

Equipo:

Eddie Lozada 175938
Elizabeth Viveros 161224
Gerardo Mathus 191459
Karla Alfaro 137314
Leonardo Marín 175903

Github

<https://github.com/gemathus/dpa-2020>

Problema:

No existe una medida confiable para saber qué tan seguro es un establecimiento. Esto es un problema pues los padres dejan a sus hijos en centros sin tener información sobre las violaciones sanitarias de los mismos.

Objetivo principal:

Proveer a los padres de familia de información con respecto a la “seguridad sanitaria” de un establecimiento. Además, puede crearse un ranking por zona geográfica de los establecimientos con base en esta medida de seguridad sanitaria.

Objetivos particulares:

- Clasificar los tipos de violaciones sanitarias de la variable “regulation summary” para poder generar la medida de seguridad sanitaria.
- Crear un ranking de establecimientos.
- Predecir la probabilidad de que se arregle el problema sanitario en el tiempo especificado y por lo tanto de mantener el registro. Esto es una medida del cumplimiento y compromiso del establecimiento.

Métricas de desempeño:

La métrica de desempeño a utilizar para la clasificación de los tipos de violaciones sanitarias será accuracy.

Datos requeridos:

El conjunto de datos a utilizar es ‘*DOHMH Childcare Center Inspections*’. Son datos sobre inspecciones realizadas a los centros de cuidados de niños en NYC para verificar el cumplimiento del reglamento del cuidado de niños. La frecuencia de actualización de los datos es diaria y se tiene histórico a partir de mayo de 2016.

<https://data.cityofnewyork.us/Health/DOHMH-Childcare-Center-Inspections/dsg6-ifza>

Diagrama de solución 1:

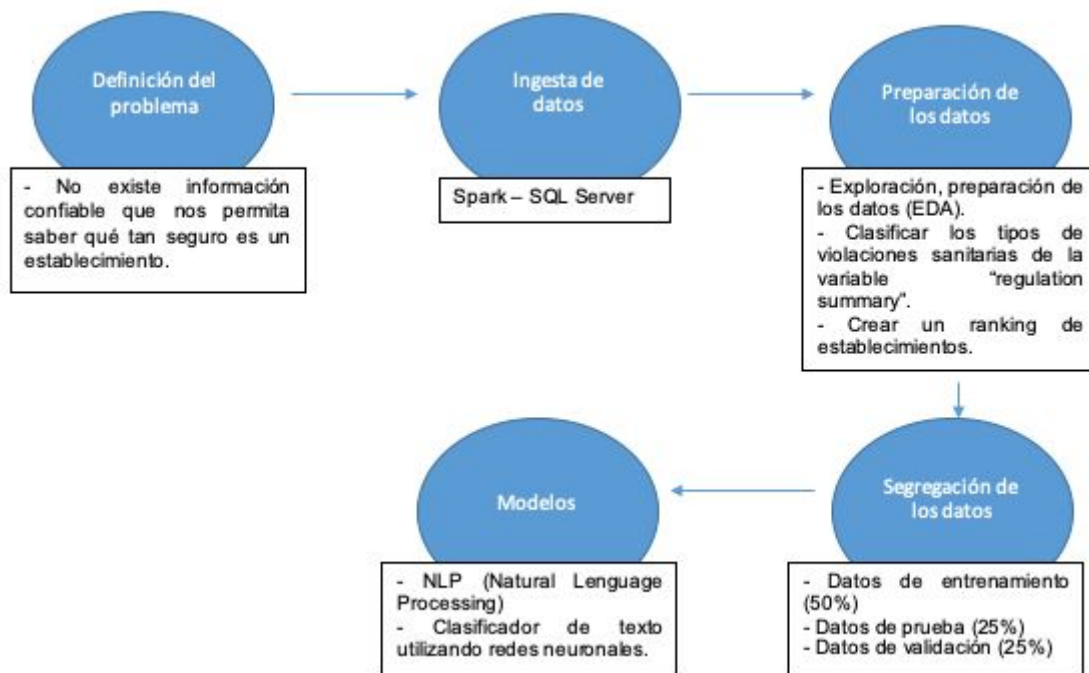
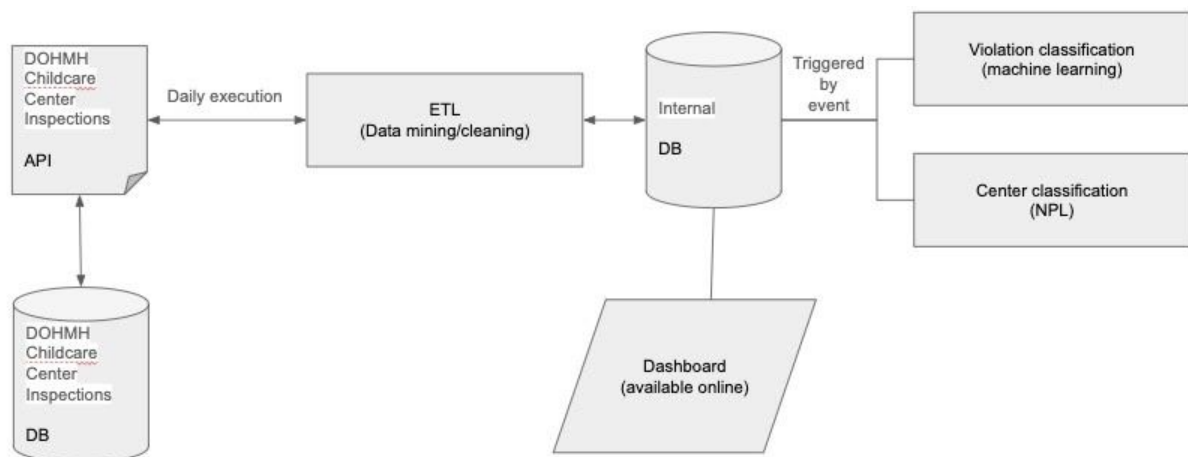


Diagrama de solución 2:



Algoritmos:

- Algoritmos de aprendizaje no supervisado para análisis de texto:
 - Analizar la variable de "regulation summary" y hacer una clasificación de las violaciones
- Algoritmos de aprendizaje no supervisado para clasificar las violaciones.
- Algoritmos de aprendizaje supervisado para predecir la probabilidad de que se arregle el problema sanitario en el tiempo especificado y por lo tanto de mantener el registro.

Entregable:

- Dashboard que incluya:
 - Lista de los establecimientos.
 - Las categorías de las violaciones y las violaciones por establecimiento.
 - El score de probabilidad de compromiso.
 - El número de ranking en el ¿municipio? del establecimiento.