Apoyo al diagnóstico de pacientes:

Text analytics

Sofia Álvarez
Brenda Barahona
Alvaro Plata

Agenda

01
Objetivos

02
Análisis de datos

03
Modelos de ML

04
Tablero de control

05
Conlusiones

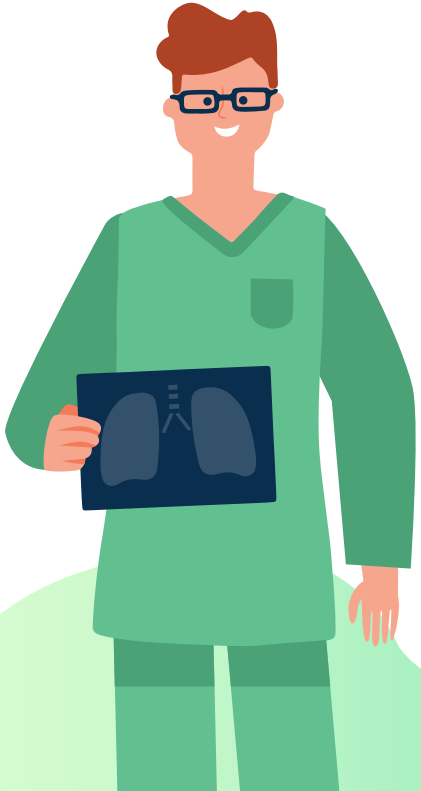
01

Objetivos

Objetivos

Crear un clasificador

Deseamos crear un modelo que sea capaz de clasificar el tipo de problema que puede estar sufriendo el paciente a partir de la descripción médica



Oportunidad para el negocio

Ayudar al personal médico a automatizar el proceso de clasificación de los pacientes en 5 diferentes tipos de condiciones o enfermedades a partir de las descripciones escritas por ellos

02

Análisis de datos

Datos de entrada

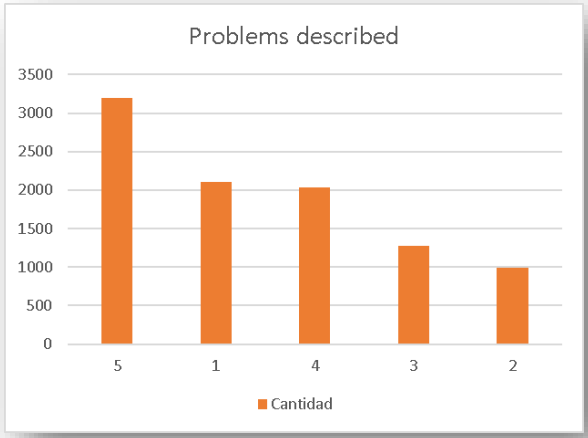
Clasificación

1. Neoplasms
2. Digestive system diseases
3. Nervous system diseases
4. Cardiovascular diseases
5. General pathological conditions

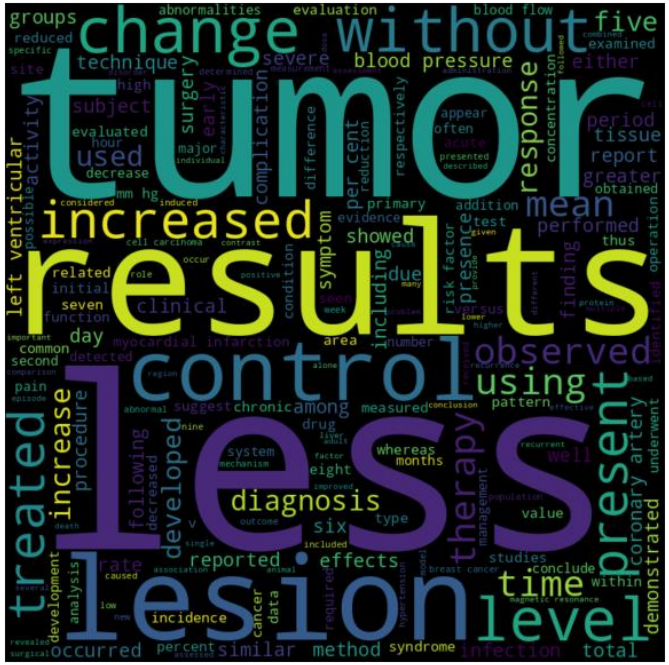
Cantidad

Problems described	Cantidad
5	3194
1	2103
4	2029
3	1280
2	994

Distribución

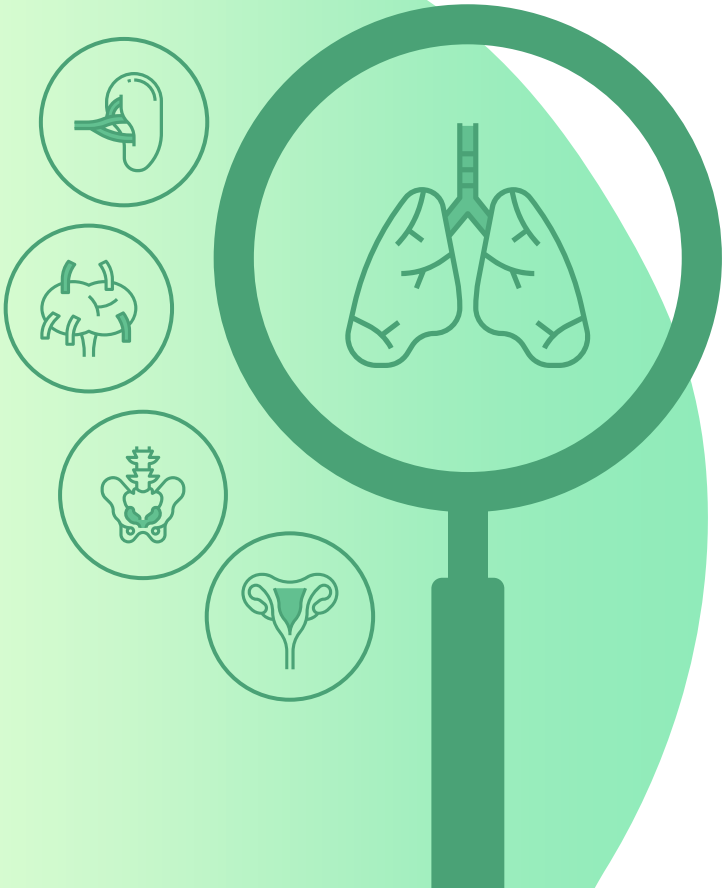


Palabras mas frecuentes



03

Modelos de ML



Modelos de clasificación

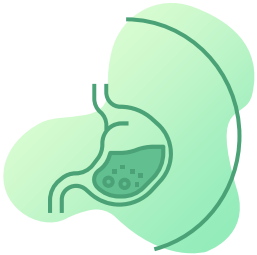
Realizaremos una tarea de **aprendizaje supervisado**

La tecnica que usamos fue la **Clafisicación multiclase**

Donde nuestra métrica será la **precisión**.

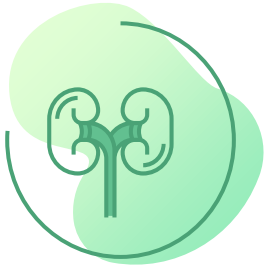
Algoritmos que usamos

Multinomial Naive Bayes.



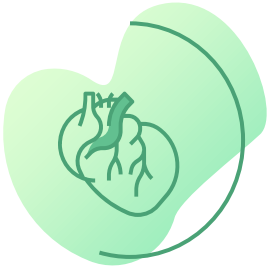
56%

One Vs Rest



71%

LSTM



64%

Modelo escogido: LSTM

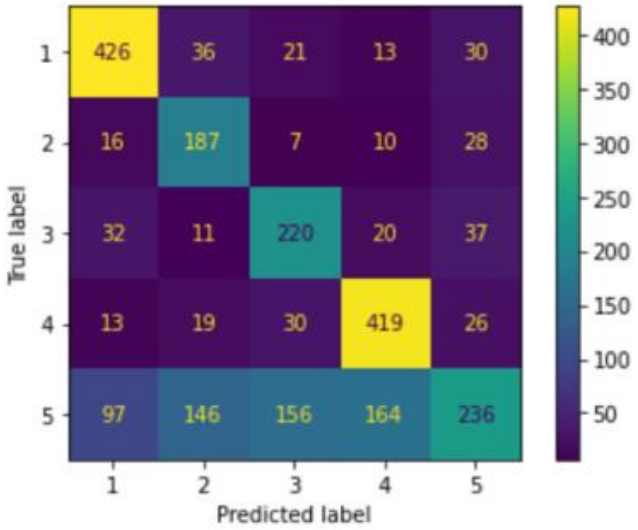
Métricas

```
---Reporte para el modelo construido---
              precision    recall  f1-score   support

           1       0.73      0.81      0.77       526
           2       0.47      0.75      0.58       248
           3       0.51      0.69      0.58       320
           4       0.67      0.83      0.74       507
           5       0.66      0.30      0.41       799

    accuracy                           0.62      2400
   macro avg       0.61      0.67      0.62      2400
weighted avg       0.64      0.62      0.60      2400
```

Matriz de confusión

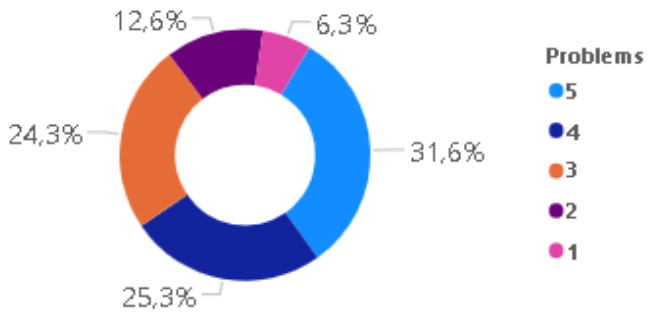


04

Tablero de control

Tablero de control

Distr. Problems described



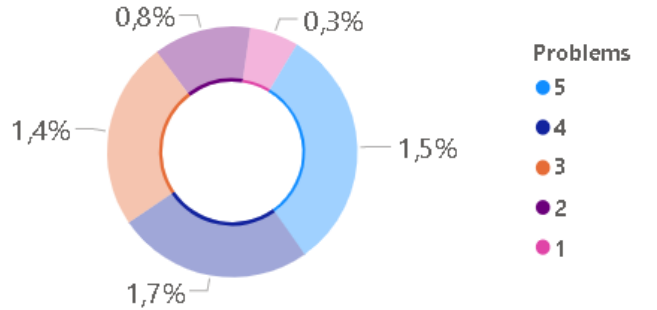
Cantidad de Datos

12 mil

probl...	medical_abstracts
1	Zollinger-Ellison syndrome: prospective assessment of abdominal US in the localization of gastrinomas. The ability of abdominal ultrasound (US) to help localize gastrinom prospectively studied in 79 patients with Zollinger-Ellison syndrome. The results were assessed by means of laparotomy, autopsy, or percutaneous liver biopsy. For hepatic gastrinoma, US had a sensitivity of 63% and a specificity of 100%, with a positive predictive value of 100% and a negative predictive value of 89%. US was slightly less sensit detecting gastrinoma in the liver than were computed tomography (CT) (66%) and selective angiography (78%). For detection of extrahepatic gastrinoma, US had a sensitiv a specificity of 94%, a positive predictive value of 100%, and a negative predictive value of 25%. US enabled detection of tumor in eight cases not detected with CT and inf detected with angiography. Specificity for extrahepatic gastrinoma was similar for all three modalities (89%-95%). CT and US were equally effective for the detection of ext gastrinoma, and angiography was significantly more effective than both US and CT (P less than .01). The authors conclude that US, although of low sensitivity, remains usefu initial imaging modality in patients with Zollinger-Ellison syndrome.
3	Zollinger-Ellison syndrome. Relation to Helicobacter pylori-associated chronic gastritis and gastric acid secretion. Since Helicobacter pylori infects the gastric mucosa in mc patients with chronic duodenal ulcer, infection with this organism has been implicated in the pathogenesis of this common disease. We postulated that if H. pylori is pathog the usual type of duodenal ulcer, it should be less common when duodenal ulcer has another specific etiology, such as Zollinger-Ellison syndrome. Gastric mucosa was com

Tablero de control

Distr. Problems described



Cantidad de Datos

12 mil

probl...	medical_abstracts
	octretide would alter hepatic bile composition and because gallbladder stasis, thereby increasing gallbladder bile solute concentrations. Fourteen control prairie dogs recei saline injections, whereas 10 animals received 1 micrograms of octreotide subcutaneously three times per day for 5 days. Cholecystectomy and common bile duct cannulati then performed. Octreotide increased hepatic bile concentrations of bilirubin monoglucuronide (p less than 0.05), total bilirubin (p less than 0.05), and total protein (p less t Rsa, an index of gallbladder stasis, was decreased (p less than 0.01) in the octreotide group. Gallbladder bile total calcium (p less than 0.05), bilirubin monoglucuronide (p l 0.05), total bilirubin (p less than 0.01), total protein (p less than 0.05), and total lipids (p less than 0.05) were increased in the octreotide group. Animals receiving octreotide decreased hepatic (p less than 0.05) and gallbladder (p less than 0.001) bile pH. No differences in cholesterol saturation index were observed. These data suggest that in the dog, octreotide (1) alters hepatic bile composition, (2) causes gallbladder stasis, and (3) increases gallbladder bile calcium, bilirubin, protein, lipid, and hydrogen ion concer We conclude that octreotide causes alterations in gallbladder bile composition that increase the likelihood of cholesterol and calcium bilirubinate precipitation.
4	The leukocyte count: a predictor of hypertension. In an exploratory study of 1031 persons observed to progress from normotension to essential hypertension and 1031 matc subjects who remained normotensive, the initial leukocyte count (WBC) was found to be related to the development of hypertension, with risk increased 40% (95% confiden 12-82%) in persons in the highest as compared to the lowest quartile of WBC. This relationship proved to be largely independent of body mass index, body fat distribution,

05

Conclusiones

Conclusiones

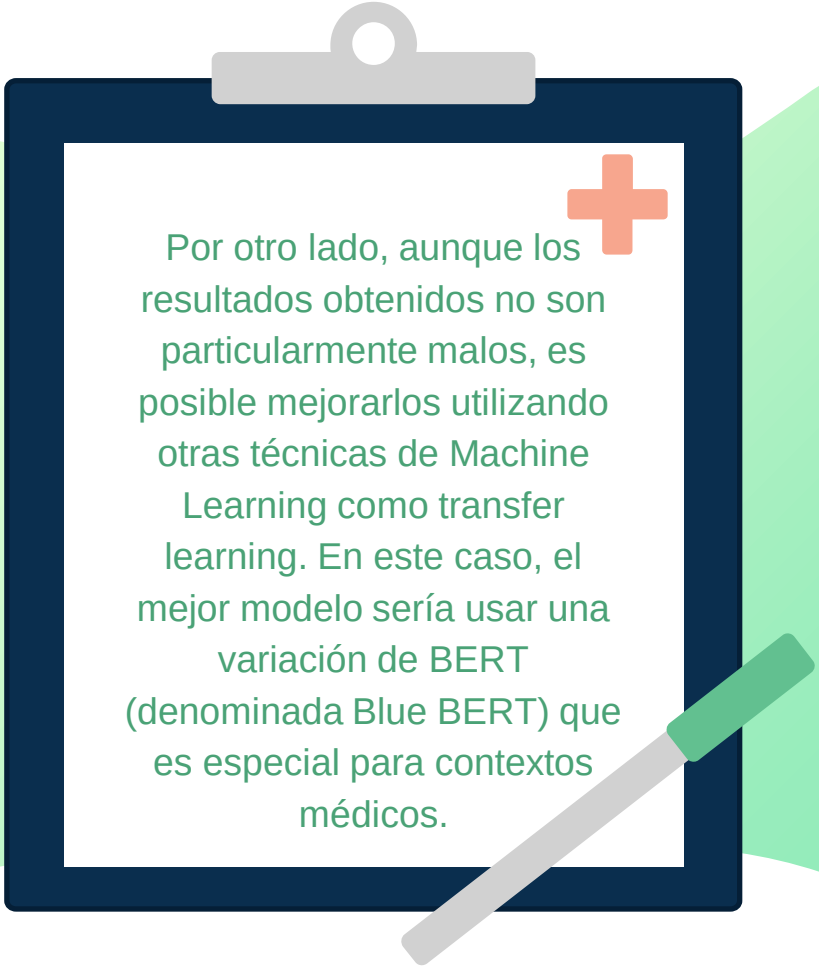
Los textos de la categoría 5 son muy amplios, según nuestro perfilamiento. Esto significa que pueden caer en muchas de las otras clases, arruinando las métricas de precisión. Por esto, los algoritmos tienen problemas al momento de clasificarlos. Se recomienda eliminar la categoría 5 o intentar distribuirla en las demás enfermedades.

Conclusiones

La precisión para la clase minoritaria (la clase 2) es mucho más baja que en los demás (47%). Por lo que se recomienda a los médicos revisar los textos que se predigan de esta clase, pues muchos estarían mal clasificados.

.

Conclusiones



Por otro lado, aunque los resultados obtenidos no son particularmente malos, es posible mejorarlos utilizando otras técnicas de Machine Learning como transfer learning. En este caso, el mejor modelo sería usar una variación de BERT (denominada Blue BERT) que es especial para contextos médicos.

Gracias!