

# 01 – Entendimiento del Negocio



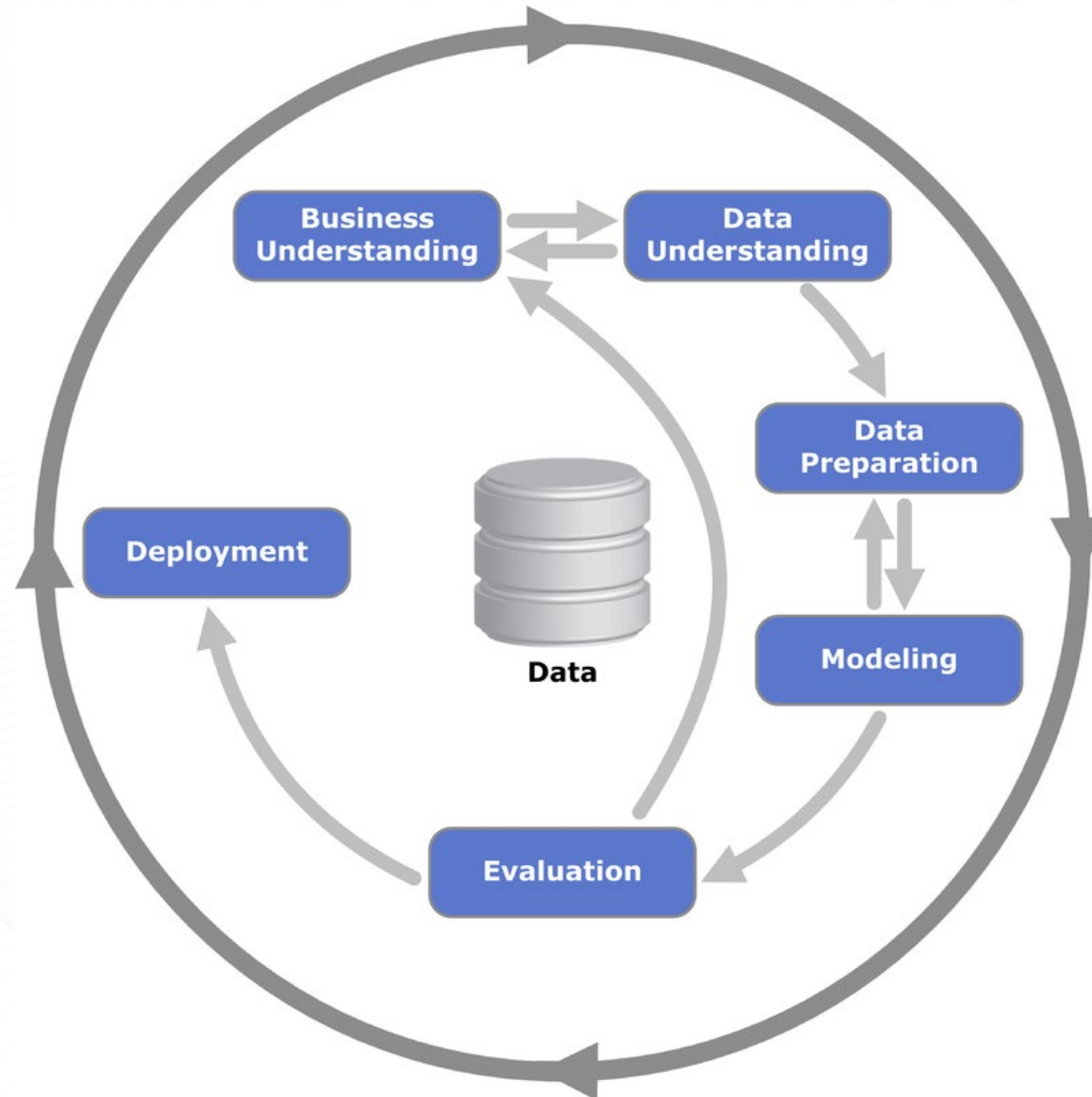
Centro de Servicios y Gestión Empresarial  
SENA Regional Antioquia



[www.sena.edu.co](http://www.sena.edu.co)

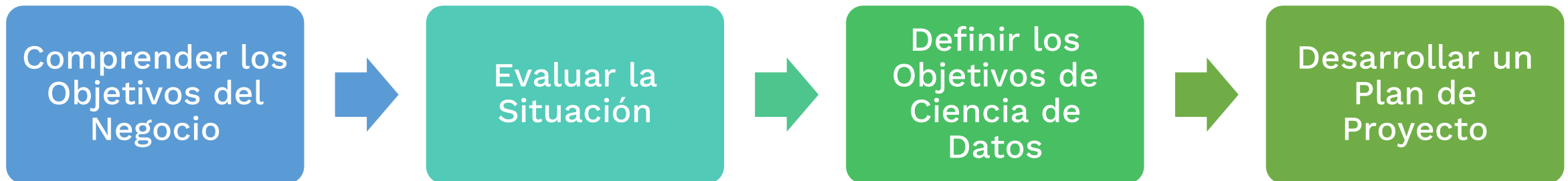
# Entendimiento del Negocio

# Entendimiento del Negocio



# Entendimiento del Negocio

El entendimiento del negocio es la fase inicial en la metodología CRISP-DM, donde se busca comprender los objetivos y requisitos del negocio desde una perspectiva de alto nivel.



# Entendimiento del Negocio

## 1. Comprender los Objetivos del Negocio:

- Identificar las metas y objetivos clave del negocio.
- Definir los problemas que se quieren resolver mediante el análisis de datos.

## 2. Evaluar la Situación:

- Entender el contexto y el entorno del negocio.
- Identificar los recursos disponibles, las restricciones y los factores que podrían influir en el proyecto.

# Entendimiento del Negocio

## **3. Definir los Objetivos de Ciencia de Datos:**

- Traducir los objetivos del negocio en objetivos específicos de ciencia de datos.
- Definir claramente lo que se espera lograr con el análisis de datos.

## **4. Desarrollar un Plan de Proyecto:**

- Crear un plan detallado para el proyecto de ciencia de datos.
- Establecer los pasos a seguir, los recursos necesarios y los plazos.



# Contexto Medico

## “Hepatitis”

# Entidades de Salud - Enfermedades hepáticas

El sector salud enfrenta el reto de diagnosticar oportunamente enfermedades hepáticas como la hepatitis, la cual afecta el hígado y puede tener desenlaces graves si no se detecta a tiempo. Los profesionales médicos suelen depender de análisis clínicos y criterios subjetivos, lo que puede dificultar la identificación temprana de pacientes en riesgo.

Por lo anterior, surge la necesidad de apoyar la toma de decisiones médicas mediante técnicas de ciencia de datos, capaces de analizar variables clínicas y biológicas (edad, bilirrubina, albúmina, enzimas hepáticas, entre otras) para predecir la probabilidad de supervivencia o mortalidad de un paciente diagnosticado con hepatitis.



# Fase 01 - *Entendimiento del Negocio*

# Entendimiento del Negocio

## **1. Comprender los Objetivos del Negocio:**

Desarrollar un modelo predictivo que permita estimar el riesgo de mortalidad o supervivencia en pacientes con diagnóstico de hepatitis, con el fin de mejorar el proceso de diagnóstico y priorización de tratamientos médicos.

### **Objetivos específicos del negocio:**

- ✓ Identificar los factores clínicos más relevantes asociados a la supervivencia.
- ✓ Facilitar la toma de decisiones médicas mediante un sistema de predicción.
- ✓ Reducir la incertidumbre diagnóstica y apoyar la asignación de recursos clínicos.

# Entendimiento del Negocio

## 2. Evaluar la Situación:

- **Actores involucrados:** médicos hepatólogos, analistas clínicos, investigadores en salud y autoridades sanitarias.
- **Datos disponibles:** registros de pacientes, variables médicas.
- **Recursos:** dataset abierto de la UCI, herramientas Python (pandas, scikit-learn, pycaret).
- **Restricciones:** tamaño limitado del dataset y posible desbalance entre clases (supervivientes vs. fallecidos).
- **Factores de éxito:** interpretabilidad del modelo, exactitud en las predicciones y facilidad de integración en entornos médicos.

# Entendimiento del Negocio

## 3. Definir los Objetivos de Ciencia de Datos:

1. Construir un modelo de clasificación supervisada que prediga la supervivencia del paciente.
2. Determinar las variables con mayor peso en la predicción (feature importance).
3. Evaluar el desempeño del modelo mediante métricas como accuracy, recall, precision y F1-score.
4. Presentar resultados en un tablero de visualización (dashboard) comprensible para personal médico.

# 4. Desarrollar un Plan de Proyecto:

Etapa	Descripción	Herramientas	Tiempos
<b>1. Recolección y comprensión de datos</b>	Cargar el dataset hepatitis desde ucimlrepo y analizar su estructura.	pandas, numpy	3 semanas
<b>2. Preparación de datos</b>	Limpieza, codificación de variables categóricas, tratamiento de valores nulos y balanceo de clases.	pandas, imblearn	1 semana
<b>3. Modelado predictivo</b>	Entrenamiento y comparación de modelos (Logistic Regression, Random Forest, XGBoost).	scikit-learn, pycaret	2 semanas
<b>4. Evaluación</b>	Validar métricas, curvas ROC y matriz de confusión.	matplotlib, seaborn	1 semana
<b>5. Comunicación de resultados</b>	Presentar conclusiones, recomendaciones y visualizaciones interpretables.	Power BI, Plotly Dash o Streamlit	1 semana





# GRACIAS

Presentó: Alvaro Pérez Niño  
Instructor Técnico

Correo: [aperezn@misena.edu.co](mailto:aperezn@misena.edu.co)

<http://centrodeserviciosygestionempresarial.blogspot.com/>

Línea de atención al ciudadano: 01 8000 910270

Línea de atención al empresario: 01 8000 910682



@SENAComunica

[www.sena.edu.co](http://www.sena.edu.co)