



PATROCINADOR
ORO+



Ciencia de Datos y Calidad: Detección automática de tendencias en calidad de procesos



Sebastián Murillo
Daniela Arroyo
HRGS PSO Quality & Analytics

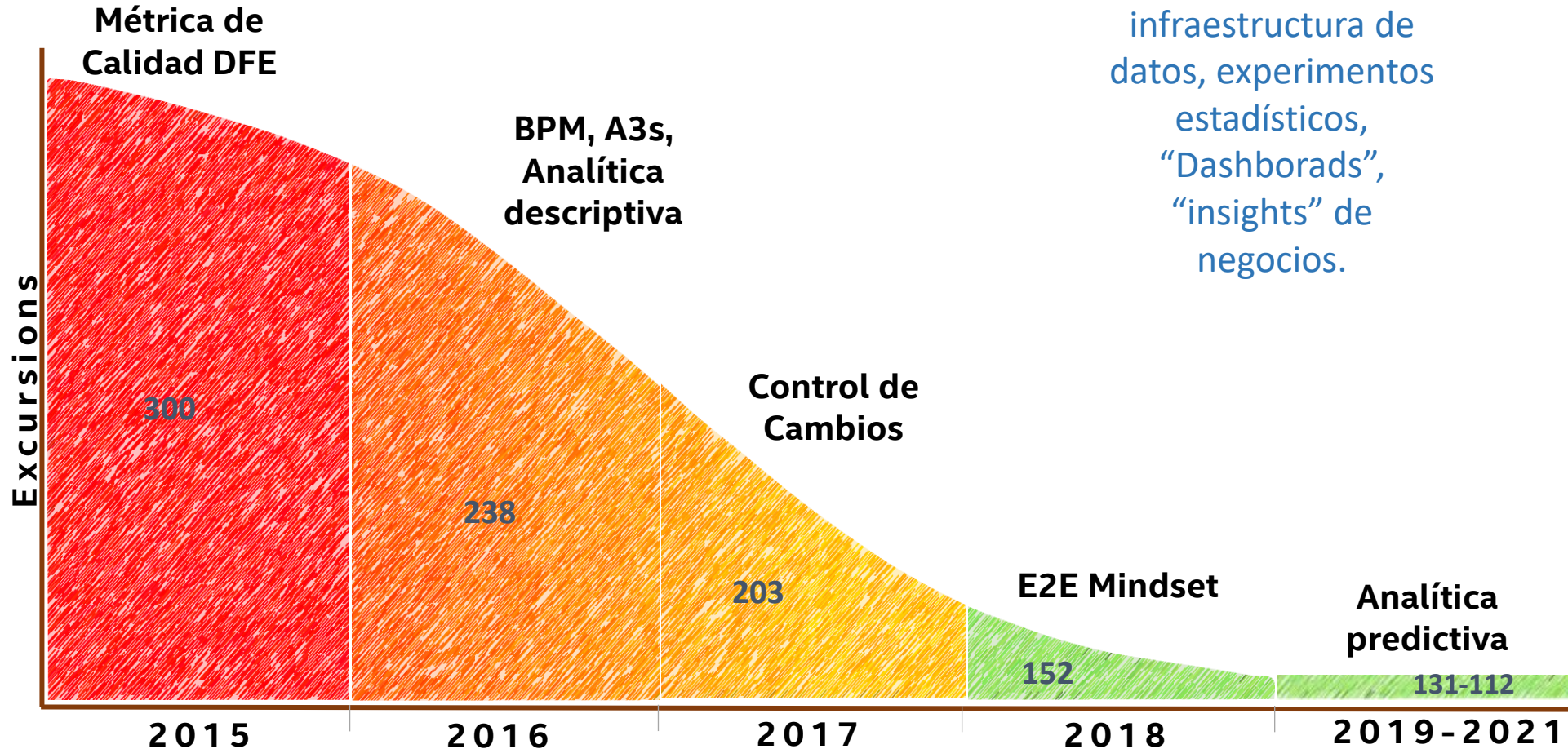


Agenda:

1. Ciencia de Datos e Inteligencia Artificial en la estrategia de calidad de Intel
2. Teoría:
 - La calidad de procesos y la ciencia de datos
 - Principales retos
3. Caso de estudio:
 - Tendencias de incidentes repetidos de calidad
 - Nuestros enfoques: Aprendizaje Automático y Procesamiento de Lenguaje Natural
4. Actividad en grupos:
 - Debate ML vs NLP
 - Votación
5. Resultados y Demo



1. Ciencia de Datos e Inteligencia Artificial como estrategia para la calidad en Intel



Desarrollo de
infraestructura de
datos, experimentos
estadísticos,
“Dashborads”,
“insights” de
negocios.



HRGS Quality Dino

2. Teoría

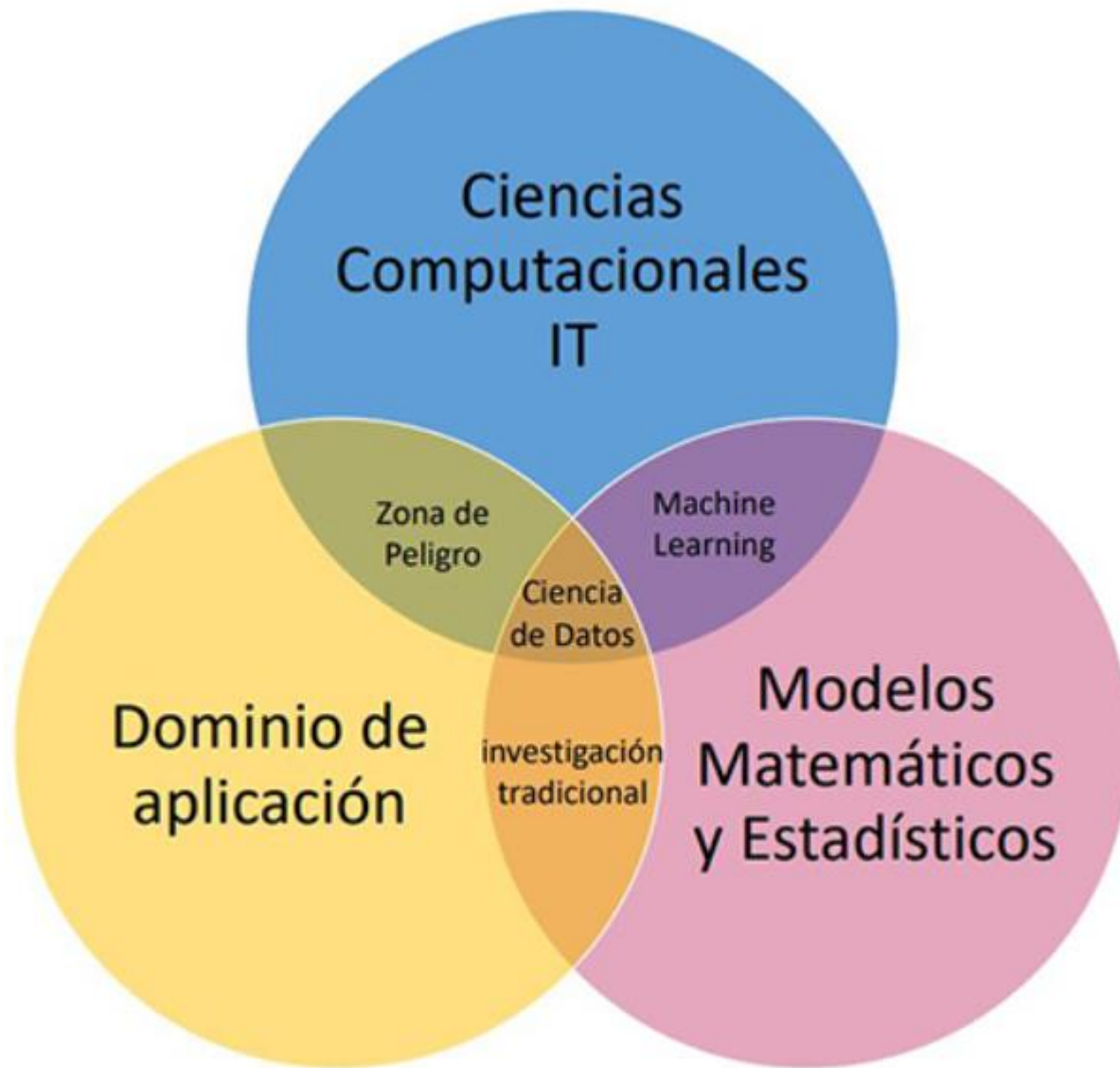
a. La calidad de procesos y la ciencia de datos

Daniela Arroyo

“Por mucho tiempo pensé que era un estadístico, interesado en inferencias desde lo particular. Pero viendo la evolución de las estadísticas matemáticas, tuve razones para dudar... y me di cuenta que mi interés central es el análisis de los datos...”

1962 - John W. Tukey
The Future of Data Analysis





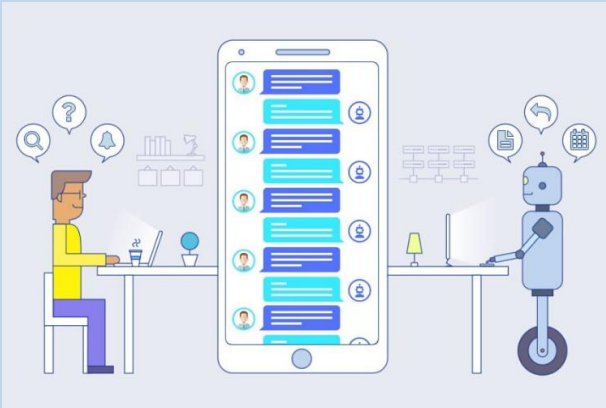
Ciencia de datos

Campo inter-disciplinario que utiliza el método científico, algoritmos y sistemas para inferir conocimiento y tomar decisiones a partir de datos estructurados y no estructurados.



Ciencia de Datos y Calidad en diferentes industrias

Farmacéutica y biomédica



- Agentes Virtuales Inteligentes
- Predicción de Demanda con Aprendizaje Automático y Profundo
- Seguimiento de participantes en ensayos clínicos
- **Mantenimiento preventivo de equipos**

Finanzas y Servicios Profesionales



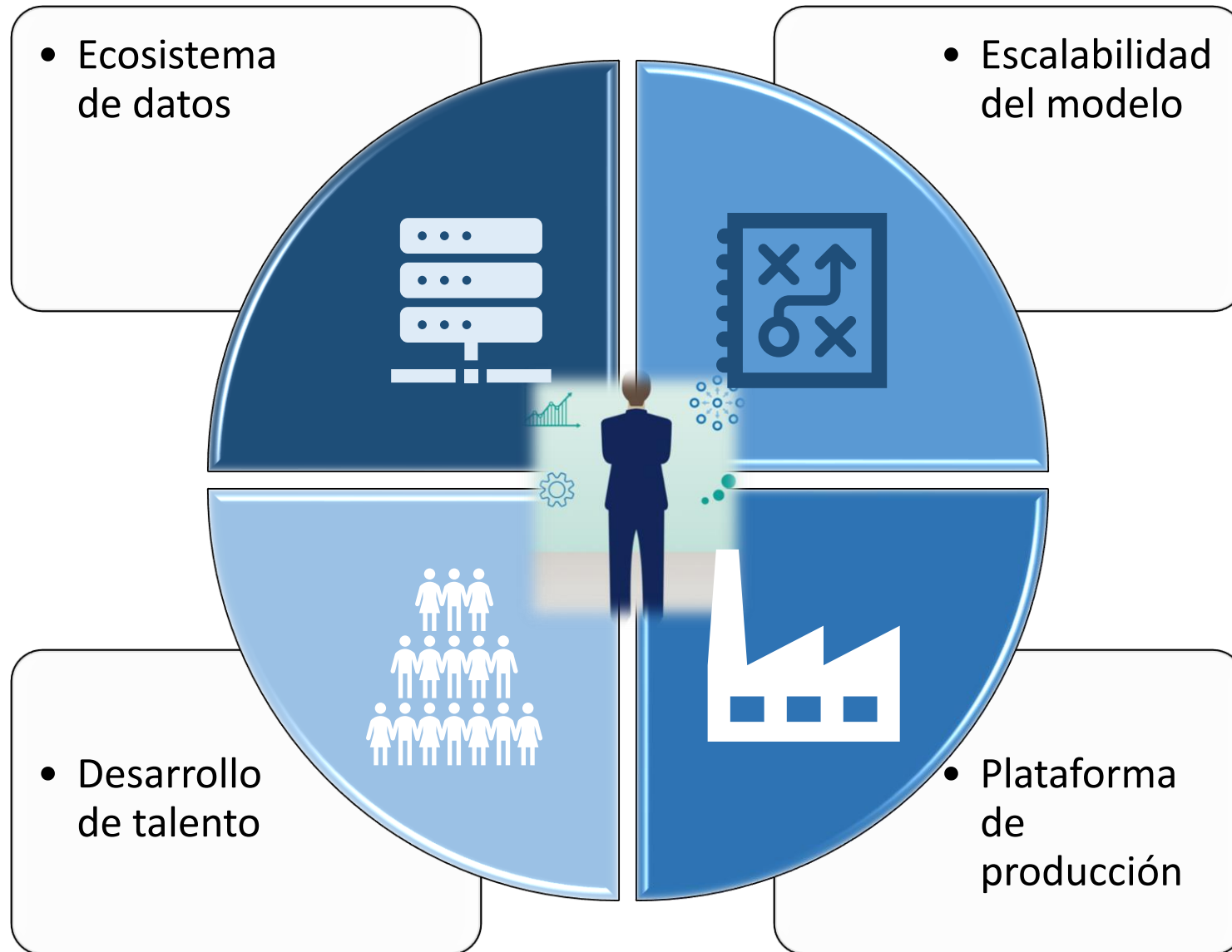
- Sistemas recomendadores de la mejor acción
- Detección de fraude
- **Clusterización automática de clientes de riesgo**
- Agentes Virtuales Inteligentes
- Marketing digital y omnicanalidad inteligente

Sector Público

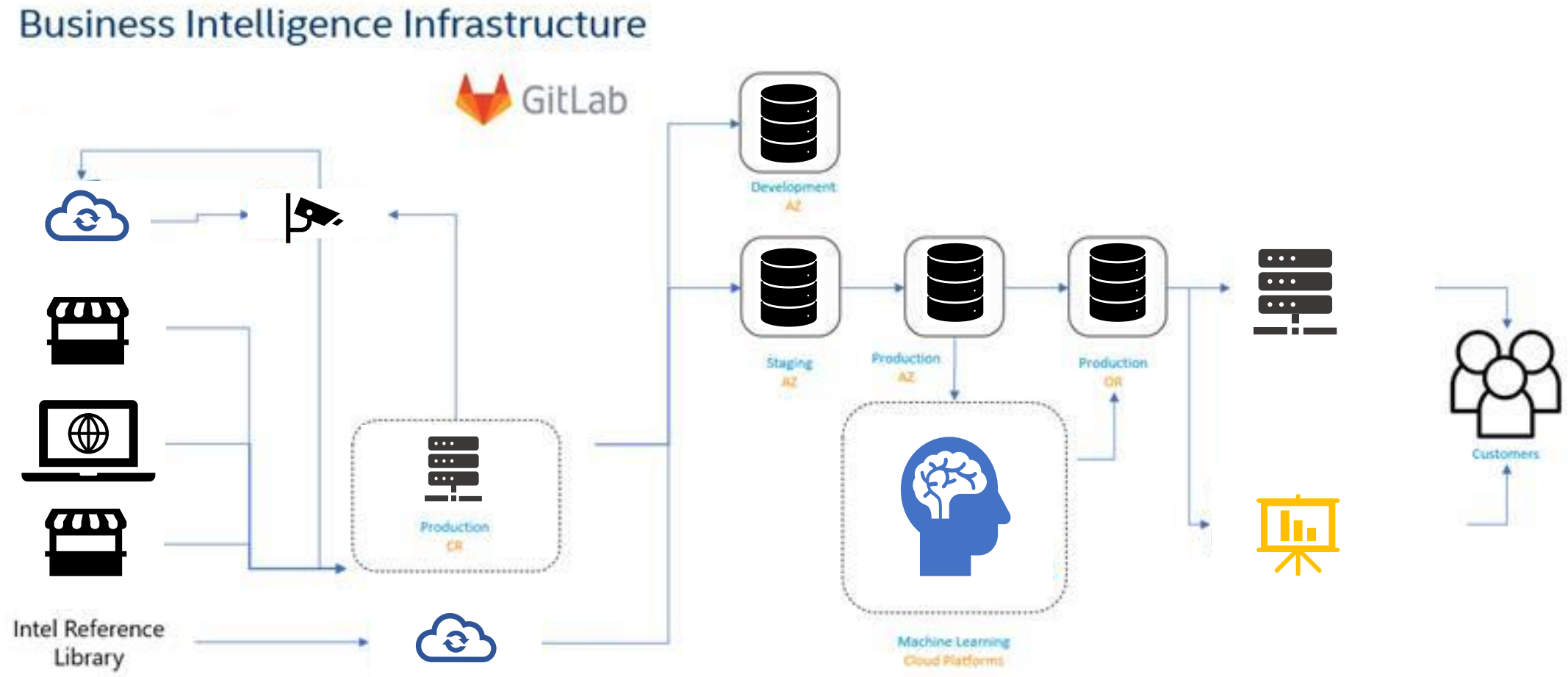


- Predicción demanda de servicios y recursos en vivo
- Predicción de enfermedades crónicas y proyección a futuro
- **Análisis de satisfacción de usuarios en evaluaciones, redes sociales, etc.**

b. Retos en la ejecución de proyectos de ciencia de datos



Ecosistemas de datos robustos facilitan el crecimiento de la ciencia de datos en su organización



3. Caso de Estudio:

Tendencias de incidentes repetidos procesos de calidad

Daniela Arroyo



Pasos para desarrollar nuestro proyecto

1. Diseño del experimento
2. Preparación y depuración de datos
3. Análisis Exploratorio
 1. Balance de clases
 2. Correlación
4. Ingeniería de variables
 1. Selección de variables
 2. Transformación de datos descriptivos (Normalización, Tokenización, Vectorización, Lematización, Eliminación de palabras no significativas)
5. Clasificación supervisada:
 1. División en muestra de entrenamiento y testeo
 2. Evaluación de algoritmos
 3. Predicción
6. Similitud por Coseno (no supervisada):
 1. Matrix de palabras vectorizadas
 2. Cálculo de TF-IDF
 3. Cálculo de Similitud por Coseno
 4. Predicción
7. Producción



4. Actividad en grupos

a. Debate ML vs NLP

Su equipo representa a la división de *Aprendizaje Automático* o bien de *Procesamiento de Lenguaje Natural* de un departamento de Ciencia de Datos.

Su trabajo es convencer al encargado de calidad de que su enfoque es el mejor para identificar incidentes recurrentes. Discuta por qué su división tiene las mejores herramientas para detectar incidentes de calidad recurrentes.

Cada equipo tiene 15 minutos para discutir sus ideas y escoger un representante.

Los líderes de cada equipo tienen 5 minutos para convencer al encargado

b. Votación

¿Cuál enfoque escogería usted si fuera el tomador de decisión?



5. Resultados y Demo



Resultados

- ✓ Procesamiento de Lenguaje Natural con Similitud de Coseno es más robusto: puede detectar nuevas categorías de similitud
- ✓ Produce algunos falsos positivos: frases similares pero incidentes diferentes. Es una buena herramienta de detección temprana, que ofrece alertas preventivas
- ✓ Casos similares son reportados en Reporte de Calidad Semanal para investigación
- ✓ Detección semanal se integra a bases de datos de incidentes. Se han encontrado 8 incidentes similares que demuestran que las causas de raíz de los incidentes deben ser revisitadas.





PATROCINADOR
ORO+

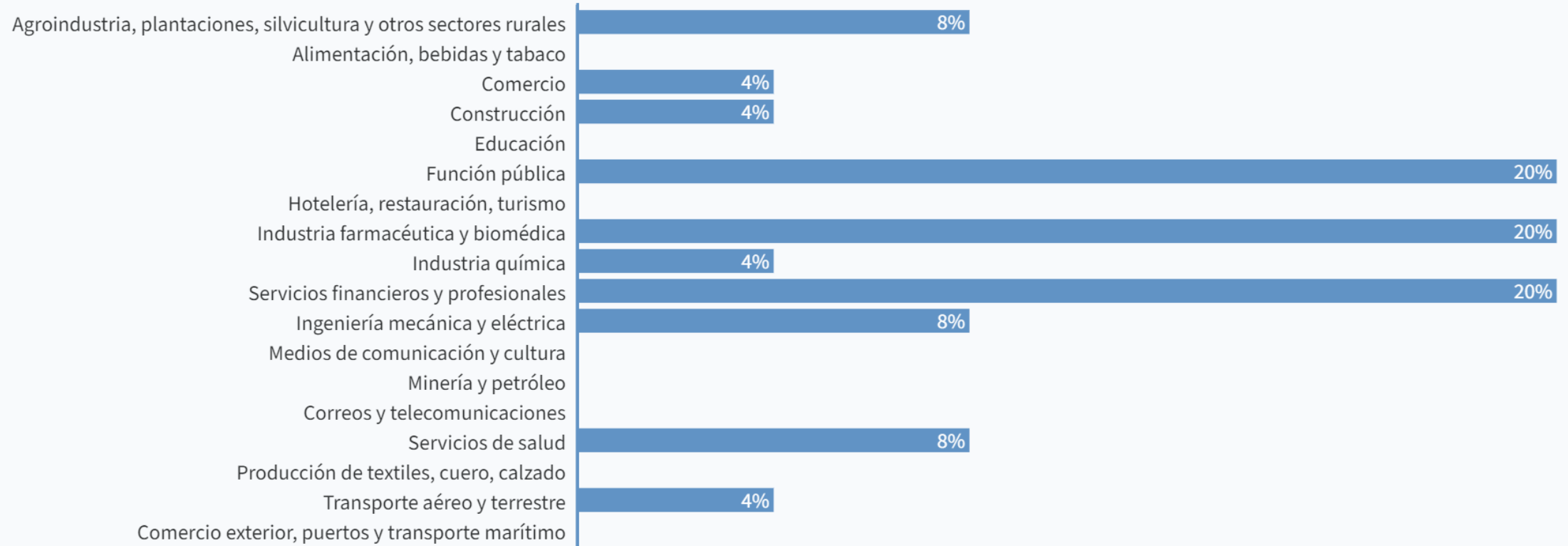


¡MUCHAS GRACIAS!

Backup

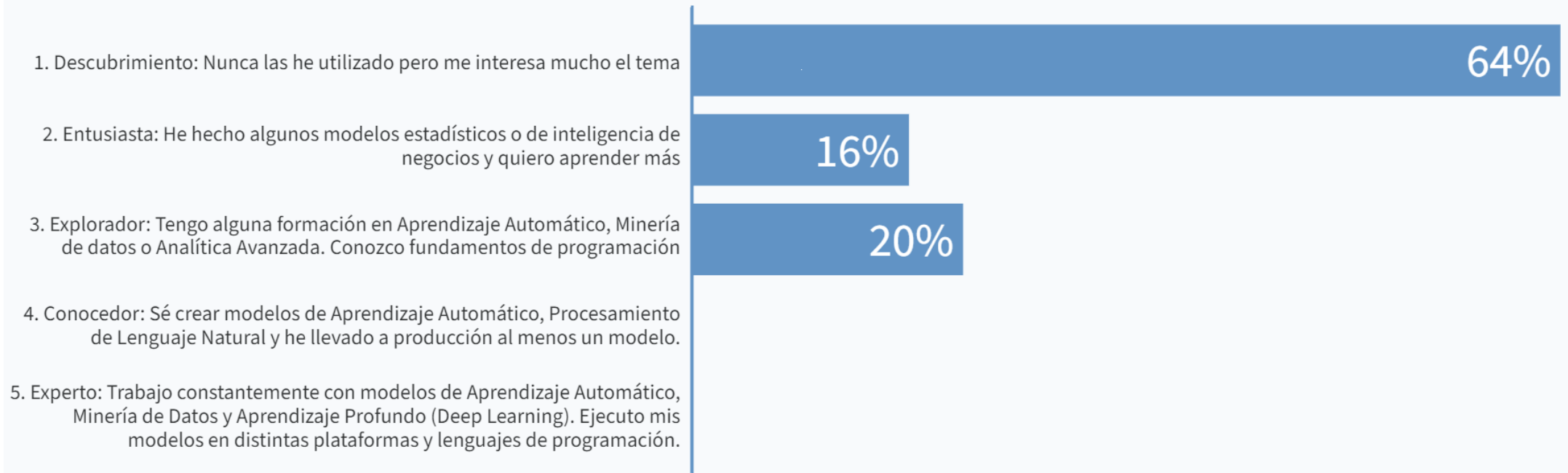


Indique el sector económico en el que usted se desempeña



Powered by  Poll Everywhere

Indique su nivel de familiaridad con técnicas de Inteligencia Artificial



Powered by  Poll Everywhere

