



# FEATURE

---

ENGINEERING

Miriam Lanabere

---

# **SOBRE MÍ**

1. Licenciada en Comercio Internacional
2. Científica de Datos
3. Fundadora de Comex Cerca
4. Seminario de Big Data y Comercio Internacional
5. Seminario Bases de Datos para Comercio Exterior
6. Analista de datos



/miriamlanabere/



**01**

---

## **PROCESO DATA SCIENCE**

Pasos y objetivos del  
proceso de ciencia  
de datos

**02**

---

## **FEATURE ENGINEERING**

Importancia y  
técnicas para  
creación de features

**03**

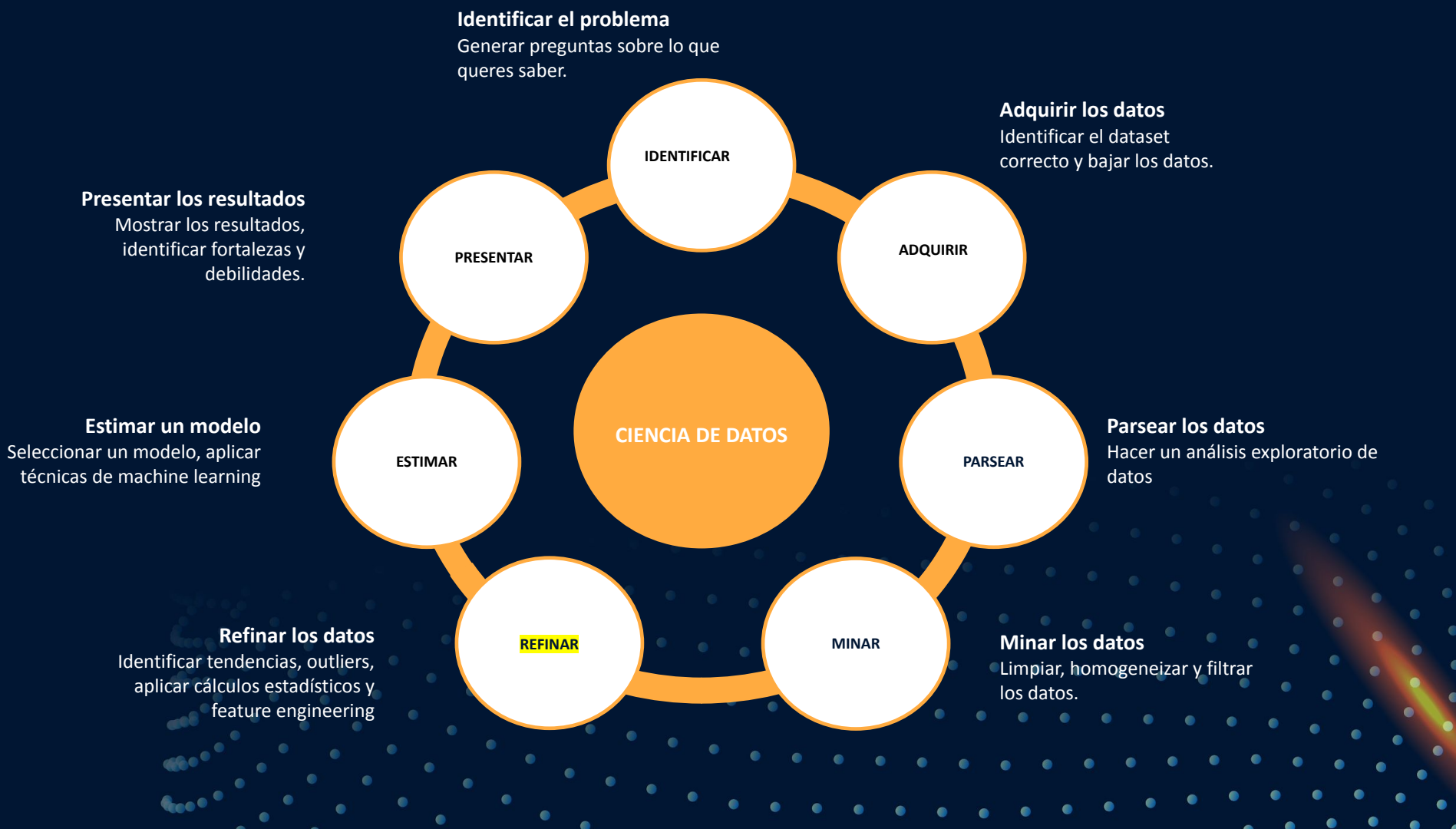
---

## **FEATURE SELECTION**

Métodos y criterios  
de selección de  
features



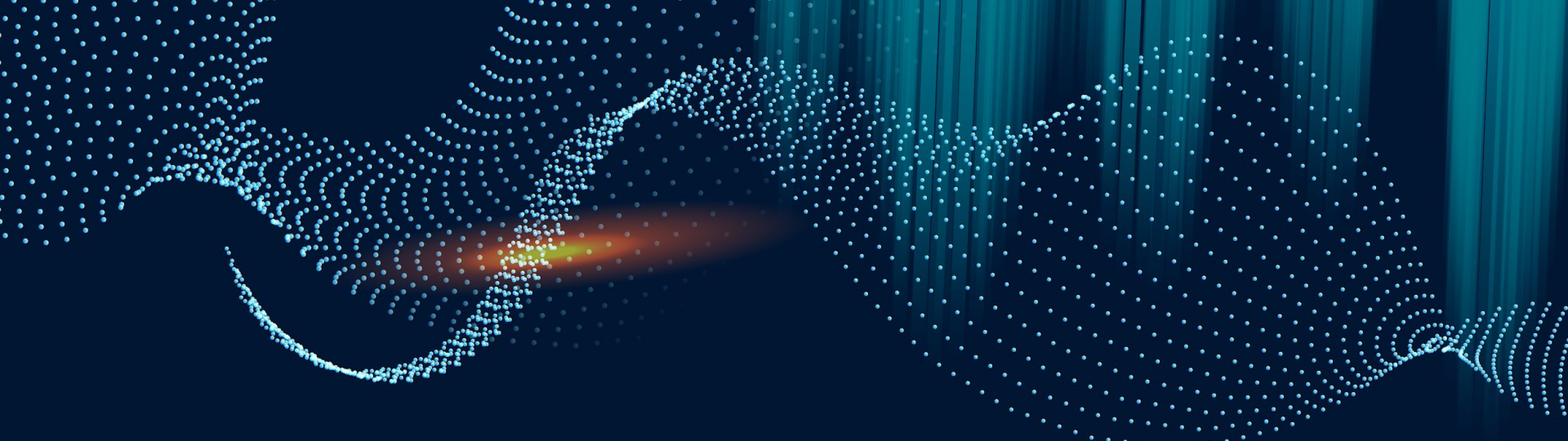
# **01** | **PROCESO DE DS**





**“El Data Scientist le  
dedica más tiempo al  
procesamiento de los  
datos que a la aplicación  
del modelo”**

**El proceso de “Refinar los datos”  
itera constantemente sobre el  
resultado de los modelos para  
encontrar la mejor combinación de  
variables.**



# **02** | **FEATURE ENGINEERING**



---

# DESCRIPCIÓN

1. Es el proceso de creación de nuevas features a partir de datos sin procesar, que combinados con nuestro conocimiento, ayudan a robustecer el modelo de machine learning.
2. Consiste en capturar información adicional que no es fácilmente detectada en el dataset original.
3. Este proceso se aplica primero para generar características adicionales y luego se realiza el proceso de Feature Selection para eliminar características irrelevantes, redundantes o altamente correlacionadas.



---

## OBJETIVO

1. Mejorar la performance del los modelos predictivos
2. Mejorar la interpretabilidad de los resultados

---

# FEATURES

1. Numérica
2. Volumen
3. Relacionadas con el objetivo

---

# HERRAMIENTAS

1. Conocimiento del negocio
2. Estadísticas
3. Librerías de Python

---

# TIPO DE ANÁLISIS

1. Imputation
2. Date-Time Decomposition
3. Binning
4. Numeric Transformation
5. Categorical Encoding
6. Text mining





# **03** | **FEATURE SELECTION**

---

## DESCRIPCIÓN

Es el proceso de detectar las features que más se relacionan con la variable Target a predecir, y en consecuencia, eliminar aquellas que menos contribuyen al modelo.



---

## **BENEFICIOS**

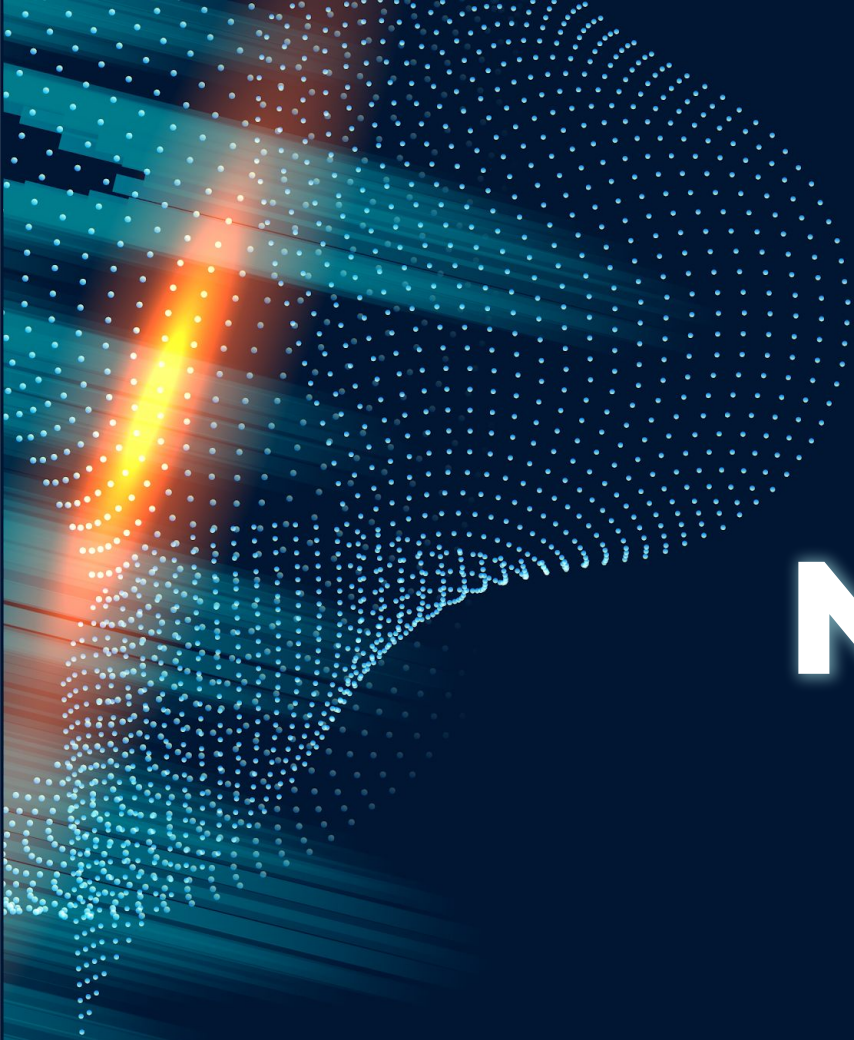
- 1. Reduce el overfitting**
- 2. Mejora el accuracy**
- 3. Reduce el tiempo de procesamiento**

---

# HERRAMIENTAS

1. SelectKBest
2. Feature Importance (Tree Model Classifier)
3. Heatmap





**NOTEBOOK!**



# GRACIAS!

---

---

CREDITS: This presentation template was created by Slidesgo, including icons by Flaticon, and infographics & images by Freepik.

