




Universidad del Desarrollo

# **El Chile que queremos: Relación entre las necesidades del país versus las necesidades personales y familiares**

18 Diciembre, 2023

# Tabla de Contenidos

- 
- ¿Que es 'El Chile que queremos'?
  - Hipótesis y objetivos
  - Metodología e Investigación
  - Resultados
  - Conclusión y Próximos Pasos

# ¿Que es 'El Chile que queremos'?

## Resumen de 'El Chile que queremos'



- 'El Chile que queremos' es una iniciativa del gobierno de Chile para construir un país más justo, inclusivo y próspero.
- Se centra en el desarrollo de políticas públicas que aborden desafíos en educación, salud, empleo, medio ambiente, entre otros.
- La iniciativa busca la participación activa de la ciudadanía en la toma de decisiones y la construcción de un futuro sostenible.
- A través de 'El Chile que queremos', el gobierno se compromete a cumplir con metas y objetivos definidos en un plan de acción concreto.

# Hipótesis y Objetivos

## Hipótesis

1. Las percepciones de necesidades a nivel nacional están significativamente correlacionadas con las necesidades personales o familiares de los ciudadanos chilenos.
2. Las necesidades y preocupaciones de los ciudadanos varían considerablemente en función de su ubicación geográfica y características demográficas.

## Objetivos

### General:

Explorar y entender cómo las percepciones de necesidades nacionales y personales/familiares en Chile están interrelacionadas y cómo estas varían en diferentes regiones y entre distintos grupos demográficos.

### Específico:

1. Identificar y clasificar las principales necesidades expresadas por los participantes de 'El Chile que Queremos'.
2. Determinar la relación entre las necesidades nacionales y personales/familiares mediante técnicas de análisis de datos.
3. Examinar cómo las percepciones de necesidades difieren entre distintas regiones y demografías en Chile.

# Metodología de Investigación

## Contenido

- Recopilación de Datos
- Preprocesamiento de los datos
- Word Embeddings
- Clustering

# Metodología de Investigación: Recopilación de Datos

## Recopilación de Datos

El dataset fue extraído del Github del Ministerio de Ciencia.

El dataset cuenta con datos demográficos y las respuestas a las preguntas.

Consta de 13.325 registros y 289 columnas.

## Categoría de Preguntas

- **P1\_A:** Emociones / Sentimientos / Sensaciones
- **P1\_B:** Explique lo mencionado
- **P2\_A:** Necesidad que enfrenta el país
- **P2\_B:** Explique lo mencionado
- **P2\_C:** Urgencia
- **P3\_A:** Necesidades que enfrentó personalmente o que existen en mi hogar o familia
- **P3\_B:** Explique lo mencionado
- **P4\_A:** Necesidades del país identificadas
- **P4\_B:** Rol del Estado (Describa)
- **P4\_C:** Actor social (empresa, organizaciones sociales, medios de comunicación, comunidad, etc.)
- **P4\_D:** Rol del actor social (Describa)
- **P5:** ¿Cuál es mi contribución personal para construir el Chile que queremos?

### Necesidad País:

- **P2\_A:** Necesidad que enfrenta el país
- **P2\_B:** Explique lo mencionado

### Necesidad Personal:

- **P3\_A:** Necesidades que enfrentó personalmente o que existen en mi hogar o familia
- **P3\_B:** Explique lo mencionado

# Metodología de Investigación: Preprocesamiento

## Lematización

- Luego de seleccionar las preguntas y reorganizar el dataset, utilizaremos un modelo pre entrenado de spaCy para lematizar las respuestas a las preguntas P2\_A Necesidades Pais y P3\_A Necesidades Personales.

### Desafíos en esta etapa

- Definición del uso de las Part of Speech (POS) que se utilizarían:  
VERB, ADJ, NOUN, PROPN, SYM, AUX, etc.

P3_A	P3_A Lematizado	P3_A Tokens
oportunidades	oportunidad	['oportunidad']
pensiones	pension	['pension']
mejorar pensiones	pension	['pension']
mejores empleos	empleo	['empleo']
• acceso a la vivienda	acceso vivienda	['acceso', 'vivienda']
• tener derechos laborales.	derecho	['derecho']
falta ingresos	ingreso	['ingreso']
orden y seguridad	orden seguridad	['orden', 'seguridad']
justicia social	justicia	['justicia']
transporte publico	transporte	['transporte']
economia familiar	economia	['economia']
salud deficiente	salud	['salud']
pocos beneficios sociales	beneficio	['beneficio']

En esta etapa el texto ya queda estandarizado para pasar a la siguiente etapa.

## Tokenización y eliminación de Stop Words

- Hacemos un segundo filtro eliminando las stopwords y generamos una columna con los tokens para el Word Embedding.

# Metodología de Investigación: Word Embeddings

## Word Embedding con FastText

- Este modelo se destaca por su eficiencia al manejar palabras fuera de vocabulario (OOV) a través de subunidades de texto (subpalabras).
- FastText pondera los vectores no solo considerando las palabras completas, si no, los n-gramas de caracteres asignados.

### Cómo Funciona

Pongamos de ejemplo la palabra "salud" que sería representada por los  $n$ -gramas:

*<sa, sal, alu, lud, salud>*

Cada palabra del  $n$ -grama se presenta como un vector de 300 dimensiones y posteriormente se promedian, generando un único vector que captura la semántica y relación con otras palabras

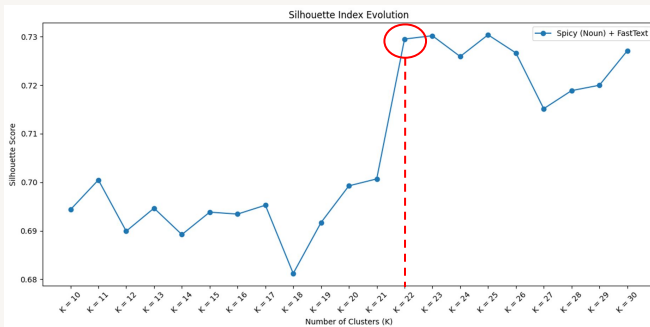
P3_A	P3_A Lematizado	P3_A Tokens
oportunidades	oportunidad	['oportunidad']
pensiones	pension	['pension']
mejorar pensiones	pension	['pension']
mejores empleos	empleo	['empleo']
• acceso a la vivienda	acceso vivienda	['acceso', 'vivienda']
• tener derechos laborales.	derecho	['derecho']
falta ingresos	ingreso	['ingreso']
orden y seguridad	orden seguridad	['orden', 'seguridad']
justicia social	justicia	['justicia']
transporte publico	transporte	['transporte']
economia familiar	economia	['economia']
salud deficiente	salud	['salud']
pocos beneficios sociales	beneficio	['beneficio']



# Metodología de Investigación: Clustering

## Clustering

- Para la categorización se optó el uso del algoritmo Kmeans de Sklearn, este método fue elegido por su habilidad para manejar grandes volúmenes de datos y por su eficiencia en revelar patrones subyacentes en conjuntos de datos complejos.
- Se configuró el modelo KMeans para categorizar en 22 clústeres distintos, basándose en la evaluación previa de la estructura de los datos con silhouette score.



Lemma	Spicy	Spicy	Spicy	Spicy
POS	(NOUN, SYM)	(NOUN)	(NOUN, VERB)	(NOUN, VERB)
Word Embedding	FastText	Spicy	Spicy	Spicy
pipeline		es_core_news_md		es_core_news_lg
Cluster	Score	Score	Score	Score
K = 10	0,694	0,642	0,639	0,376
K = 11	0,700	0,643	0,640	0,372
K = 12	0,690	0,644	0,641	0,384
K = 13	0,695	0,646	0,649	0,385
K = 14	0,689	0,651	0,651	0,397
K = 15	0,694	0,652	0,659	0,401
K = 16	0,693	0,662	0,659	0,403
K = 17	0,695	0,662	0,659	0,407
K = 18	0,681	0,663	0,660	0,417
K = 19	0,692	0,663	0,655	0,418
K = 20	0,699	0,669	0,661	0,412
K = 21	0,701	0,665	0,669	0,416
K = 22	<b>0,730</b>	0,671	0,666	0,420
K = 23	0,730	0,670	0,670	0,432
K = 24	0,726	0,672	0,682	0,422
K = 25	0,730	0,687	0,684	0,420
K = 26	0,727	0,690	0,681	0,428
K = 27	0,715	0,692	0,683	0,432
K = 28	0,719	0,693	0,686	0,428
K = 29	0,720	0,694	0,686	0,433
K = 30	0,727	0,692	0,691	0,425

# Resultados: Resumen hasta ahora

**22**

Número de clusters designados

**66%**

Porcentaje de datos del total de encuestas

**0.73**

Silhouette Score

**2**

Clases de POS utilizados en la lematización



Después de eliminar cadenas vacías post preprocesamiento



Se seleccionaron NOUN y SYM

# Resultados

## Calidad del Cluster

- Se aprecia una relación en los principales insights en los cluster de necesidades del país versus personales.
- Los temas de mayor frecuencia y relevancia, reflejados en los clusters como pensiones, vivienda, salud y educación, resaltan la importancia de estas áreas en la vida cotidiana y la demanda pública para su consideración en agendas políticas y sociales.
- Algunos cluster tienden a agrupar muchos temas por lo que no se logra definir el tema principal del mismo.

### Ejemplo Cluster 15 Necesidades Personales:

Este cluster agrupa 6.868 respuestas de varios temas como servicios básicos, medio ambiente, tiempo libre, calidad de los caminos y temas como espacios (Públicos, Casa, Oportunidad).

### Necesidades Personales

- \* **Cluster 0:** transporte, diálogo
- \* **Cluster 1:** salud, salud salud
- \* **Cluster 2:** dinero, falta trabajo
- \* **Cluster 3:** cambio, cambio afp
- \* **Cluster 4:** educacion, edad
- \* **Cluster 5:** pension, pension pension
- \* **Cluster 6:** vivienda, vivir
- \* **Cluster 7:** justicia, acceso vivienda
- \* **Cluster 8:** sueldo, sueldo sueldo
- \* **Cluster 9:** seguridad, estabilidad
- \* **Cluster 10:** afp
- \* **Cluster 11:** medicamento, endeudamiento
- \* **Cluster 12:** recurso, curso
- \* **Cluster 13:** respeto, hora
- \* **Cluster 14:** trabajo, trabajo mujer
- \* **Cluster 15:** servicio, medio ambiente
- \* **Cluster 16:** empleo, desempleo
- \* **Cluster 17:** agua, ingreso
- \* **Cluster 18:** salario, sala
- \* **Cluster 19:** jubilación, alimentación
- \* **Cluster 20:** igualdad, igualdad género
- \* **Cluster 21:** locomoción, locomoción ser

### Necesidades País

- Cluster 0:** transporte, diálogo
- Cluster 1:** salud, salud salud
- Cluster 2:** paz, abuso
- Cluster 3:** cambio, cambio ley
- Cluster 4:** educacion, edad
- Cluster 5:** pension, pension pension
- Cluster 6:** vivienda, vivienda salud
- Cluster 7:** justicia, delincuencia
- Cluster 8:** sueldo, sueldo afp
- Cluster 9:** seguridad, equidad
- Cluster 10:** afp, afps
- Cluster 11:** medicamento, aumento
- Cluster 12:** recurso, recurso estado
- Cluster 13:** respeto, confianza
- Cluster 14:** trabajo, trabajo sueldo
- Cluster 15:** medio ambiente, sistema
- Cluster 16:** empleo, desempleo
- Cluster 17:** agua, educación calidad
- Cluster 18:** salario, salario salario
- Cluster 19:** jubilacion, sistema pension
- Cluster 20:** igualdad, igualdad género
- Cluster 21:** locomoción

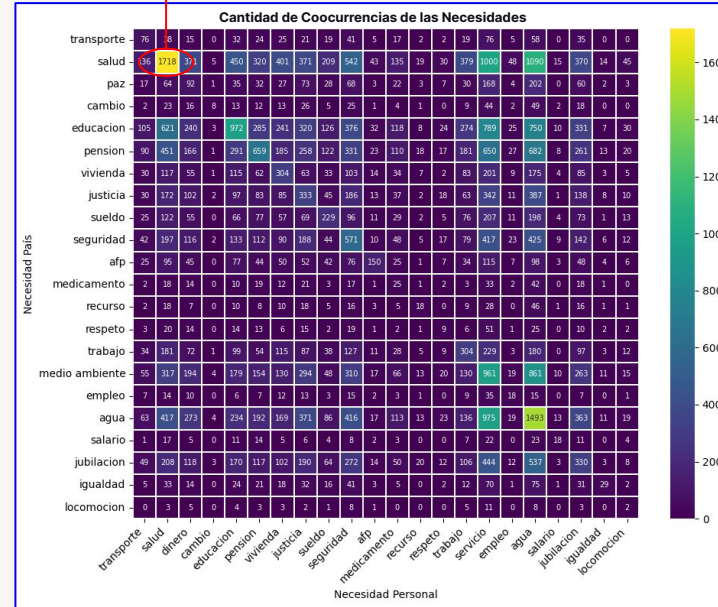
Podemos ver que en el 95% de los cluster hay al menos un insight que se repite en ambos cluster.

# Resultados

## Análisis específico

- A nivel personal vemos que las personas que identificaban temas como salud, educación, pensión o agua como una necesidad personal también las identifican como una necesidad del país.
- Temas como servicio o salud tienen una correlación adyacente, lo que podría indicar que las preocupaciones, ejemplo, sobre la salud, están vinculadas a problemas medioambientales y de acceso al agua.
- Vemos en la tabla que del total de ocurrencias solo un 10,4% identifican la misma necesidad tanto personal como país.

## Cluster 1 Salud



Cluster	Coocurrencias	Total Ocurrencias	Porcentaje
0	76	1.233	6.2%
1	1.718	10.857	15.8%
2	92	2.848	3.2%
3	8	300	2.7%
4	972	7.757	12.5%
5	659	6.217	10.6%
6	304	3.261	9.3%
7	333	4.655	7.2%
8	229	2.369	9.7%
9	571	5.781	9.9%
10	150	1.231	12.2%
11	25	1.098	2.3%
12	18	344	5.2%
13	9	413	2.2%
14	304	3.339	9.1%
15	961	9.978	9.6%
16	18	407	4.4%
17	1.493	11.346	13.2%
18	18	245	7.3%
19	330	5.212	6.3%
20	29	525	5.5%
21	2	267	0.7%
<b>Total</b>	<b>8319</b>	<b>79683</b>	<b>10.4%</b>

# Resultados

## Resultados por Región

### Necesidades País

**Región Antofagasta:** salud, educación, pensión  
**Región Arica y Parinacota:** salud, educación, pensión  
**Región Atacama:** salud, educación, pensión  
**Región Aysén del Gral. Carlos Ibáñez del Campo:** salud, educación, pensión  
**Región Coquimbo:** salud, educación, pensión  
**Región Los Ríos:** salud, educación, pensión  
**Región Magallanes:** salud, educación, pensión  
**Región Metropolitana:** salud, educación, pensión  
**Región Tarapacá:** salud, educación, pensión  
**Región Valparaíso:** salud, educación, pensión  
**Región de la Araucanía:** salud, educación, pensión  
**Región de los Lagos:** salud, educación, pensión  
**Región del BioBío:** salud, educación, pensión  
**Región del Libertador Gral. Bernardo O'higgins:** salud, educación, pensión  
**Región del Maule:** salud, educación, pensión  
**Región del Ñuble:** salud, educación, pensión

### Necesidades Personales

**Región Antofagasta:** salud, educación, vivienda  
**Región Arica y Parinacota:** salud, educación, vivienda  
**Región Atacama:** salud, educación, vivienda  
**Región Aysén del Gral. Carlos Ibáñez del Campo:** salud, educación, vivienda  
**Región Coquimbo:** salud, educación, trabajo  
**Región Los Ríos:** salud, pensión, educación  
**Región Magallanes:** salud, pensión, educación  
**Región Metropolitana:** salud, educación, sueldo  
**Región Tarapacá:** salud, educación, pensión  
**Región Valparaíso:** salud, educación, pensión  
**Región de la Araucanía:** salud, educación, trabajo  
**Región de los Lagos:** salud, educación, vivienda  
**Región del BioBío:** salud, educación, pensión  
**Región del Libertador Gral. Bernardo O'higgins:** salud, educación, pensión  
**Región del Maule:** salud, educación, pensión  
**Región del Ñuble:** salud, educación, trabajo

- Existen variaciones en las necesidades percibidas entre regiones, como 'vivienda' y 'trabajo' en el ámbito personal, sugiriendo diferencias locales en las condiciones económicas y sociales.
- Tanto a nivel personal como a nivel país, 'salud' y 'educación' son prioridades consistentes en todas las regiones, reflejando su importancia transversal en Chile.

# Conclusiones

## 01

Se lograron avances significativos hacia el entendimiento de las necesidades nacionales y personales/familiares en el país, temas tales como Salud, Educación y Pensiones son temas críticos en las necesidades de los chilenos.

## 02

El análisis de los datos de “El Chile que Queremos” nos permitió categorizar las necesidades expresadas por los ciudadanos, no obstante, el análisis podría beneficiarse probando modelos más robustos de word embedding o seguir probando diferentes combinaciones de POS en la lematización.

## 03

El estudio demuestra que tan solo un **10,4%** de las personas expresa las mismas necesidades tanto a nivel personal como a nivel país, siendo el tema Salud el con mayor índice de porcentual de coocurrencia con un 15,8%.

## 04

A nivel regional se destaca la salud y la educación como los temas más importantes a nivel país, por otra parte, existen variaciones en temas como vivienda y trabajo que sugieren diferencias locales en las condiciones económicas y sociales. RM es la única región que dentro de sus principales necesidades personales considera el sueldo, esto puede sugerir un alto nivel de precio de las viviendas, comida o arriendo.

# Próximos Pasos

1

Analizar y comparar más modelos de Word Embedding como BERT o GloVe.

2

Explorar otras características demográficas como la edad o el nivel educacional pueden permitir una mayor comprensión de los datos.

3

Analizar la relación de los resultados con estadísticas nacionales estudios regionales y políticas actuales para proporcionar un contexto más rico.

**Muchas Gracias**