



Fundamentos de Machine y Deep Learning

Ejercicio N° 3 (Análisis de emociones en texto no-estructurado (NLP))
08 / 10 / 2023

INTEGRANTES

- Camila Constanza Aguilera Bustamante
- Wladimir Richard Parada Rebolledo
- Néstor Patricio Rojas Ríos
- Ramiro Andrés Uribe Garrido



Pregunta 1.1

¿Cuánto mejora el rendimiento de los modelos de análisis de sentimiento (implementados en la siguiente sección) al agregar las técnicas de normalización? La idea es comparar el rendimiento con y sin las técnicas de normalización, analizando cuantitativamente las posibles mejoras en cada uno de los dos modelos (a nivel de sus indicadores de desempeño).

Al comparar el modelo de **Árbol de decisiones** con el corpus de vocabulario con y sin normalizar apenas se observan diferencias a favor del primero: la *exactitud* se mantiene invariable y, al analizar los sentimientos, vemos que en **alegría** la *sensibilidad* y la *especificidad* suben y bajan respectivamente, pero menos de 1 punto porcentual cada uno, logrando apenas mejora en la diferencia de ambos indicadores; para **enojo** *sensibilidad* disminuye y *especificidad* aumenta, disminuyendo la gran diferencia que se observa entre ambos indicadores, pero también con diferencias menores a 1 punto porcentual en cada variación; finalmente, **sorpresa** no es detectado con ninguno de los 2 corpus, así que sus valores no cambian.

Por otro lado, al evaluar el modelo de **Red neuronal con 1 capa intermedia** sorpresivamente el **corpus sin normalización presenta mejor desempeño**: la *exactitud* sube 1,35 puntos porcentuales; en **alegría** la *sensibilidad* sube y la *especificidad* baja, disminuyendo la diferencia entre ambas, con variaciones cercanas a los 2 puntos porcentuales cada uno; para **enojo** y **sorpresa** tanto *sensibilidad* como *especificidad* mejoran levemente. Ver resultados en tabla 1.

Pregunta 1.2

¿Cuál es el modelo que mejora más? Analice y describa qué características de este modelo justifican ser el que mejora con las técnicas de normalización, más que el otro.

Primero que nada, queremos destacar el hecho de que para la **Red neuronal de 1 capa intermedia** el corpus sin normalización tuvo mejor desempeño; lo más probable es que el número que escogimos como semilla (`set.seed(1234)`) para estandarizar la ejecución de procesos aleatorios del código tenga relación con este fenómeno. De hecho, este fue el modelo con mejor rendimiento de esta comparación. Ver resultados en tabla 1.

Dicho esto, el modelo que mejoró su desempeño al aplicar las técnicas de normalización fue el **Árbol de decisiones**. Al parecer esta mejora no tiene relación con el modelo en sí, sino que con la generación de los datasets de validación, pues al comparar los árboles de ambos corpus se produce la misma ramificación, cambiando sólo la palabra entre su forma original y la forma normalizada. Ver las figuras 1 y 2.



Pregunta 2.1

¿Cuál es el valor usado en `removeSparseTerm()` para lograr mejores resultados en los modelos de clasificación? Realice una comparación de 3 valores: 0,990, 0,995, 0,998.

Tal como se observa en la tabla 2, el valor de *dispersión* (*Sparse*) que presenta mejor resultados en las métricas es **0,998**, tanto para el **Árbol de decisiones** como para la **Red neural de 1 capa intermedia**. En el caso del **Árbol de decisiones** se observa el mayor valor de *exactitud* sin una gran variación entre las diferencias de *sensibilidad* y *especificidad* para los tres sentimientos. Por otro lado, en el caso de **Red neuronal de 1 capa intermedia** si bien se sacrifica un poco de *exactitud*, la brecha existente entre *sensibilidad* y *especificidad* disminuye notoriamente para los tres sentimientos, por lo que lo consideramos el mejor modelo en esta comparación.

Pregunta 2.2

¿Cómo se interpreta en este caso que mejore o empeore con más y a veces con menos cantidad de palabras (o columnas del dataset)?

En nuestra experiencia a mayor cantidad de palabras mejor desempeño presentaron los modelos:

- Sparse 0,990: 71 palabras.
- Sparse 0,995: 195 palabras.
- Sparse 0,998: 625 palabras.

Es probable que esta diferencia se explique dado que si el dataset es mayor los modelos tendrán más elementos para identificar la complejidad asociada a la expresión verbal de emociones, aprendiendo la mejor manera de clasificar los mensajes de Twitter en base al sentimiento.

Pregunta 3.1

Considere cambiar los parámetros de funcionamiento de la red neuronal (*size*, *maxit*) buscando su mejor desempeño e indique los valores de estos dos parámetros que resultaron en mejores resultados.

Para encontrar el mejor desempeño para el modelo de **Red neuronal con 1 capa intermedia** consideramos los 2 factores anteriormente estudiados: el **corpus de vocabulario no normalizado** y una **dispersión de 0,995**. Teniendo esto presente es que se probó con valores de *nodos* entre 5 y 50 e *iteraciones* entre 10 y 100, obteniendo el mejor desempeño del modelo con **45 nodos** y **20 iteraciones** dado que se obtuvo una de las puntuaciones más altas en *exactitud* entre todas las combinaciones probadas, presentando además una menor diferencia entre *sensibilidad* y *especificidad* considerando los tres sentimientos estudiados. Ver los resultados en la tabla 3.



Pregunta 3.2

Teniendo este resultado se ve que de los dos modelos, hay uno con accuracy mejor que el otro, pero en forma más completa y precisa ¿cuál de los dos se puede considerar el mejor modelo al mirar su desempeño más completo y por qué?

Los 2 modelos que se utilizaron como base comparativa fueron el **Árbol de decisiones** con *sparse 0,998* y *corpus normalizado* y la **Red neuronal con 1 capa intermedia, 45 nodos, 20 iteraciones, sparse 0,995** y *corpus sin normalizar*. En base a los resultados obtenidos al momento de ejecutar ambos modelos, si se considera solamente el valor de la *exactitud*, **Red neuronal con 1 capa intermedia** fue superior al **Árbol de decisiones**.

Poniendo el foco en *especificidad* y *sensibilidad* hay que hacer la diferencia para cada uno de los sentimientos: para **alegría** el **Árbol de decisiones** prácticamente no la logra identificar (*sensibilidad* muy baja) mientras que es muy bueno excluyéndola (*especificidad* muy alta); en cambio, la **Red neuronal** presenta un mejor balance entre estos indicadores estando ambos sobre el 50%. Con **enojo** el **Árbol de decisiones** presenta un escenario inverso al anterior pues es muy bueno detectando cuando está presente (*alta sensibilidad*) pero muy malo excluyéndolo (*baja especificidad*); por otro lado, la **Red neuronal** muestra también un comportamiento más balanceado con poca diferencia entre *sensibilidad* y *especificidad*, estando ambos también sobre el 50%. Por último, para **sorpresa** el **Árbol de decisiones** es incapaz de detectarla en todos los escenarios probados; en contraste, la **Red neuronal**, si bien presenta *baja sensibilidad* y *muy alta especificidad*, si logra detectar algo de su existencia.

Pregunta 3.3

Se puede observar que el Árbol de Decisión tiene un muy mal desempeño en reconocer una de esas clases ¿cuál y por qué podría darse esto? (En otras palabras ¿qué característica de este modelo puede provocar que en este caso una de las clases no sea factible de predecir?)

El mal desempeño del **Árbol de decisión** puede estar relacionado a lo complejo que puede ser el análisis de las emociones. La clase **sorpresa** no fue identificada por ninguno de los árboles generados durante todo el desarrollo de este trabajo (obtuvo **0 verdaderos positivos**). Es posible que este suceso esté relacionado con la forma en que se produce la división de ramas en el modelo, pues lo que se intenta es buscar la mayor *homogeneidad* (menor *impureza*) en los nodos resultantes. Pues bien, independiente de la fórmula que utilicemos para calcular la *impureza* (Gini, entropía, etc.), todas ponderan su cálculo con la proporción en que las clases a predecir están presentes. La clase **sorpresa**, al ser poco frecuente en el dataset, termina teniendo poca relevancia en el cálculo final y, por consiguiente, no sería considerada para hacer las divisiones.



Anexos

ÁRBOL DE DECISIÓN								
Normalización del corpus	Dispersión (Sparse)	Exactitud (%)	Sensibilidad (%)			Especificidad (%)		
			alegría	enojo	sorpresa	alegría	enojo	sorpresa
<i>Sí</i>	0,995	43,84	3,82	99,77	0,00	99,83	2,76	1,00
<i>No</i>	0,995	43,84	3,70	99,89	0,00	99,92	2,68	1,00
RED NEURONAL								
Normalización del corpus	Dispersión (Sparse)	Exactitud (%)	Sensibilidad (%)			Especificidad (%)		
			alegría	enojo	sorpresa	alegría	enojo	sorpresa
<i>Sí</i>	0,995	45,68	49,36	53,71	14,59	63,20	53,43	92,94
<i>No</i>	0,995	47,03	51,56	54,05	16,41	61,86	56,03	93,57

Tabla 1. Comparativa de métricas con corpus normalizado y sin normalizar, para ambos modelos

ÁRBOL DE DECISIÓN								
Normalización del corpus	Dispersión (Sparse)	Exactitud (%)	Sensibilidad (%)			Especificidad (%)		
			alegría	enojo	sorpresa	alegría	enojo	sorpresa
<i>Sí</i>	0,990	43,84	3,82	99,77	0,00	99,83	2,76	1,00
<i>Sí</i>	0,995	43,84	3,82	99,77	0,00	99,83	2,76	1,00
<i>Sí</i>	0,998	45,63	3,14	99,89	0,00	99,76	2,43	1,00
RED NEURONAL								
Normalización del corpus	Dispersión (Sparse)	Exactitud (%)	Sensibilidad (%)			Especificidad (%)		
			alegría	enojo	sorpresa	alegría	enojo	sorpresa
<i>Sí</i>	0,990	45,87	32,72	73,20	7,60	77,61	34,67	95,92
<i>Sí</i>	0,995	45,68	49,36	53,71	14,59	63,20	53,43	92,94
<i>Sí</i>	0,998	45,53	49,58	49,02	25,16	61,92	57,43	90,57

Tabla 2. Comparativa de métricas con distintos niveles de dispersión (Sparse), para ambos modelos



Sentimiento	Nodos	Iteraciones	Exactitud (%)	Sensibilidad (%)	Especificidad (%)
<i>Enojo</i>	20	60	44,91	46,41	62,40
<i>Alegría</i>	20	60	44,91	50,87	62,11
<i>Sorpresa</i>	20	60	44,91	25,23	86,51
<i>Enojo</i>	50	20	47,37	50,63	60,05
<i>Alegría</i>	50	20	47,37	55,95	57,13
<i>Sorpresa</i>	50	20	47,37	16,11	94,49
<i>Enojo</i>	10	30	46,74	53,14	55,44
<i>Alegría</i>	10	30	46,74	51,10	64,26
<i>Sorpresa</i>	10	30	46,74	18,24	91,96
<i>Enojo</i>	20	30	47,03	54,05	56,03
<i>Alegría</i>	20	30	47,03	51,56	61,86
<i>Sorpresa</i>	20	30	47,03	16,41	93,57
<i>Enojo</i>	45	20	48,38	54,96	57,79
<i>Alegría</i>	45	20	48,38	55,14	59,29
<i>Sorpresa</i>	45	20	48,38	13,07	95,75
<i>Enojo</i>	15	30	44,52	53,36	53,18
<i>Alegría</i>	15	30	44,52	44,86	65,59
<i>Sorpresa</i>	15	30	44,52	20,06	89,95
<i>Enojo</i>	35	20	47,66	56,44	54,52
<i>Alegría</i>	35	20	47,66	51,45	63,10
<i>Sorpresa</i>	35	20	47,66	14,29	94,49
<i>Enojo</i>	5	10	47,46	54,39	53,85
<i>Alegría</i>	5	10	47,46	57,80	56,47
<i>Sorpresa</i>	5	10	47,46	1,82	99,31
<i>Enojo</i>	20	20	48,09	60,89	50,17
<i>Alegría</i>	20	20	48,09	48,67	66,42
<i>Sorpresa</i>	20	20	48,09	12,46	95,69

Tabla 3. Muestra de la comparativa de métricas de Redes neuronales según Nodos e Iteraciones

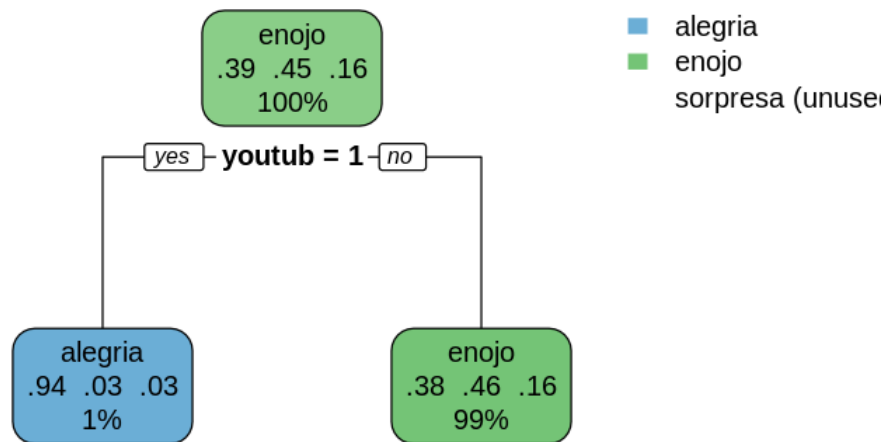


Figura 1. Resultados Árbol de Decisión con Corpus normalizado y Sparse 0,995

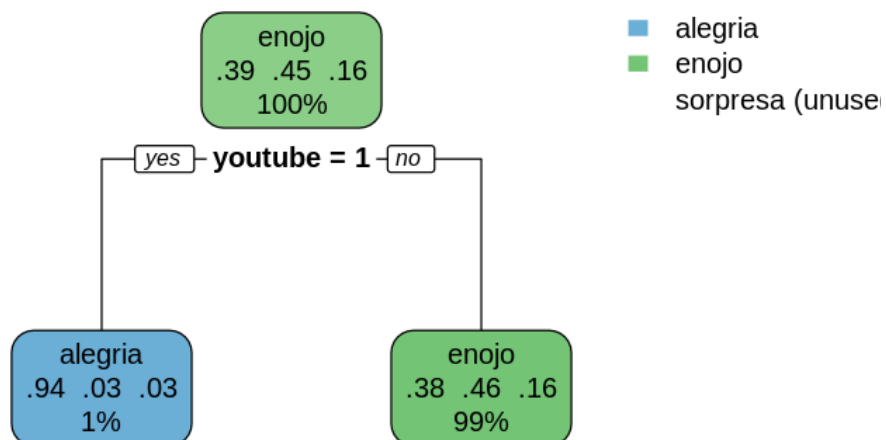


Figura 2. Resultados Árbol de Decisión con Corpus sin normalizar y Sparse 0,995