



Universidad Austral de Chile

Conocimiento y Naturaleza

Tratamiento automático del lenguaje

Proyecto de análisis del discurso de los medios de prensa

Integrante: Gerson Sebastian Andrade Meza

Carrera: Ingeniería Civil Informática

Código asignatura: INFO279

Profesor: Matthieu Vernier

Fecha: 01-01-2023

Introducción

El Gobierno Regional de Los Lagos (GORE) desea monitorear indicadores territoriales para facilitar sus procesos de toma de decisión, a corto y mediano plazo (“decisiones tácticas”) y a largo plazo (“decisiones estratégicas”).

En este sentido, el GORE está interesado en transformar datos de medios de prensa regionales en indicadores territoriales. Estos indicadores complementarán indicadores cuantitativos ya a su disposición.

Para esto, se programó una rutina que, a través del motor de búsqueda **sophia2**, se tomaron **2182 noticias** de medios de prensa de la región de “Los Lagos”, entre los periodos **01-01-2020 al 01-02-2020**, siendo en total un mes.

Con estas noticias y procesos de tratamiento del lenguaje se cumplieron una serie de objetivos, entre ellos están:

1. Diseñar e implementar un protocolo de análisis de datos que responda a ciertas solicitudes del GORE Los Lagos.
2. Utilizar técnicas de tratamiento automático del lenguaje para transformar datos de medios de prensa en indicadores territoriales.
3. Sintetizar los resultados de análisis a través del presente informe.

Desarrollo del proyecto

1. Búsqueda de las noticias:

Para esta parte, se utilizó el motor de búsqueda de sophia2 para recolectar las 2182 noticias pertenecientes a la región de Los Lagos entre los periodos:

01-01-2020 al 01-02-2020. (Esta rutina se guardó en Import_Data.ipynb)

Parámetros de búsqueda

```
#Parámetros de búsqueda

country="chile"

from_="2020-01-01"

to_="2020-02-01"

media_outlets=["elllanquihue","elaustral","laestrelladechiloe","elheraldoaustrial","radiosago",

               "elrepuertero","elvacanudo","elhuemul","seminariolocal","elquellonino",
               "elinsular",

               "radiopudeto","radioacogida","elcalbucano","segundos33",

               "prensadelestuario","fresiaahora","soychiloe"]

simple_keyword=False
```

2. Extracción de tópicos con LDA.

Latent Dirichlet Allocation (LDA), es básicamente un algoritmo que, a partir de una cantidad de textos, descubre los tópicos que los componen.

En el proyecto, con la ayuda de la función **compute_coherence_values** se definieron la cantidad de tópicos óptima a tomar, luego de eso se construyó el modelo LDA y se guardó en la carpeta modelo.

```
lda_model = gensim.models.ldamodel.LdaModel(corpus=dataset,

                                             id2word=id2word,

                                             num_topics=8,

                                             random_state=100,

                                             update_every=1,

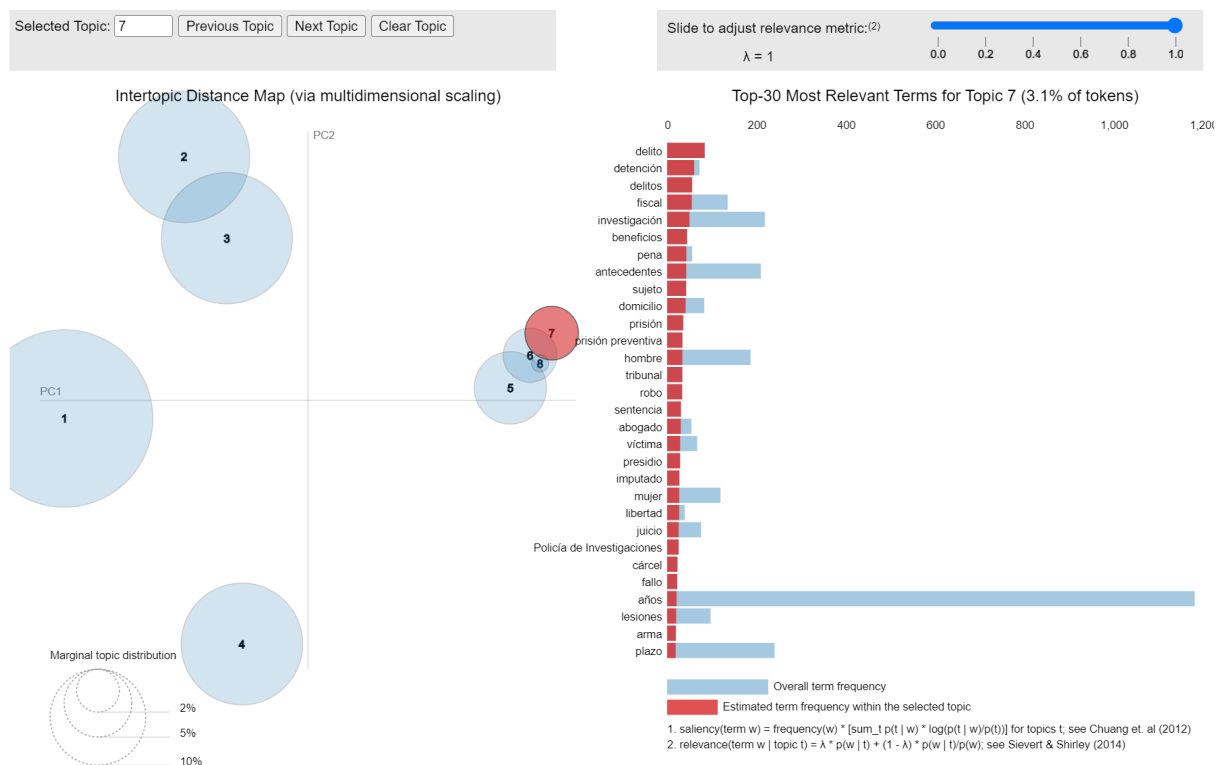
                                             chunksize=100,

                                             passes=10,

                                             alpha='auto',

                                             per_word_topics=True)
```

Se encontraron 8 tópicos diferentes, donde la división entre estos es bastante lógica y servirá para el futuro del proyecto.



3. Análisis de positividad de noticias con Sentiment Analysis

Se midió la positividad de las noticias anteriormente extraídas, asignándole un valor positivo (POS), neutro (NEU) o negativo (NEG) según correspondía respecto al contenido de estas noticias.

```
# --- Cargamos el modelo ---

model_name = "finiteautomata/beto-sentiment-analysis"

model = AutoModelForSequenceClassification.from_pretrained(model_name)

tokenizer = AutoTokenizer.from_pretrained(model_name)

nlp = pipeline("sentiment-analysis", model=model, tokenizer=tokenizer)

df_dominant_topic['title_label'] = ""

df_dominant_topic['title_score'] = ""


for index, row in tqdm(df_dominant_topic.iterrows(), desc='Análisis de
sentimiento', total=df_dominant_topic.shape[0]):

    # Analizamos su sentimiento en inglés

    sentiment_value = nlp(row['title'])

    # Insertamos en dataframe

    df_dominant_topic.at[index, "title_label"] =
sentiment_value[0].get('label')

    df_dominant_topic.at[index, "title_score"] =
sentiment_value[0].get('score')
```

4. Gráfica de mapas en base a parámetros asignados

Se graficaron mapas con la librería GeoPandas para que se entienda de mejor manera la información.

5. Creación de Reportes

Se crearon “Reportes” y se respondieron en base a la información obtenida con la rutina programada en el proyecto.

Indicadores, reportes y secciones

Reporte 1: ¿Qué comunas son las más nombradas en los medios de prensa a lo largo de la región de Los Lagos?

Ventana de tiempo: 1 mes.

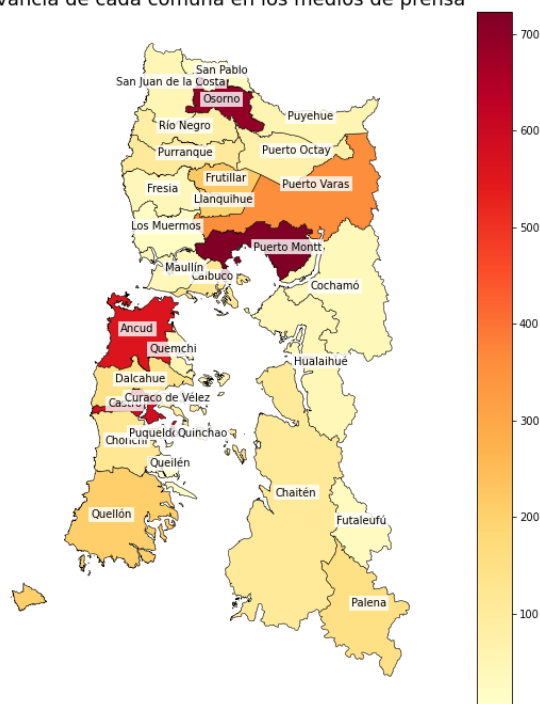
Mapa: Región de Los Lagos.

Relevancia de comunas: En base a la cantidad de noticias en las que se nombran cada comuna, se le asignó una “relevancia”, donde mientras más veces sea nombrada, más relevante son y más sucesos ocurren en ellas.

Las comunas más nombradas en medios de prensa son:

Comuna	N° de veces nombrado en noticias
Puerto Montt	723
Osorno	694
Castro	584
Ancud	560
Puerto Varas	362

Relevancia de cada comuna en los medios de prensa



Los Lagos, Enero 2020 a Febrero 2020, medios locales

Reporte 2: ¿Cuál es la distribución de Positividad, Neutralidad y Negatividad de las noticias en la región de Los Lagos?

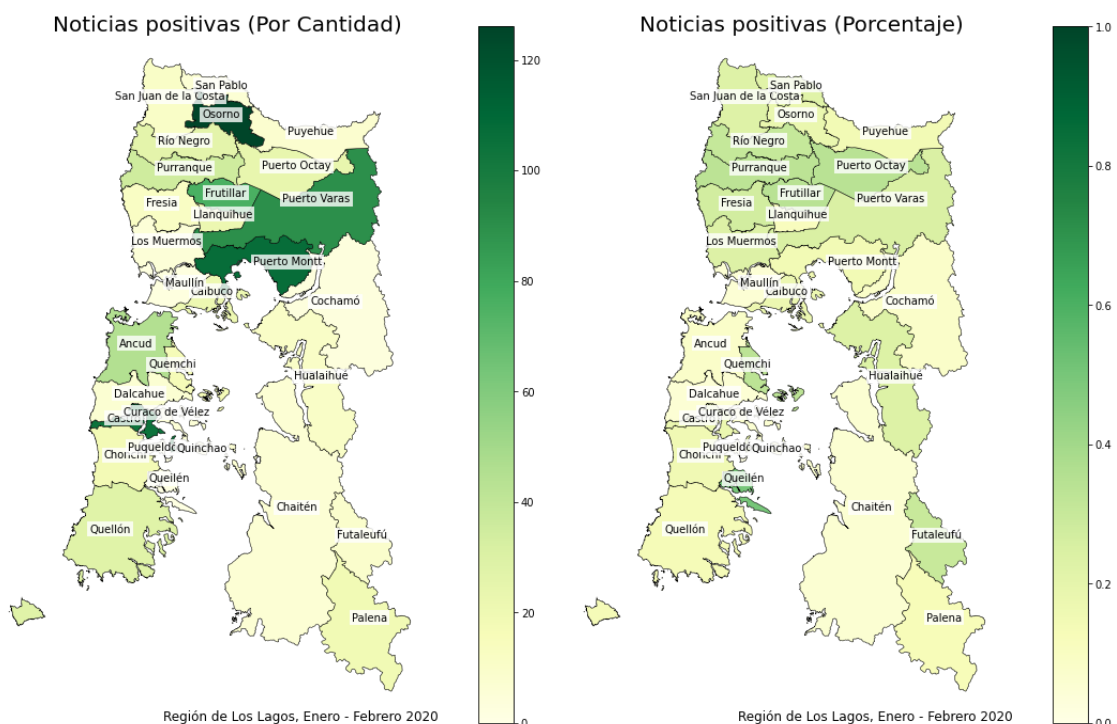
Ventana de tiempo: 1 mes.

Mapa: Región de Los Lagos.

Sección 1: ¿Cuáles son las comunas con más noticias positivas?

5 comunas con más noticias positivas y su porcentaje respecto al total de noticias de la comuna:

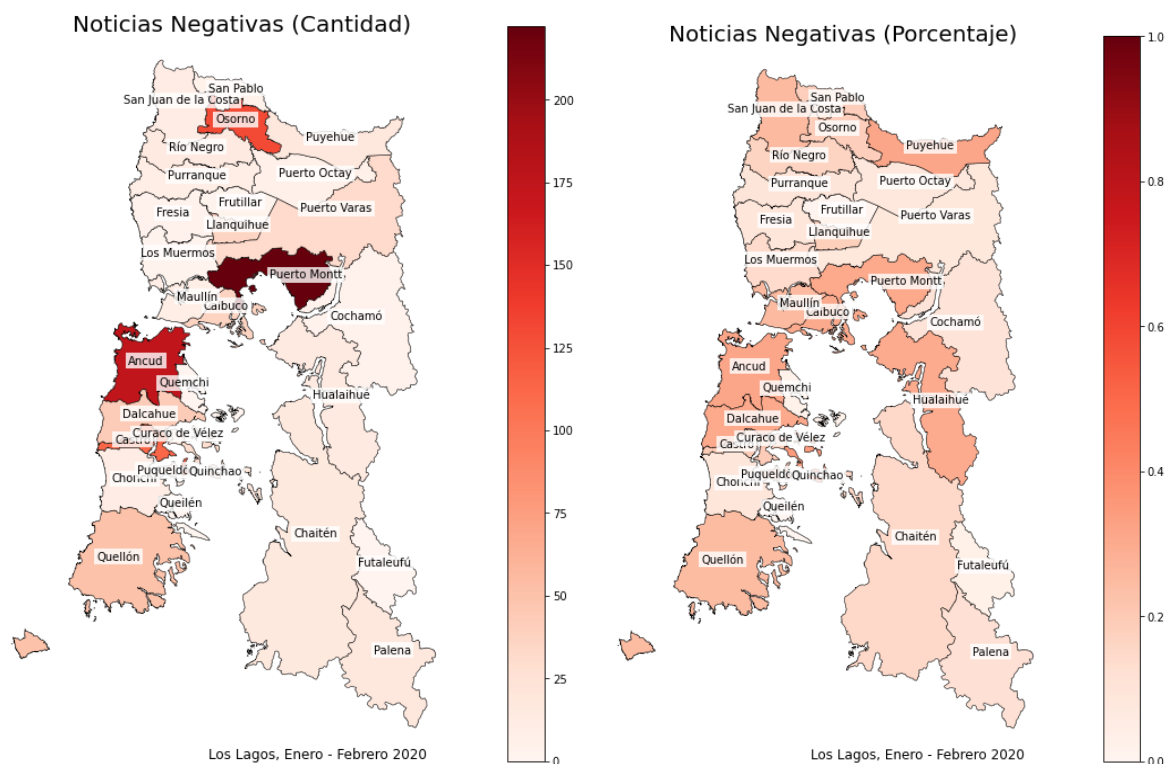
Comuna	N° noticias positivas	% noticias positivas
Osorno	126	18.1556
Puerto Montt	107	14.7994
Castro	104	17.8082
Puerto Varas	90	24.8619
Frutillar	76	34.2342



Sección 2: ¿Cuáles son las comunas con más noticias negativas?

5 comunas con más noticias negativas y su porcentaje respecto al total de noticias de la comuna:

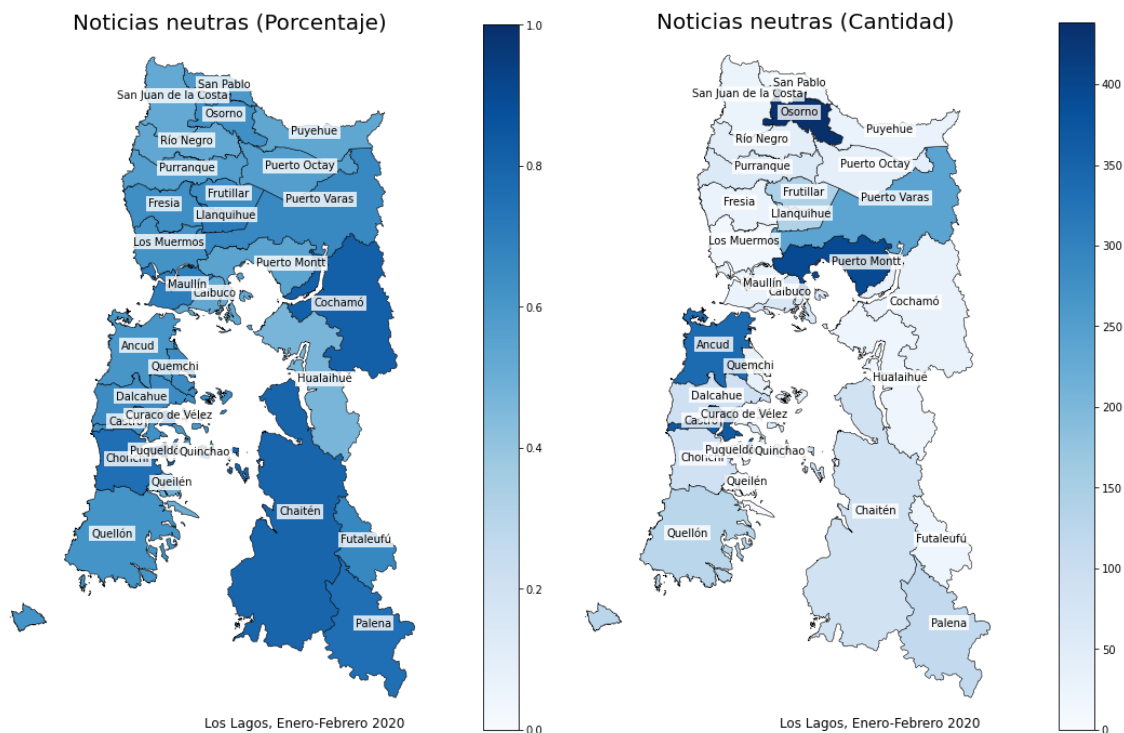
Comuna	N° noticias negativas	% noticias negativas
Puerto Montt	222	30.7054
Ancud	175	31.25
Osorno	130	18.7320
Castro	113	19.3493
Quellón	51	24.7573



Sección 3: ¿Cuáles son las comunas con más noticias neutras?

5 comunas con más noticias neutras y su porcentaje respecto al total de noticias de la comuna:

Comuna	N° noticias neutras	% noticias neutras
Osorno	438	63.1124
Puerto Montt	394	54.4952
Castro	367	62.8425
Ancud	340	60.7143
Puerto Varas	240	66.2983



Sección 4: ¿Cuál es el total de noticias Positivas, Neutras y Negativas?

N° noticias positivas	N° Noticias negativas	N° Noticias Neutras
848	1033	3141

Reporte 3: ¿Cuántos delitos han ocurrido en este periodo en la región de los Lagos?

Ventana de tiempo: 1 mes.

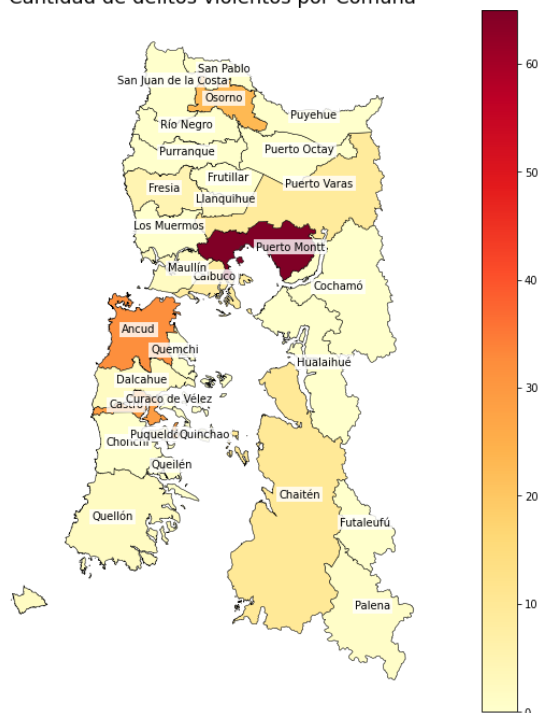
Mapa: Región de Los Lagos.

Para este caso, se usó el tópico 4, ya que este tiene las siguientes keywords: Delito, detención, delitos, investigación, penal, antecedentes, sujeto, prisión, etc

Sección 1: ¿Cuáles son las 5 comunas con más delitos violentos en el periodo del 01-01-2020 al 01-02-2020?

Comuna	Cantidad de delitos violentos
Puerto Montt	65
Ancud	32
Castro	29
Osorno	23
Chaitén	10

Cantidad de delitos violentos por Comuna

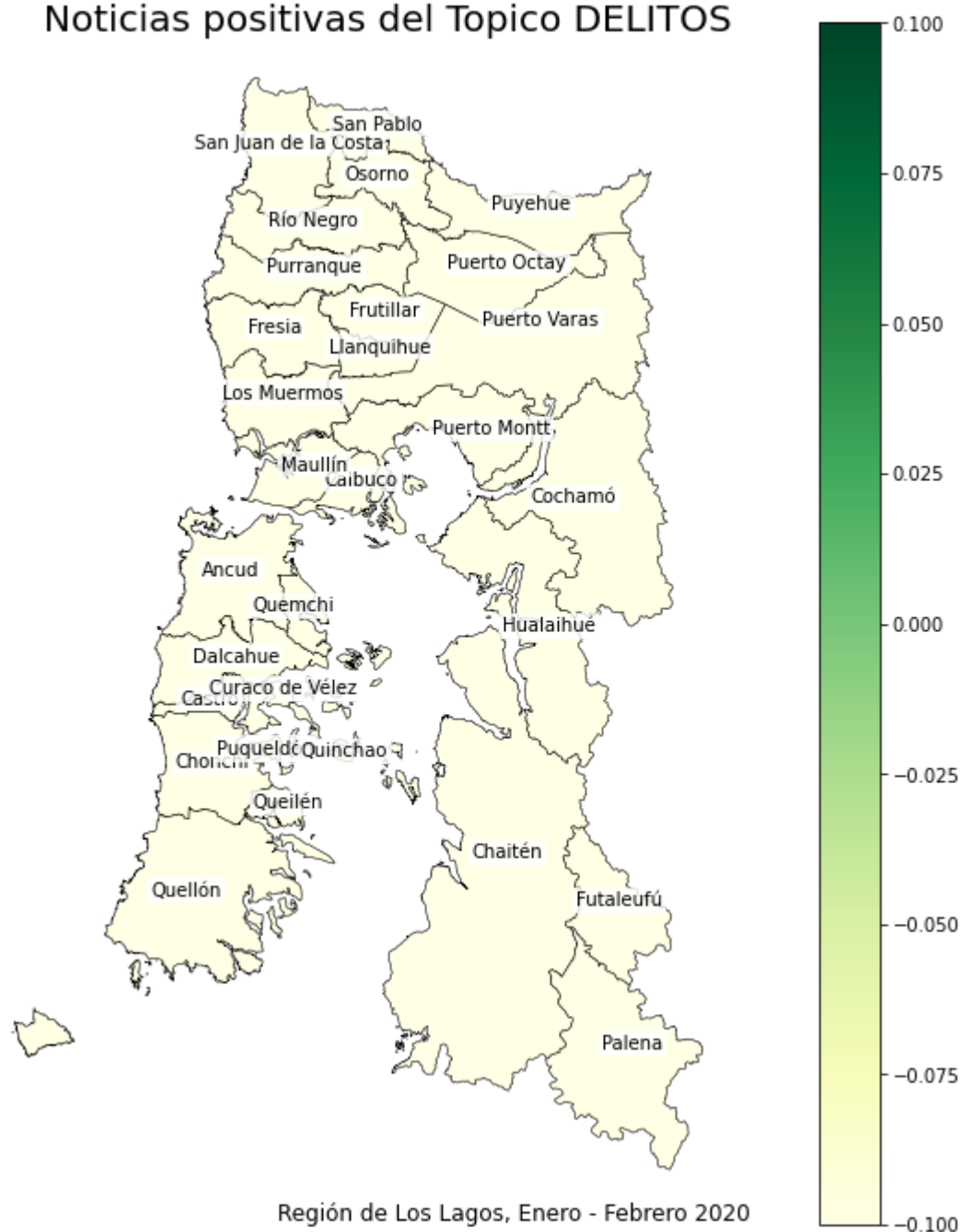


Los Lagos, Enero 2020 a Febrero 2020, medios locales

Sección 2: ¿Cuáles son las comunas con más noticias positivas?

En este caso, al ser todos delitos, no hay ninguna noticia relacionada al tópico 4 que sea positiva.

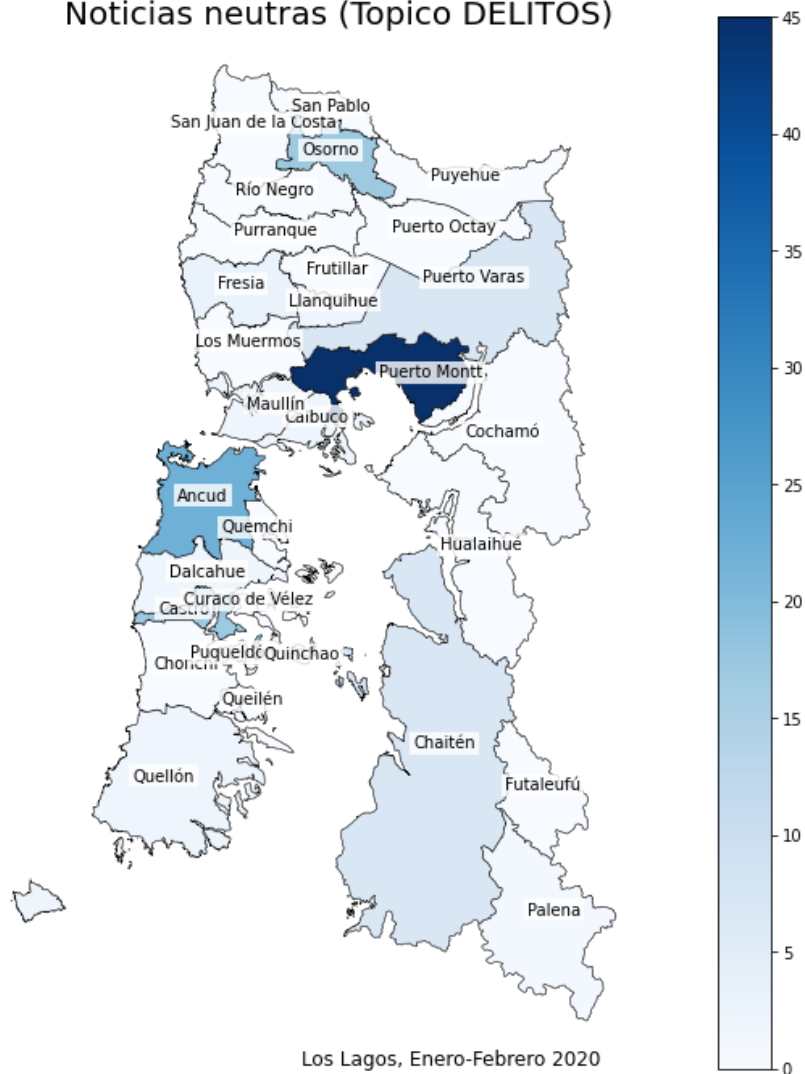
Noticias positivas del Topico DELITOS



Sección 3: ¿Cuáles son las 5 comunas con más noticias neutras?

Comuna	N° de noticias neutras
Puerto Montt	45
Ancud	22
Castro	17
Osorno	17
Chaitén	7

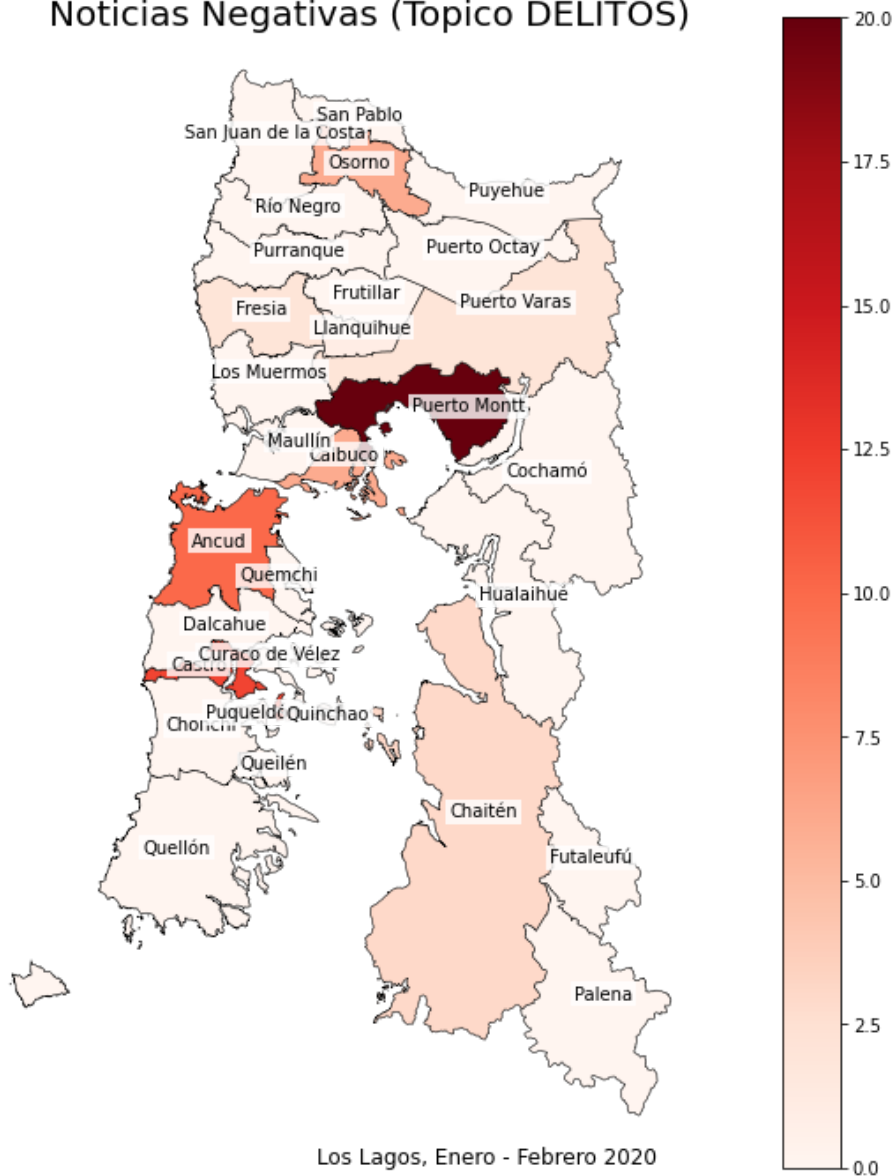
Noticias neutras (Topico DELITOS)



Sección 4: ¿Cuáles son las 5 comunas con más noticias negativas?

Comuna	N° noticias negativas
Puerto Montt	20
Castro	12
Ancud	10
Calbuco	6
Osorno	6

Noticias Negativas (Topico DELITOS)



Sección 5: 5 ejemplos de noticias relacionadas a delitos

-Caso Sophie: defensa del acusado Francisco Ríos busca anular la condena acusando vicios en la sentencia

-Hombre acusado por femicidio en Calbuco es declarado culpable

-Condenan a 9 personas, entre ellos dos gendarmes por formar una red de narcotráfico al interior de la cárcel de Puerto Montt

-Entre cinco y 11 años condenan a asaltantes a servicentros en Puerto Montt

-Hoy se inició en Puerto Montt el juicio oral por femicidio registrado en Calbuco

Conclusiones

Como podemos observar en los reportes, las comunas más nombradas son las principales de la región, (Puerto Montt, Castro, Osorno, etc), así mismo estas son también las que contienen las noticias más positivas, neutras y negativas.

En base al reporte número 3, podemos rescatar información muy importante, ya que estas nos muestran las comunas donde ocurren más delitos, lo que en un futuro, o para una institución como el GORE, podría ayudar a tomar decisiones respecto a futuros proyectos para el cuidado de la integridad de la ciudadanía.

Así mismo, del mismo reporte en un futuro se podría indagar cuales son los delitos más comunes, la hora en que ocurren y los días, esta información, en conjunto a inteligencias artificiales podría ayudar a predecir y evitar futuros delitos, lo que sería un gran avance en la seguridad de la ciudadanía.