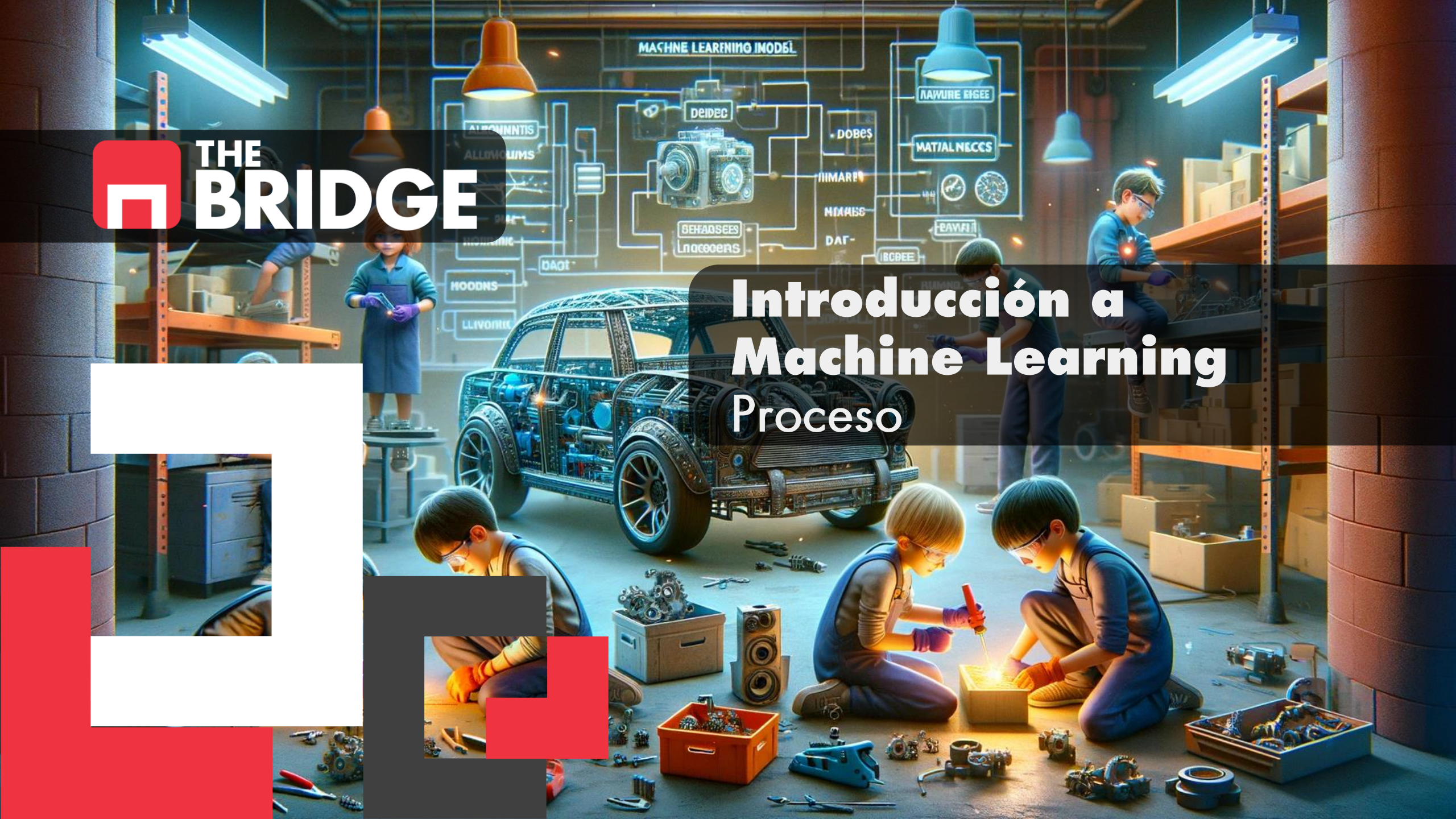


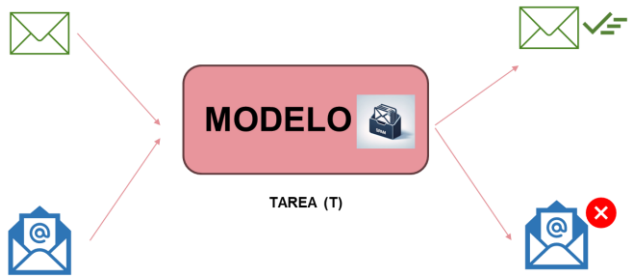
THE BRIDGE

Introducción a Machine Learning

Proceso

Modelos: Recordatorio

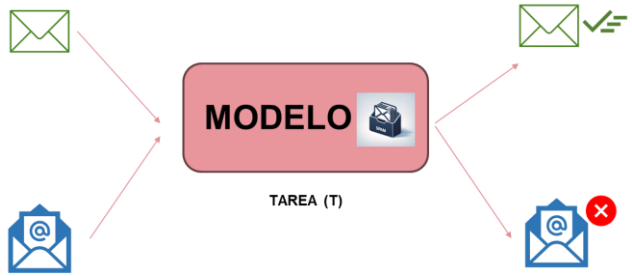
- Conceptualmente: Un constructo matemático que pone en relación unas variables de entrada (features) con una variable de salida (target)



Modelos: Recordatorio

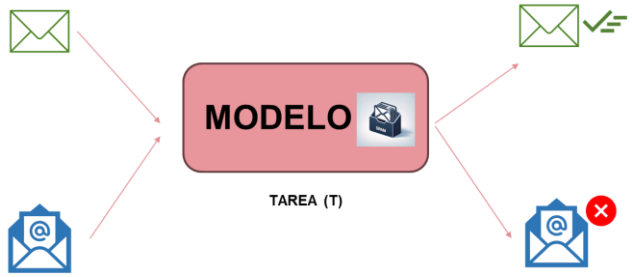
- Conceptualmente: Un constructo matemático que pone en relación unas variables de entrada (features) con una variable de salida (target)

- Físicamente: Una pieza de código que permite el entrenamiento (ajustando unos parámetros internos) y la predicción (que dada unas features de entrada genera una valor para la variable de salida target)



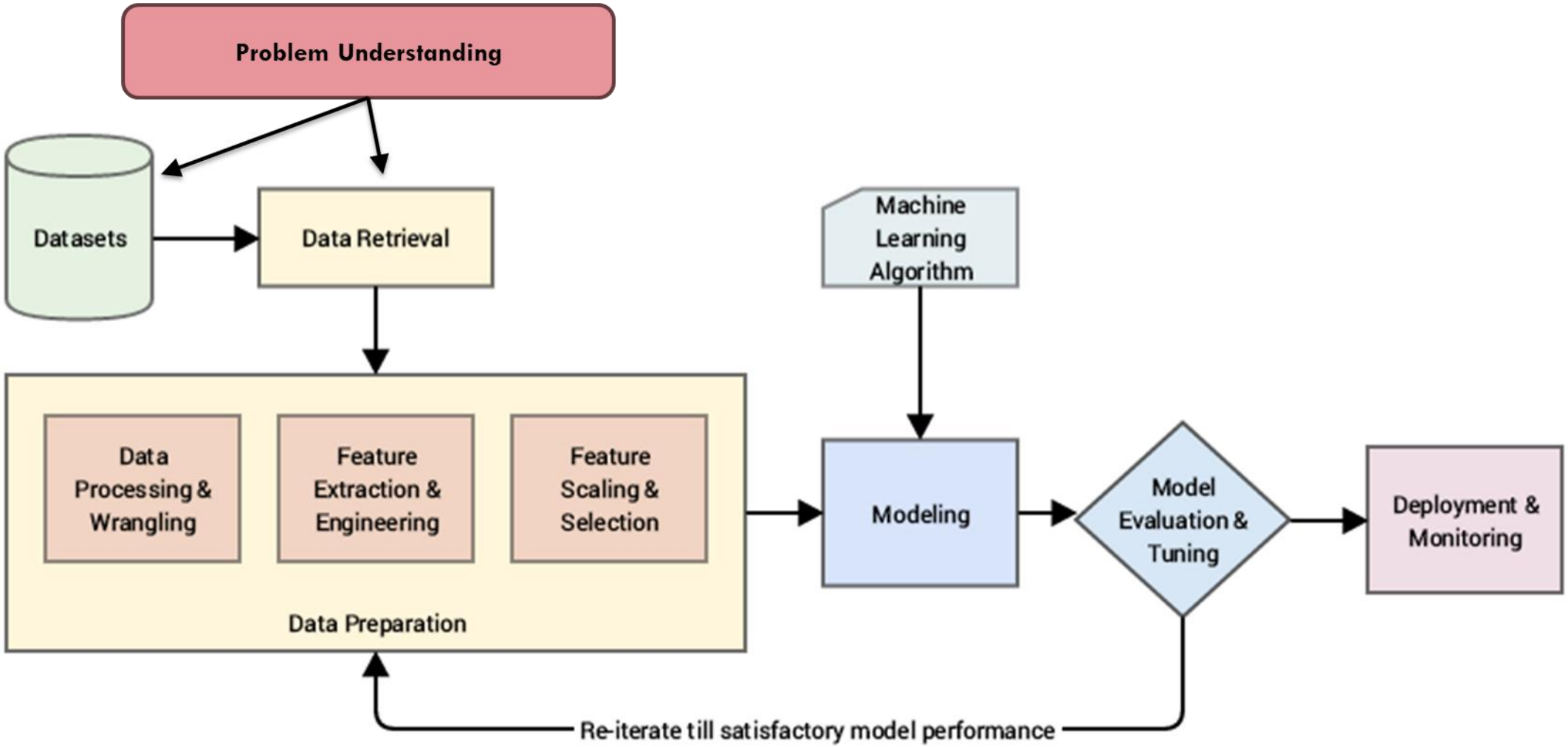
Modelos: Recordatorio

- Conceptualmente: Un constructo matemático que pone en relación unas variables de entrada (features) con una variable de salida (target)
- Físicamente: Una pieza de código que permite el entrenamiento (ajustando unos parámetros internos) y la predicción (que dada unas features de entrada genera una valor para la variable de salida target)

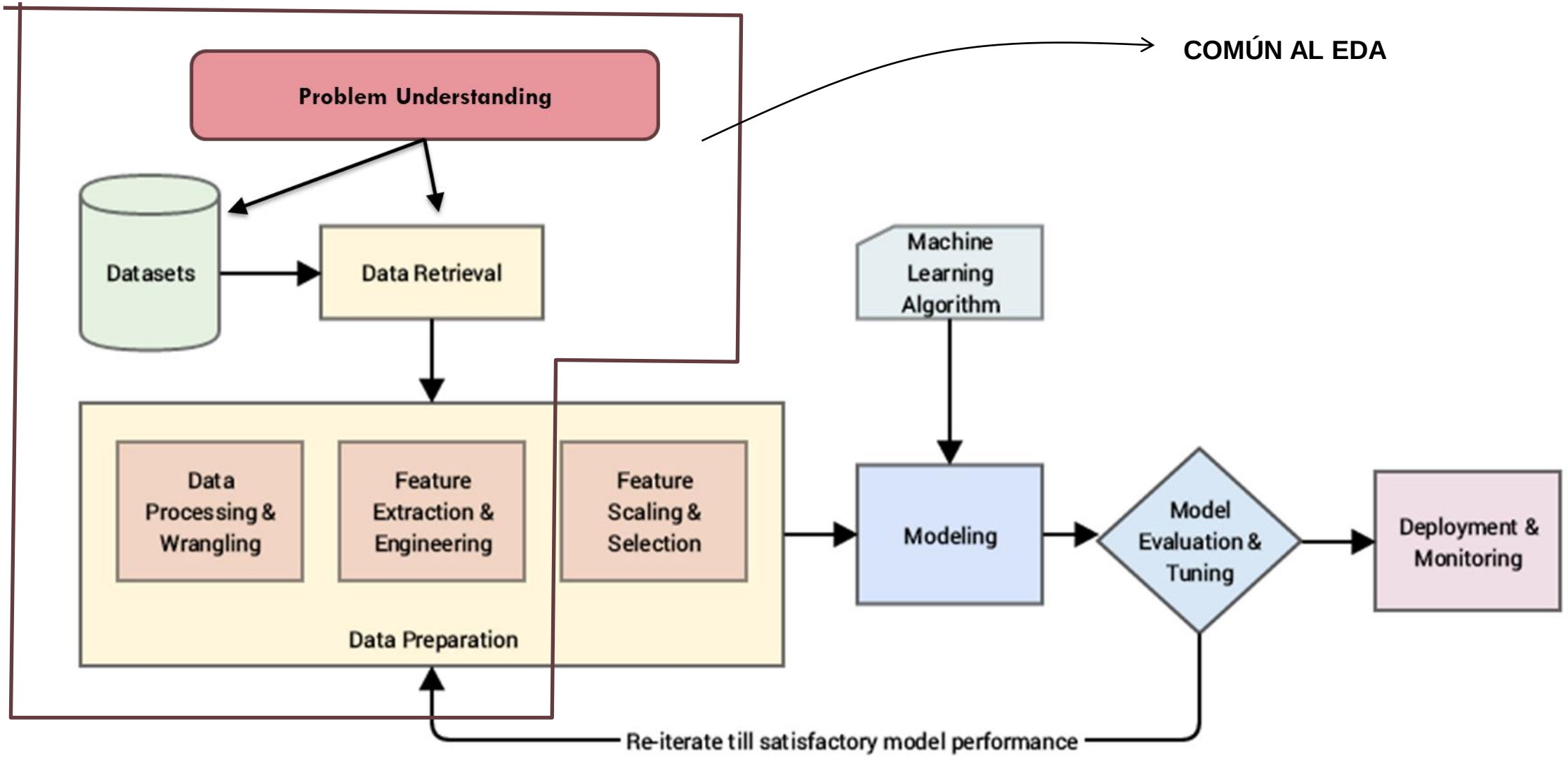


Parece que un modelo es algo que me sirve para hacer predicciones a partir de ciertos inputs. Ahora bien, ¿cómo se crea uno de estos?

Proceso de Modelado

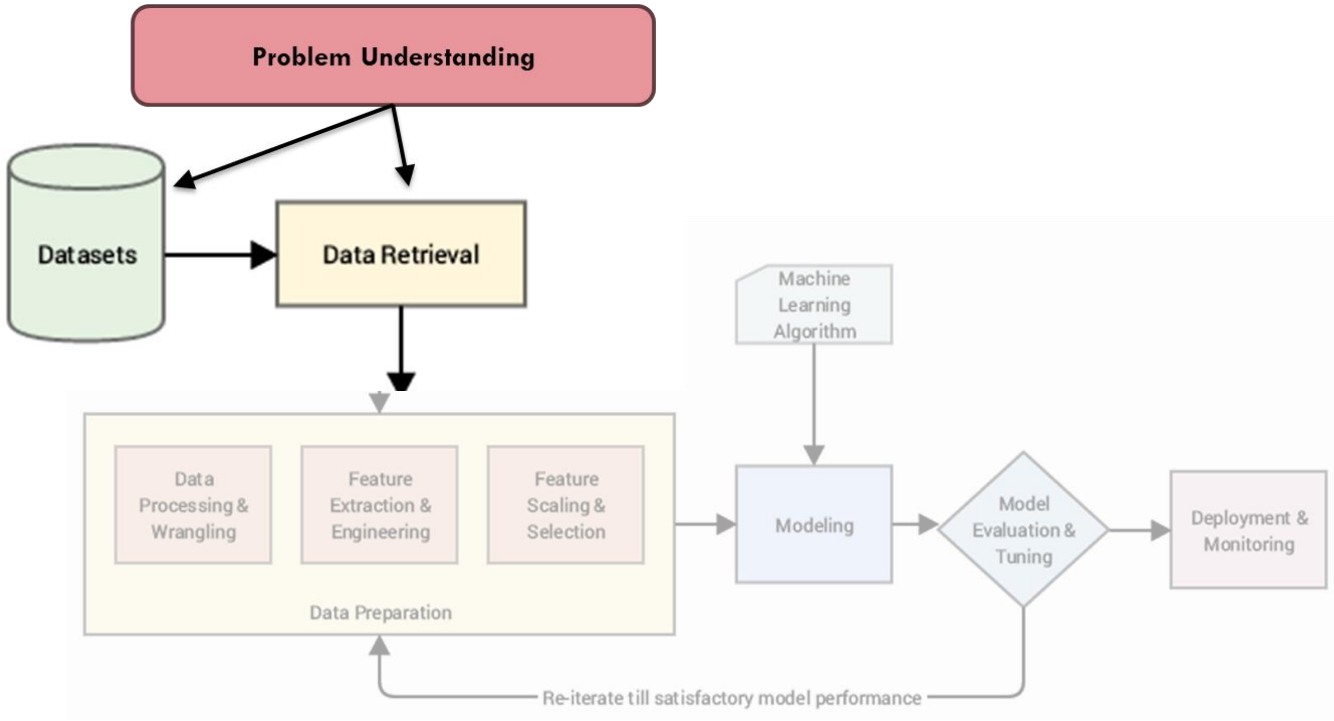


Proceso de Modelado: Relación con el EDA



Proceso de Modelado

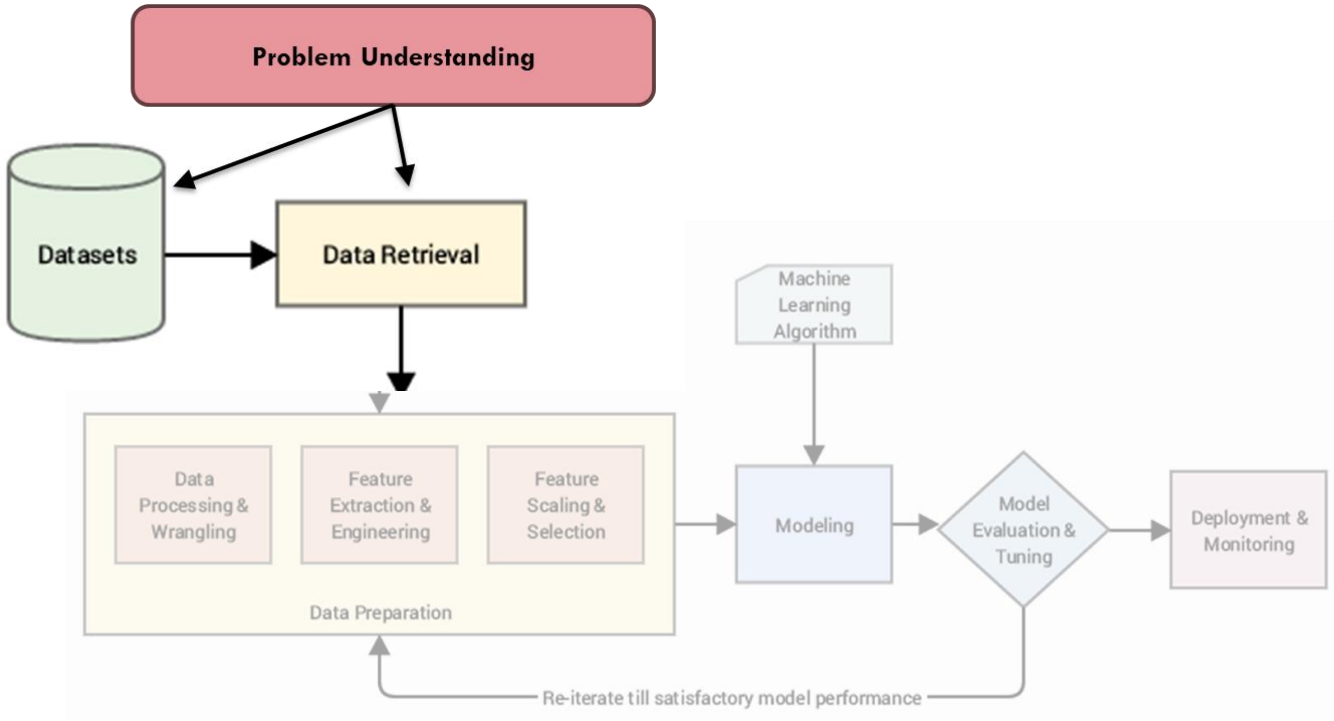
1



- Lo primero es saber qué queremos resolver:
Problema de negocio vs Problema técnico

Proceso de Modelado

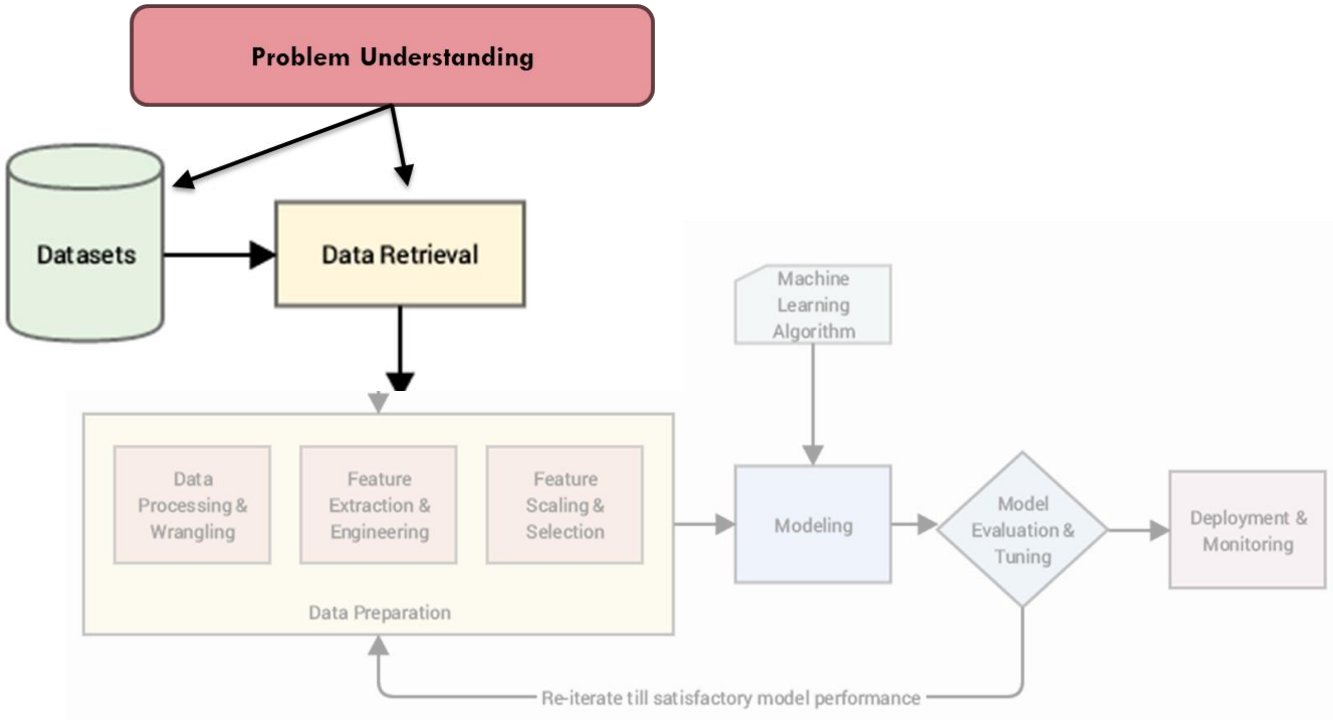
1



- Lo primero es saber qué queremos resolver: Problema de negocio vs **Problema técnico**

Proceso de Modelado

1



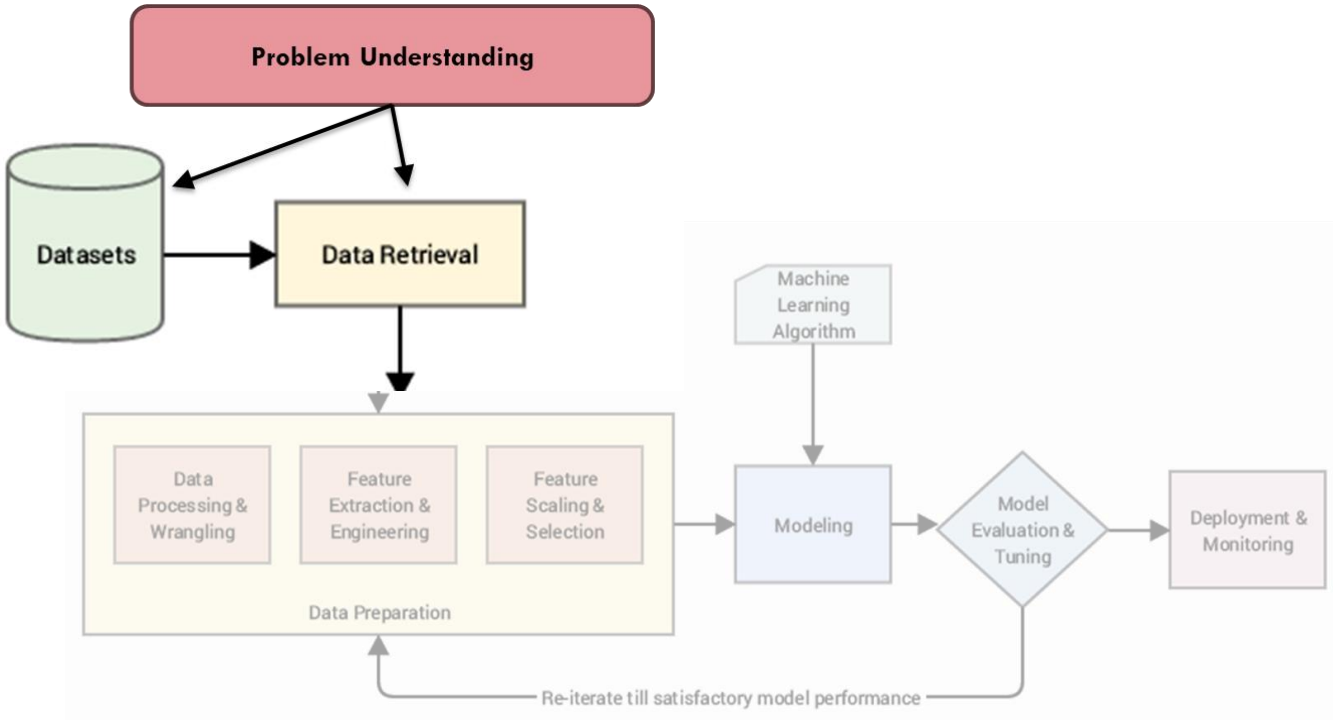
- Lo primero es saber qué queremos resolver: Problema de negocio vs Problema técnico

- Lo siguiente es entender que datos tenemos y podemos tener

- Distinguir entre feature y target

Proceso de Modelado

1



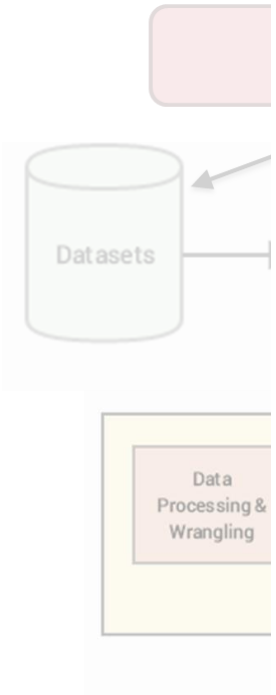
- Lo primero es saber qué queremos resolver: Problema de negocio vs Problema técnico

- Lo siguiente es entender que datos tenemos y podemos tener

- Distinguir entre feature y target

- **SEPARAR ENTRE TRAIN Y TEST (y a veces en VALIDACION)**

Proceso de Modelado: Train, Validación, Test



- No podemos utilizar todos los datos en entrenamiento porque necesitamos poder evaluar la capacidad del modelo con datos que no haya visto en entrenamiento.

- Tenemos que evaluar la capacidad de generalización

- Normalmente, vamos a dividir en dos conjuntos: Entrenamiento (Train) y Test. Lo haremos con una separación aleatoria (es decir como si obtuviéramos dos muestras de una población)

- En otras ocasiones además haremos una tercera separación: Validación. Este grupo nos servirá para comparar diferentes modelos entre sí, entrenados sobre los mismos datos.

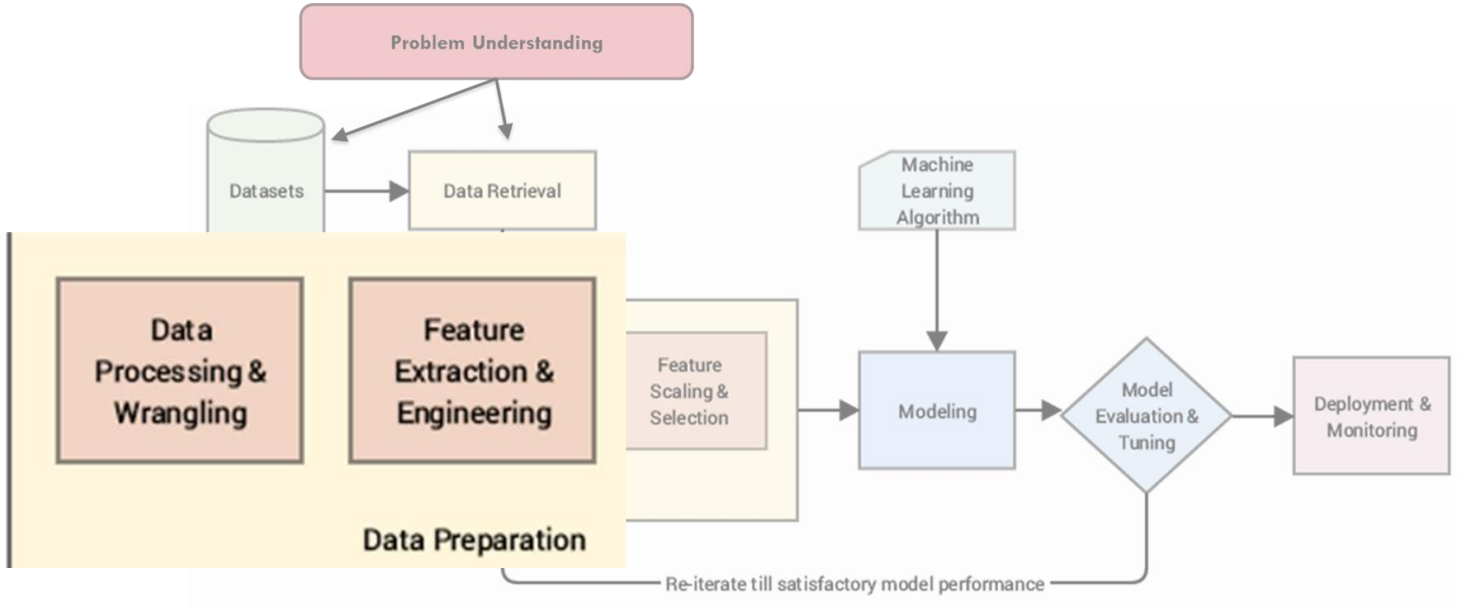
- En general, utilizaremos sólo train y test y para validar y compara modelos una técnica denominada cross-validation.

Con cross-validation

Con set de Validacion

Proceso de Modelado: "EDA"

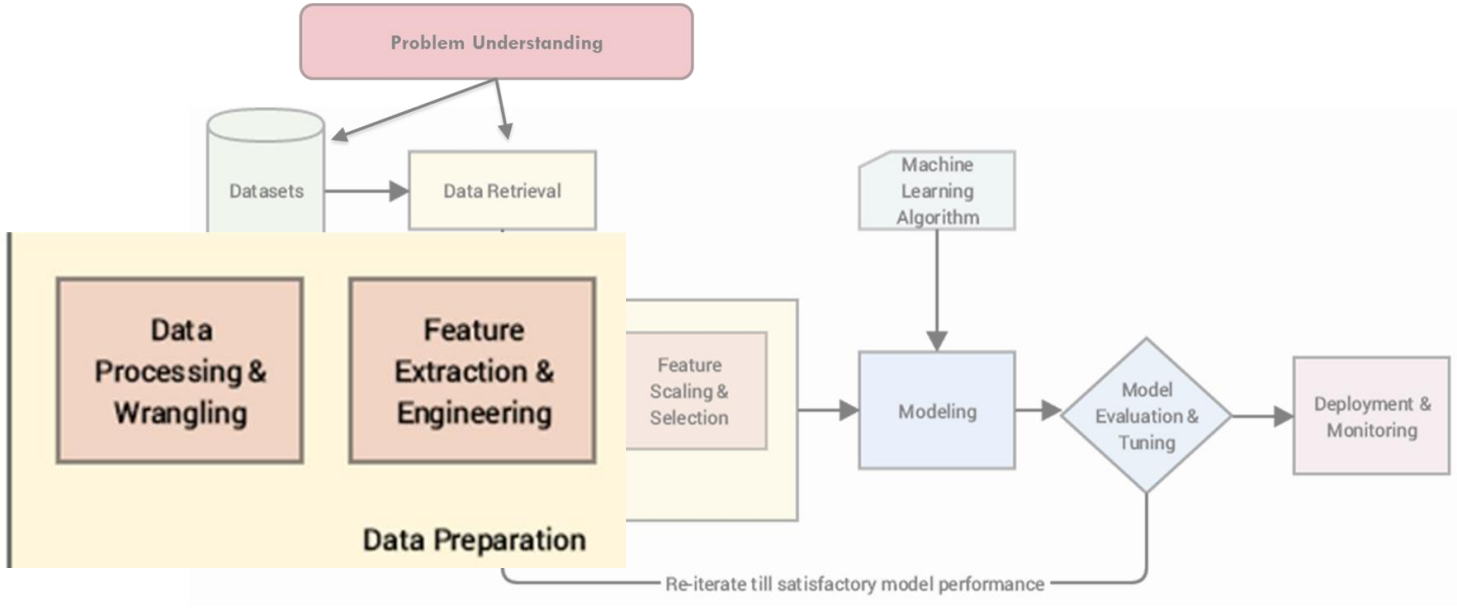
2



Proceso de Modelado: "EDA"

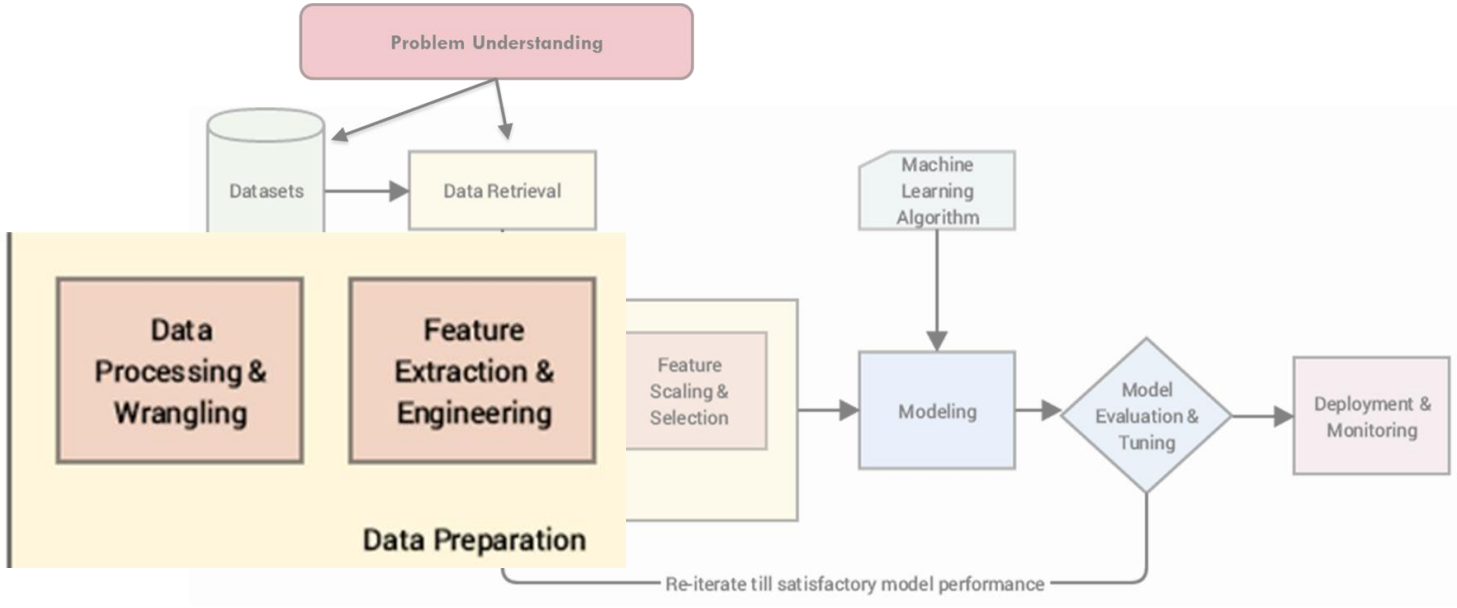
2

- **Se limpia el dataset de train**



Proceso de Modelado: "EDA"

2

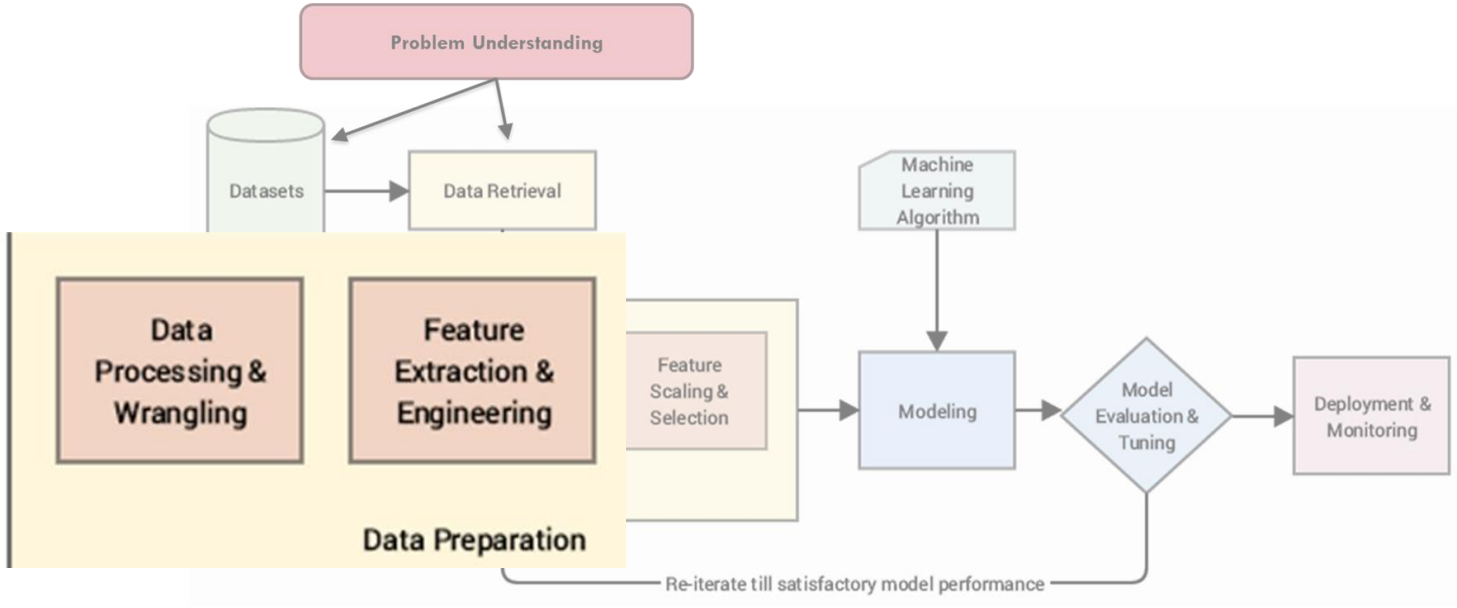


- Se limpia el dataset de train

- **Análisis univariante y de correlación.**

Proceso de Modelado: "EDA"

2



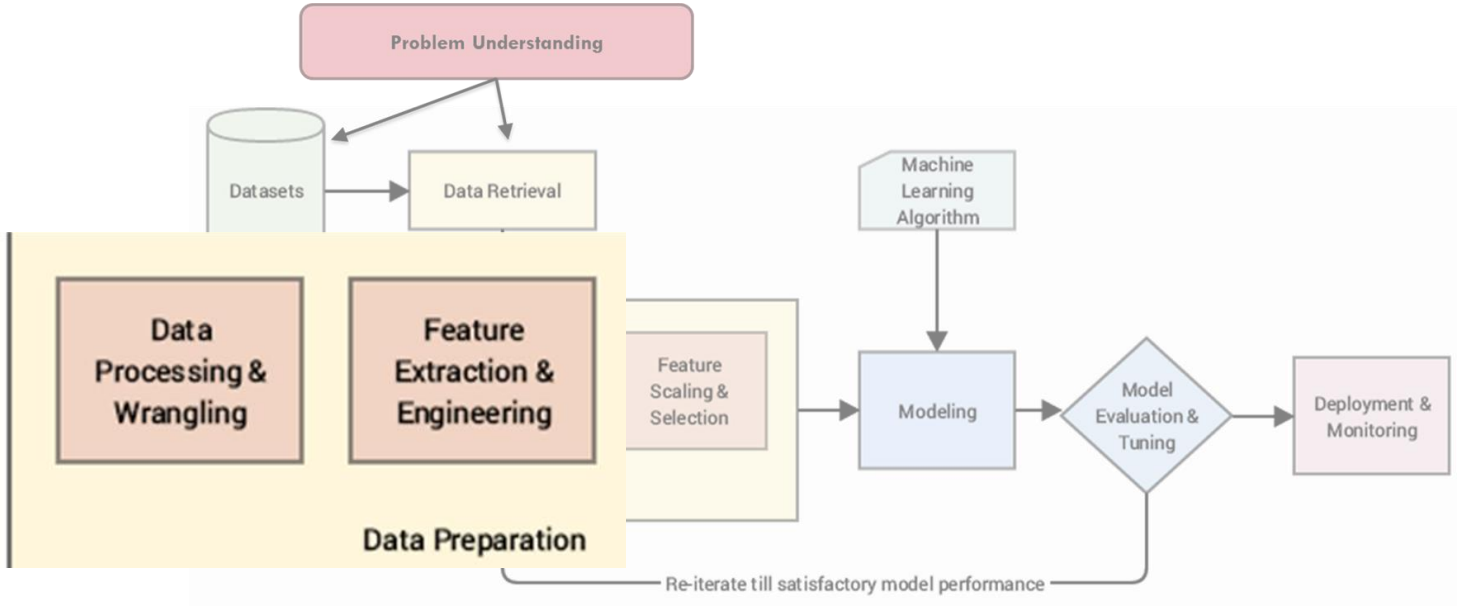
- Se limpia el dataset de train

- Análisis univariante y de correlación.

- **Análisis bivariante con la variable target**

Proceso de Modelado: "EDA"

2



- Se limpia el dataset de train

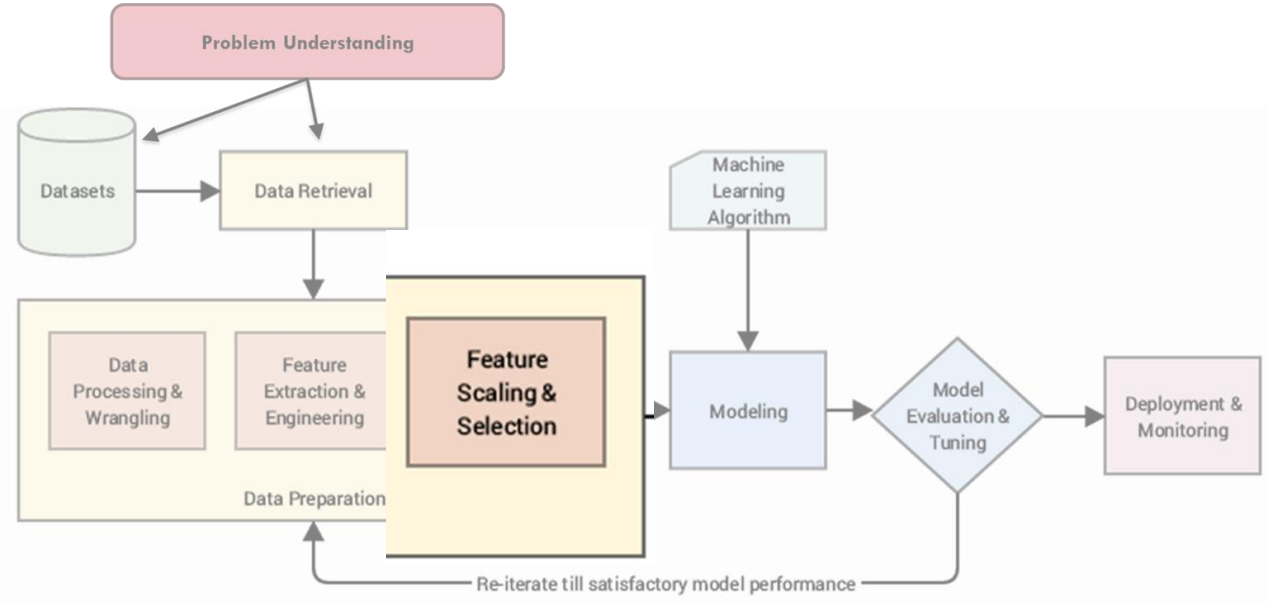
- Análisis univariante y de correlación.

- Análisis bivariante con la variable target

- **Generación de nuevas variables y análisis respecto al target**

3

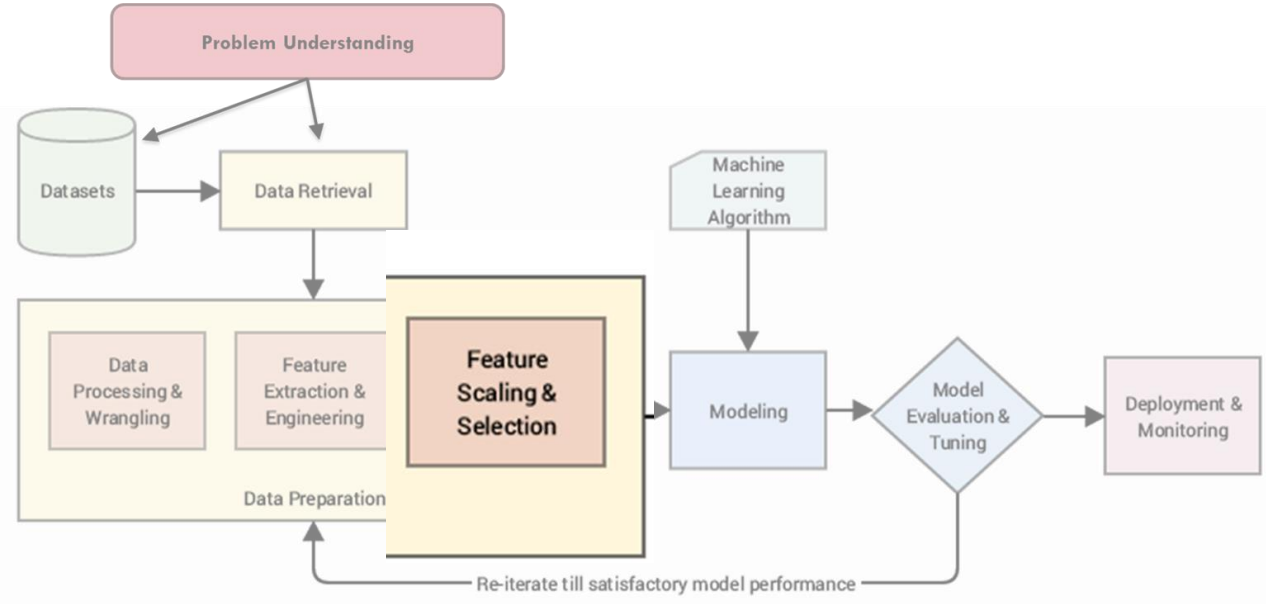
Proceso de Modelado: Preparación y Selección



- **Escogemos las variables relevantes para ser "features" del dataset de entrenamiento**

Proceso de Modelado: Preparación y Selección

3



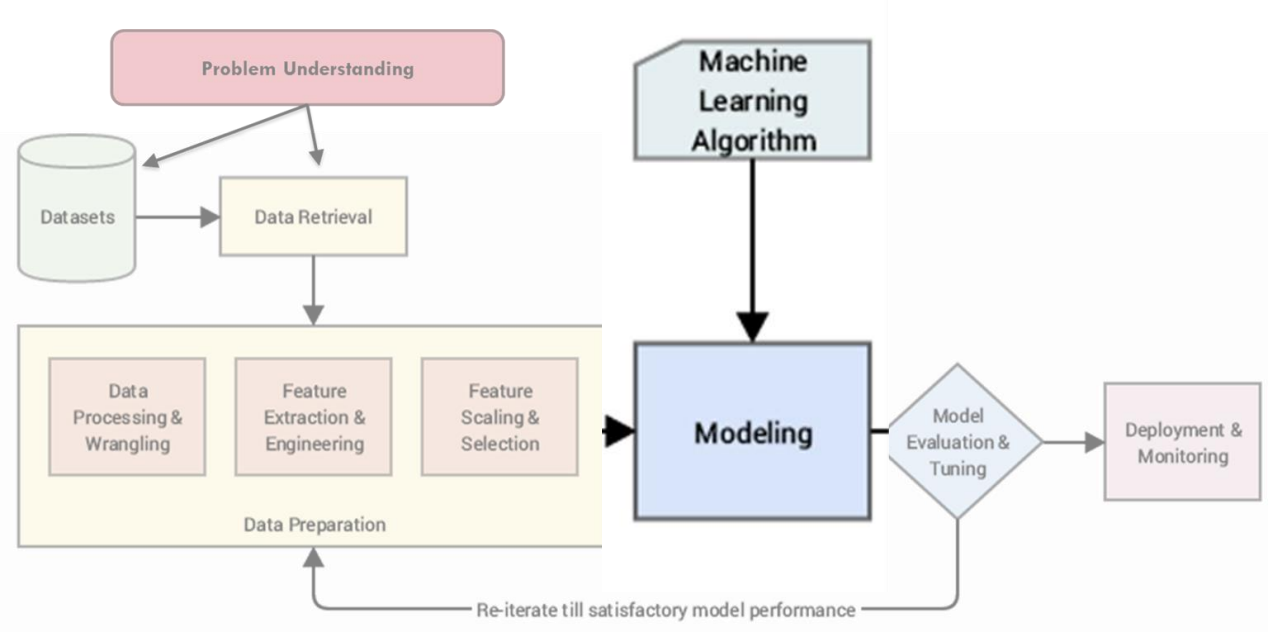
- Escogemos las variables relevantes para ser “features” del dataset de entrenamiento

- Transformar las variables restantes a números y limitar los rangos de variación de las numéricas continuas (escalado y normalización)

Proceso de Modelado: Creación de modelos

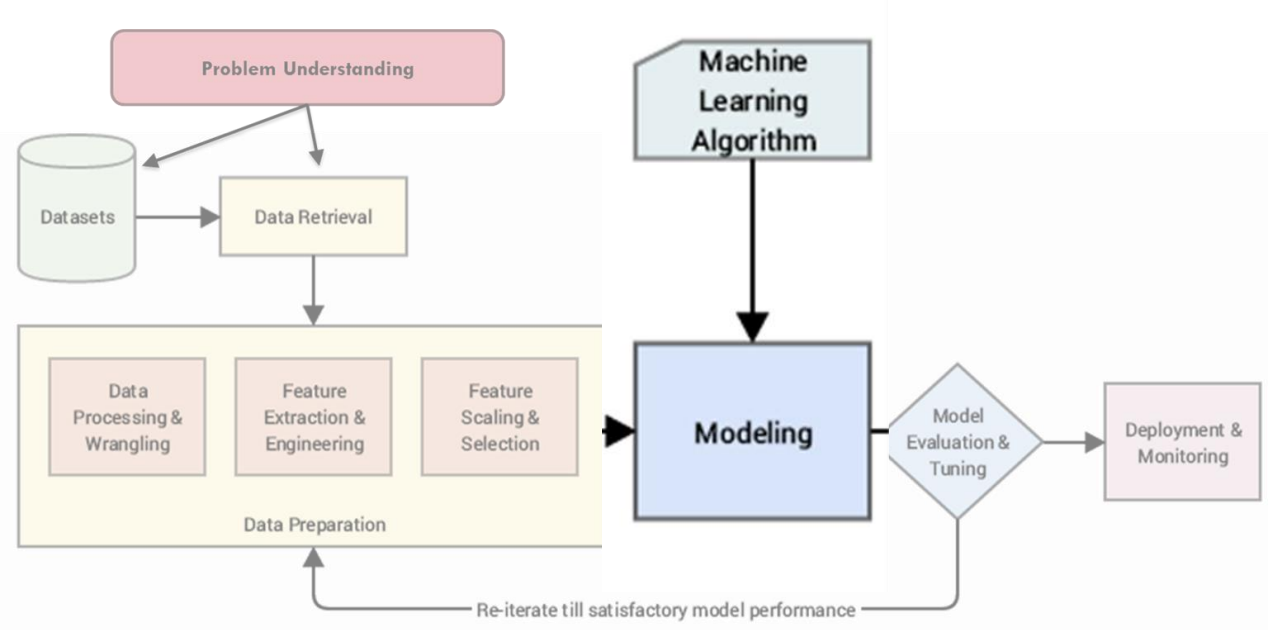
4

- **Escoger algoritmos de creación: Cargar las librerías Python necesarias**



Proceso de Modelado: Creación de modelos

4

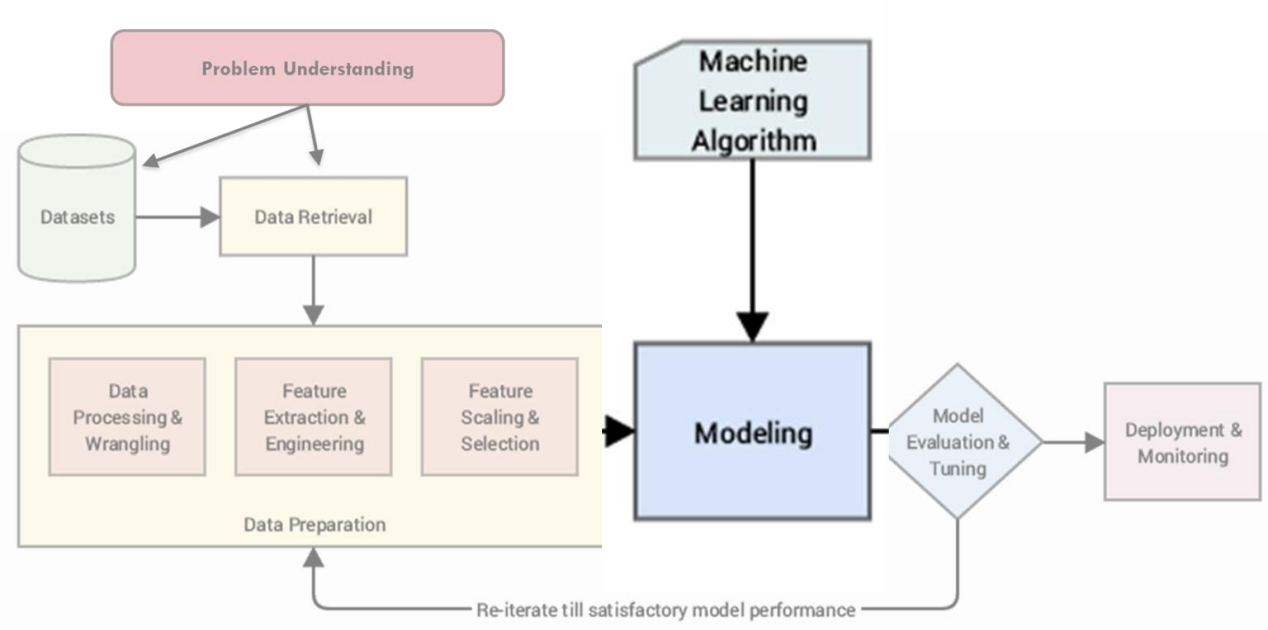


- Escoger algoritmos de creación: Cargar las librerías Python necesarias

- Definir la baseline o línea de comparación

Proceso de Modelado: Creación de modelos

4



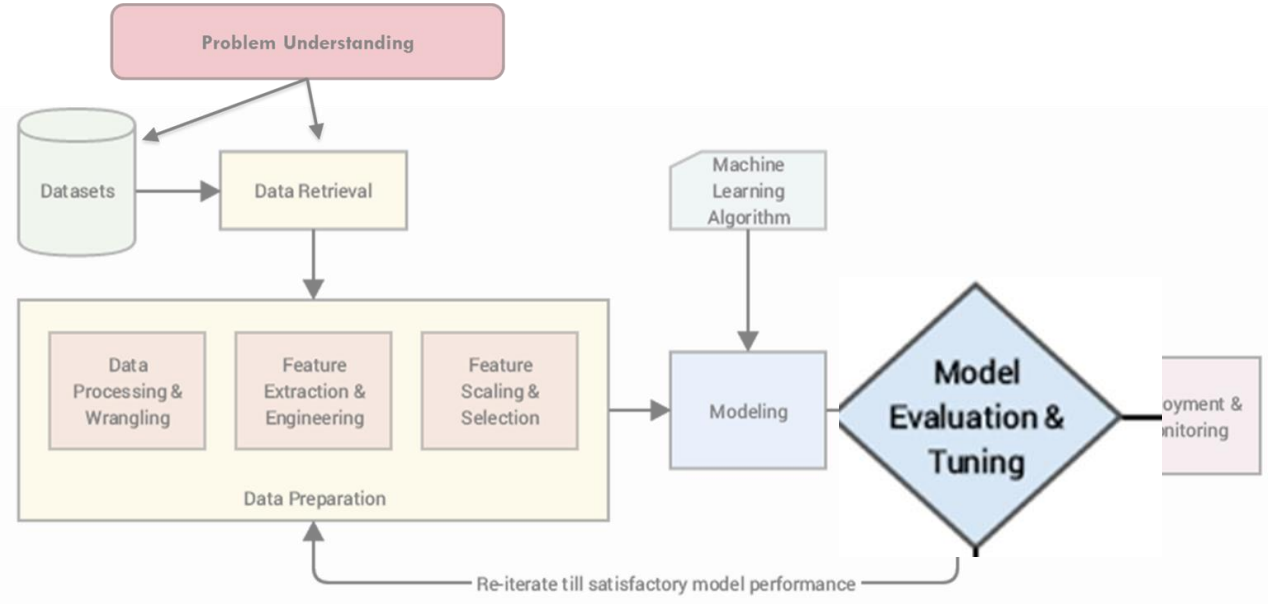
- Escoger algoritmos de creación: Cargar las librerías Python necesarias

- Definir la baseline o línea de comparación

- Generar los modelos

Proceso de Modelado: Evaluación y Ajuste

5



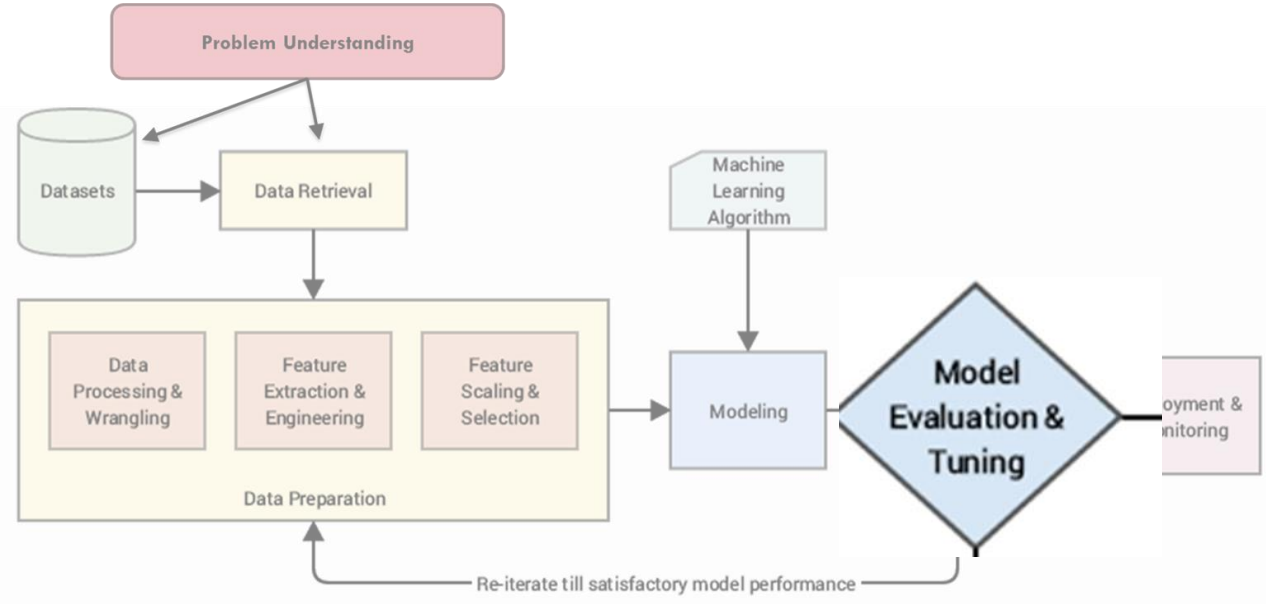
- **Se escogen las métricas de evaluación**:

Clasificación: Accuracy, Precision, Recall, AuROC,…

Regresión: MSE, MAE, etc

Proceso de Modelado: Evaluación y Ajuste

5



- Se escogen las métricas de evaluación:

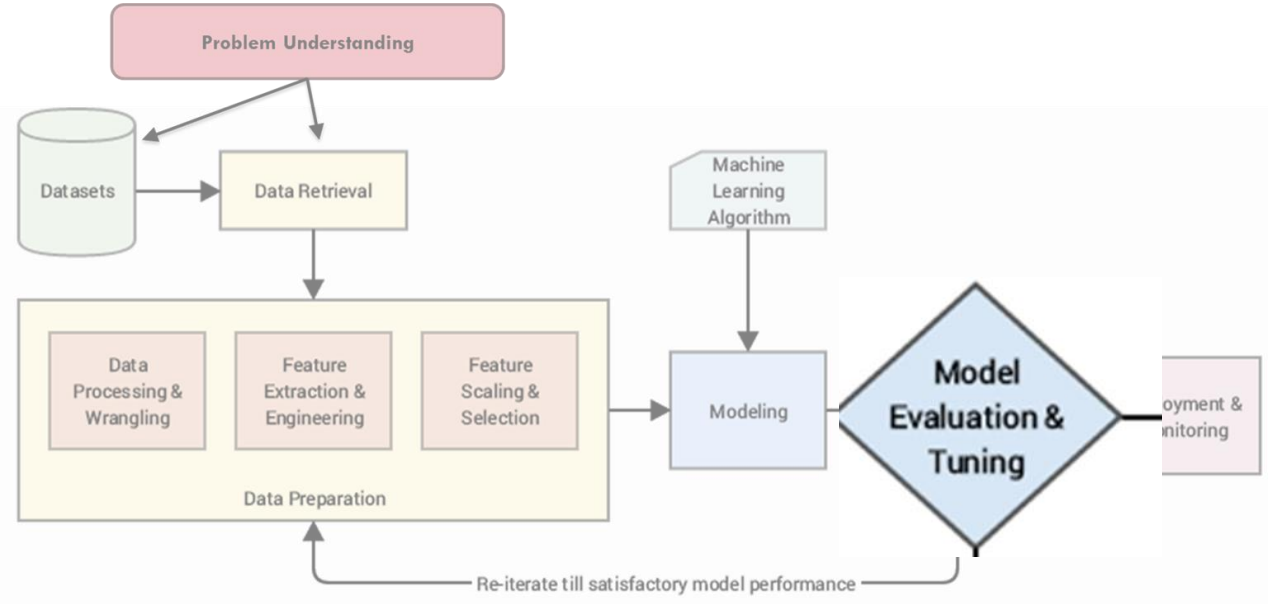
Clasificación: Accuracy, Precision, Recall, AuROC,…

Regresión: MSE, MAE, etc

- Si existen varios modelos: se comparan utilizando el dataset de validación o la técnica de cross-validation

Proceso de Modelado: Evaluación y Ajuste

5



- Se escogen las métricas de evaluación:

Clasificación: Accuracy, Precision, Recall, AuROC,…

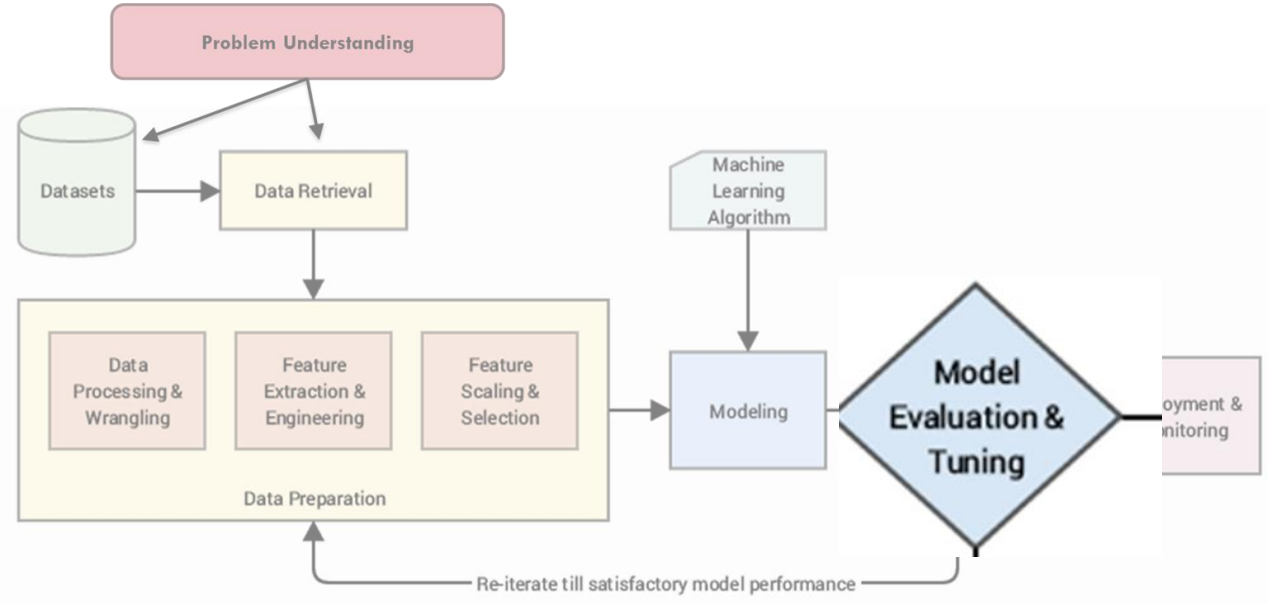
Regresión: MSE, MAE, etc

- Si existen varios modelos: se comparan utilizando el dataset de validación o la técnica de cross-validation

- **Ajustaremos los hiperparámetros del modelo escogido: GridSearch, Random Search, Optimización Bayesiana…**

Proceso de Modelado: Evaluación y Ajuste

5



- Se escogen las métricas de evaluación:

Clasificación: Accuracy, Precision, Recall, AuROC,…

Regresión: MSE, MAE, etc

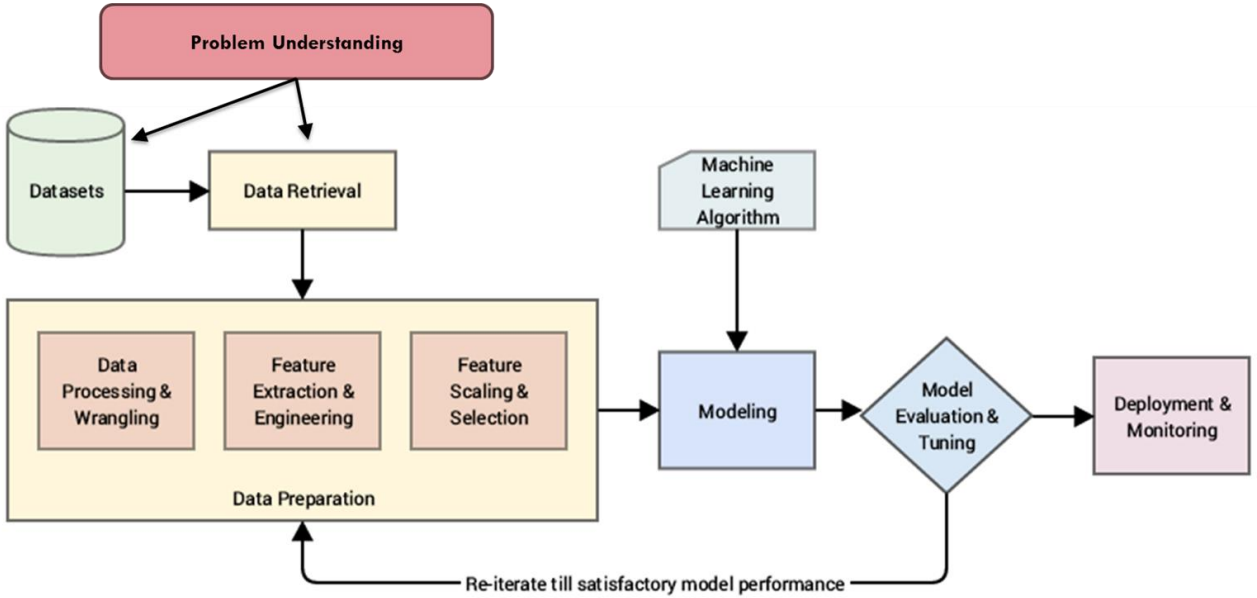
- Si existen varios modelos: se comparan utilizando el dataset de validación o la técnica de cross-validation

- Ajustaremos los hiperparámetros del modelo escogido: GridSearch, Random Search, Optimización Bayesiana…

- **Probaremos en el examen final el nivel de generalización: Prueba con el dataset de Test y Análisis de errores**

Proceso de Modelado: Reentrenamiento y recalibrado

-

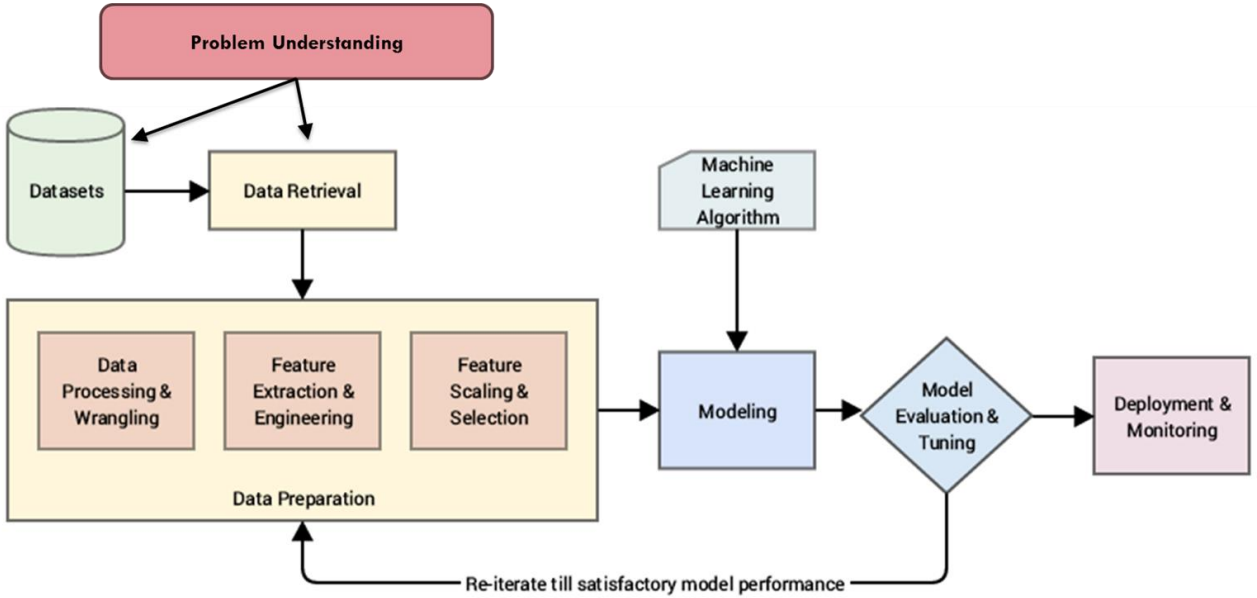


- Con el tiempo una vez el modelo esté funcionando:

- Tendremos datos nuevos

- El rendimiento puede decaer

Proceso de Modelado: Reentrenamiento y recalibrado



- Con el tiempo una vez el modelo esté funcionando:

- Tendremos datos nuevos

- El rendimiento puede decaer


THE
BRIDGE