

INFO279

Tratamiento automático del lenguaje

Proyecto de análisis del discurso de los medios de prensa

Estudiante:

José Nicolás Aillapi Gomez

Carrera:

Ingeniería Civil Informática

Profesor:

Matthieu Vernier

Código asignatura:

INFO279

Índice

Introducción	- 3
RESULTADOS	- 5
Análisis de noticias generales	- 5
Análisis de noticias “Positivas”	- 10
Análisis de noticias “Negativas”	- 13
Análisis de noticias con palabra clave “SALUD”	- 16
Análisis de noticias positivas/negativas con palabra clave “SALUD”	- 19
Conclusiones y fuente	- 20

Introducción

En el presente informe usaremos técnicas de tratamiento automático del lenguaje (**TAL**) para transformar datos de medios de prensa regionales en indicadores territoriales de interés para el Gobierno Regional de Los Lagos (**GORE**)

En el análisis de los indicadores propuestos vamos a recolectar dos cuerpos o datasets de noticias a partir del motor de búsqueda **Sophia**, el primer dataset tendrá noticias generales de la región de los lagos y el segundo estará conformado de noticias de la región de los lagos que hablan de la palabra clave elegida “Salud”. La recolección del dataset de noticias pertenece al periodo del tercer cuatrimestre de 2021 (2021/09/01 hasta el 2021/12/31), la palabra clave elegida puede ser reemplazada por cualquier otro término de interés para el **GORE**.

Procedimiento

Crearemos un análisis separado para:

- El dataset de noticias en **general**
- El dataset de noticias con palabra clave “SALUD”
 - Los datasets de positividad/negatividad derivadas de los datasets mencionados arriba

En el análisis de **noticias generales**, se recolectaron 4.257 noticias en el tercer cuatrimestre de 2021 que mencionan comunas de la región de los lagos, en el dataset se utilizarán las técnicas TAL vistas en clase, tales como extracción de conceptos claves, entidades, análisis de tópicos, word embedding y sentiment analysis.

Con la ayuda de PhraseMatcher de la librería **spaCy** y **geoPandas** resaltaremos en un mapa las comunas de la región de los lagos, la intensidad del color significa la cantidad de menciones de noticias sobre la comuna en cuestión, mediante **consultas SQL** sobre datasets, crearemos tablas con el ranking de comunas por menciones de noticias, además del ranking de medios de prensa por noticias publicadas, además, utilizando análisis de tópicos (**LDA**) representaremos mediante

graficos los 6 tópicos más recurrentes con sus asociados términos, en este caso el modelo óptimo es de 12 tópicos para optimizar la métrica de coherencia, pero decidí utilizar 6 tópicos para ejemplificar tal función del proyecto. Posteriormente utilizaremos estos análisis en los datasets derivados de positividad y negatividad de noticias, además del dataset de noticias con palabra clave SALUD.

Para la creación de los datasets de **positividad y negatividad** de las **noticias generales** utilizaremos Sentiment analysis mediante el uso de redes neuronales transformers, este proceso concluirá un “sentimiento” a cada noticia las cuales pueden ser positiva (POS), neutra (NEU) o negativa (NEG) de acuerdo al contenido de la noticia, cabe mencionar que el sentimiento que encuentre la red neuronal no es la más precisa, ya que es muy posible que un ser humano encuentre un sentimiento más coherente en varios casos, pero para un gran dataset de noticias este proceso es más eficaz que un humano en términos de esfuerzo. Con la inclusión del fichaje de positividad y negatividad en el dataset general, podremos crear mediante consultas SQL un dataset para las noticias que tengan un sentimiento **POS** o **NEG**, siguiendo el análisis empleado en el dataset principal se consigue representar el ranking de las comunas en términos de positividad / negatividad de noticias, las menciones totales de prensa y para cada dataset crear un **ranking top 2** de tópicos principales en las noticias positivas o negativas para facilitar la ejemplificación.

Para el análisis de **noticias con palabra clave “SALUD”**, recolectamos el dataset desde el motor de búsqueda Sophia con una palabra clave como parámetro de búsqueda, en este caso se encontraron 3092 noticias que referencian “SALUD” en el tercer cuatrimestre de 2021, luego utilizaremos el mismo análisis realizado en los datasets anteriores, además de crear un dataset de positividad/negatividad. para representar el estado de las menciones de noticias a comunas en “SALUD”.

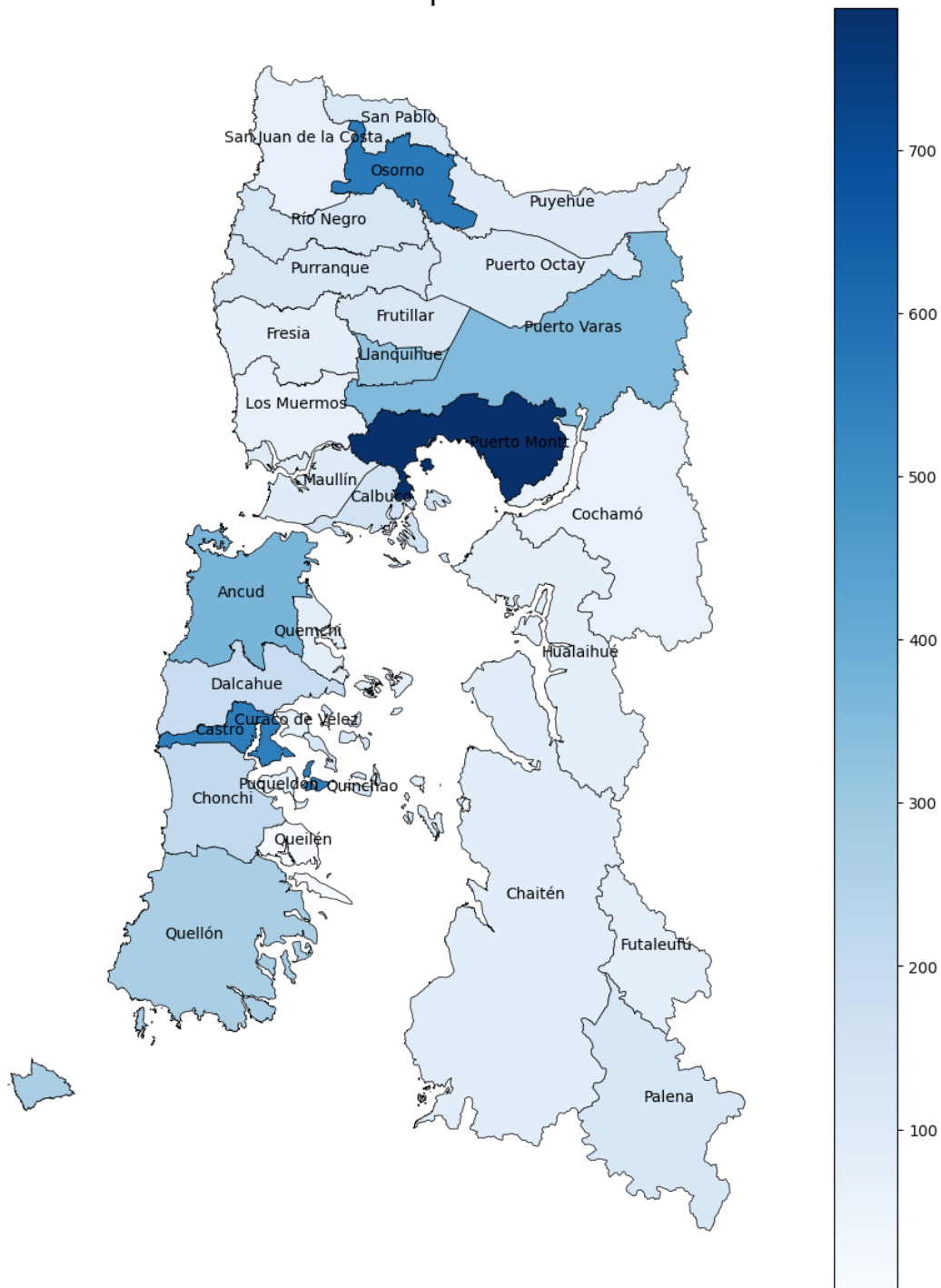
RESULTADOS:

Análisis de noticias generales

Utilizando 4.257 noticias de la región de los lagos del tercer trimestre de 2021 (09/01 hasta el 12/31)

Mapa general

Menciones totales por comuna



RegionLosLagos_2021-09-01_2021-12-31

Ranking de comunas por noticias

Ranking	Comuna	Valor
0	Puerto Montt	787
1	Osorno	566
2	Castro	554
3	Ancud	368
4	Puerto Varas	354
5	Llanquihue	315
6	Quellón	266
7	Chonchi	209
8	Dalcahue	187
9	Calbuco	153
10	Río Negro	132
11	Quinchao	129
12	Palena	125
13	Frutillar	124
14	Purranque	123
15	San Pablo	121
16	Puerto Octay	112
17	Puyehue	102
18	Mauñín	94
19	Chaitén	88
20	Quemchi	83
21	Futaleufú	79
22	Puqueldón	78
23	Fresia	77
24	Curaco de Vélez	74
25	Hualaihué	74
26	San Juan de la Costa	67
27	Los Muermos	61
28	Cochamó	60
29	Queilén	2

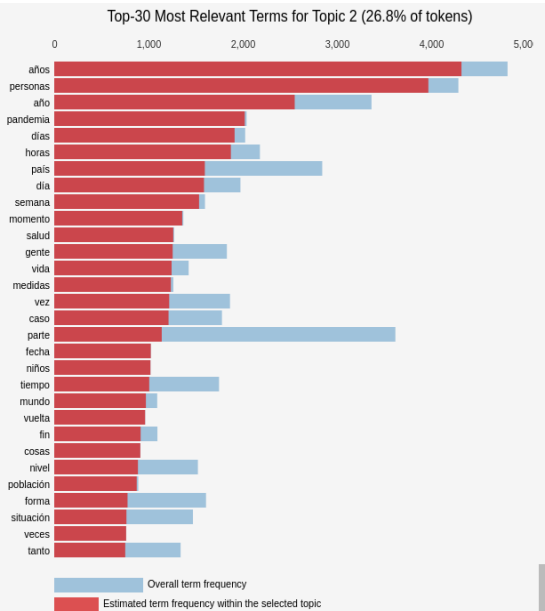
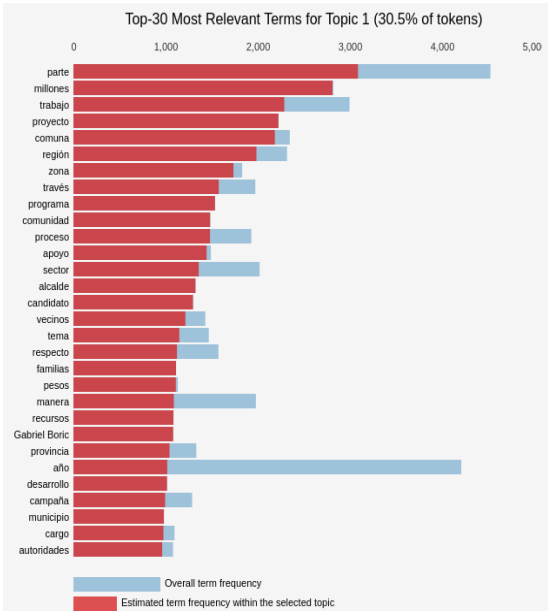
Ranking de medios de prensa en base a noticias publicadas

Ranking	Prensa	Noticias Publicadas
0	elaustral	1358
1	laestrelladechiloe	1320
2	elllanquihue	1277
3	radiosago	339
4	elheraldoaustral	156
5	seminariolocal	70
6	elcalbucano	23
7	elvacanudo	8
8	elrepuertero	8

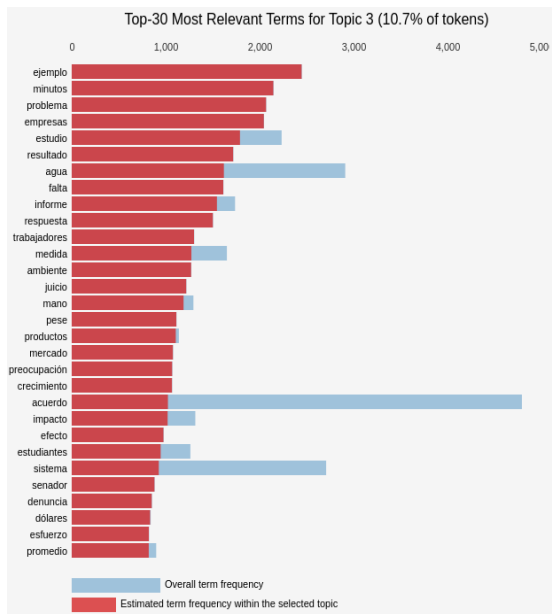
Ranking TOP 6 de tópicos presentes en las noticias y sus términos asociados

Los 30 términos más destacadas del tópico 1

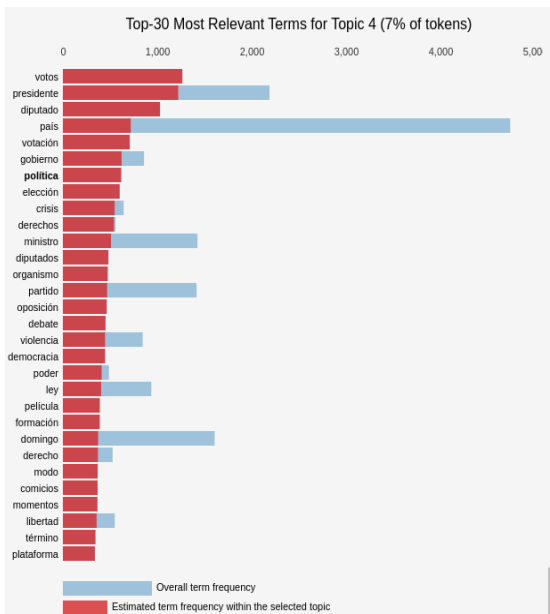
Los 30 términos más destacadas del tópico 2



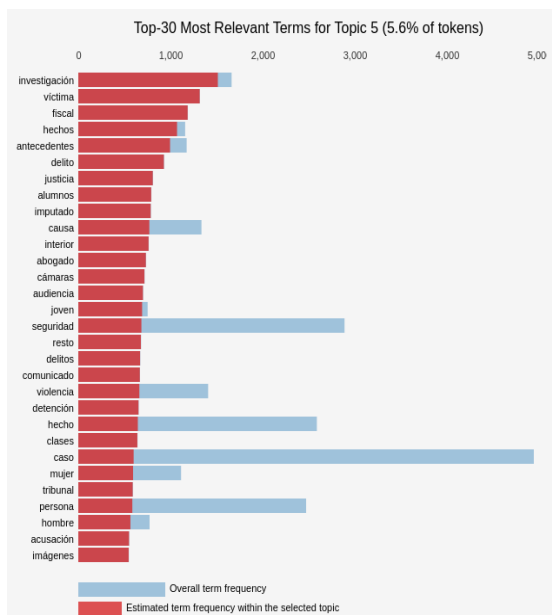
Los 30 términos más destacadas del tópico 3



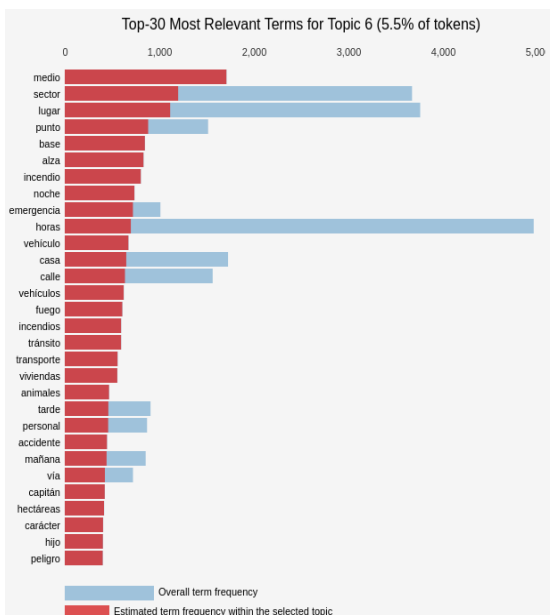
Los 30 términos más destacadas del tópico 4



Los 30 términos más destacadas del tópico 5



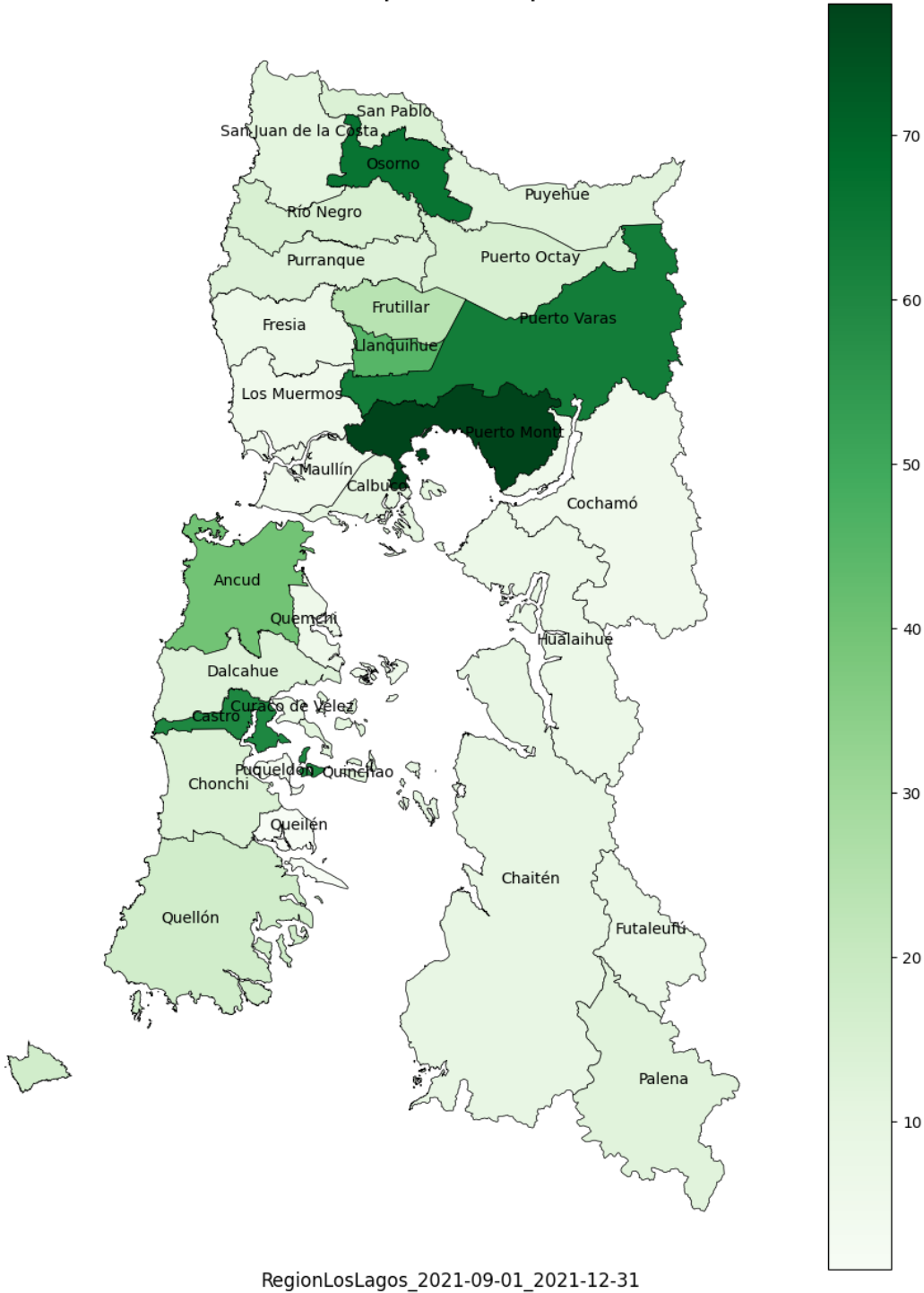
Los 30 términos más destacadas del tópico 6



Análisis de noticias “Positivas”

Mapa de noticias positivas

Menciones de noticias positivas por comuna



Ranking de comunas por noticias positivas

Ranking	Comuna	Valor
0	Puerto Montt	78
1	Osorno	66
2	Puerto Varas	63
3	Castro	60
4	Llanquihue	45
5	Ancud	40
6	Frutillar	24
7	Quellón	17
8	Puerto Octay	15
9	Río Negro	15
10	Chonchi	14
11	San Pablo	14
12	Dalcahue	13
13	Purranque	13
14	Palena	12
15	Puyehue	12
16	Quinchao	11
17	San Juan de la Costa	11
18	Calbuco	10
19	Chaitén	9
20	Curaco de Vélez	8
21	Futaleufú	8
22	Fresia	7
23	Hualaihué	7
24	Quemchi	6
25	Cochamó	5
26	Los Muermos	5
27	Mauñín	5
28	Puqueldón	3
29	Queilén	1

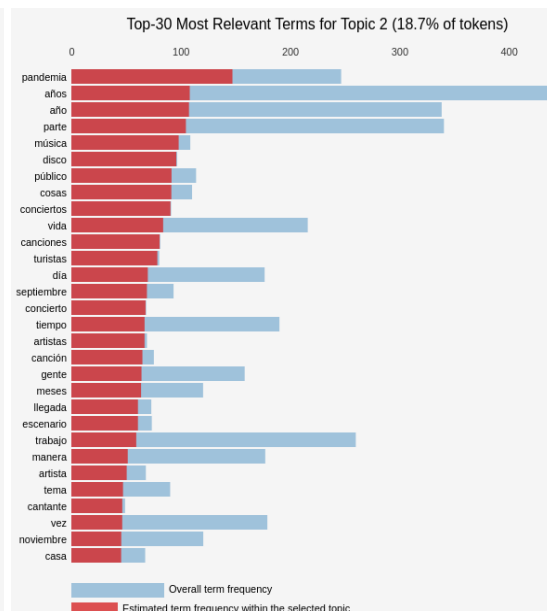
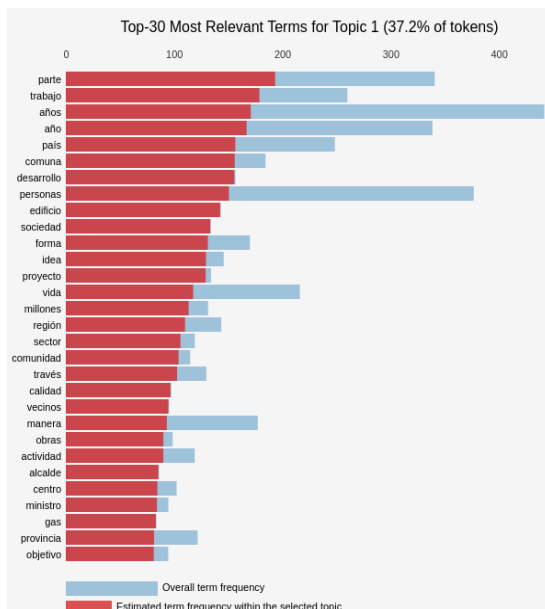
Ranking de medios en base a noticias POSITIVAS publicadas

Ranking	Prensa	Noticias Positivas publicadas
0	laestrelladechiloe	196
1	elaustral	175
2	elllanquihue	159
3	elheraldoaustral	42
4	radiosago	16
5	seminariolocal	11
6	elcalbucano	6
7	elvacanudo	2
8	elrepuertero	2

Ranking TOP 2 de tópicos presentes en las noticias POSITIVAS y sus términos asociados

Los 30 términos más destacadas del tópico 1

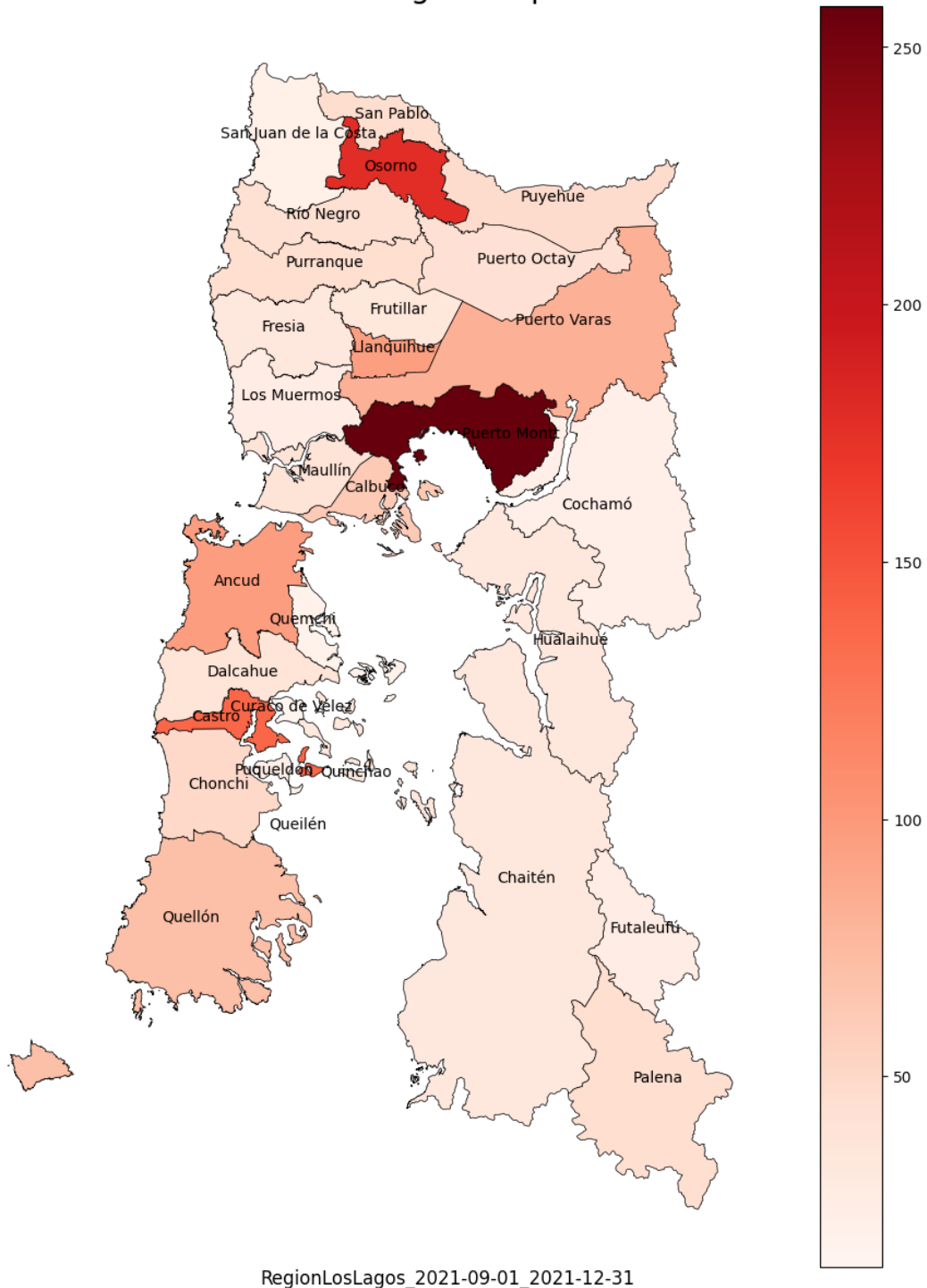
Los 30 términos más destacadas del tópico 2



Análisis de noticias “Negativas”

Mapa de noticias negativas

Menciones de noticias negativas por comuna



Ranking de comunas por noticias negativas

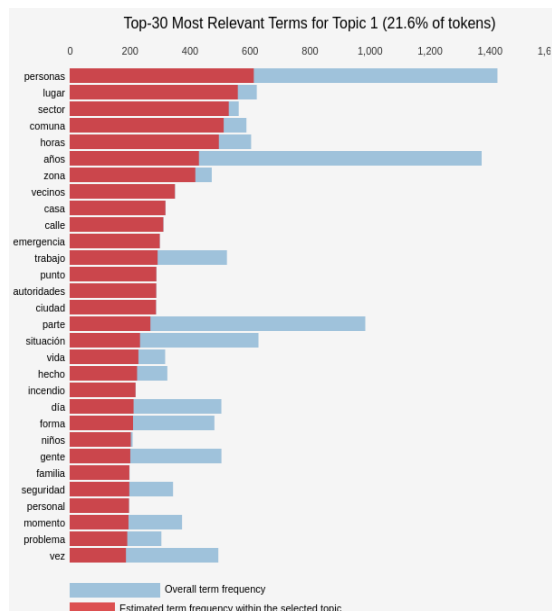
Ranking	Comuna	Valor
0	Puerto Montt	258
1	Osorno	178
2	Castro	138
3	Ancud	97
4	Llanquihue	96
5	Puerto Varas	82
6	Quellón	71
7	Calbuco	62
8	Chonchi	50
9	Puyehue	47
10	San Pablo	46
11	Palena	45
12	Purranque	45
13	Río Negro	43
14	Puerto Octay	42
15	Maullín	39
16	Dalcahue	37
17	Frutillar	35
18	Chaitén	34
19	Fresia	34
20	Hualaihué	32
21	Quinchao	31
22	Futaleufú	26
23	Los Muermos	25
24	Cochamó	22
25	Puqueldón	22
26	San Juan de la Costa	21
27	Quemchi	18
28	Curaco de Vélez	13
29	Queilén	0

Ranking de medios en base a noticias NEGATIVAS publicadas

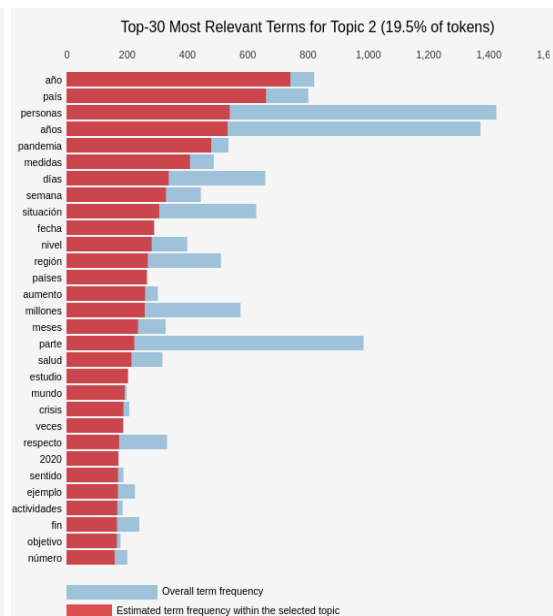
Ranking	Prensa	Noticias negativas publicadas
0	elaustral	420
1	elllanquihue	391
2	laestrelladechiloe	283
3	radiosago	126
4	elheraldoaustral	12
5	seminariolocal	8
6	elvacanudo	2
7	elrepuertero	2
8	elcalbucano	1

Ranking TOP 2 de tópicos presentes en las noticias NEGATIVAS y sus términos asociados

Los 30 términos más destacadas del tópico 1



Los 30 términos más destacadas del tópico 2

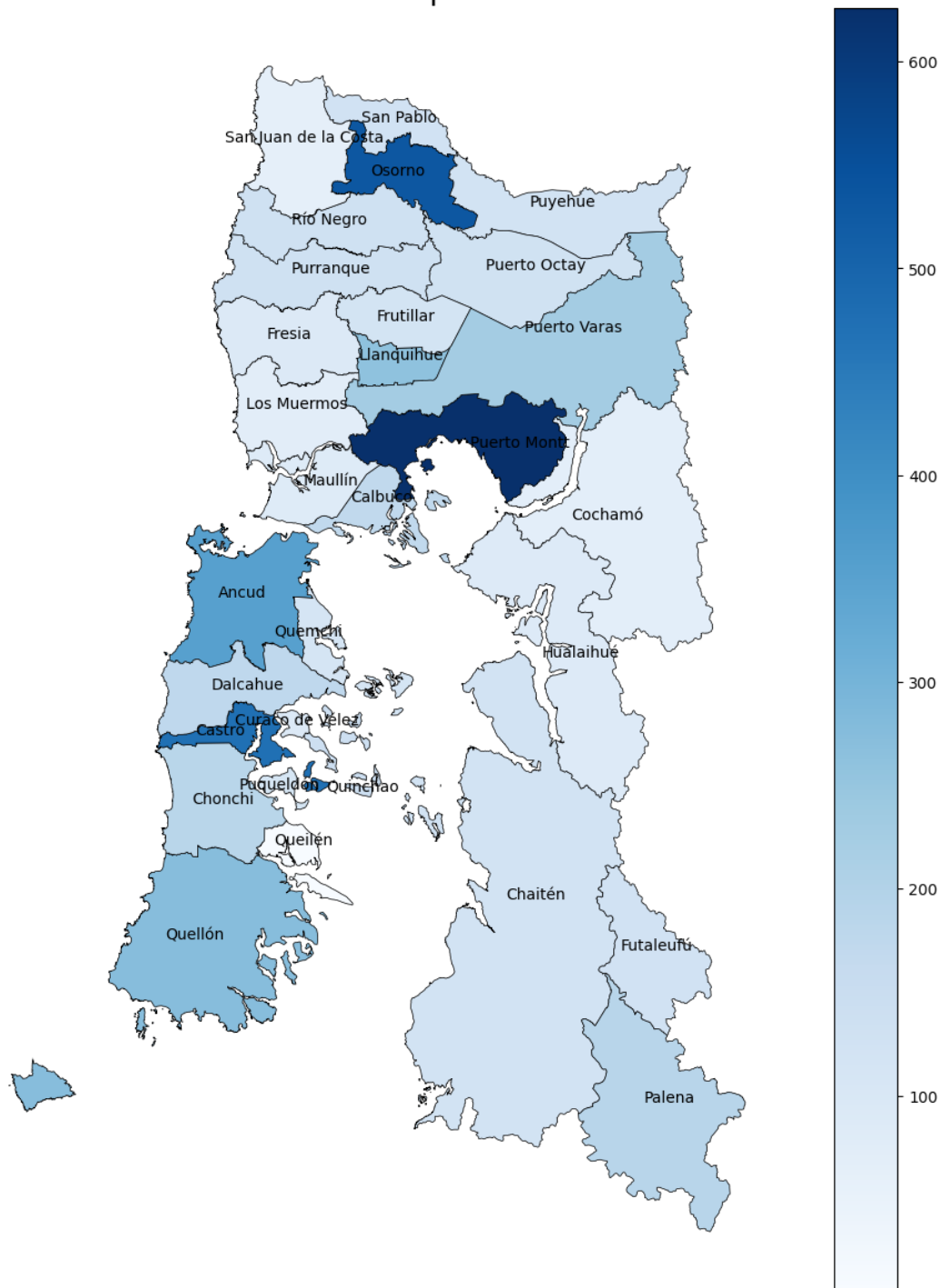


Análisis de noticias con palabra clave: Salud

Utilizando 3092 noticias

Mapa de comunas por noticias referenciando Salud

Menciones totales por comuna



RegionLosLagos_2021-09-01_2021-12-31

Ranking de comunas por noticias con palabra clave SALUD

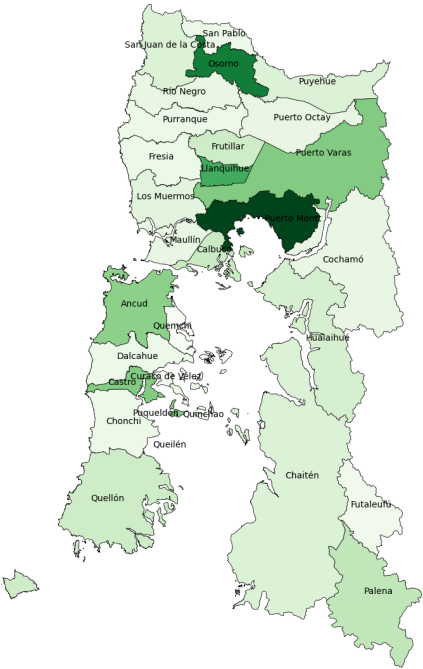
Ranking	Comuna	Valor
0	Puerto Montt	626
1	Osorno	533
2	Castro	473
3	Ancud	353
4	Quellón	274
5	Llanquihue	261
6	Puerto Varas	227
7	Palena	190
8	Chonchi	187
9	Dalcahue	174
10	Calbuco	172
11	Quinchao	140
12	Purranque	133
13	Río Negro	132
14	Puerto Octay	129
15	Futaleufú	125
16	San Pablo	125
17	Chaitén	124
18	Puyehue	123
19	Frutillar	119
20	Quemchi	114
21	Puqueldón	109
22	Curaco de Vélez	92
23	Fresia	92
24	Hualaihué	87
25	Mauñín	84
26	Los Muermos	73
27	Cochamó	67
28	San Juan de la Costa	62
29	Queilén	6

Ranking de medios de prensa en base a noticias publicadas con palabra clave SALUD

Ranking	Prensa	Noticias publicadas
0	elaustral	60
1	laestrelladechiloe	56
2	elllanquihue	55
3	elheraldoaustral	24
4	elvacanudo	14
5	elrepuertero	14
6	seminariolocal	10
7	radiosago	10
8	elinsular	10
9	radiopudeto	7
10	radioacogida	6
11	elhuemul	6
12	elquellonino	4
13	elcalbucano	4
14	fresiaahora	2

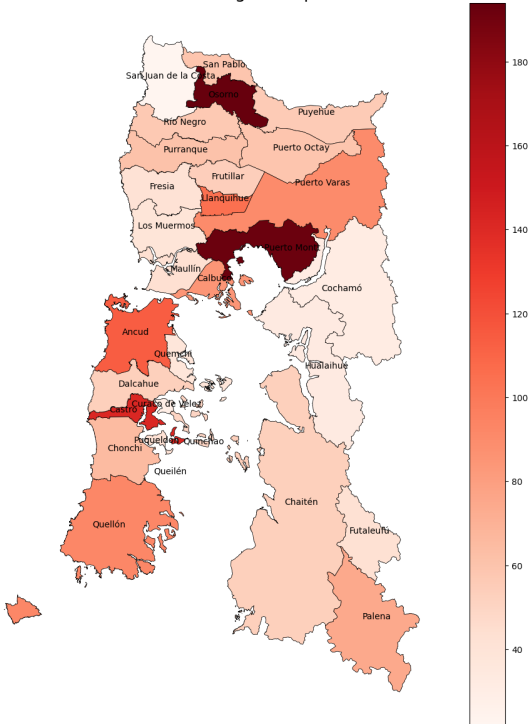
Mapa de comunas por noticias positivas/negativas palabra clave SALUD

Menciones de noticias positivas por comuna



RegionLosLagos_2021-09-01_2021-12-31

Menciones de noticias negativas por comuna



RegionLosLagos_2021-09-01_2021-12-31

Análisis de noticias positivas/negativas con palabra clave SALUD

Ranking comunas de noticias positivas de SALUD

Ranking	Comuna	Valor
0	Puerto Montt	39
1	Osorno	32
2	Llanquihue	25
3	Castro	19
4	Puerto Varas	19
5	Ancud	18
6	Palena	12
7	Calbuco	11
8	Frutillar	10
9	Quellón	10
10	Hualaihué	9
11	Chaitén	8
12	Puyehue	8
13	San Juan de la Costa	8
14	Los Muermos	7
15	Cochamó	6
16	Mauullín	6
17	Puerto Octay	6
18	Río Negro	6
19	Chonchi	5
20	Dalcahue	5
21	Fresia	5
22	Purranque	5
23	Quinchao	5
24	San Pablo	5
25	Futaleufú	4
26	Curaco de Vélez	3
27	Quemchi	2
28	Puqueldón	0
29	Queilén	0

vs

Ranking comunas de noticias negativas de SALUD

Ranking	Comuna	Valor
0	Osorno	194
1	Puerto Montt	192
2	Castro	142
3	Ancud	114
4	Llanquihue	105
5	Quellón	92
6	Puerto Varas	90
7	Calbuco	85
8	Palena	75
9	Chonchi	65
10	Purranque	61
11	Puerto Octay	59
12	San Pablo	59
13	Río Negro	57
14	Puyehue	56
15	Frutillar	54
16	Chaitén	53
17	Dalcahue	53
18	Quinchao	50
19	Mauullín	44
20	Futaleufú	43
21	Fresia	42
22	Puqueldón	42
23	Los Muermos	39
24	Quemchi	38
25	Hualaihué	33
26	Cochamó	32
27	Curaco de Vélez	30
28	San Juan de la Costa	22
29	Queilén	0

Conclusiones

El uso de Sentiment Analysis sobre medios de comunicación tales como el área de prensa ayuda en el proceso de creación de indicadores interesantes, tal y como denotan en los resultados del proyecto durante el tercer cuatrimestre de 2021 sobre los medios de prensa con mención a la región de los lagos, podemos encontrar tópicos principales referentes a candidatura política o sobre la pandemia del covid-19 además de otros tópicos de interés para el GORE. Esta información obtenida mediante técnicas de tratamiento automático de lenguaje es útil para mantener un seguimiento detallado sobre los indicadores mencionados, y para un mejor monitoreo de urgencias en caso de seleccionar un periodo corto para el análisis.

Repositorio: <https://github.com/FederalITN/INFO279-PROYECTO-TAL>

Fuente:

Clases profesor Matthieu vernier: https://github.com/matthieuvernier/INFO279_2022