



Atendóni Puerta Automática



CONTEXTO

¿ Quiénes somos ?

¿QUÉ ES ADL?

ADL es un **Laboratorio Digital** que le apuesta a construir un **Ecosistema** robusto integrando a los Bancos Aval, Entidades y Aliados, a través de metodologías, capacidades y frentes de trabajo para generar una **oferta de valor diferencial** que empodere digitalmente a los usuarios.

Bancos Aval

Banco AV Villas
Banco de Bogotá
Banco de Occidente
Banco Popular

Bancos Aval

Entidades

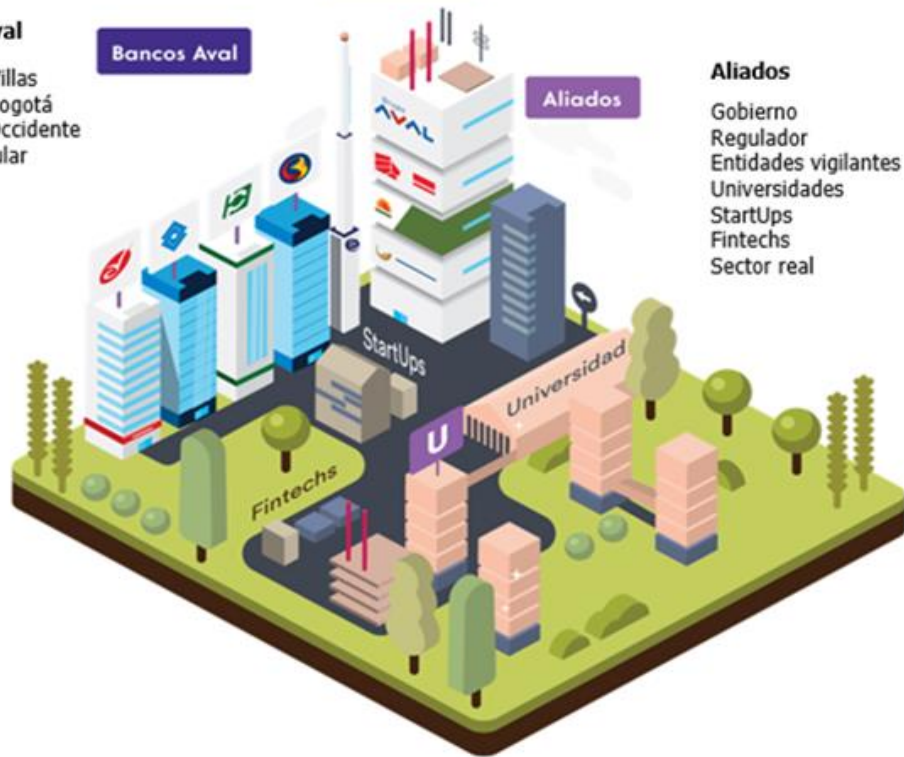
Entidades Aval y OLCSAL

Porvenir
Seguros Alfa
CEET
Entidades Grupo Aval

Aliados

Aliados

Gobierno
Regulador
Entidades vigilantes
Universidades
StartUps
Fintechs
Sector real



Analítica Avanzada en ADL





ESTRATEGIA DE DATOS

Conoce nuestro
equipo

Líderes Estrategia de Datos y Analítica



Tito Neira

Chief Strategy Data Officer



Lina Osorio

Líder Analítica Core



Monica Carvajal

Líder Analítica Digital



Sofia Orbes

Líder Centro de Excelencia Analítica
Avanzada



Liliana Betancourt

Líder Ingeniería de Datos

Equipo Analítica Avanzada ADL



20 ingenieros de datos

Que garantizan la calidad, gobernabilidad, unicidad de los datos y puesta en producción de soluciones



Centro de excelencia analítica

Alineado con políticas grupo AVAL, el COE de analítica es experto en buenas prácticas de analítica avanzada y Digital, manejo de información en nube y manejo de data



5 Equipos soporte

Devops, arquitectura, seguridad, finanzas, estrategia entre otros



13 Científicos de datos

Con experiencia en desarrollar modelos analíticos y dar solución a problemas de negocio



10 Científicos digitales e ing de implementación

Expertos en manejo de datos digitales, taggeo, digital analytics, visualización near real time y similares



Scrum Master

Que garantiza el desarrollo de la metodología ágil aplicada a la analítica



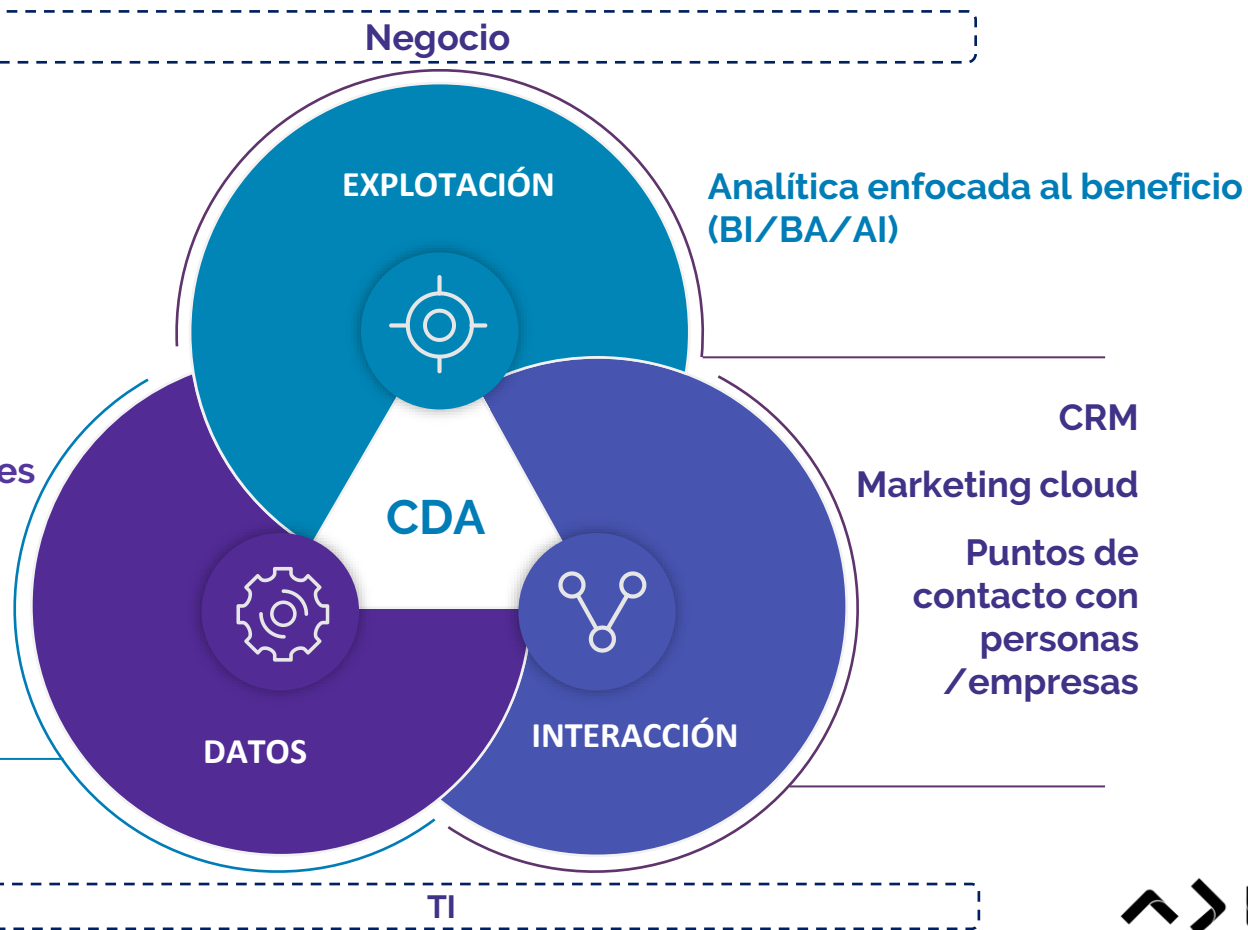
200 Usuarios de soluciones

Entre comerciales, directivos, células de digitalización, células de casos de uso

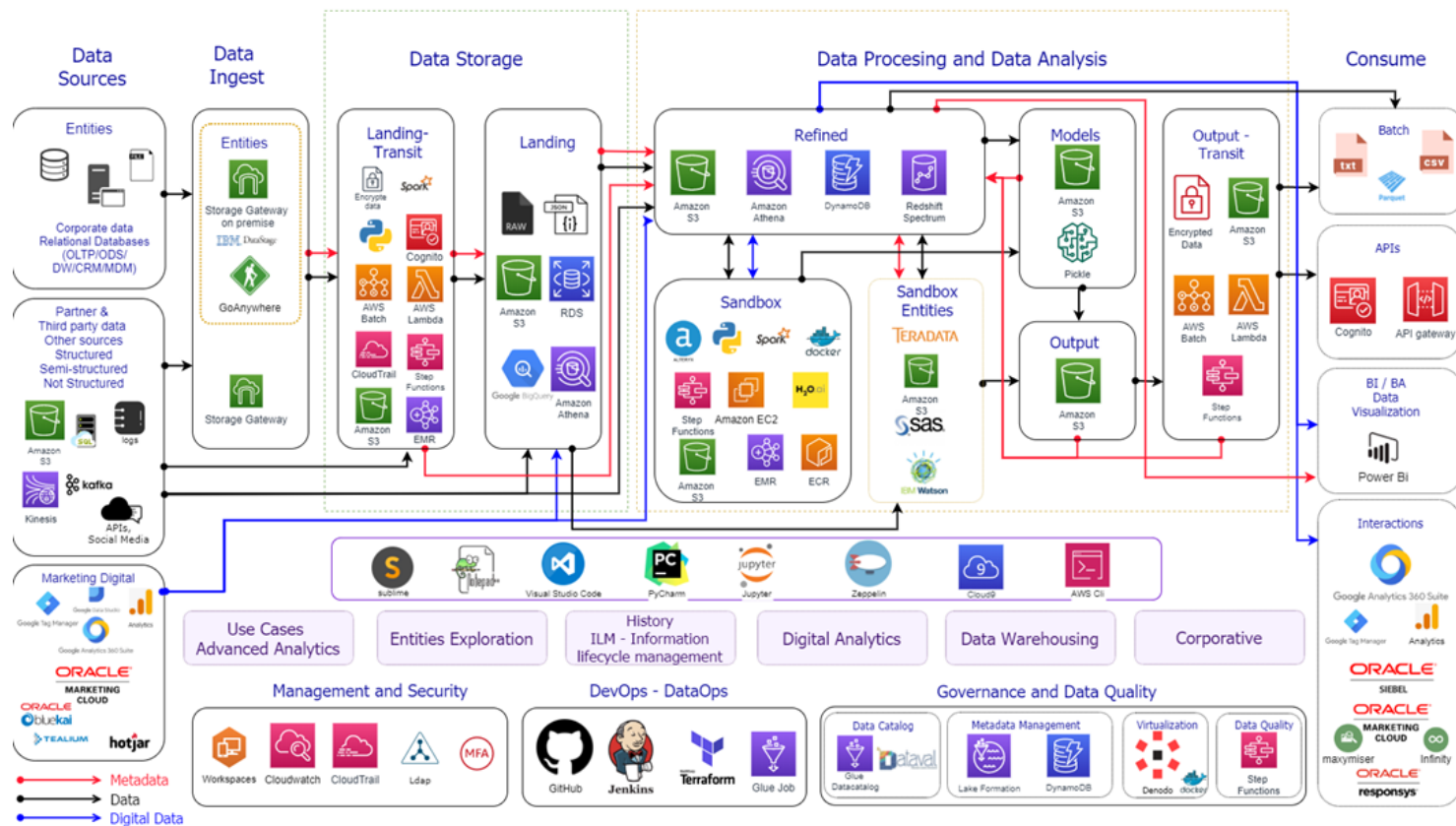
Este es nuestro modelo de trabajo ADL - Analítica

A ayudamos a extraer el valor de los datos
diseñando el camino en su transformación
digital, de manera que tomen decisiones a
partir de los insights extraídos.

Fuente única de clientes
Calidad, Legalidad,
Complejidad
Gobierno de datos
Disponibilidad



¿En dónde lo hacemos?

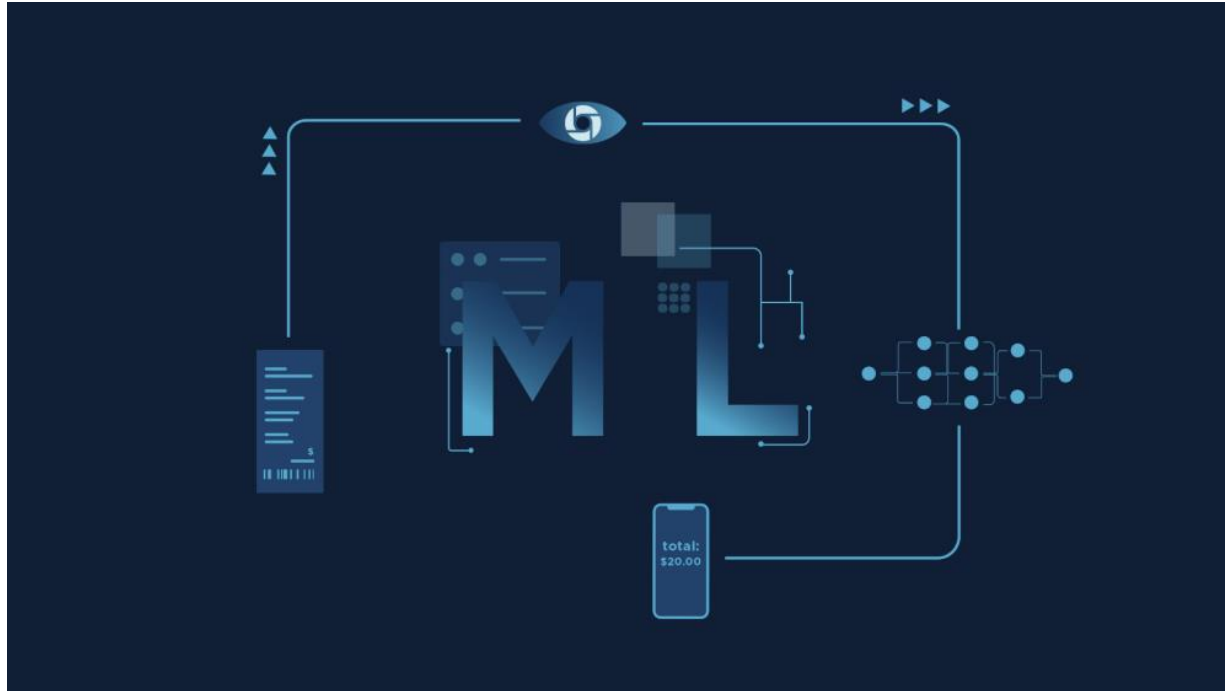


Plataforma de Datos **Augusta**

V 1.3

Machine Learning en la vida real

Introducción



¿Quiénes dictarán las sesiones?



Johanna Cardona
Científica de Datos ADL

Estadística

Universidad Nacional de Colombia

Maestría en Economía y Negocios

Universidad de Groningen - Holanda

Más de 8 años de experiencia en desarrollo y consultoría de proyectos de analítica avanzada e inteligencia de negocios.

Experiencia en sector financiero riesgo de crédito, mercadeo.



Harry Torres
Científico de Datos ADL

Lic en matemáticas

Universidad Pedagógica Nacional

Estadístico

Universidad Nacional de Colombia

Maestría en ingeniería de información

Universidad de los Andes

Maestría en tecnologías de información para el negocio

Universidad de los Andes

Experiencia en banca desarrollando modelos para las áreas de riesgo de crédito y mercadeo.

Experiencia en sistemas de recomendación, análisis de imágenes satelitales e ingeniería del conocimiento.



Guillermo Rojas
Científico de Datos ADL

Estadístico

Universidad Nacional de Colombia

Experiencia más de 5 años en desarrollo de modelos analíticos sobre lenguaje abierto Python, R y Spark. Énfasis en banca. Experiencia en modelamiento de algoritmos de refuerzo y NPL

Sesión 1

Machine Learning en la vida real

Introducción



Johanna Cardona
Científica de Datos ADL

Estadística

Universidad Nacional de Colombia

Maestría en Economía y Negocios

Universidad de Groningen - Holanda

Más de 8 años de experiencia en desarrollo y consultoría de proyectos de analítica avanzada e inteligencia de negocios.

Experiencia en sector financiero
riesgo de crédito, mercadeo.

Agenda

1

ANALÍTICA Y MACHINE LEARNING

2

MACHINE LEARNING DENTRO DE UNA ORGANIZACIÓN: ETAPAS Y RETOS

3

METODOLOGÍAS MACHINE LEARNING

4

MACHINE LEARNING: MITOS

Agenda

1

ANALÍTICA Y MACHINE LEARNING

2

MACHINE LEARNING DENTRO DE UNA
ORGANIZACIÓN: ETAPAS Y RETOS

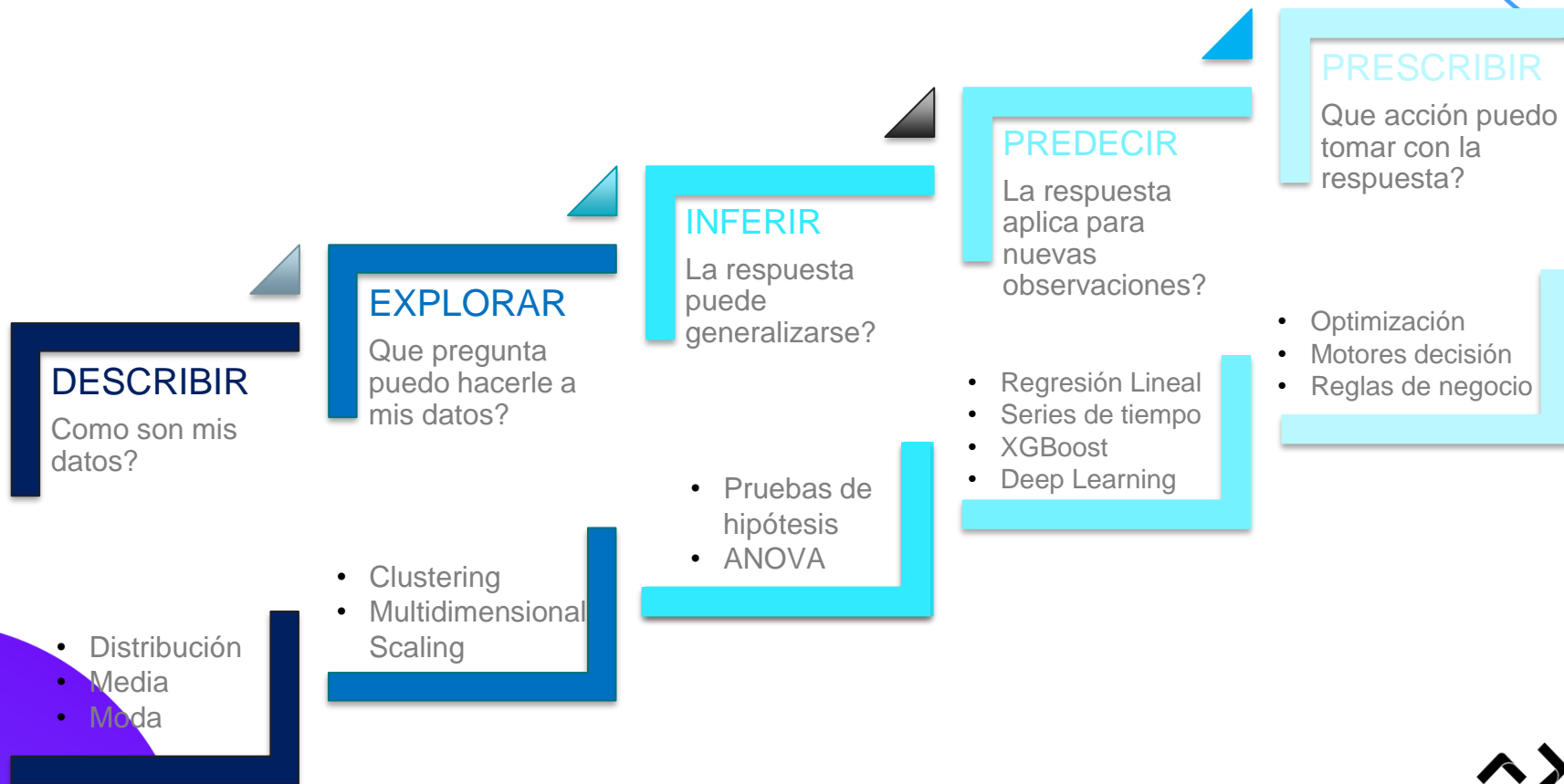
3

METODOLOGÍAS MACHINE LEARNING

4

MACHINE LEARNING: MITOS

Tipos de Analítica



Estadística



Machine Learning



Evolución de la Analítica



ANALITICA 1.0

La era del Business Intelligence

ANALITICA 2.0

La era del Big Data

ANALITICA 3.0

La era de las ofertas basadas
en data

Evolución de la Analítica



Data

Información interna:
Ventas, producción
Small Data
Data estructurada



Software

Costoso
No apto para grandes
volúmenes de datos



Almacenamiento

Data Warehouse



ANALITICA 1.0

La era del Bussines Intelligence



Decisiones

De experto basadas
en indicadores históricos
y reportes



Rol dentro de la organización

Área de reportería



Tipo de Analítica

Descriptiva y exploratoria

Evolución de la Analítica



Data

Mejor Información interna
Información externa
Big data
Data No estructurada



Software

Libre
Procesamiento paralelo
(Hadoop)



Almacenamiento

En la nube



ANALITICA 2.0
La era del Big Data



Decisiones

Informadas con data histórica
y predicciones.



Rol dentro de la organización

Fuente importante para
toma de decisiones



Tipo de Analítica

Descriptiva, exploratoria
predictiva y prescriptiva

Evolución de la Analítica



Fuente: <https://hbr.org/2013/12/analytics-30>

Agenda

1

ANALÍTICA Y MACHINE LEARNING

2

MACHINE LEARNING DENTRO DE UNA ORGANIZACIÓN: ETAPAS Y RETOS

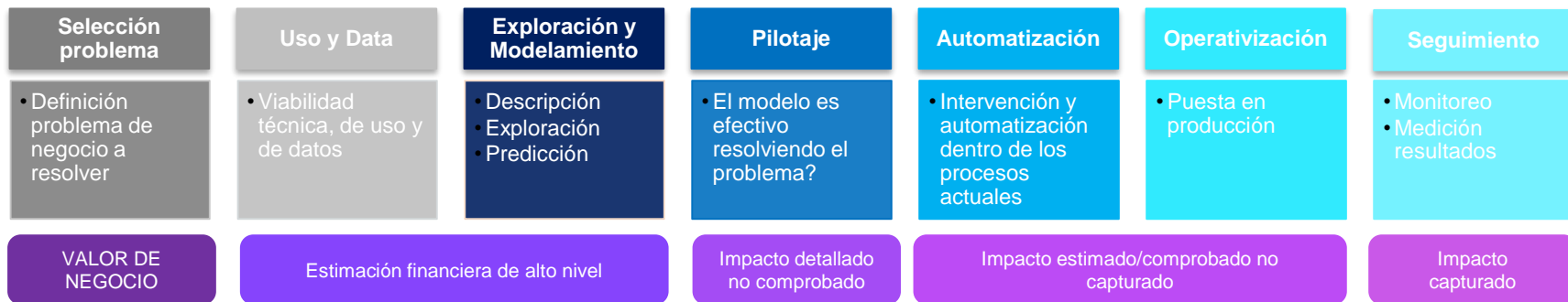
3

METODOLOGÍAS MACHINE LEARNING

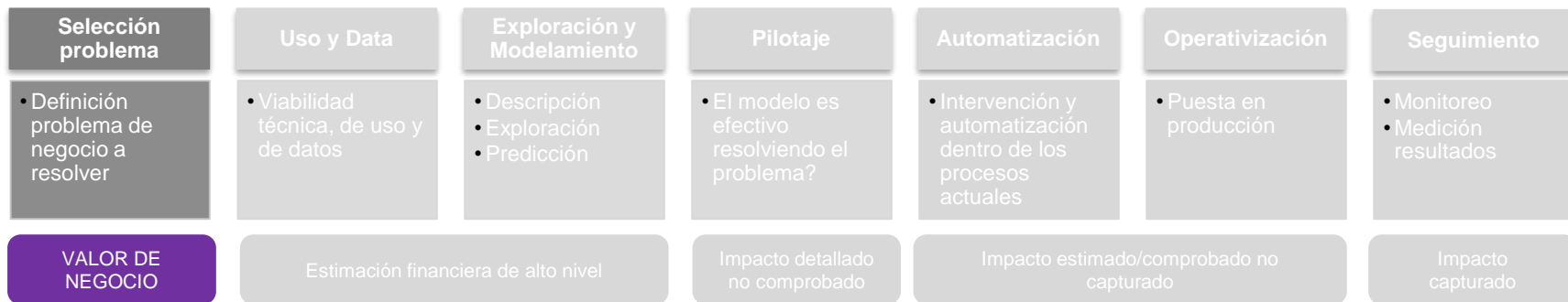
4

MACHINE LEARNING: MITOS

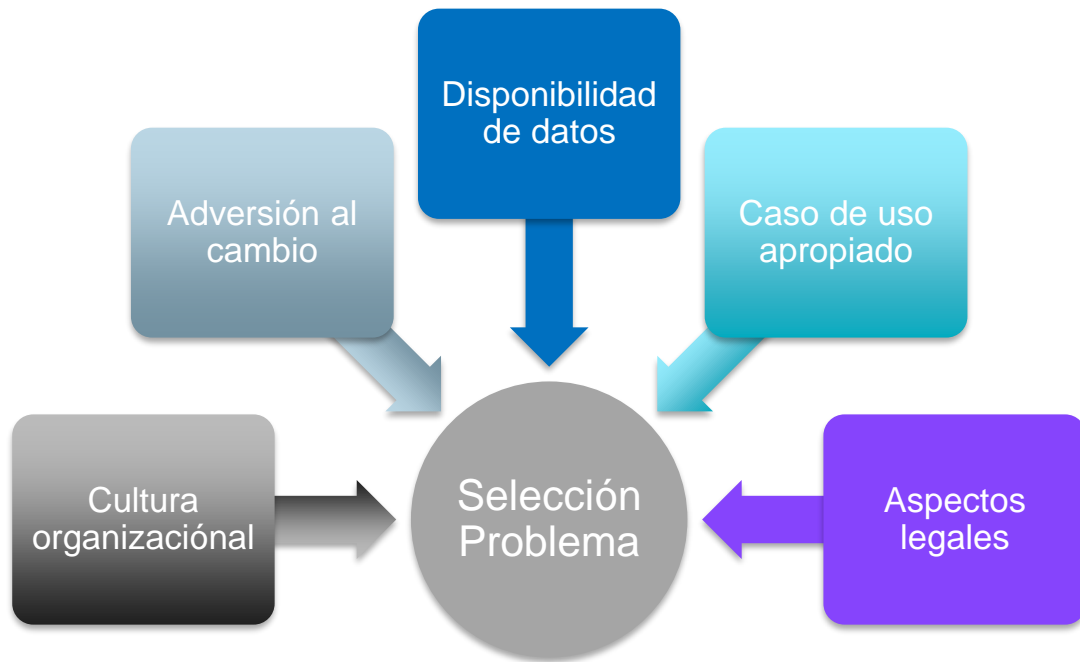
Proceso end-to-end ML dentro de una organización



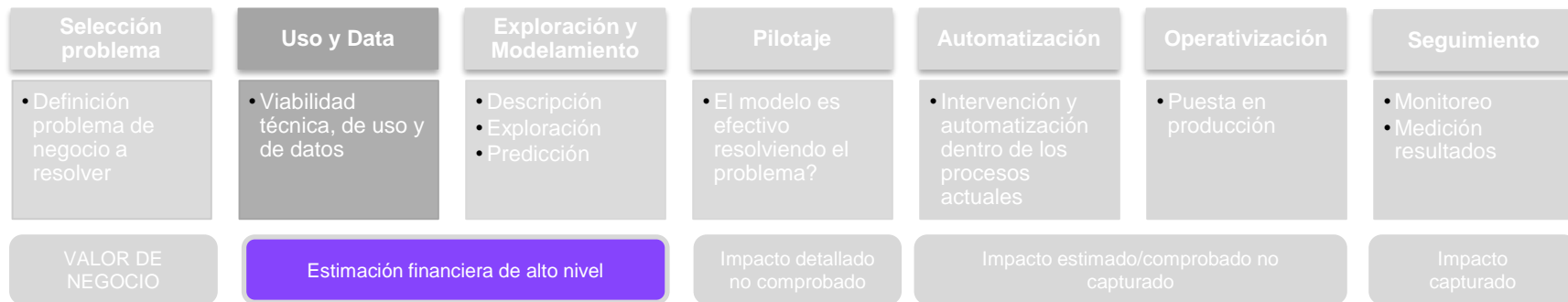
Retos : Proceso end-to-end ML



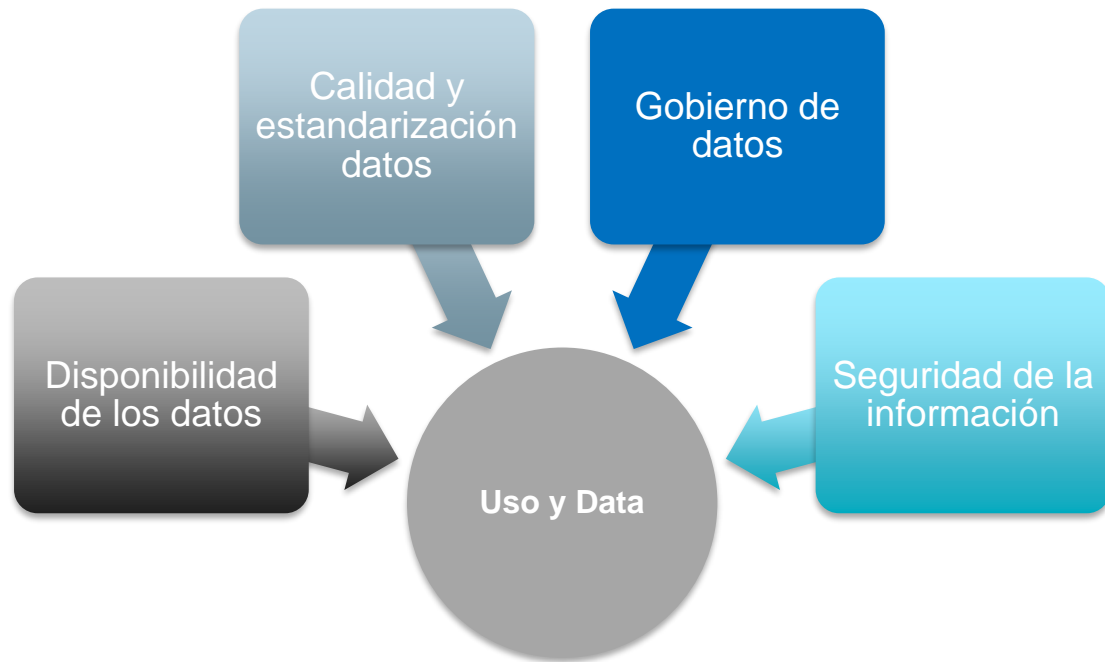
Retos : Proceso end-to-end ML



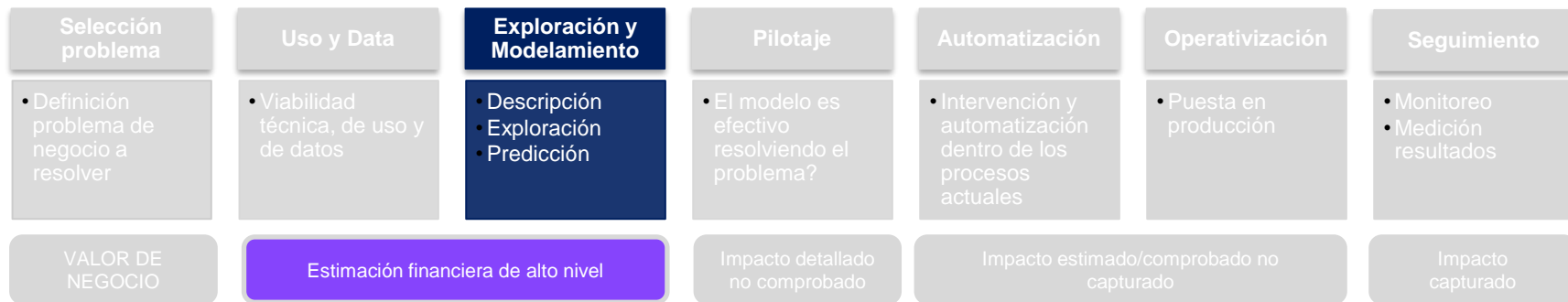
Retos : Proceso end-to-end ML



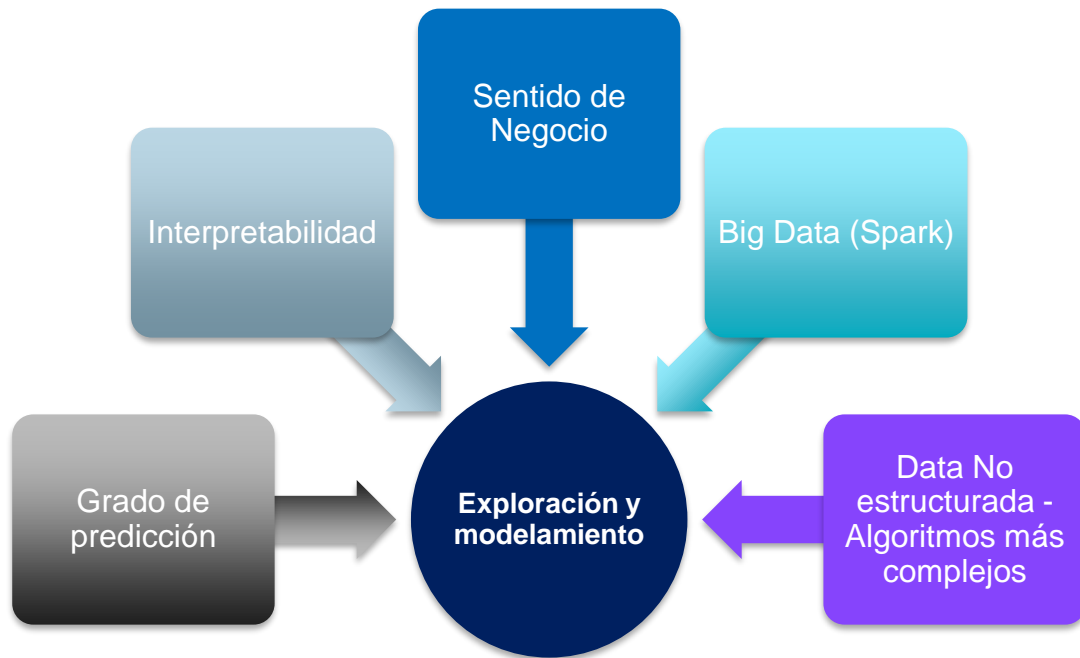
Retos : Proceso end-to-end ML



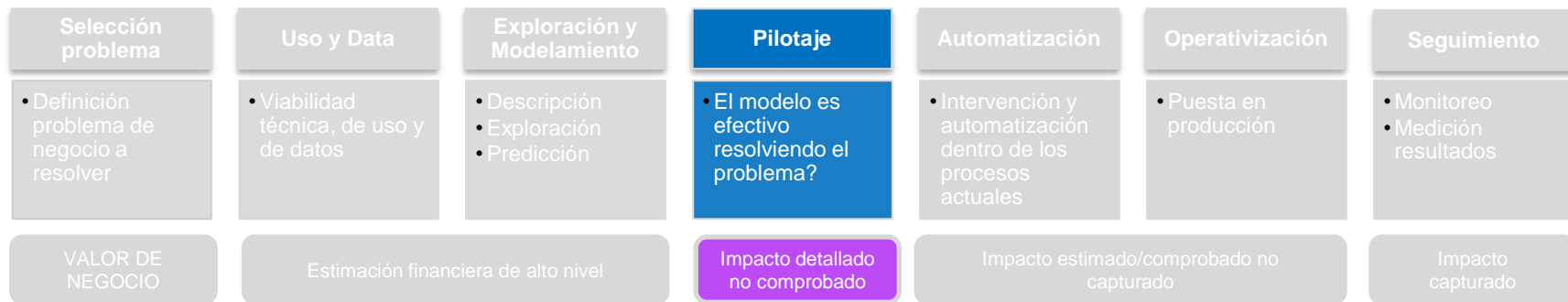
Retos : Proceso end-to-end ML



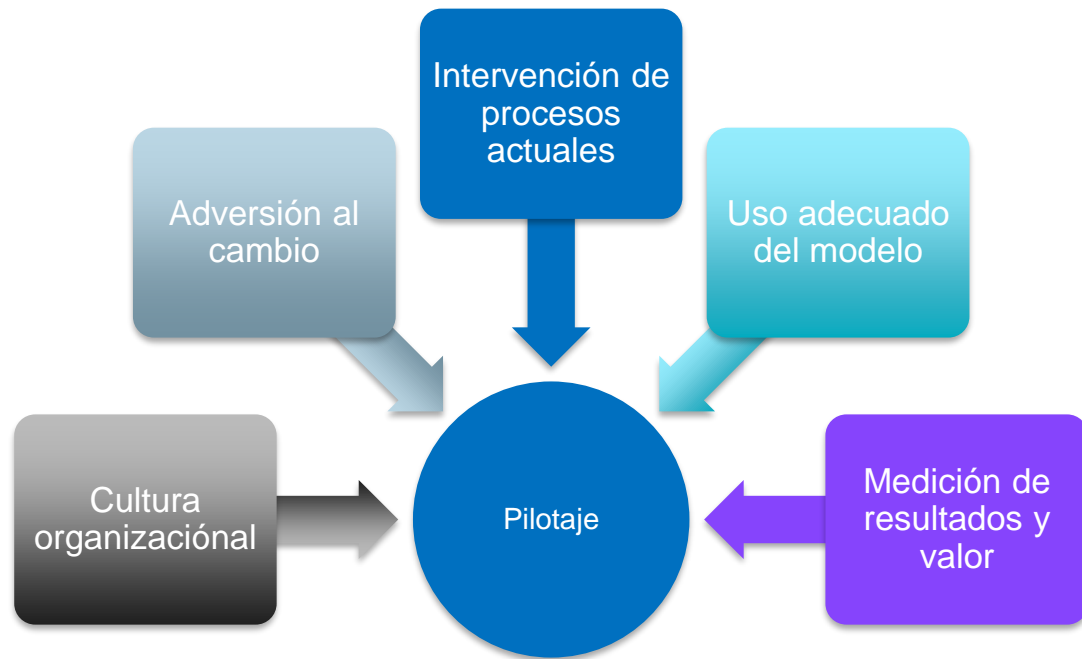
Retos : Proceso end-to-end ML



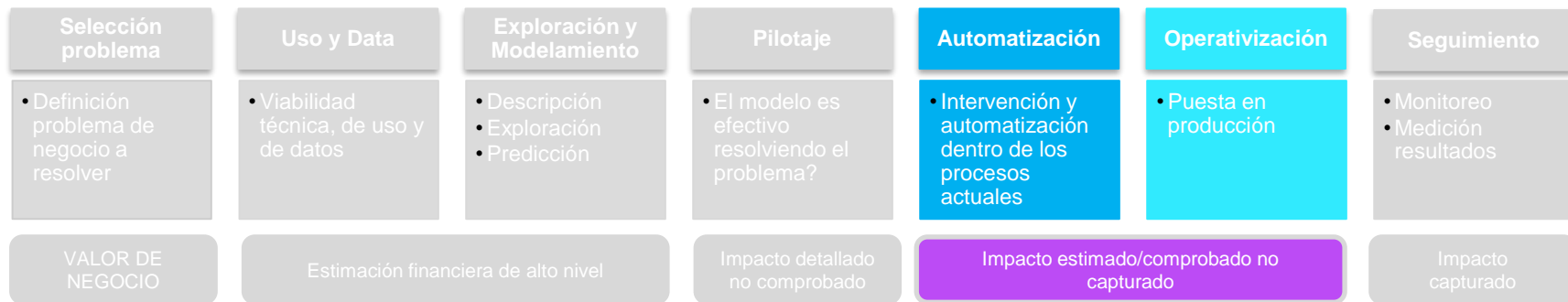
Retos : Proceso end-to-end ML



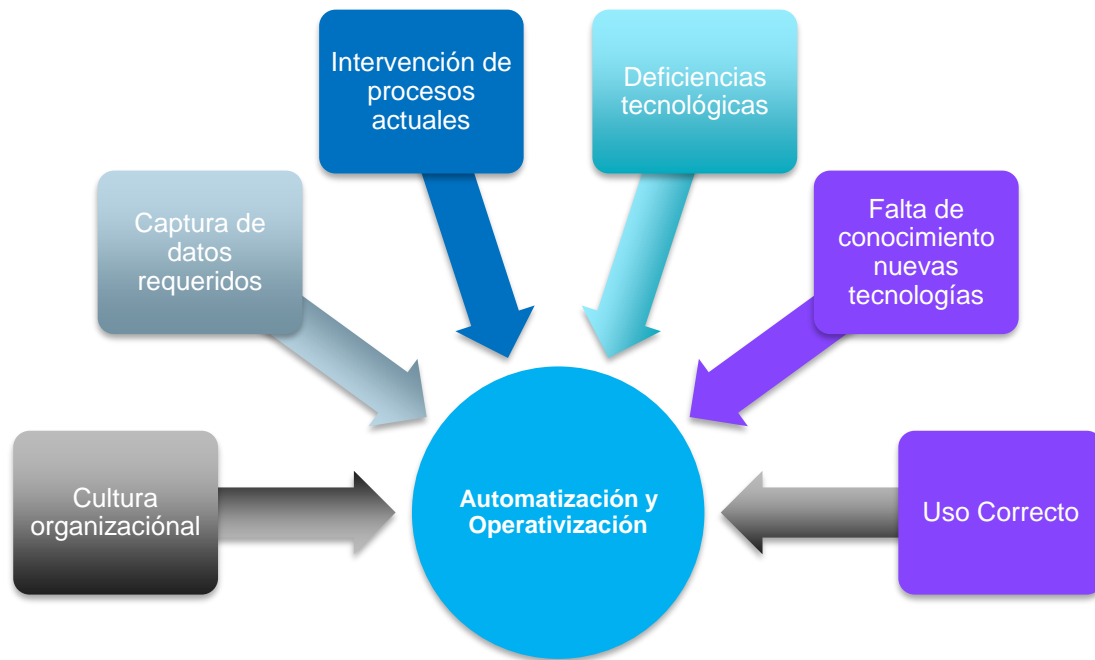
Retos : Proceso end-to-end ML



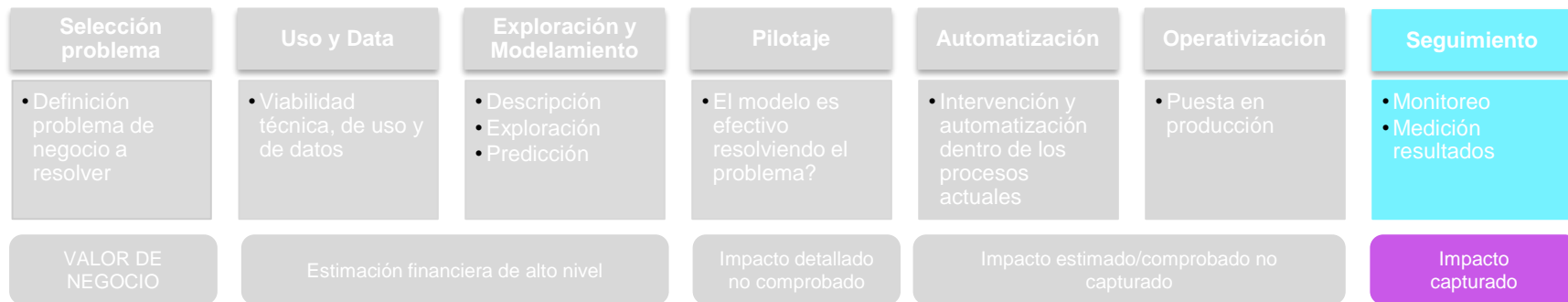
Retos : Proceso end-to-end ML



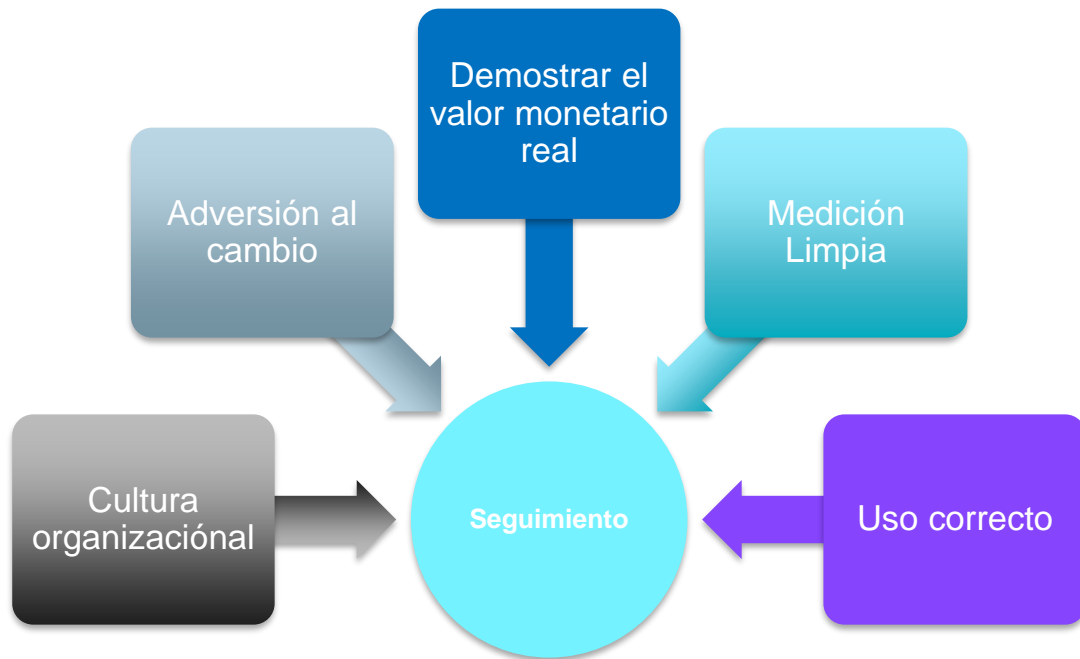
Retos : Proceso end-to-end ML



Retos : Proceso end-to-end ML



Retos : Proceso end-to-end ML



Agenda

1

ANALÍTICA Y MACHINE LEARNING

2

MACHINE LEARNING DENTRO DE UNA ORGANIZACIÓN: ETAPAS Y RETOS

3

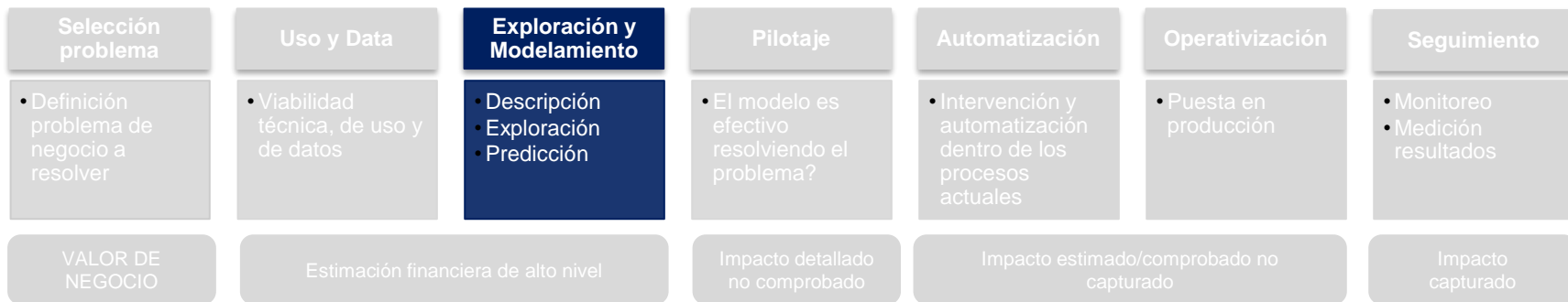
METODOLOGÍAS MACHINE LEARNING

4

MACHINE LEARNING: MITOS

Proceso end-to-end

Exploración y Modelaje



Metodologías Machine learning

Algoritmos No supervisados

K – means
Jerárquico Ward
Dbscan
Agrupamiento gaussiano
Reducción de dimensiones
(PCA, MCA)
Association rules
Anomaly detection



Algoritmos supervisados

Support vector machine
Xgboost
LightGBM
Clasificación Naive bayes
Ensemble
Árboles de decisión (C1 - CART bayesianos)
ADABOOST
Cat- boost
Reg Lineal / Logística / Bayesiana



Algoritmos Aprendizaje profundo NLP y Refuerzo

RNA (Red neuronal artificial)
Deep learning
NLP
Reinforcement learning



Para qué sirve?

Identifica perfiles
Asociaciones
Agrupar similitudes
Identifica irregularidades

Genera predicción de un fenómeno a partir de patrones de datos

Aprendizaje de comportamientos humanos/ tareas programadas/ respuestas

Los modelos son construidos en lenguaje Python/spark, sobre nube AWS y deployment en nube.

Otras metodologías Machine learning

Modelo/Algoritmo

¿Qué hace ?

¿En qué se usa?

Linear Optimization

Genera la decisión numérica óptima para un problema con restricciones dadas

Cuando se busca optimizar un recurso, ejemplo Maximizar rentabilidad de un portafolio de inversión dadas condiciones de gasto, mercado etc.

Markov Chains

Determina la probabilidad de ocurrencia de una serie de eventos dependientes

Pruebas en procesos de causación, toma de decisión en juicios legales.

Fuzzy logic model

Asigna un juicio de valor según un sistema de referencia a un evento

Se usa en problemas donde la decisión se adapta a la lógica humana, ejemplo: Sistemas lenguaje predictivo ó simulación de un criterio experto

Uplift models

Modela impactos incrementales de un tratamiento a una población

Pruebas médicas, Pruebas de marketing, Pruebas de impacto sobre procesos

Conjoint analysis models

Encuentra el valor que le dan las personas a atributos de productos

Pricing de productos, re-branding de productos, rediseños de marca

Machine Learning Algorithms Cheat Sheet

```
graph TD
    START([START]) --> UR([Unsupervised Learning: Dimension Reduction])
    START --> SL([Supervised Learning: Regression])
    START --> SC([Supervised Learning: Classification])
    START --> UL([Unsupervised Learning: Clustering])

    UR --> DR([Dimension Reduction])
    DR -- YES --> TM([Topic Modeling])
    DR -- NO --> HR([Have Responses])
    TM -- YES --> P([Probabilistic])
    P -- YES --> LDA([Latent Dirichlet Analysis])
    P -- NO --> SVD([Singular Value Decomposition])
    TM -- NO --> PCA([Principal Component Analysis])

    HR -- YES --> PN([Predicting Numeric])
    HR -- NO --> UL

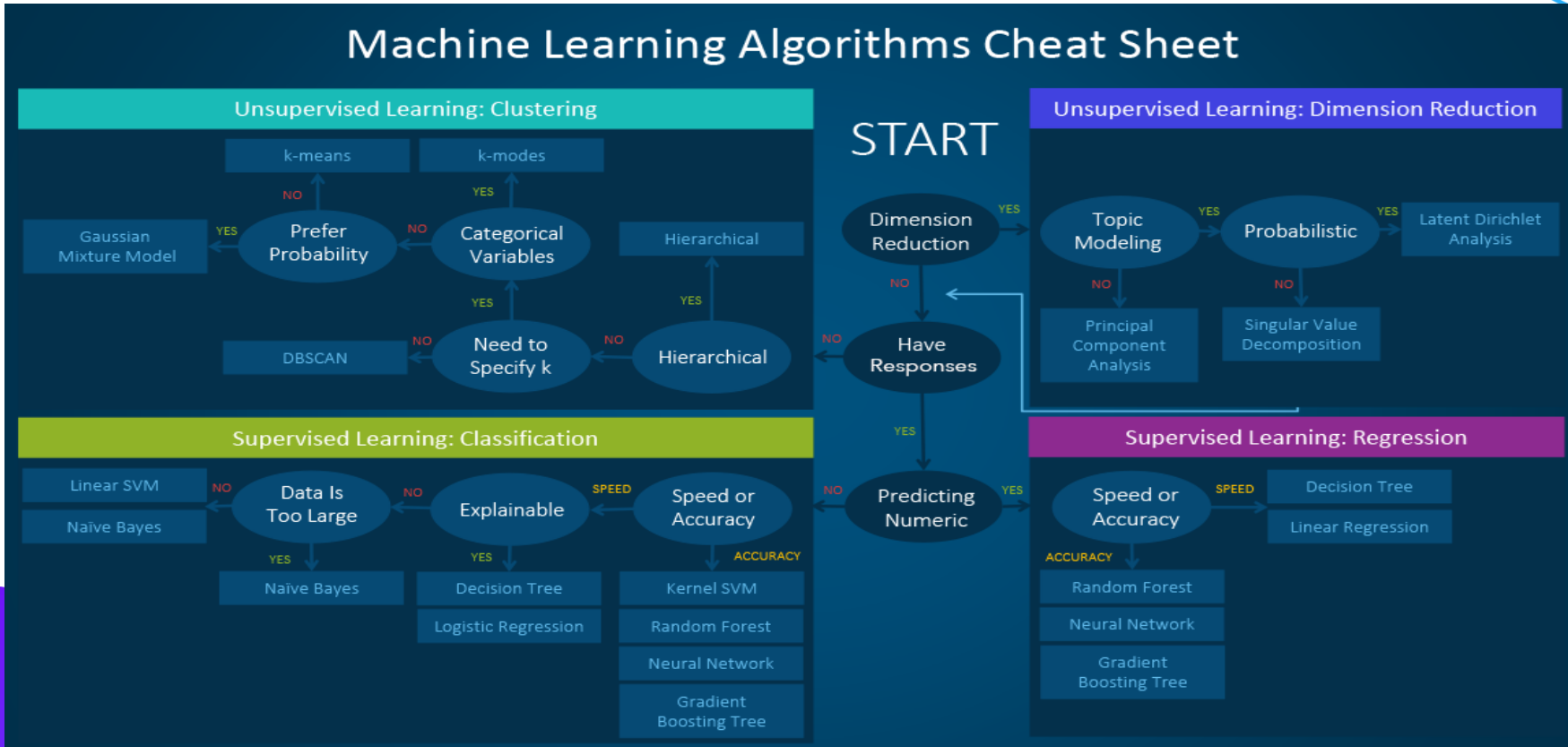
    SL --> PN
    PN -- YES --> SA([Speed or Accuracy])
    SA -- SPEED --> DT([Decision Tree])
    SA -- ACCURACY --> RF([Random Forest])
    RF --> NN([Neural Network])
    NN --> GB([Gradient Boosting Tree])

    SC --> PN
    PN -- NO --> SC
    SC --> DA([Data Is Too Large])
    DA -- YES --> NB1([Naïve Bayes])
    DA -- NO --> EX([Explainable])
    EX -- YES --> DT2([Decision Tree])
    EX -- NO --> EX
    EX -- SPEED --> SA2([Speed or Accuracy])
    SA2 -- ACCURACY --> K([Kernel SVM])
    K --> RF2([Random Forest])
    RF2 --> NN2([Neural Network])
    NN2 --> GB2([Gradient Boosting Tree])
    SA2 -- SPEED --> EX

    UL --> PCV([Prefer Probability])
    PCV -- YES --> GMM([Gaussian Mixture Model])
    PCV -- NO --> CV([Categorical Variables])
    CV -- YES --> KM([k-modes])
    CV -- NO --> PCV
    CV -- YES --> NSK([Need to Specify k])
    NSK -- YES --> H1([Hierarchical])
    H1 -- YES --> H2([Hierarchical])
    H2 -- YES --> H3([Hierarchical])
    H3 -- NO --> NSK
    NSK -- NO --> DBSCAN([DBSCAN])
    NSK -- YES --> KM
```

The cheat sheet is organized into four main sections, each with a color-coded header:

- Unsupervised Learning: Clustering** (Teal header):
 - Starts with a decision point: **Prefer Probability**.
 - If **YES**, leads to **Gaussian Mixture Model**.
 - If **NO**, leads to **Categorical Variables**.
 - Categorical Variables**:
 - If **YES**, leads to **k-modes**.
 - If **NO**, loops back to **Prefer Probability**.
 - If **YES**, leads to **Need to Specify k**.
 - Need to Specify k**:
 - If **YES**, leads to **Hierarchical**.
 - If **NO**, leads to **DBSCAN**.
 - Hierarchical**:
 - If **YES**, leads to **Hierarchical** (repeated).
 - If **NO**, loops back to **Need to Specify k**.
- Unsupervised Learning: Dimension Reduction** (Purple header):
 - Starts with **Dimension Reduction**.
 - If **YES**, leads to **Topic Modeling**.
 - If **NO**, leads to **Have Responses**.
 - Topic Modeling**:
 - If **YES**, leads to **Probabilistic**.
 - If **NO**, leads to **Principal Component Analysis**.
 - Probabilistic**:
 - If **YES**, leads to **Latent Dirichlet Analysis**.
 - If **NO**, leads to **Singular Value Decomposition**.
 - Have Responses**:
 - If **YES**, leads to **Predicting Numeric**.
 - If **NO**, loops back to **Dimension Reduction**.
- Supervised Learning: Classification** (Yellow-green header):
 - Starts with **Predicting Numeric**.
 - If **YES**, leads to **Speed or Accuracy**.
 - If **NO**, loops back to **Predicting Numeric**.
 - Speed or Accuracy**:
 - If **ACCURACY**, leads to **Kernel SVM**, **Random Forest**, **Neural Network**, and **Gradient Boosting Tree**.
 - If **SPEED**, leads to **Explainable**.
 - Explainable**:
 - If **YES**, leads to **Decision Tree** and **Logistic Regression**.
 - If **NO**, leads to **Data Is Too Large**.
 - Data Is Too Large**:
 - If **YES**, leads to **Naïve Bayes**.
 - If **NO**, loops back to **Explainable**.
- Supervised Learning: Regression** (Pink header):
 - Starts with **Speed or Accuracy**.
 - If **SPEED**, leads to **Decision Tree** and **Linear Regression**.
 - If **ACCURACY**, leads to **Random Forest**, **Neural Network**, and **Gradient Boosting Tree**.



Agenda

1

ANALÍTICA Y MACHINE LEARNING

2

MACHINE LEARNING DENTRO DE UNA
ORGANIZACIÓN: ETAPAS Y RETOS

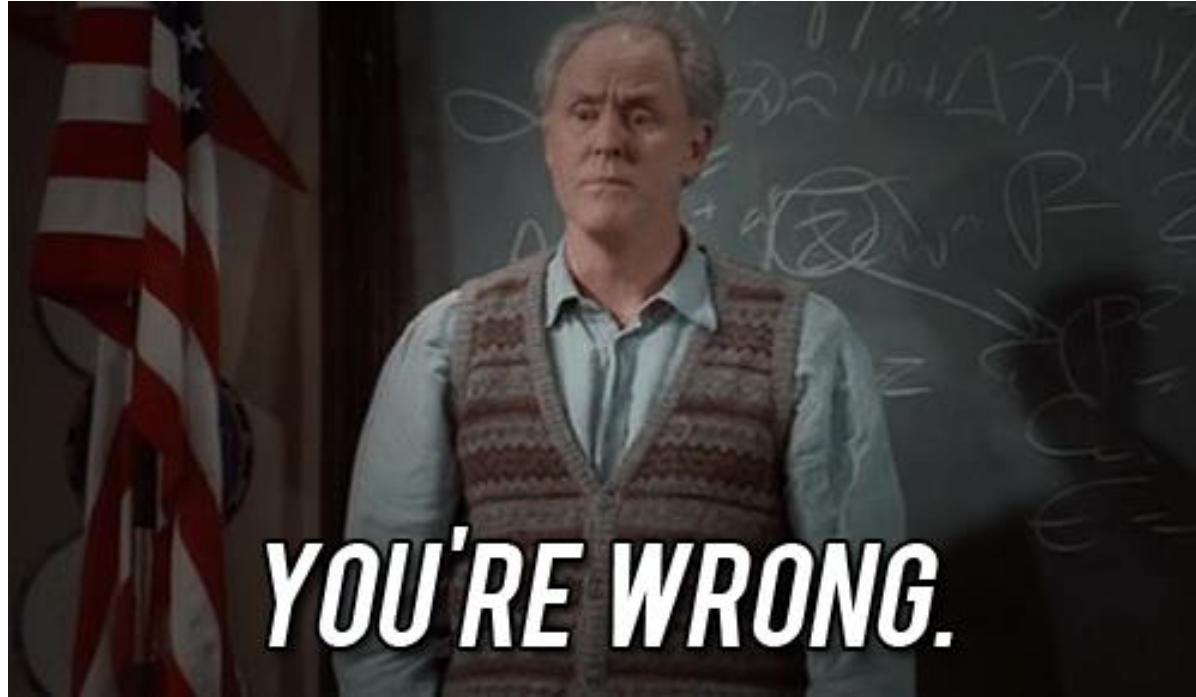
3

METODOLOGÍAS MACHINE LEARNING

4

MACHINE LEARNING: MITOS

ML: Mitos



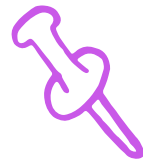
ML: Mitos

$$E=mc^2$$

ML siempre debe ser interpretable

$$1+1=2$$

El 1 a 1 siempre tiene sentido



No requiere intervención de distintas áreas



ML es perfecto y nunca se equivoca



Solo es útil en industrias como la robótica



Solo es útil cuando se tiene un gran volumen de datos



Si hay datos siempre se puede predecir



ML es una varita mágica

Tarea próxima sesión

- 1** Identifique por lo menos una necesidad dentro de su organización que pueda ser resuelta a través de un modelo de machine learning.
- 2** ¿Por qué tendría alto impacto para el negocio?
- 3** ¿Qué retos o impedimentos tendría?
- 4** ¿Qué mitos sería necesario romper para poder implementar el modelo?
- 5** ¿Desarrollaría un modelo supervisado o no supervisado y por qué?

https://docs.google.com/forms/d/e/1FAIpQLScGINBiTK3McnKizORkkc_3u7M_JXyhDp8M618a9_U4jVSVpA/viewform?usp=sf_link

¿Qué veremos en las próximas sesiones?



Harry Torres
Científico de Datos ADL

Sesión 2
Machine Learning en la
vida real
Modelos Supervisados



Guillermo Rojas
Científico de Datos ADL

Sesión 3
Machine Learning en la
vida real
Modelos No Supervisados

Gracias ;)