by its formal inclusion in over 80% of data and analytics strategies and change management programs.
- By 2022, 30% of CDOs will partner with their CFO to formally value the organization's information assets for improved information management and benefits.
- By 2023, 60% of organizations with more than 20 data scientists will require a professional code of conduct incorporating ethical use of data and AI.
- By 2022, more than half of major new business systems will incorporate continuous intelligence that uses real-time context data to improve decisions.

Predicts 2019: Analytics and BI Solutions

- Through 2020, 80% of AI projects will remain alchemy, run by wizards whose talents will not scale in the organization.
- Through 2022, only 20% of analytic insights will deliver business outcomes.
- By 2021, proof-of-concept analytic projects using quantum computing infrastructure will have outperformed traditional analytic approaches in multiple domains by at least a factor of 10

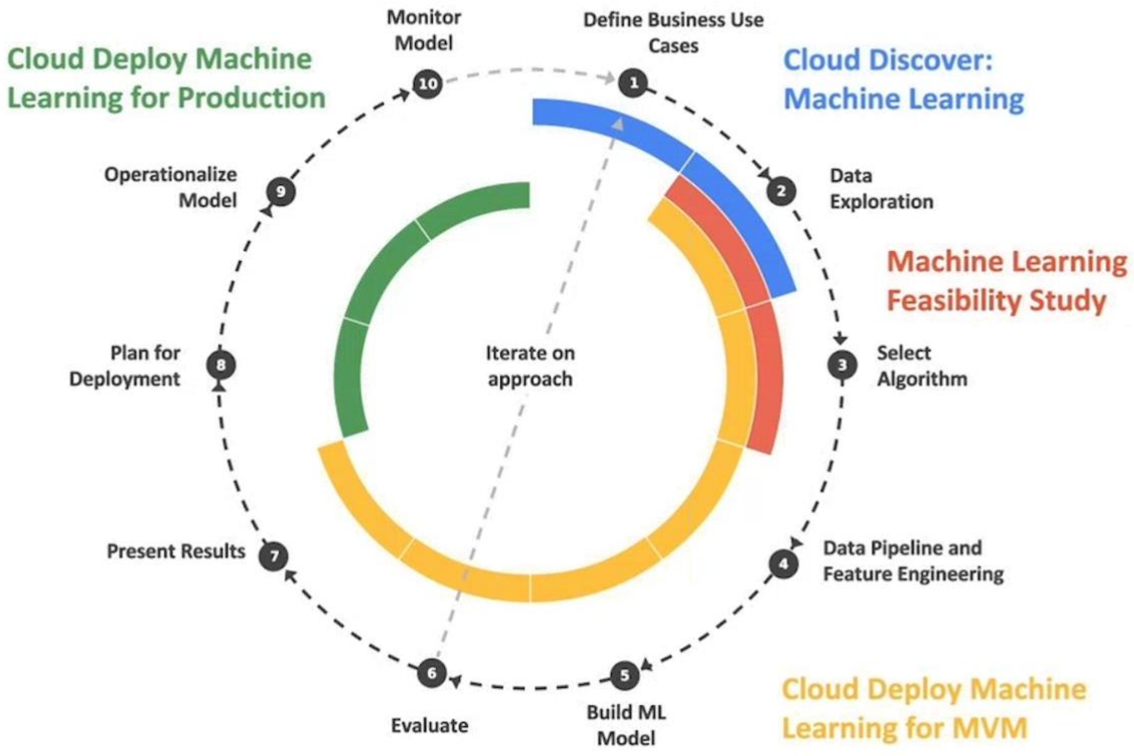
Hasta 2020, el 80% de los proyectos de IA seguirán siendo alquimia, a cargo de magos, cuyos talentos no escalarán en la organización.

Hasta 2022, solo el 20% de los insights analíticos generarán resultados comerciales.

https://blogs.gartner.com/andrew_white/2019/01/03/our-top-data-and-analytics-predicts-for-2019/

Soluciones AutoML

¿Qué cubren la mayoría de soluciones AutoML?



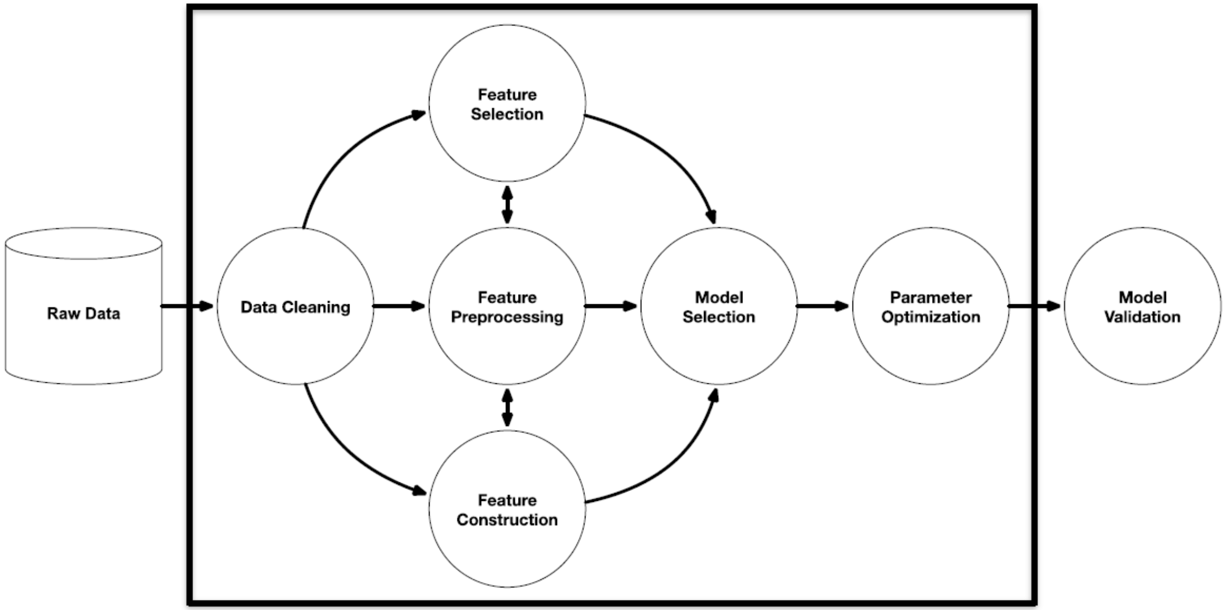


Image source: R. Olson *et. al.* (2016) “Evaluation of a Tree-based Pipeline Optimization Tool for Automating Data Science.”

Soluciones AutoML

Open Source – Predictive Modeling

- AutoSklearn: https://automl.github.io/auto-sklearn/master/
- AutoWeka: https://github.com/automl/autoweka
- H2O AutoML: https://github.com/h2oai/h2o-3
- Ludwig: https://github.com/ludwig-ai/ludwig
- MLBox (WIP): https://mlbox.readthedocs.io
- PyCaret: https://github.com/pycaret/pycaret
- Tpot: https://http://epistasislab.github.io/tpot
- AutoGluon: https://auto.gluon.ai/stable/index.html

teradata.

Soluciones AutoML

Open Source – Otras categorías

Deep Learning:

- AutoKeras: https://autokeras.com
- AutoPytorch: https://github.com/automl/Auto-PyTorch

Reinforcement Learning:

- LEARNA: https://github.com/automl/learna
- SEARL: https://github.com/automl/SEARL

Forecasting:

- Facebook Prophet: https://facebook.github.io/prophet/

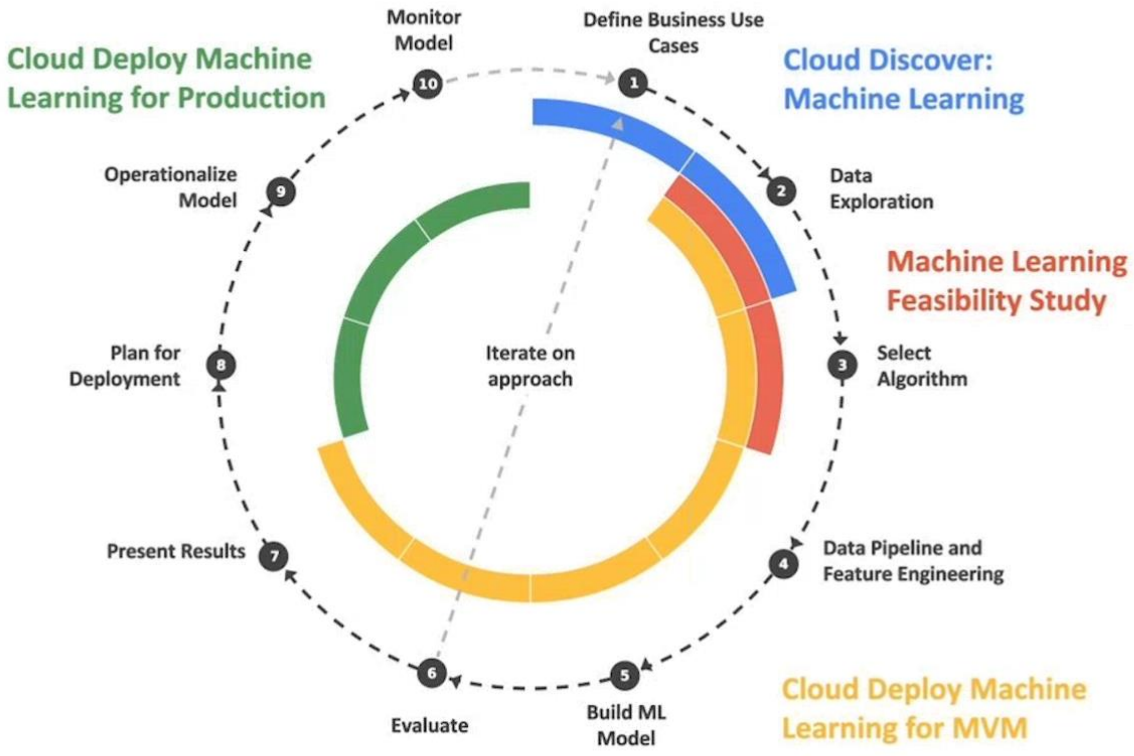
Soluciones AutoML

Soluciones Comerciales As a Service

- SAS Model Studio: https://support.sas.com/en/software/model-studio-support.html
- DataRobot: https://www.datarobot.com/platform/automated-machine-learning/
- H2O Driverless AI: https://www.h2o.ai/products/h2o-driverless-ai/
- Google AutoML: https://cloud.google.com/automl/
- Azure AutoML: https://ml.azure.com/
- Amazon Sagemaker Autopilot: https://aws.amazon.com/sagemaker/
- Dataiku: https://doc.dataiku.com/dss/latest/machine-learning/auto-ml.html
- BigML OptiML: https://bigml.com/releases/winter-2018
- Databricks AutoML: https://databricks.com/product/automl
- IBM Watson AutoAI: https://developer.ibm.com/learningpaths/explore-autoai/nextgen-automl-watson-autoai/
- Salesforce Einstein: https://www.salesforce.com/mx/products/einstein/overview/
- TiMi Modeler: https://timi.eu/timi/timi-modeler/

¿Hacia dónde va AutoML?

¿Qué cubren la mayoría de soluciones AutoML?



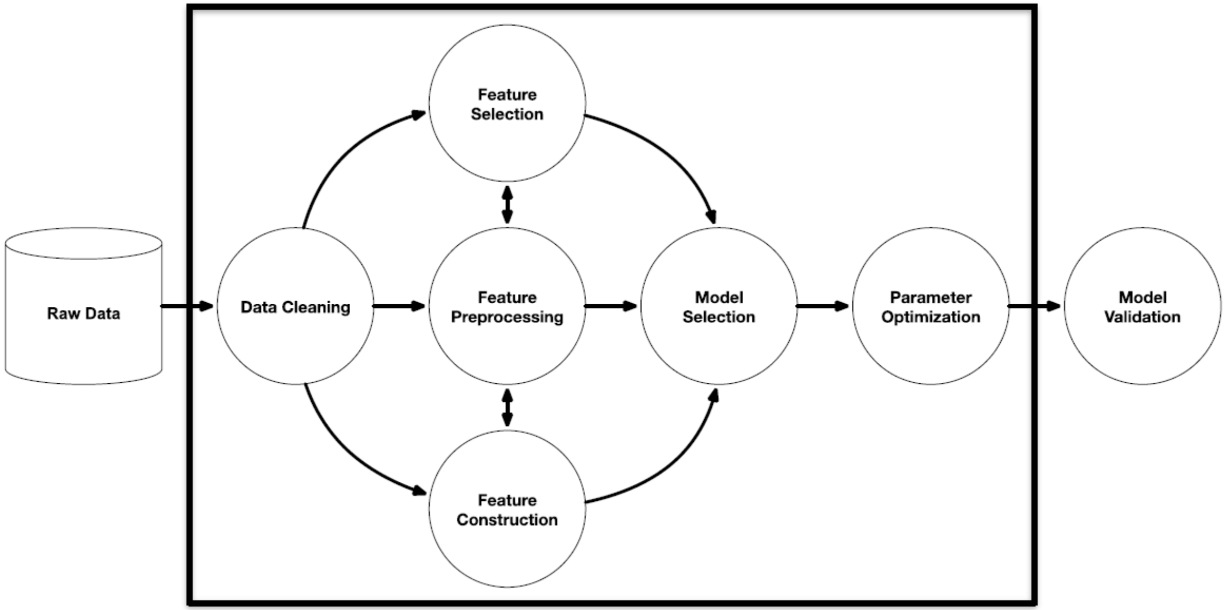


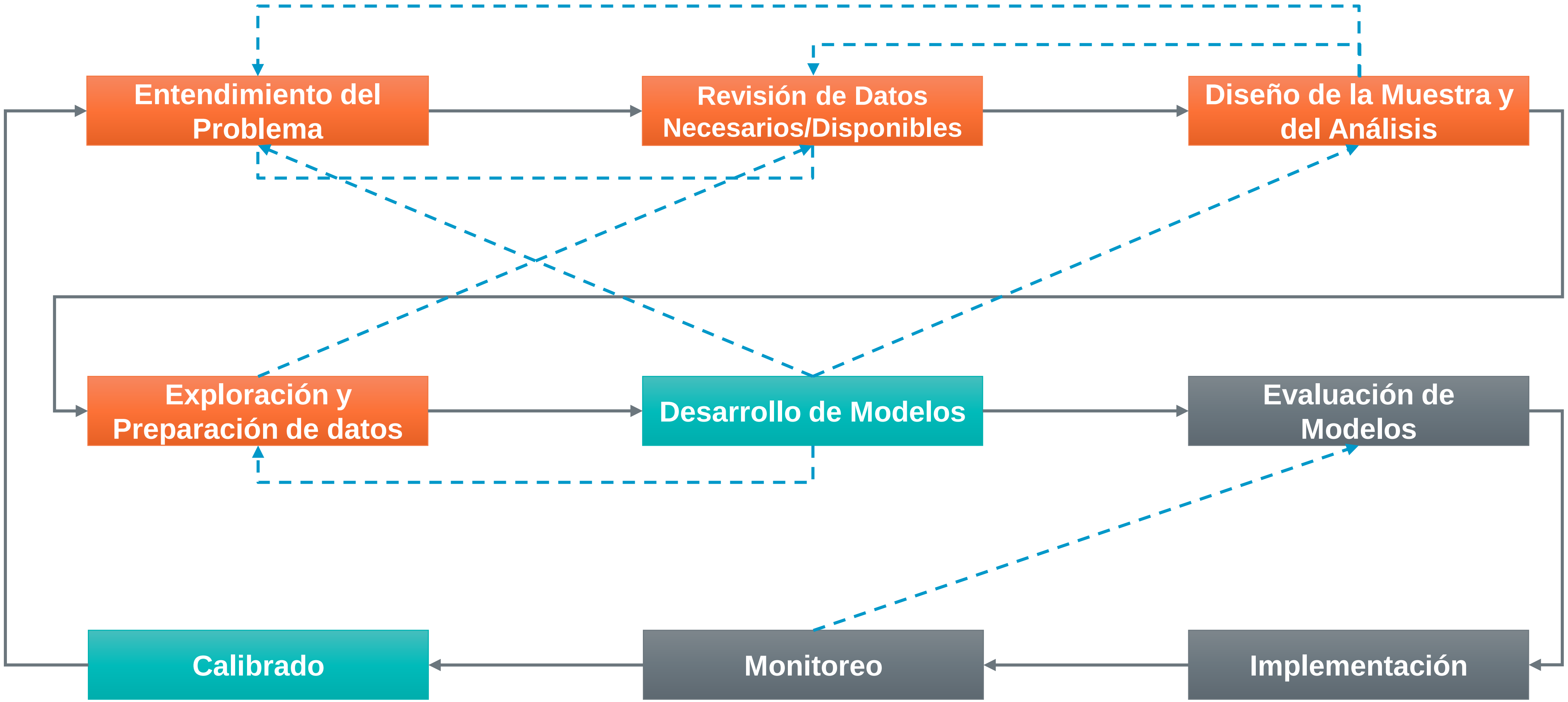
Image source: R. Olson *et. al.* (2016) "Evaluation of a Tree-based Pipeline Optimization Tool for Automating Data Science."

Pero…

- El Aprendizaje Supervisado no es un proceso secuencial, es iterativo
- ¿Y el Aprendizaje No Supervisado?
  - ✓ Clustering
  - ✓ Afinidad/Recomendación
  - Path Analysis
  - Optimización Estocástica/Programación Lineal
- ✓ ¿Y la Analítica de Texto?
- ✓ ¿Y las Series de Tiempo?
- ¿Y el Análisis de Redes Sociales/Relaciones?

teradata.

Revisemos el Ciclo de Vida de un Modelo



teradata.

Entendimiento del Problema y Revisión de Datos

- ✓ Prioridades del Negocio (Business Outcomes)
- Entendimiento de la problemática
- ✓ Revisión de la información disponible
- ✓ Definiciones de Negocio y procesos implicados
- ✓ Búsqueda de variables relevantes
- Planteamiento del Caso

Entendimiento del Problema y los Datos

Industry

- Healthcare
- Retail
- Oil & Gas
- Telco
- Finance
- Manufacturing
- Automotive
- Transportation
- Online Commerce
- Travel & Hospitality
- Entertainment
- Public Sector

Outcomes

- Increased Revenue
- Reduced Cost
- Increased Subs
- Reduced Calls to Care
- Increased ARPU
- Retained Customers
- Reduced Maintenance Costs
- Increased Market Share

Use Case

- Customer Segmentation
- Customer Journey
- Churn Prediction
- Predictive Asset Maintenance
- Supply-Chain Forecasting
- Commodity Price Prediction
- Fraud Detection

Domain

- Clustering
- Classification
- Anomaly Detection
- Estimation
- Forecasting
- Path Analysis
- Computer Vision
- Text Analysis
- Optimization
- Graph Analysis
- Dimensionality Reduction

Algorithm

- K-Means
- Hierarchical
- Spectral
- GLM
- XGBoost
- Random Forest
- SVM
- nPath
- ARIMA
- Moving Avg
- Naïve Bayes
- Jaccard
- RNN/CNN
- GAN
- cFilter
- Betweenness

teradata.

Título	Autor(es)	Pá	Técnicas	Tip	Añ
A framework for identification of high-value customers by including social network ba	Abbasimehr	17	Tree NN MLP RBF	Paper	2012
Customer churn analysis: Churn determinants and mediation effects of partial defecti	Ahn	17	Logit Factor	Paper	2006
Churn prediction in the mobile telecommunications industry: An application of Surviv	Alberts	47	Survival Tree Cart	Tesis	2006
Determinants of consumer retention in cellular industry of Pakistan	Ali, J., Ali, I., and Re	7		Paper	2010
Diseño e implementación de un modelo predictivo para detectar patrones de fuga en	Alvarado, J.	96	Tree NN	Tesis	2011
An SVM based churn detector in prepaid mobile telephony	Archaux, C.	4	SVM	Paper	2004
Modeling churn and usage behavior in contractual settings	Ascarza	26	Bayes MarkovDyna	Paper	2009
A Joint Model of Usage and Churn in Contractual Settings	Ascarza, E., and Ha	69	RFM MarkovChain	Paper	2013
A novel evolutionary Data Mining algorithm with applications to churn prediction	Au	14	AIGen	Paper	2003
Identificação e caracterização de situações de "churn" em sistemas de telecomunic	Azevedo, J.	109	Tree Chaid NN RBF	Tesis	2009
The relevant length of customer event history for churn prediction: How long is long e	Ballings	11	Tree Cart Chaid Log	Paper	2012
Modelo de gerenciamento de serviços, utilizando o valor do cliente no tempo: Uso d	Baraniuk, J.	296	Tree LTV	Tesis	2009
Diseño e implementación de una metodología de predicción de fuga de clientes en ui	Barrientos	310	Tree C45 NN Naivel	Tesis	2011
Aplicación de Minería de Datos para Predecir Fuga de Clientes en la Industria de las	Barrientos, F., and	35	Tree C45 NN Naivel	Paper	2013
Semi-supervised learning: A comparative study for web spam and telephone user chu	Benczúr, A., Csalo	8	Tree C45 Bagging	Paper	2007
Customer churn prediction based on the decision tree in personal handyphone syster	Bin	5	Tree	Paper	2007
Comparing complete and partial classification for identifying customers at risk	Bloemer	15	Tree Cart C5 MixRe	Paper	2003
A dynamic model of the duration of the customer's relationship with a continuous se	Bolton	51	Survival CoxReg	Paper	1998
A dynamic model of customers' usage of services: Usage as an antecedent and con:	Bolton, R. N.	16	Panel Tobit	Paper	1999
Prevedere il churn: Un approccio longitudinale	Bonetto, M.	98	Tree Cart Logit	Tesis	2007
Diagnosing and predicting individual customer defection in a contractual setting	Bonfrer	49	MovimientoBrowni	Paper	2007
Hybrid models using unsupervised clustering for prediction of customer churn	Bose	6	Kmeans Kmedioids	Paper	2009
Modelagem de probabilidade de churn	Botelho	15	Logit	Paper	2010
Modeling customer lifetimes with multiple causes of churn	Braun, M., and Sch	38	LTV BayesJer	Paper	2010
CRM at a Pay-TV company: Using analytical models to reduce customer attrition by t	Burez	11	Logit RandomFores	Paper	2007
Separating financial from commercial customer churn: A modeling step towards resc	Burez	30	Survival RandomFc	Paper	2007
Handling class imbalance in customer churn prediction	Burez, J., and Van c	11	Logit RandomFores	Paper	2009
Applying Data Mining to telecom churn management	Chang	11	AIGen	Paper	2009
Goal-oriented sequential pattern for network banking churn analysis	Chiang	9	Apriori	Paper	2003
Toward a hybrid Data Mining model for customer retention	Chu	15	SOM	Paper	2006
Analysis of marketing data to extract key factors of telecom churn management	Chueh, H-E.	6	FuzzyCorrelation	Paper	2011
Mineração de dados para a análise de atrito em telefonia móvel	Cister	167	NN	Tesis	2005
Churn prediction in subscription services: An application of support vector machines	Coussement	55	Logit RandomFores	Paper	2006
Um modelo de risco de cancelamento de clientes de telefonia fixa: A aplicação da Re	Da Cruz, M.	128	Logit	Tesis	2009
Retencao de clientes ao luz do gerenciamiento de churn: Um estudio no setor de tele	Dare	163		Tesis	2007
Social ties and their relevance to churn in mobile telecom networks	Dasgupta, K., Singh	10	Tree C45 SNA	Paper	2008
Domain knowledge integration in Data Mining for churn and customer lifetime value r	De Oliveira, E.	240	LTV	Tesis	2009
Regressão Logística: Um modelo de risco de cancelamento de clientes	De Almeida, K.	98	Logit	Tesis	2006
An empirical evaluation of rotation-based ensemble classifiers for customer churn pı	DeBock, K., and Va	25	RotBoost AdaBoo	Paper	2011
Reconciling performance and interpretability in customer churn prediction using ense	DeBock, K., and Va	33	GAM	Paper	2012
Estimating the effect of word of mouth on churn and cross-buying in the mobile phon	Dierkes, T., and Bic	33	MarkovLogicNet	Paper	2011
Modeling network effects with Markov Logic networks for churn prediction in the tele	Dierkes, T., Bichler	3	MarkovLogicNet	Paper	2009
Logistic model trees with AUC split criterion for the KDD cup 2009 small challenge	Doetsch, P., and B	12	NN MLP SVM LMT	Paper	2009
Churn predictive analytics	Dominissini, D., an	1	Tree Cart Chaid C4!	Paper	2010
New Evidence on the Reasons for Switching Service Providers	East	8		Paper	2007
Determinants of customer loyalty in the wireless telecommunications industry	Eshghi	14	EQS	Paper	2007
Uma aplicação de mineração de dados no gerenciamento do churn em serviços de b	Fernandes, A., Car	21	LinReg	Paper	2008
Mining and understanding wireless churn	Ferreira	14	AIGen NeuroFuzzy	Paper	2004
Data Mining techniques on the evaluation of wireless churn	Ferreira	6	Tree C45 NN MLP /	Paper	2004
Mineração de dados na retenção de clientes em telefonia celular	Ferreira	93	Tree C45 NN MLP /	Tesis	2005
Satisfação, lealdade e retenção: Um pré-experimento aplicado à telefonia móvel	Ferreira, J., Morigu	14	EQS Factor	Paper	2008
Survival analysis models to estimate Customer Lifetime Value	Figini, S., Giudici, P	11	LTV CoxReg	Paper	2005
Uma analise de cancelamentos em telefonia utilizando mineração de dados	Andrade, D.	74	Tree NN MLP Logit	Tesis	2007
Championing LTV at LTC	Freeman, E., and M	7	LTV	Paper	2005
Identification of churn routes in the Brazilian telecommunications market	Garcia, D., Vellido,	6	GTM SOM	Paper	2007
Enhanced customer relationship management using Fuzzy Clustering	Gayathri A., and Mı	5	FuzzyClustering Km	Paper	2011
Customer retention, loyalty, and satisfaction in the german mobile cellular telecommu	Gerpott, T., Rams,	20		Paper	2000
Modeling churn using Customer Lifetime Value	Glady	16	Tree NN MLP Logit	Paper	2009
Customer churn time prediction in mobile telecommunication industry using Ordinal F	Gopal, R., and Meh	6	OrdinalReg	Paper	2008
Customer duration in non-life insurance industry	Gustafsson, E.	53	Survival CoxReg	Tesis	2009
Design and analysis of the KDD cup 2009: Fast scoring on a large Orange customer c	Guyon, I., Lemaire,	9	NaiveBayes	Paper	2010

Título	Autor(es)	Pá	Técnicas	Tip	Añ
Churn prediction: Does technology matter?	Hadden, J., Tiwari, .	7	Tree Cart NN MLP I	Paper	2006
The analysis of logistic Regression in customers' churn of vip electronic mailbox	Han, J., Zhang, L., S	5	Logit	Paper	2007
Customer Churn Prediction in Telecommunication A Decade Review and Classificat	Hashmi, N., Butt, N	12		Paper	2013
Applying Data Mining to telecom churn management	Hung	15	Tree NN BP	Paper	2004
An LTV model and customer segmentation based on customer value: A case study c	Hwang	8	LTV	Paper	2004
Churn Prediction in Telecommunication Using Data Mining Technology	Jadhav, R., and Pa	3	NN BP	Paper	2011
Churn management in the telecom industry of Pakistan: A comparative study of Ufor	Jahanzeb, S., and J	10		Paper	2007
Improving the diagnosis and prediction of customer churn: A heterogeneous hazard r	Jamal	13	Survival Weibull	Paper	2006
Customer segmentation and customer profiling for a mobile telecommunications co	Jansen, S.	76	Cluster Fuzzy KMea	Tesis	2007
Modelado de Influencia de un Comportamiento en una Red de Telefonía Celular Mec	Jimenez, J.	15	SNA SPA	Paper	2011
The research on applying Data Mining to telecom churn management	Jin, S., Meng, J., Fa	7	Tree C45 NN BP	Paper	2012
The application of AdaBoost in customer churn prediction	Jinbo	6	AdaBoost	Paper	2007
Research on customer classification based on Logistic Regression analysis	Jing, Z., and Xing-H	4	Logit	Paper	2008
A Survey on Churn Prediction Techniques in Communication Sector	Kamalraj, N., and M	4		Paper	2013
Data Mining via cellular neural networks in the GSM sector	Karahoca, A.	6	NN	Paper	2004
Comparing clustering techniques for telecom churn management	Karahoca, A., and I	6	DBSCAN Kmeans	Paper	2006
Benchmarking the Data Mining algorithms with adaptive neuro-fuzzy inference systen	Karahoca, A., Kara	15	ADTree BayesNet	Paper	2009
Churn in Social Networks: A Discussion Boards Case Study	Karnstedt, M.	8	SNA	Paper	2010
Data Mining as a tool to Predict the Churn Behaviour among Indian bank customers	Kaur, M., Singh, K.,	6	Tree NaiveBayes S\	Paper	2013
Applying Data Mining to customer churn prediction in an internet service provider	Khan, A., Jamwal, S	7	Tree Cart NN BP Lc	Paper	2010
Intelligent Churn prediction for Telecommunication Industry	Khan, I., Usman, I., I	6	Kmeans AIGen SVN	Paper	2013
Customer segmentation and strategy development based on customer lifetime value	Kim	7	Tree LTV	Paper	2006
Determinants of subscriber churn and customer loyalty in the Korean mobile telephor	Kim	15	Logit	Paper	2004
Predicting Customer Churn in Mobile Telephony Industry Using Probabilistic Classifi	Kirui, C., Hong, L., (	8	Tree C45 NN Bayes	Paper	2013
Modeling Data Mining applications for prediction of prepaid churn in telecommunicat	Kraljevic, G., and G	9	Tree NN Logit	Paper	2010
Predicting credit card customer churn in banks using Data Mining	Kumar	25	Tree C45 NN MLP I	Paper	2008
Extending traditional telecom churn prediction using social network data	Kusuma, P. D.	49	Tree CHAID Logit S	Tesis	2013
Combining customer attribute and social network mining for prepaid mobile churn pre	Kusuma, P., Rados	9	SNA	Paper	2013
Churn Prediction	Lazarov, V., and Ca	5	Logit Tree Cart NN I	Paper	2008
Measuring the impact of Data Mining on churn management	Lejeune, M.	13		Paper	2001
Bagging and Boosting classification trees to predict churn	Lemmens	40	Tree C45 Logit Boo	Paper	2006
Applying fuzzy Data Mining to telecom churn management	Liao, K., and Chueh	5	Fuzzy	Paper	2011
An ensemble of three classifiers for KDD cup 2009: Expanded linear model, heteroge	Lo, H-Y., Chang, K-	8	AdaBoost	Paper	2009
Modeling partial customer churn in the portuguese fixed telecommunications industr	Lopes, S.	203	Survival CoxReg	Tesis	2010
Modeling customer lifetime value using survival analysis - an application in the teleco	Lu, J.	6	Survival	Paper	2008
Predicting customer churn in the telecommunications industry – an application of sur	Lu, J.	6	Survival	Paper	2007
Subscriber churn in the australian ISP market	Madden	14	RegLin	Paper	1999
Churn prediction and management system	Maga, M., Canale,	27		Paper	2007
Statistics and Data Mining techniques for Lifetime Value modeling	Mani, D., Drew, J., E	10	LTV NN CoxReg PH	Paper	1999
Predicción de fugas de Clientes en una compañía de seguros utilizando redes neuror	Martinez, C.	93	NN	Tesis	2010
CHAMP: A prototype for automated cellular churn prediction	Masand, B., Datta,	6		Paper	1999
A Logit model of customer churn as a way to improve the customer retention strateg	Menezes, R., and F	9	Logit	Paper	2009
Hierarchical Neural Regression Models for Customer Churn Prediction	Mohammadi	10	Cmeans NN SOM C	Paper	2013
Aplicación de un modelo predictivo de fuga de clientes utilizando data mining en VTR	Moreno, M., and O	248	Logit Kmeans	Tesis	2011
Analysing customer churn in insurance data - A case study	Morik	12	Tree C45 SVM Naiv	Paper	2004
Churn reduction in the wireless industry	Mozer	7	NN Logit	Paper	2000
Predicting subscriber dissatisfaction and improving retention in the wireless telecomi	Mozer, M., and Wo	14	Tree C5 NN MLP Lc	Paper	2000
Customer churn analysis - A case study	Mutanen, T.	19	Logit LTV	Paper	2006
Analyzing the structure and evolution of massive telecom graphs	Nanavati, A., Singh	16	SNA	Paper	2008
Customer churn analysis in the wireless industry: A Data Mining approach	Nath	19	NaiveBayes	Paper	2003
Winning the KDD cup Orange challenge with ensemble selection	Niculescu-Mizil, A.	12		Paper	2009
Data Mining in churn analysis model for telecommunication industry	Oseman, K., Mohd	9	Tree C45	Paper	2010
Churn models for prepaid customers in the cellular telecommunication industry using	Owczarczuk, M.	3	Tree Logit	Paper	2010
Extracting dense communities from telecom call graphs	Pandit, V., Modani,	8	SNA	Paper	2008
Genetic algorithm based neural network approaches for predicting churn in cellular wi	Pendharkar	7	AIGen NN	Paper	2009
Proyecto de minería de datos para el análisis del comportamiento de los clientes de	Peralta, D.	160	NN	Tesis	2009
Modelo de Mineração de Dados para clasificação de clientes em telecomunicacoes	Petermann	164	Tree C45 NN RBF E	Tesis	2006
Prediction of Subscriber Churn Using Social Network Analysis	Phadke	63	SNA Tree	Paper	2013
On the use of continuous duration models to predict customer churn in the ADSL ind	Portela, S.	7	Survival CoxReg	Paper	2009
Detecting customer defections: An application of continuous duration models	Portela, S., and Me	9	Survival CoxReg	Paper	2011
Modeling customer churn: An application of duration models	Portela, S., and Me	9	Survival	Paper	2009

Diseño de la Muestra y del Análisis

- Diseño de investigación
- Criterios de filtrado/inclusión
- Horizontes temporales
- Tamaño de muestras
- Corrección de posibles sesgos

teradata.

Diseño de la Muestra y del Análisis

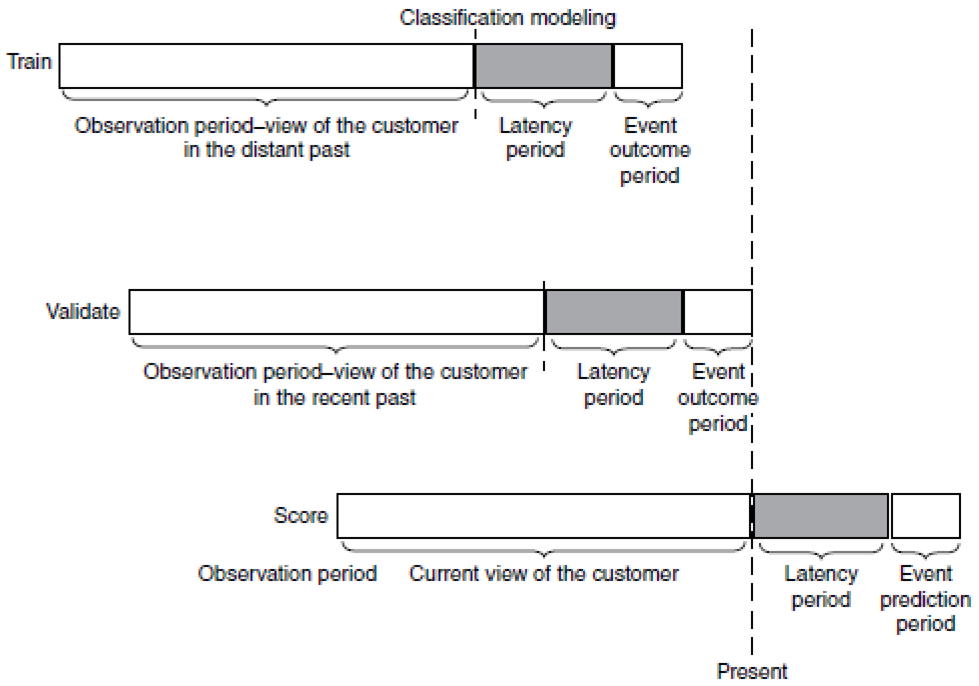


Figure 2.1 The data setup and time frames in a classification model trained on historical data. Source: Tsiptsis and Chorianopoulos (2009). Reproduced with permission from Wiley

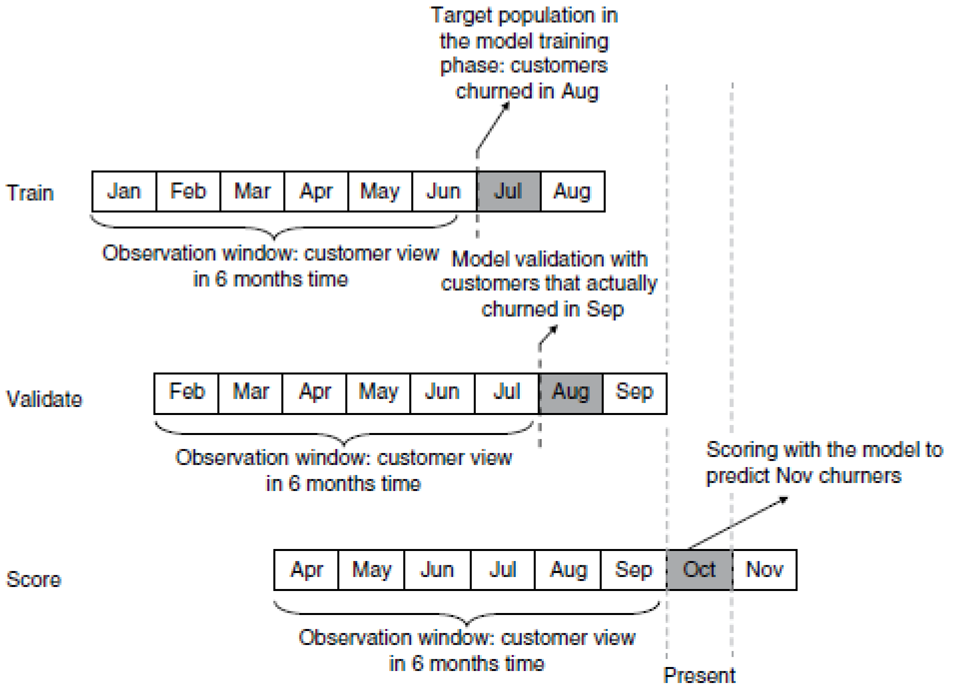


Figure 2.2 The data setup and time frames in a churn model. Source: Tsiptsis and Chorianopoulos (2009). Reproduced with permission from Wiley

Exploración y Preparación de datos

- ✓ Identificar Tipos de datos / Escala
- ✓ Imputar Valores
- ✓ Identificar Anomalías
- ✓ Encoding
- ✓ Discretizar
- ✓ Análisis Exploratorio
- ✓ Visualización

Exploración y Preparación de datos

- ✓ Reducción de Dimensionalidad
- ✓ Transformación de Variables
- ✓ Selección de Variables
- ✓ Construcción de Variables Derivadas

Desarrollo de Modelos

- ✓ Optimización de Hiperparámetros
- ✓ Entrenamiento
- ✓ Selección
- ✓ Ensamblaje
- ✓ Testing Automatizado

Implementación, Monitoreo y Calibrado

- ✓ Exportar Modelos
- ✓ Implementación de Modelos
- ✓ Champion-Challenger
- ✓ Monitoreo de Modelos
- ✓ Calibrado

Accionamiento de los Modelos

Factores a tener en cuenta

Canal de contacto preferido	Costo de Retención (Regalo, Premio)	Comportamiento de Compra
Propuesta de Valor	Tipo de Cliente	Principalidad
Modelo de Atención	CLTV	Segmento de Valor

teradata.

¿Entonces, cuánto del proceso es factible de ser automatizado?

Bastante.

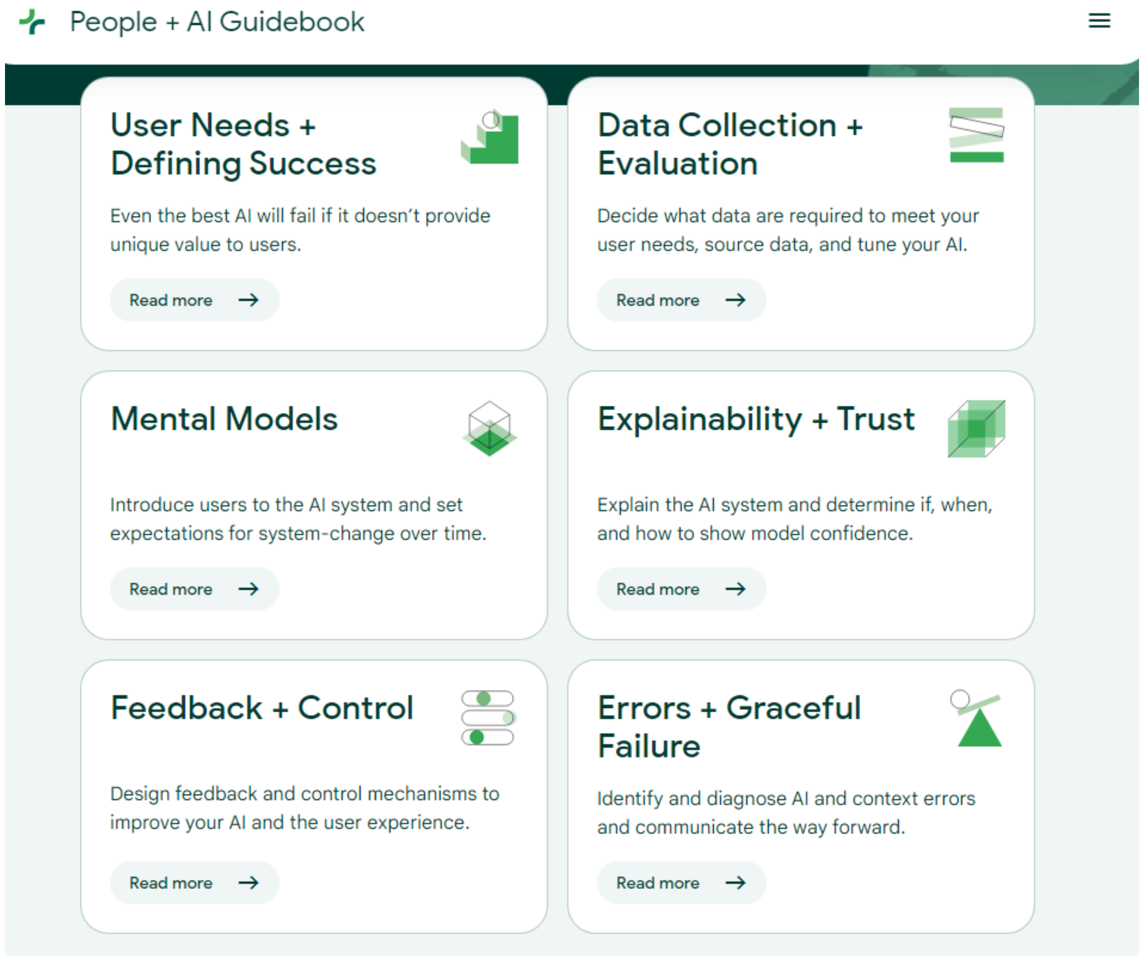
¿Y qué más está sucediendo?

Google y las Cascadas de Datos

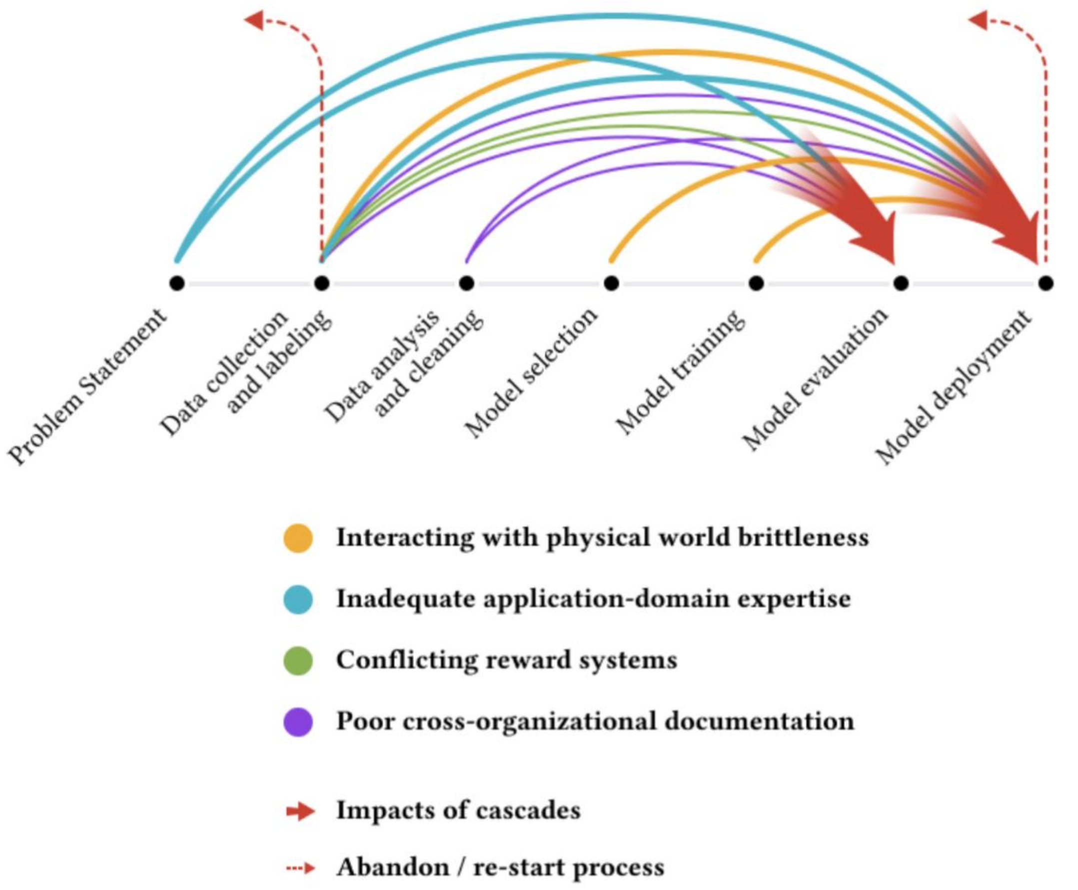
https://ai.googleblog.com/2021/06/data-cascades-in-machine-learning.html

¿Y qué más está sucediendo?

PAIR



Data Cascades in High-Stakes AI



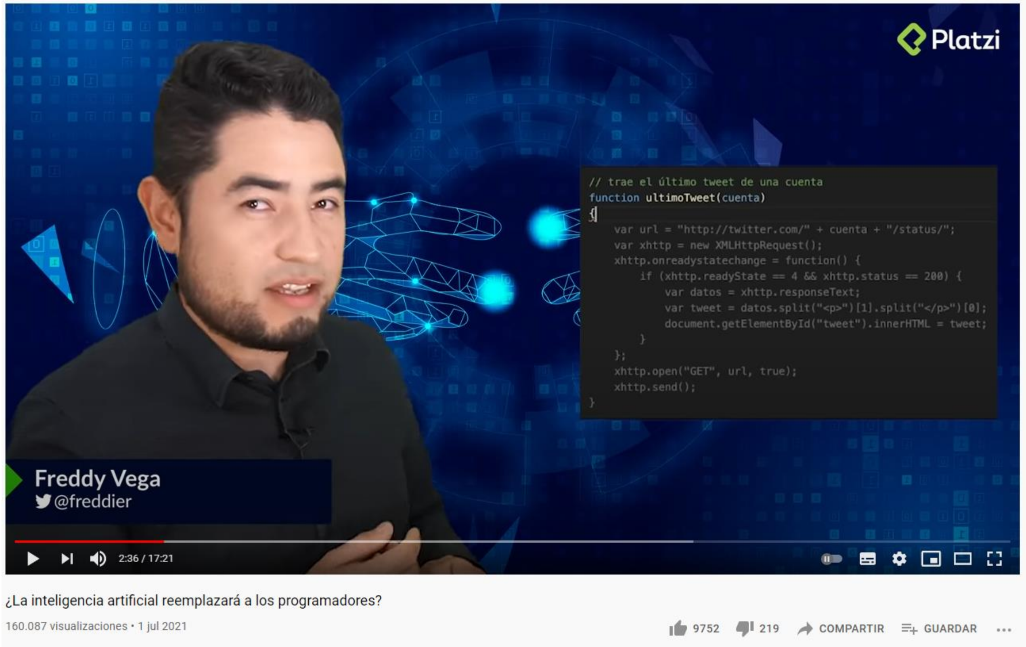
teradata.

¿Y qué más está sucediendo?

GitHub Copilot

kite

tabnine



Kite VS. TabNine: Which AI Code Autocomplete Should You Choose?

I tested two of the hottest AI code assistant in 2021

Jason Zhang Follow

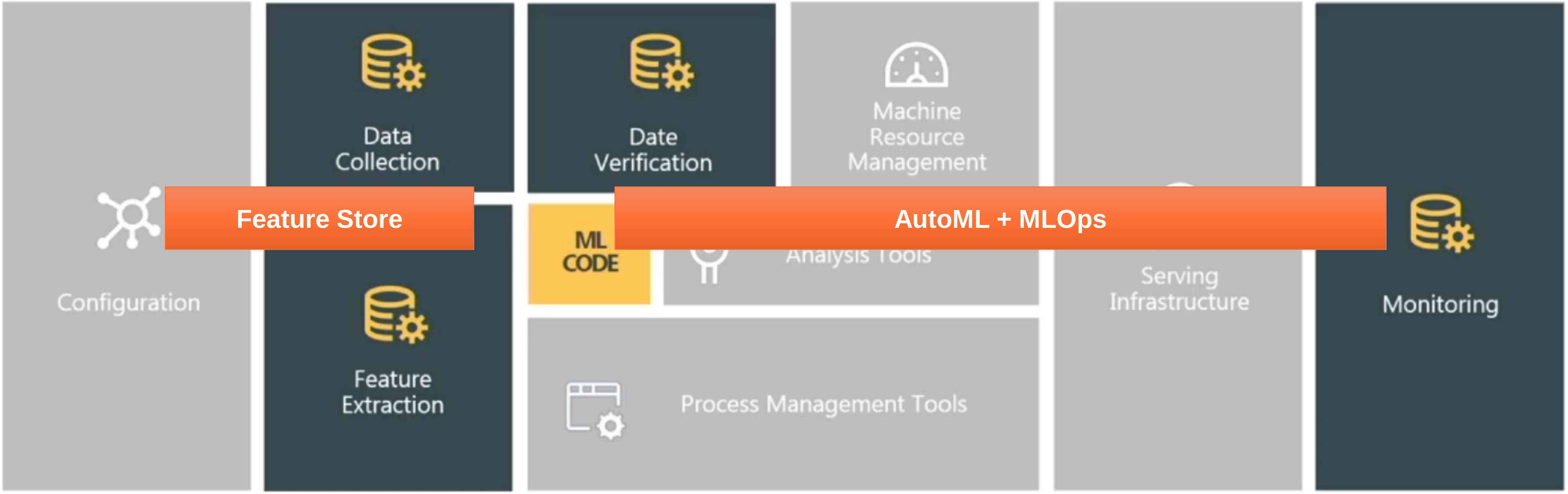
Dec 31, 2020 · 5 min read ★

teradata.

¿Y la deuda técnica?

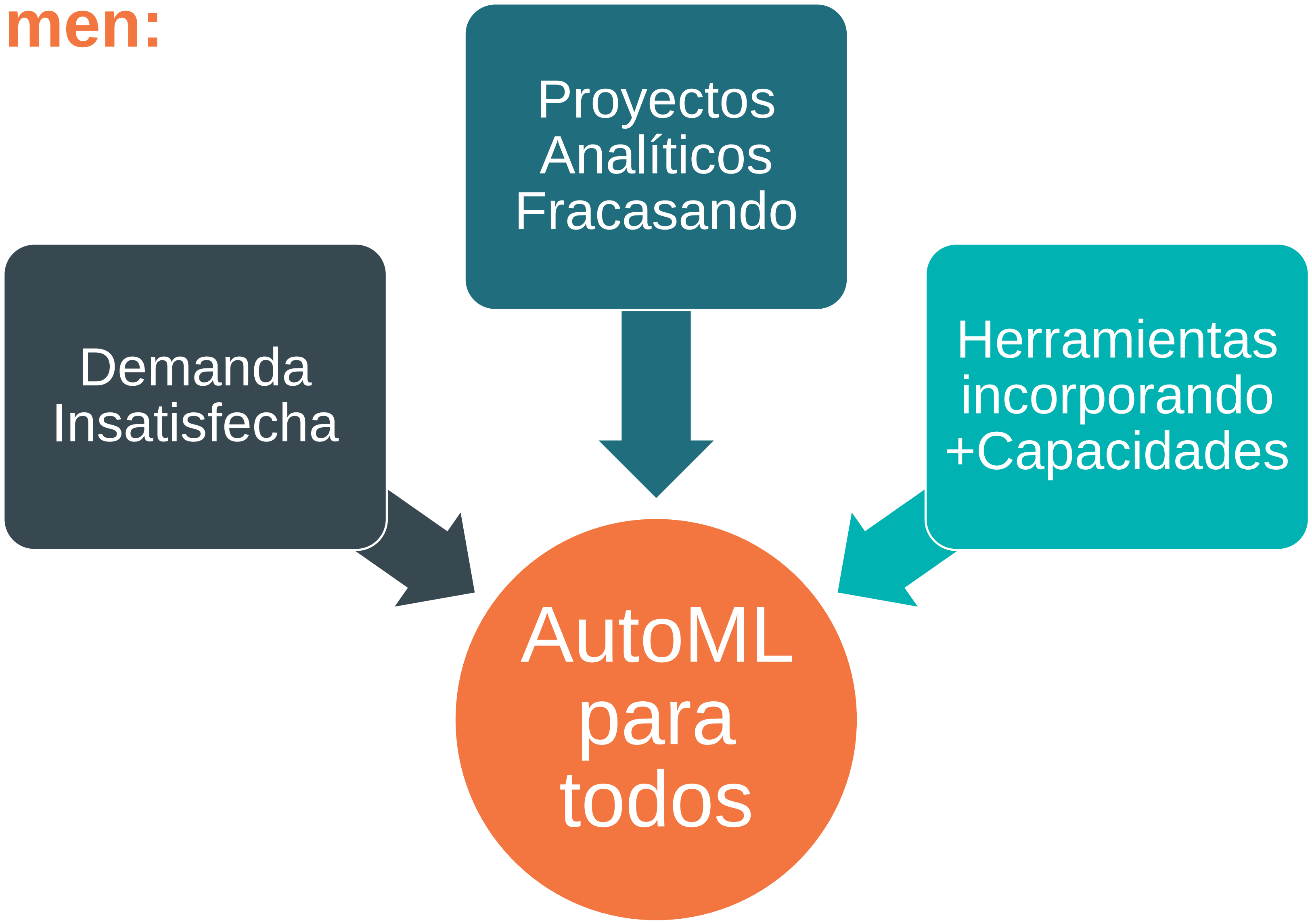
Hidden Technical Debt in Machine Learning Systems

D. Sculley, Gary Holt, Daniel Golovin, Eugene Davydov, Todd Phillips
{dsculley,gholt,dgg,edavydov,toddphillips}@google.com
Google, Inc.



Source: https://papers.nips.cc/paper/5656-hidden-technical-debt-in-machine-learning-systems.pdf

En Resumen:


Proyectos Analíticos Fracasando
Demanda Insatisfecha
Herramientas incorporando +Capacidades
AutoML para todos


Y CUANDO TODOS
SEAN SÚPER...


NADIE VA A SER
45

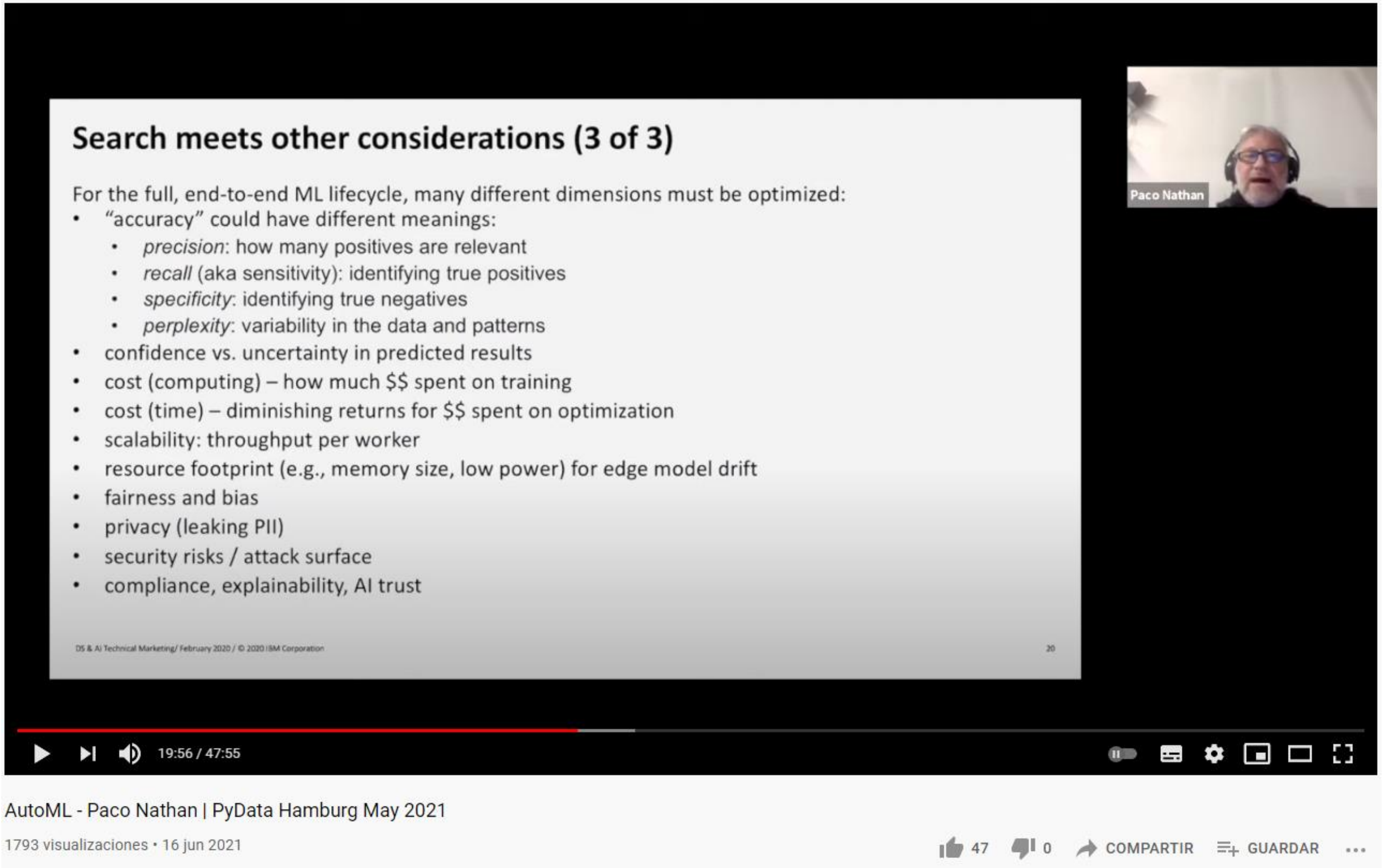
¿Y cuál es la mayor amenaza que trae AutoML?

Que los financieros se preocupen por precio y no por el valor

teradata.

Es necesario evidenciar todo lo que se necesita

Sobre todo que no se trata sólo de reducir el headcount



Recomendaciones

Recomendaciones

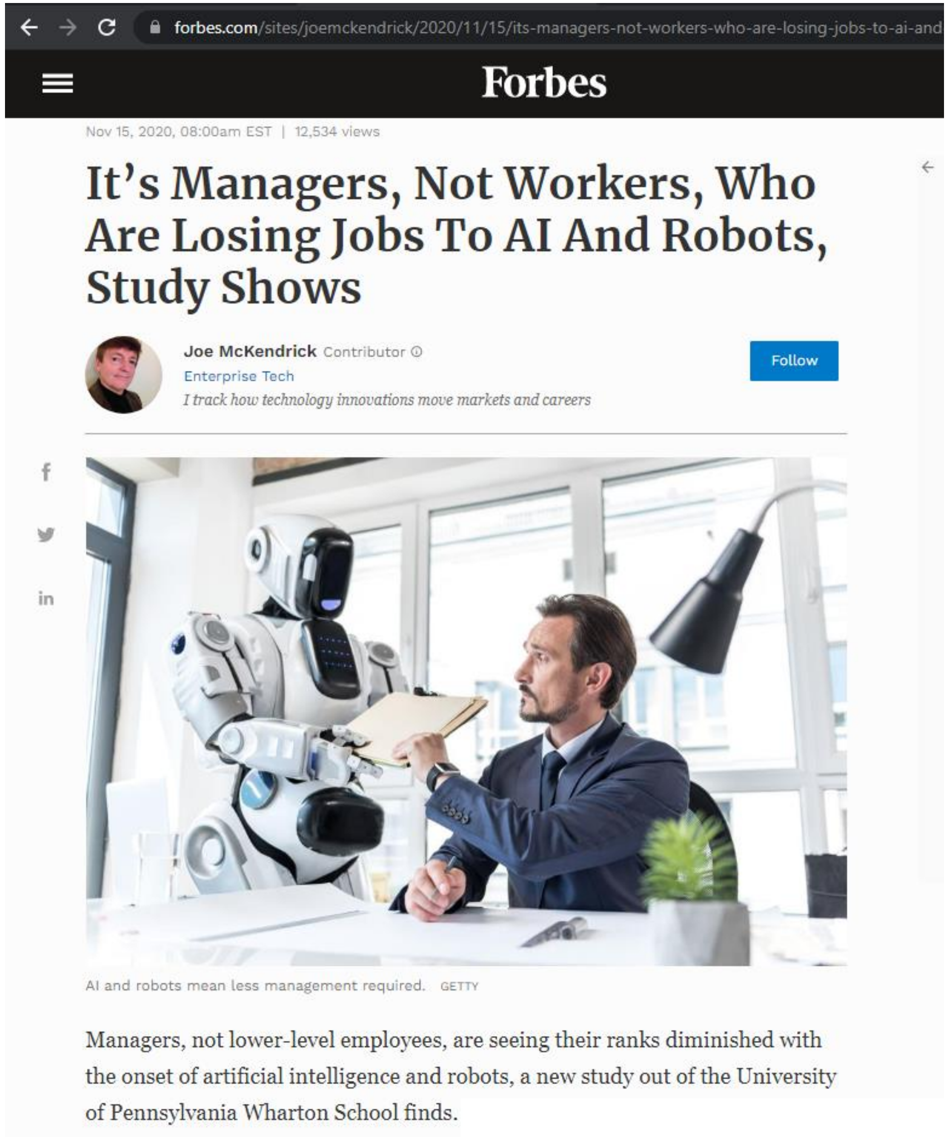
Gente que está adoptando AutoML:

- Revisar las necesidades
- Evaluar diferentes alternativas
- Priorizar capacidades requeridas
- Hacer experimentos (muchos!)
- Trabajar con distintos escenarios
- Generar conciencia de las limitaciones tecnológicas

A Futuro:

- Ser parte del cambio
- Buscar aumentar las capacidades, no reemplazar personas
- Nunca dejar de aprender y desaprender
- Reforzar las capacidades que difícilmente pueden ser automatizadas

¿Obsoletos?



forbes.com/sites/joemckendrick/2020/11/15/its-managers-not-workers-who-are-losing-jobs-to-ai-and

Forbes

Nov 15, 2020, 08:00am EST | 12,534 views

It's Managers, Not Workers, Who Are Losing Jobs To AI And Robots, Study Shows

Joe McKendrick Contributor

Enterprise Tech

I track how technology innovations move markets and careers

Follow

AI and robots mean less management required. GETTY

Managers, not lower-level employees, are seeing their ranks diminished with the onset of artificial intelligence and robots, a new study out of the University of Pennsylvania Wharton School finds.

teradata.

Para recordar:

1

Las soluciones AutoML **seguirán evolucionando** e incorporando más capacidades. **Hay muchas oportunidades claras, construyendo el futuro.**

2

Las partes del proceso menos automatizables son también **las menos técnicas.**

3

Renovarse continuamente o Desaparecer.


Let’s go!