

PARTE IA

Objetivo:

- Poder predecir futuros conflictos armados a partir de los datos generados por la parte Cloud y Big Data

Necesidad de Big Data:

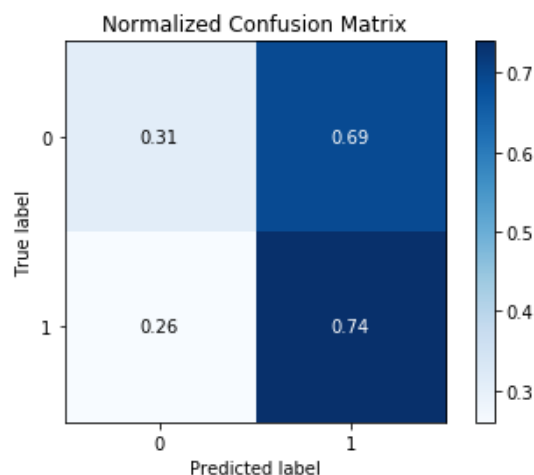
- Necesitamos una gran cantidad de datos que nos permitan tener las muestras suficientes para tener un entrenamiento correcto.

Necesidad de Cloud:

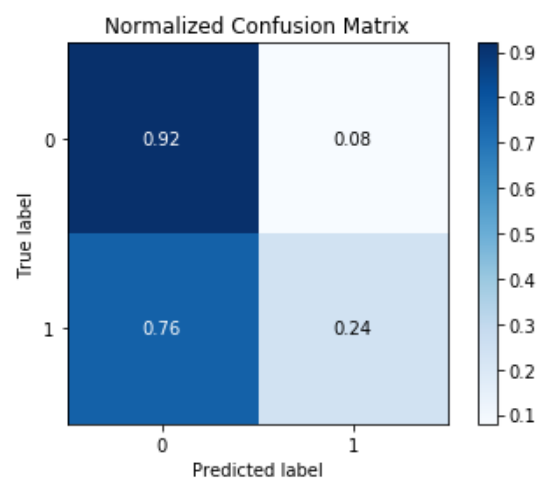
- A mayor cantidad de datos, mayor necesidad de procesamiento. La Inteligencia Artificial, depende en gran medida del algoritmo aplicado y del número de iteraciones.

Descripción de la solución:

- Está implementada por Scikit
- Los algoritmos utilizados han sido:
 - Máquina de Soporte Vectorial (SVM), *algoritmo LinearSVC*
 - El número de iteraciones ha sido de 10.000.000 → 54 min
 - El resultado del aprendizaje ha sido de un 38% de aciertos
 - Naive Bayes (NB), *algoritmo GaussianNB*
 - No hay número de iteraciones → 0.006 segundos
 - El resultado del aprendizaje ha sido de un 65'9% de aciertos



SVC



GaussianNB

- Datos gestionados a través de Pandas
- Entrenada a partir de los datos generados anteriormente, en la parte de Big Data
- El tamaño del Conjunto de Datos entrenados es del 60% respecto al total, este porcentaje se ha escogido para tener un 40% de casos 'test'
- El resultado del entrenamiento es del 66% de aciertos según la matriz de confusión

Dependencias:

- Librerías de Python
 - o Numpy
 - o Matplotlib.pyplot
- Librerías de Inteligencia Artificial
 - o Scikit
 - sklearn.metrics
 - sklearn.preprocessing
 - sklearn.model_selection
 - sklearn.metrics
 - sklearn.svm
 - o Scikitplot
- Librerías de gestión de archivos
 - o Pandas → pip install pandas

Entorno

- *Jupyter Notebook*
 - o Evitar cargar datos cada vez que utilizamos el programa
 - o Evitar entrenar la IA cada vez que queremos visualizar gráficas

Discusión de overhead

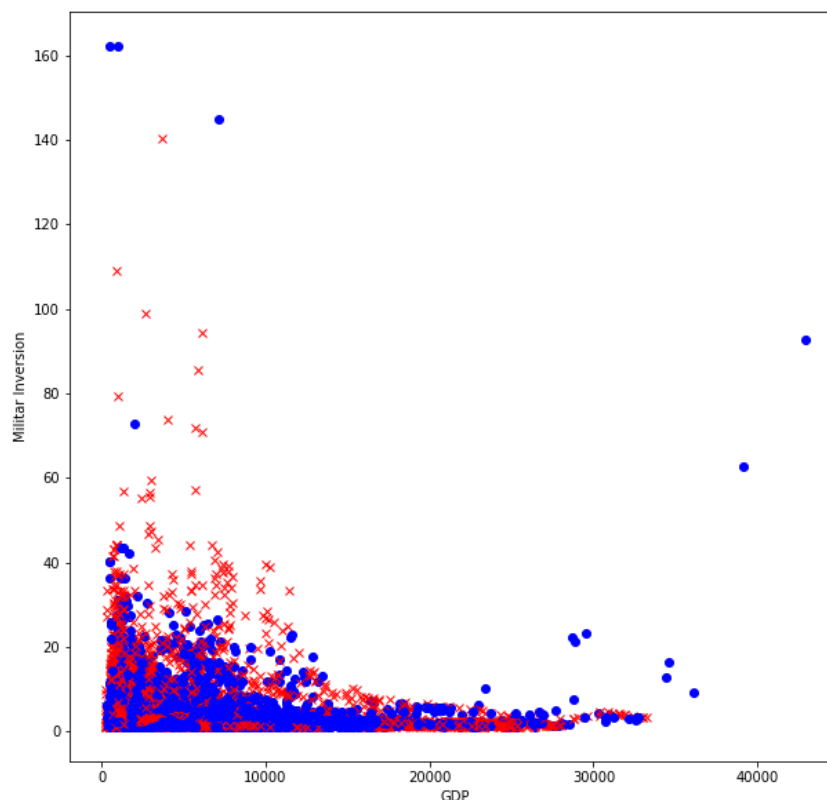
- Para entrenar a la Inteligencia Artificial y escoger un algoritmo óptimo, se necesita tiempo dado que los recursos que tenemos son limitados y no escalables. Este problema se puede solucionar gracias al uso del Cloud

Conclusiones

- Todo aquello relacionado con el Big Data y la Inteligencia Artificial debe hacerse en el Cloud, dado a que las prestaciones de los ordenadores corrientes no llegan a ser suficientes para poder programar de forma fluida
 - o Tiempos estimados en un ordenador con un i7 3770K + 8GB RAM
 - Lectura del CSV ~ 1s
 - Entrenamiento de la Inteligencia Artificial con un algoritmo SVC con ~ 1 MB de datos

Iteraciones	t
10	0.01 s
100	0.04 s
1.000	0.37 s
10.000	3.6 s
100.000	35.25 s
1.000.000	351.49 s ~ 6 min
10.000.000	54m

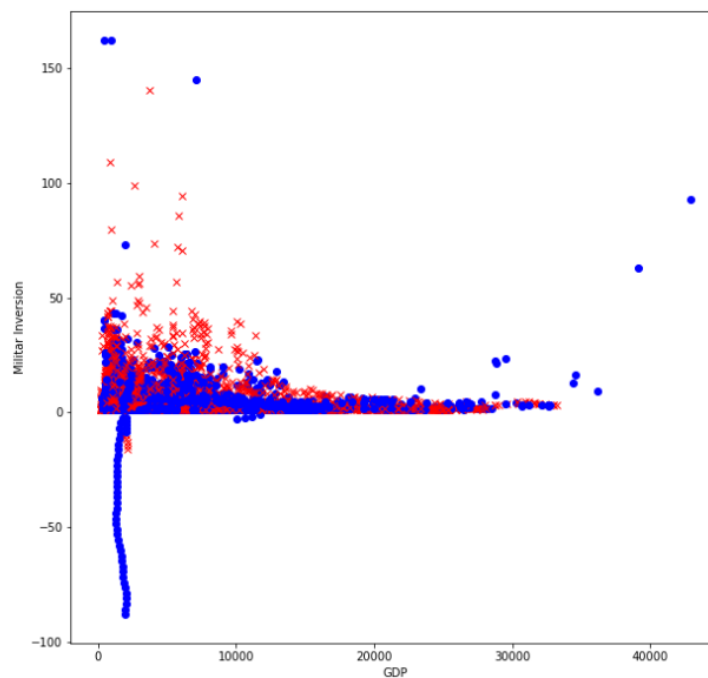
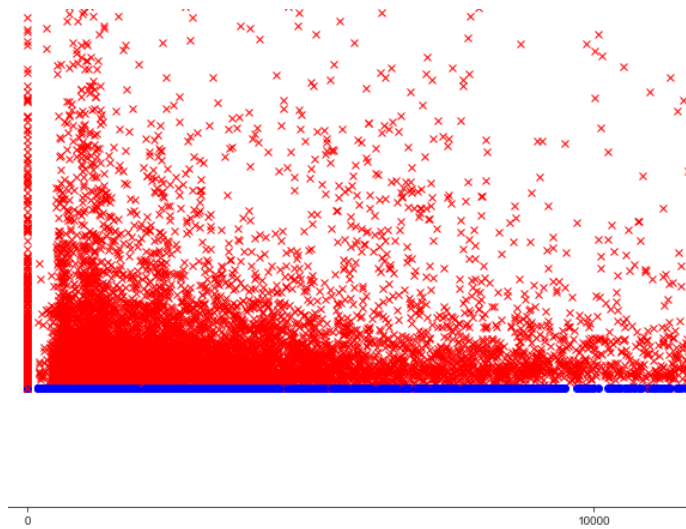
- Dibujar la gráfica con el Dataset final ~ 37 s



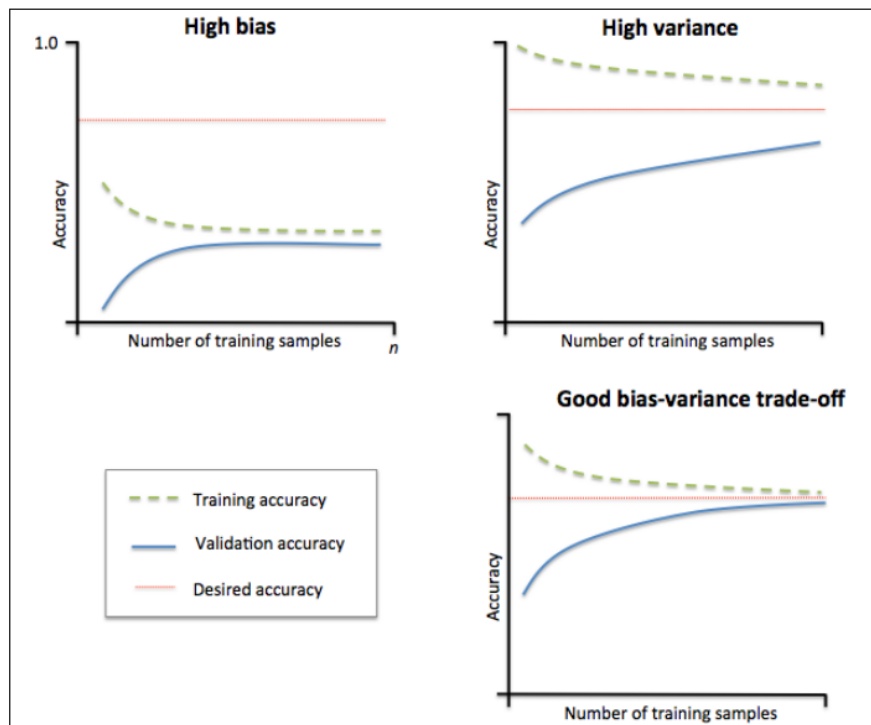
- o La elasticidad de los servicios de la nube permite procesar/gestionar gran cantidad de información en pocos segundos

Sugerencias y aprendizaje

- Dibujar la matriz de confusión
 - o ¿Por qué?
 - Relación entre los Verdaderos y Falsos Positivos (y Negativos) que ha conseguido testear la IA
- Dibujar los datos, y diferenciarlos por su resultado, en este caso, si estuvo en guerra o no.
 - o ¿Por qué?
 - Dataset generado erróneos



- Dibujar la curva de aprendizaje, para buscar una correlación correcta entre Train Set y Cross Validation Set



- Probar algoritmos distintos puede dar mejores y peores resultados (SVC, LinearSVC, NuSVC, LogisticRegression, SGDClassifier, Bayes...)
 - o Numero de iteraciones relativamente bajo para poder comparar dichos resultados