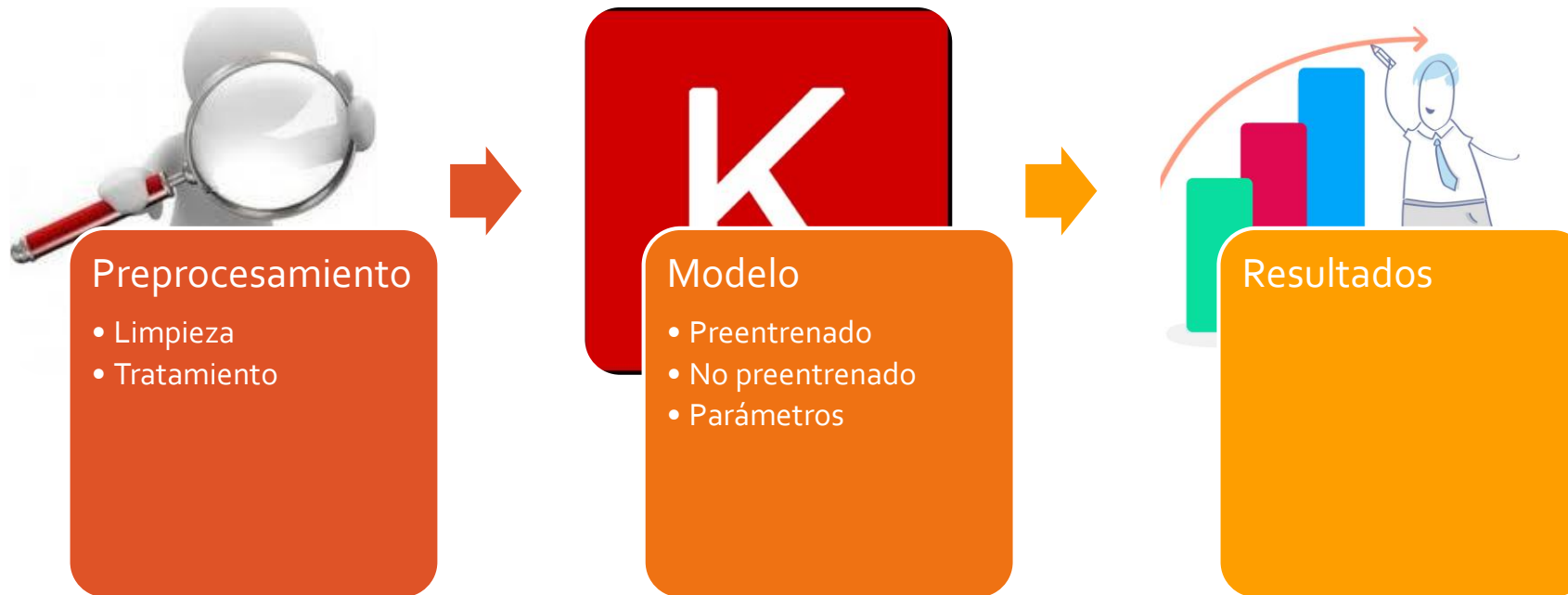


P3 –Toxicity Classification

Objetivo: Detectar la toxicidad de varias conversaciones.

Hazel Vanesa Jaime Huerfano

Sandra Yessel Tellez

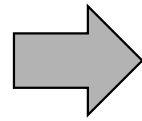


P3 –Toxicity Classification

Modelo

Limpieza

- Verificación de nulos
- Eliminar puntuación y caracteres especiales
`!"#$%&()*+,-.`
- Contracción de palabras
`{ain't: is not, arent, are not}`



Preprocesamiento

- Cada palabra será un número, un comentario es una lista de números, cuantos números queremos en la lista? El largo de la lista será una constante
- Construcción de matriz con las palabras de todo el corpus con modelos preentrenados (GloVe + fasttext) /1 palabra ->Vec 300d/
- Entrenando con todo los comentarios (train + test)
- Texts_to_sequences y Sequence Pad_sequences
- Cambio de target a False(0) or True(1)

Modelo LSTM

- Establecer parametros
- Dos Capas "Forward Order" y " Backward Order" Flexibilidad representando las variables. Costo Computacional
- Learning Rate Schedule: Función que toma la tasa actual de aprendizaje y la epoca y calcula una nueva tasa. Generalmente se usa exponencial.

P3 –Toxicity Classification

Resultados

Variación	Kaggle
Raw data	0.90940
Preprocesamiento(Diccionario de Contracciones mas amplio)	0.91086
Preprocesamiento(Diccionario de Contracciones reducido + Caracteres especiales) Sin preentrenamiento	0.91477
Modelo Preentrenado	0.93073