



CLAR
II CONGRESO
LATINOAMERICANO DE
RIESGOS

Machine Learning y su Utilización en Riesgo de Crédito

Seemant Teotia
Sr. Director International Analytics Equifax Inc.
USA

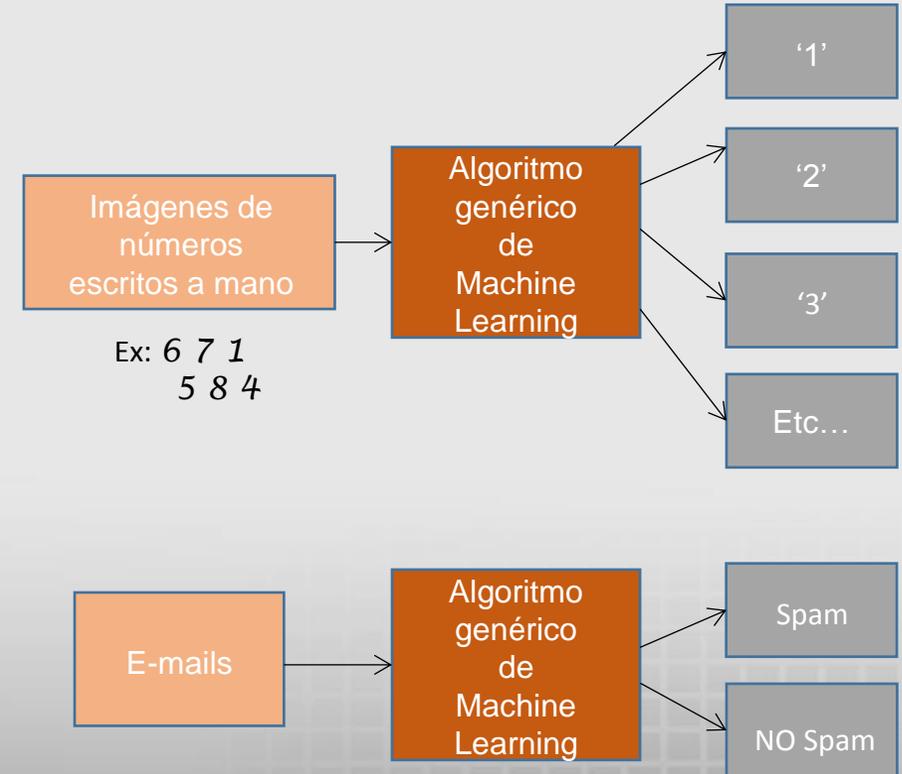
EQUIFAX[®]



- Principios de Machine Learning
 - ¿Qué es?
 - ¿Por qué lo usamos?
 - ¿Cuándo lo usamos?
- Ejemplos de Uso en Equifax
 - Datos Crudos
 - Creación de atributos
 - Segmentación del riesgo
 - Técnicas Avanzadas para Modelos de Riesgo
 - Fraude

Agenda

- Machine learning es la idea de que hay algoritmos genéricos que pueden decir algo interesante sobre un conjunto de datos sin tener que escribir un código específico
- En lugar de escribir código, se alimentan los datos al algoritmo genérico y esto construye su propia lógica basada en los datos.
- “Machine learning” es un término que cubre muchos de estos tipos de algoritmos genéricos.

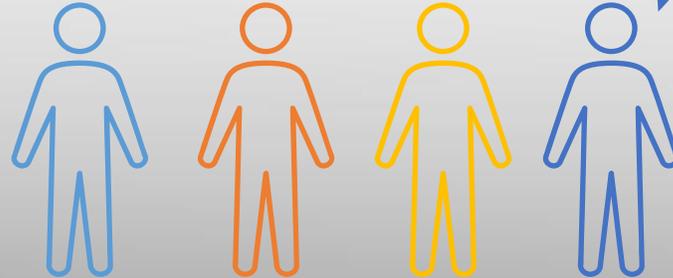


¿Qué es Machine Learning?

¿Es sólo para los coches de auto conducción y robóticos chefs?
¡NO!

Es una rama de Estadística, diseñada para el mundo de Big Data. Big Data no es sólo LARGA sino ANCHA!

Estadística tradicional: centrado en la causalidad,
Machine Learning: centrado en la predicción.

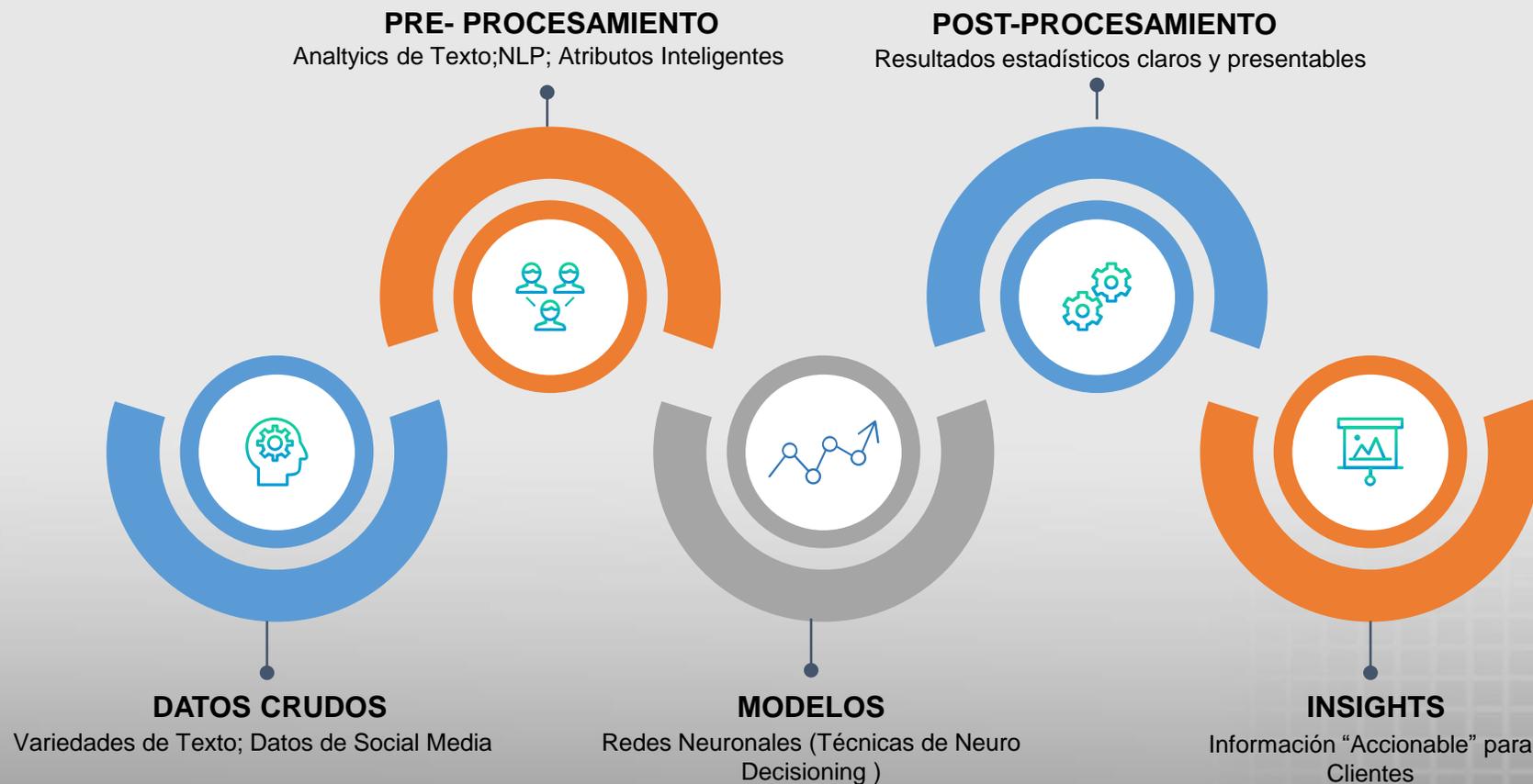


Ejemplo: Una tienda online; cuanto más clientes se obtienen más “larga” es la base de datos y cuanto más datos recopilamos respecto a un cliente más “ancha” es nuestra base de datos.

¿Por qué Machine Learning?



Aplicaciones de Machine Learning Analytics



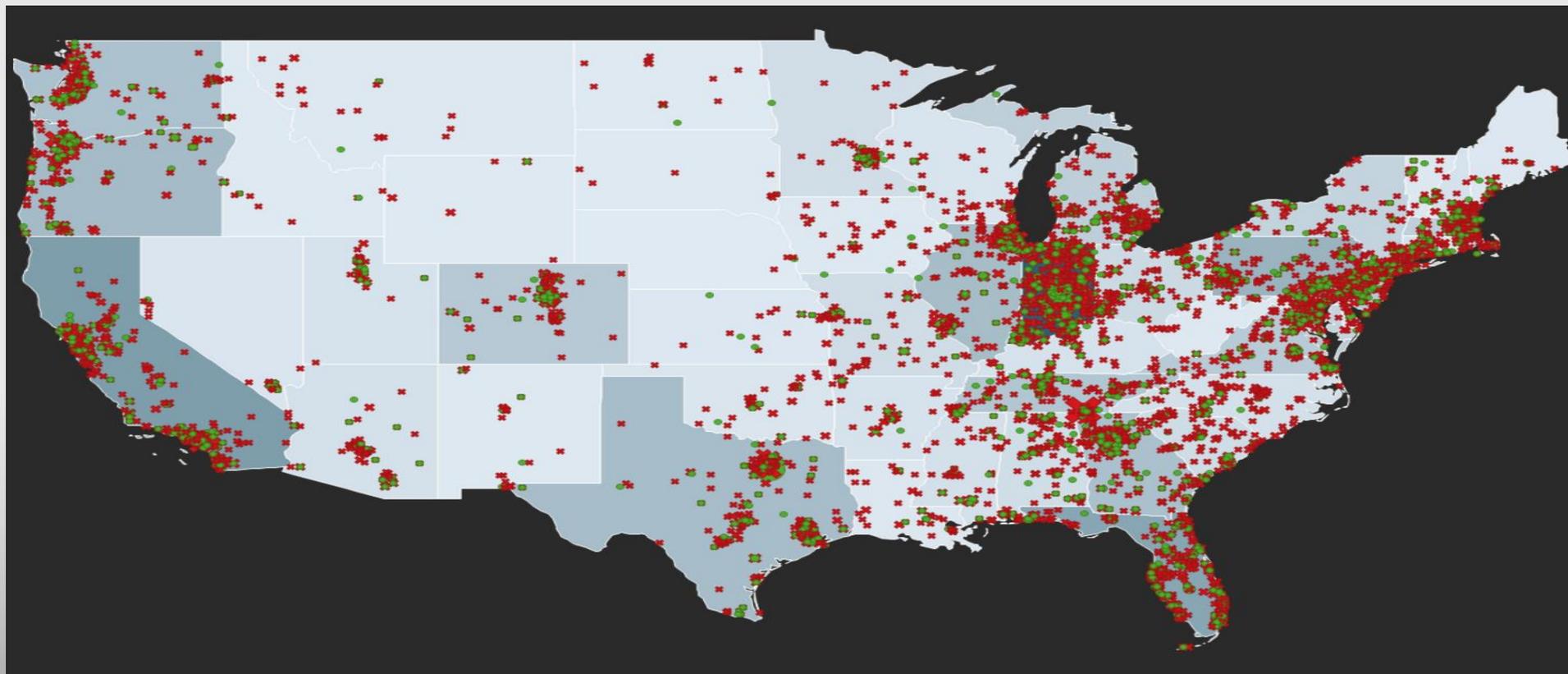
¿Dónde usamos Machine Learning en EFX?



Clasificación de Datos Crudos - "Identificar quién está molesto con su Proveedor de Internet"

- Tipo de Problema: Clasificación de texto
- Algoritmos de clasificación típicos:
 - Regresión logística, Bayes', SVM, Nearest Neighbor, Random Forest, NN/Deep Learning
- Proceso: Examinar cada tweet y tomar una decisión si el tweet expresa frustración con su servicio de Internet
- No hay datos etiquetados
- No hay diccionarios específicos para ayudar en el etiquetado de los datos

Definición y Desafíos



● Frustrado con ISP1 ■ Encantado con ISP1

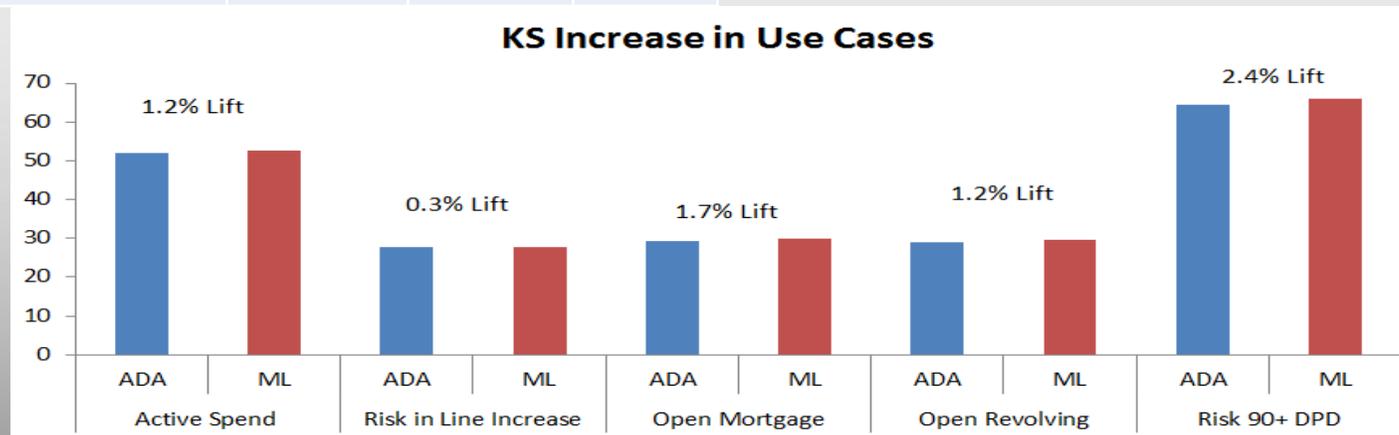
Análisis Geográfico



Atributos Inteligentes

- Machine Learning supera los métodos tradicionales
 - Combinaciones de los atributos proporcionan un lift a través de información conjunta
 - Las combinaciones de atributos eliminan los atributos originales

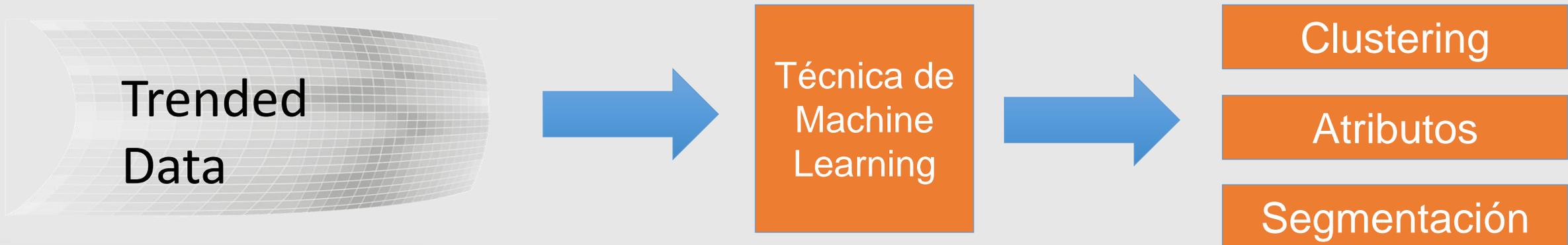
Atributo	Chi-Sq	Rank	KS		Atributo<	Chi-Sq	Rank	KS
%Cuentas siempre en buen estado / total de cuentas (A)	577	5	31	➔	(B)> 0 or (A)< 0.8	964	3	56
# Veces en estado avanzado de cobranzas en los últimos 12 meses (B)	316	17	5					



Ejemplos de Atributos



Segmentación de Riesgo

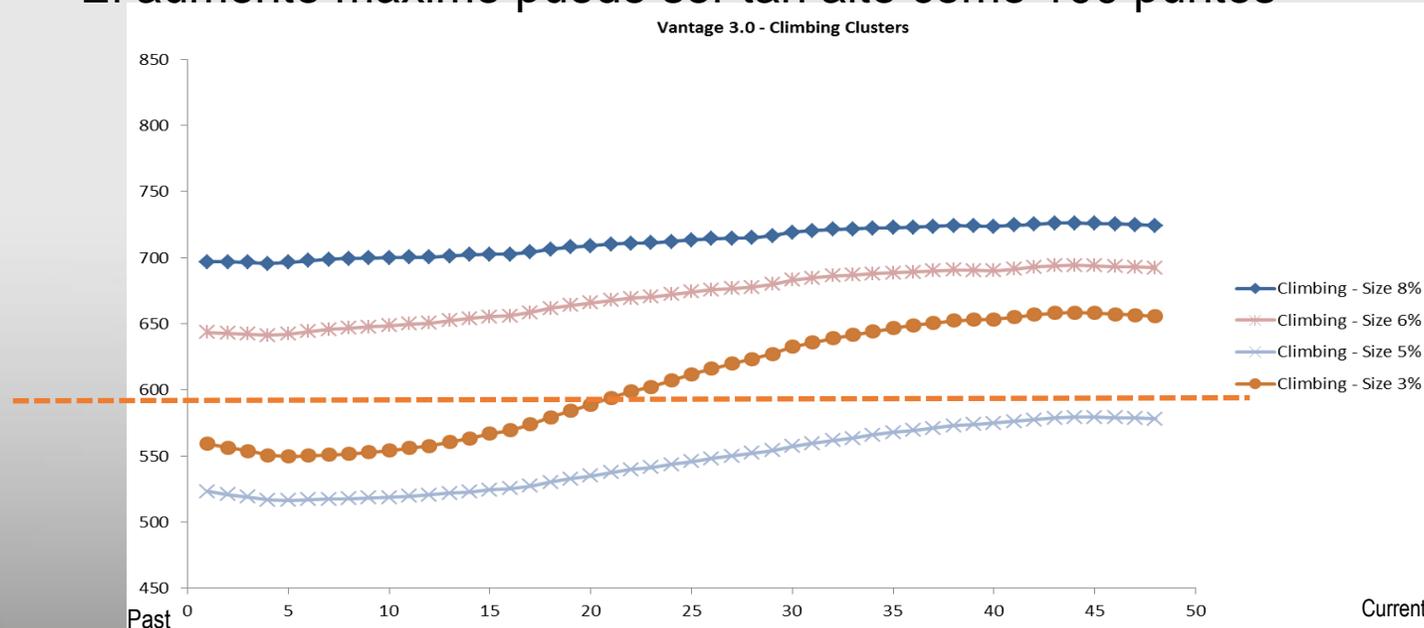


- Aprendizaje no supervisado - no hay variable objetivo



Atributos de Tendencia: generan una mejora significativa

- Se identificaron 4 Clusters con comportamiento ascendente, trayectoria de mejora crediticia. Representa el 23% de la población activa de Estados Unidos
 - ✓ **Grupo Sub Prime para Prime** se ha movido a través de la línea más amplia y están calificados para más oportunidades de crédito
 - ✓ El aumento máximo puede ser tan alto como 100 puntos



Grupo 1: Creciendo hacia Super-Prime (mejora de 20 puntos)

Grupo 2: Sub-Prime a Prime (mejora de 30 puntos)

Grupo 3: Sub-Prime Ascendente (mejora de 100 puntos)

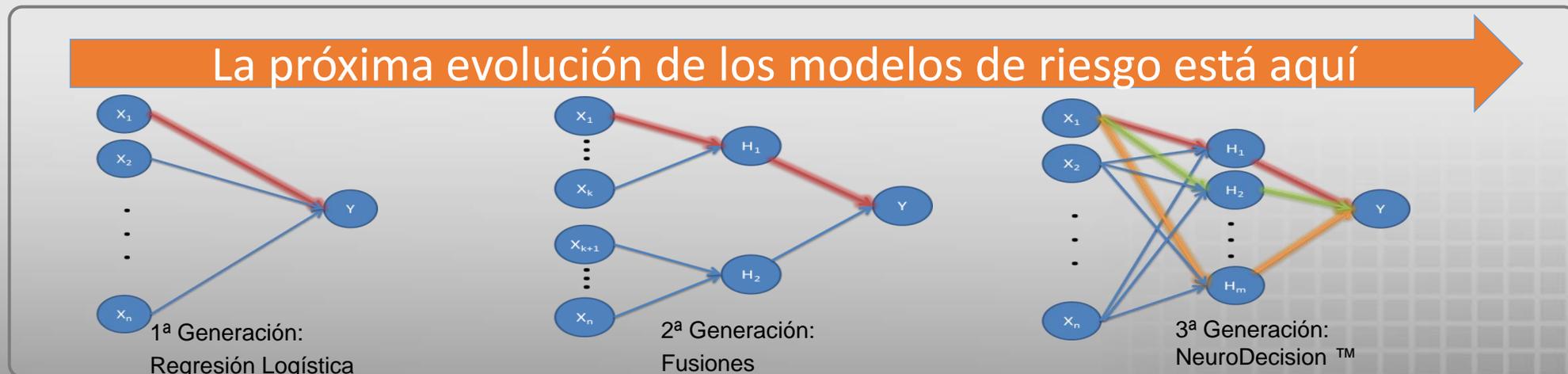
Grupo 4: Superando Situación Crítica (mejora de 40 puntos)

Score Clustering – Grupos Ascendentes



Modelaje del Riesgo de Crédito

- Tecnología exclusiva Equifax de Machine Learning (patentada) que produce redes neuronales conformes a la regulación para aplicaciones de decisiones de riesgo.
- Redes neuronales con restricciones óptimas
 - Mejora en rendimiento y precisión, asigna una puntuación de riesgo y motivos de razón, transparencia para clientes y reguladores. Permite un mejor conocimiento del comportamiento del consumidor a través de interacciones complejas de atributos no lineales
- Puede ser aplicado e implementado dondequiera que los modelos tradicionales sean apropiados
- Evaluación de modelos completamente automatizada - acelera el desarrollo



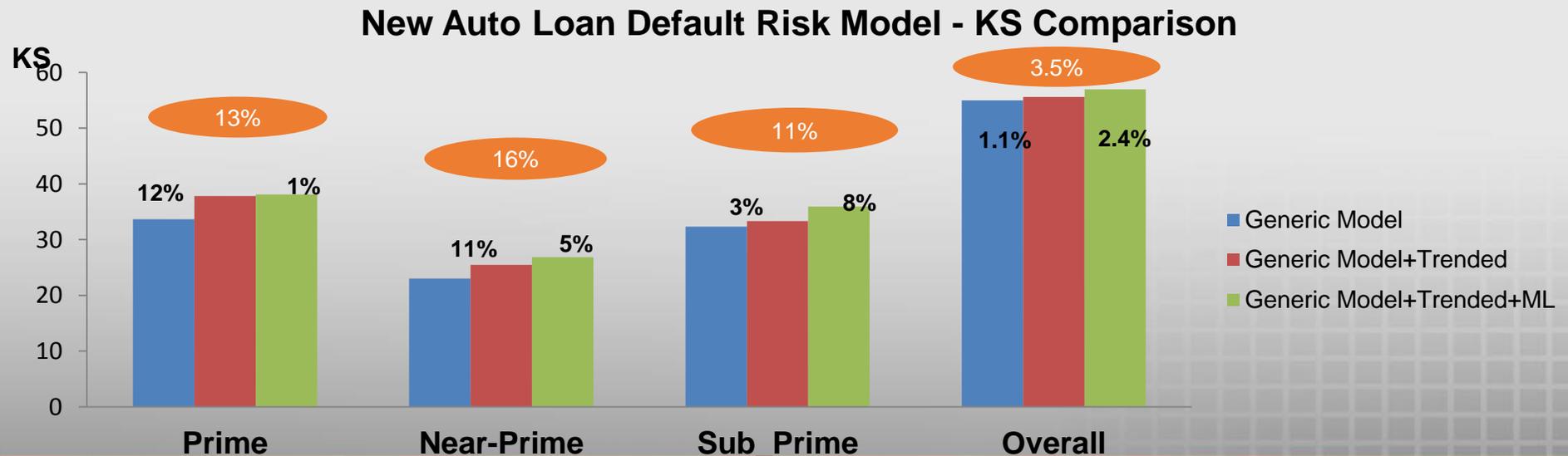
Tecnología NeuroDecision™ – Evaluación superior de riesgo con mayor transparencia para los clientes

- ¿Como Trended Data y Neurodecisioning de Equifax ayudan en la predicción de nuevas originaciones de hipoteca a incumplimiento?
- Propósito:
 - Predecir la probabilidad de que un nuevo crédito hipotecario o prendario entre en default en los próximos 24 meses
- Muestra:
 - Muestra: 5% de la población total de USA al punto de observación
- Punto de observación:
 - Marzo 2014
- Ventana de desempeño:
 - Abril 2014 - Marzo 2016 (24 meses)
- Variable dependiente:
 - 90+ días en créditos prendarios dentro de la ventana de desempeño



Caso de Éxito: Default en un nuevo crédito hipotecario o prendario

- El modelo con trended data + NDT genera una mejora de 3.5% en el estadístico KS
- Únicamente incorporando datos de tendencia se observa una ganancia de 12% en el segmento ‘Prime’ y 11% en el ‘Near Prime’
- Se observa un 8% de mejora con NDT en el segmento ‘Subprime’ . NDT es más fuerte cuando la cantidad de malos es más alta.



Trended Data y NDT llevan a una Mejora “Lift Estadístico” en todos los Segmentos Clave



Fraude

Motivación

- Exploración de las relaciones indirectas entre atributos de identidad de una persona (PII) hacia soluciones de fraude de identidad
- Graph Mining revela relaciones indirectas en la red
- La computación distribuida potencia Graph Mining a gran escala

Tecnología



Graph

- Graph basado en las consultas realizadas a la base de datos de EFX - Detección de anomalías
- Graph basado en atributos de Identidad (PII)- Nuevos atributos